RISK ANALYSIS FOR LARGE POOLS OF LOANS

JUSTIN A. SIRIGNANO AND KAY GIESECKE

Abstract. Financial institutions, government-sponsored enterprises such as Fannie Mae, and asset-backed security investors are often exposed to delinquency and prepayment risk from large numbers of loans. Examples include mortgages, credit cards, auto, commercial, real estate, student, and small business loans. Due to the size of such loan pools and the potentially long maturities of their cashflows, the measurement and management of these exposures is computationally expensive. This paper develops and tests efficient numerical methods for the risk analysis of large pools of loans. For a broad class of dynamic loan-level models of delinquency and prepayment, we develop a law of large numbers and a central limit theorem for the loss and prepayment levels in the pool. The asymptotics are then used to construct efficient Monte Carlo approximations of the loss and prepayment distributions for a large pool. The approximations aggregate the full loan-level dynamics, making it possible to take advantage of the detailed loan-level data often available. The Monte Carlo approximation allows for efficient risk analysis of loan portfolios as well as MBSs, CMOs, and other ABSs backed by pools of loans. To demonstrate the effectiveness of our approach, we implement it on a data set of over 25 million actual subprime and agency mortgages. The results show the accuracy and speed of the approximation in comparison to brute-force simulation of a pool, for a variety of pools with different risk characteristics. Computational cost is often several orders of magnitude less than brute-force simulation of the actual pool with a similar level of accuracy. Furthermore, the computational expense of our efficient Monte Carlo approximation is constant no matter the dimension of the loan-level features; this is key since loan-level feature data is often high-dimensional.

1. Introduction

Financial institutions, government sponsored enterprises (e.g., Fannie Mae and Freddie Mac), and investors have credit exposures to large pools of loans, including mortgages, credit card receivables, auto loans, student loans, and business loans. According to recent estimates of the Federal Reserve and the Securities Industry and Financial Markets Association, there is over $13.4 trillion in mortgage debt outstanding, $8.7 trillion in mortgage-related securities, $3.3 trillion in consumer credit, and over $1 trillion in student loans. In particular, mortgages have received significant attention recently. The unprecedented level of subprime mortgage defaults beginning in 2007 was largely responsible for the ensuing financial crisis. Today, there is widespread recognition amongst both market participants and regulators that better models and computational methods for loan delinquency and prepayment are needed. More accurate risk analysis can better inform capital requirements for the banking system, improve asset-backed security (ABS) ratings, and help financial institutions better manage risk for pools of loans they hold or trade. The main challenges include developing more accurate models, methods to fit these models to the massive amounts of available data, and ways to efficiently simulate the very large pools of loans common in practice.

Loan pools are computationally challenging to analyze due to their size and the potentially long maturities of their cashflows. For instance, mortgage-backed securities (MBS) typically have thousands to hundreds of thousands of mortgages, many of which might have 30 year terms. The MBS market is also relatively liquid; for instance, the daily volume for the agency MBS market is, on average, 300 billion dollars. There is often a need for computational speed; a typical mortgage trading desk at a major bank will on a daily basis need to price thousands of mortgage-backed securities and hundreds of collateralized mortgage obligations (CMOs) backed by mortgage pools. The government-sponsored enterprises (GSEs) Freddie Mac and Fannie Mae have credit exposure to roughly 25 million mortgages, either through directly owning the mortgages or providing

Date: May 25, 2015. The authors gratefully acknowledge support from the National Science Foundation through Methodology, Measurement, and Statistics Grant No. 1325031. We thank seminar participants at the International Monetary Fund for comments. Special thanks are due to Kostas Spiliopoulos for insightful comments. We would also like to thank participants at the 2014 Annual INFORMS Meeting and the 2014 SIAM Conference on Financial Mathematics and Engineering, where this paper won the 2014 SIAM Financial Mathematics and Engineering Conference Paper Prize.

credit guarantees against default for the mortgages. Major US banks can service up to ten million mortgages and, even though they do not directly own the loans, mortgage servicing cashflows are still strongly affected by default and prepayment risk. Moreover, a major US bank might directly own a million mortgages. Large pools of loans are also common for auto loans, student loans, credit cards, commercial loans, and wholesale loans. For instance, a major US bank might easily have on the order of 20,000 wholesale loans and 100,000 mid-market and commercial loans. A major credit card company can have hundreds of millions of credit card accounts. Over half of all consumer credit is eventually securitized, and each deal can consist of tens of millions of credit card accounts. These institutions need risk management tools for their large loan portfolios and investors require methods to price and analyze risk for MBSs, CMOs, and other ABSs.

This paper develops and tests efficient numerical methods for the analysis of large, heterogeneous pools of loans. We focus on a broad class of dynamic, discrete-time models of loan-by-loan delinquency and prepayment. The functions for the conditional transition probabilities are allowed to be very general and could come from a range of statistical or machine learning models such as generalized linear models (for example, logistic regression), support vector machines, neural networks, and decision trees. The transitions might depend on a vector of loan-level features such as credit score and loan-to-value ratio, a vector of common risk factors influencing multiple loans, such as the unemployment rate, and the past behavior of the loans in the pool (through a "mean-field" term). These "data-driven" models are widely used in practice and are fitted from historical loan performance data that are collected internally or acquired from data vendors. For this important class of models, we develop a law of large numbers and a central limit theorem for the pool loss and prepayment processes. The convergence results are then used to construct efficient Monte Carlo approximations of the default and prepayment distributions for a large pool. The approximations account for the full loan-level dynamics, making it possible to take advantage of the detailed loan-level data often available for loans. Very importantly, the cost of our approximation remains constant no matter the dimension of the loan-level features. This is essential since loan-level data is often high-dimensional (the dimension could be in the hundreds). Furthermore, since the law of large numbers and central limit theorem are dynamic, the approximation provides the loss and prepayment distributions across all time horizons at no extra computational cost. Given the loss and prepayment distributions for a pool, risk analysis for MBSs, CMOs, and other ABSs backed by the pool is immediate.

In order to demonstrate the accuracy and low computational expense of our approximations, we numerically test our approach on a large, loan-level mortgage data set which includes over 10 million subprime mortgages and 16 million agency mortgages. We compare the approximate distribution with the true distribution (obtained by brute-force simulation) for various pools drawn from this data set. The comparison is performed using model parameters fitted to the data set. Therefore, the numerical studies reflect the actual performance of the approximation method with real data from actual mortgage pools. The approximation's computational cost is often several orders of magnitude less than the cost of brute-force simulation of the actual pool and has a similar level of accuracy. Although the approximation's accuracy increases with the size of the pool, it is highly accurate even for a pool having as little as 500 loans.

1.1. **Literature.** The computational expense associated with loan-by-loan models for large pools is widely recognized. Prior research has analyzed several approaches to tackle this issue. [43], [51], and [53] propose and implement distributed simulation (i.e., parallel computing) of mortgage-backed securities using clusters of computers. [18], [52], [40], [30] and [32] develop top-down models of mortgage pool behavior, and [3], [23], [16], [17] and others develop top-down models of corporate credit pools. The models are formulated without reference to the constituent credits and only model the pool in aggregate. The Monte Carlo approximation that we develop for loan-by-loan models is as tractable as a top-down model while taking full advantage of the loan-level feature information.

Laws of large numbers and central limit theorems have previously been proven for pools of loans with default risk in other model frameworks; see [9], [12], [13], [15], [21], [22], [39], [42], and others. In contrast to this literature, we consider a discrete-time formulation that is natural given the data structure common in practice, where events are reported on a monthly or quarterly basis. Our formulation is well-adapted to settings with high-dimensional loan-level feature data, which are typically hard to treat using earlier approaches. Earlier model frameworks for which limiting laws have been proven do not include loan-level data as a model input. In this paper, we are able to treat heterogeneous pools whose constitutent loans have

high-dimensional features, where the features can be continuously valued. This paper also proves limiting laws for the loss from default, while previous literature only proves limiting laws for the default rate (or for the loss assuming a non-random loss given default). Finally, unlike the aforementioned papers, we allow for multiple types of events which may be mutually exclusive (e.g., default, prepayment, and potentially others). Many loans, such as subprime mortgages, are subject to both prepayment and default risk.

There are some additional differences between the current model framework in this paper compared to previous papers [21] and [42]. Those papers develop limiting laws for a continuous-time model where defaults arrive at a stochastic intensity, and the coefficient functions for the stochastic intensity process are of a particular form. In this paper, the transition probability function is allowed to be very general and, for instance, could be chosen from many potential models in machine learning. Moreover, the parameters for the transition probability function in this paper can be more tractably fitted to historical loan performance data than the parameters for the coefficient functions in [21] and [42]. Both those papers and this paper include a common stochastic process which produces correlation between different loans. The convergence results in this paper are stronger than in [21] and [42] since here convergence is proven for each path of the common stochastic process. This paper also includes a more general mean-field term than in [21] and [42], whose mean field term is simply the default rate of the pool of loans. Finally, as mentioned, [21] and [42] only prove limiting laws for the default rate, not the loss given default.

Discrete-time interacting particle models have been previously studied, especially in the context of filtering. These papers consider computationally challenging nonlinear filtering equations which are discrete-time measure-valued evolution equations. To solve these equations, they prove that a system of interacting particles weakly converges to the solution of the nonlinear filtering equation. Provided a sequence of observations, a system of particles transition according to particle filtering updates. In the limit of the number of particles tending to infinity, the empirical measure converges to the solution of the nonlinear filtering problem. In effect, they use the finite system to approximate the limiting law, while this paper uses the limiting law to tractably approximate the finite system. In [35], a nonlinear filtering equation is proven to be the limiting law of a discrete-time particle system with mean-field interaction. A central limit theorem is proven for the same system in [36]. [11] proves a convergence result for a nonlinear filtering equation using a branching and interacting particle system. Discrete-time mean field games have also been studied, see [24], [48], and [49]. A mean-field game includes both mean-field interaction and decisions being made by the agents. Finally, there is some discrete-time mean field literature on various other topics, including computer systems (e.g., [7]) and traffic (e.g., [37]).

1.2. **Structure of the paper.** The class of models we consider is described in Section 2. The law of large numbers and central limit theorem for this class of models is presented in Section 3. These limiting laws are used to develop an efficient Monte Carlo approximation for the pool loss and prepayment processes. Numerical methods to simulate the law of large numbers and central limit theorem are described in Section 4. The efficient Monte Carlo approximation is tested on actual mortgage data. Section 5 describes the mortgage data we use and Section 6 contains the various numerical studies. All proofs can be found in Appendix A. Appendix B proves a convergence rate for a simulation scheme for the limiting laws.

2. Problem Formulation

2.1. **Model Framework.** We consider a broad family of dynamic loan-by-loan models for loan delinquency and prepayment in a pool at times $t \in I = \{0, 1, \ldots, T\}$. We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an information filtration $(\mathcal{F}_t)_{t\in I}$. $\mathbb{P}$ is the actual probability measure. The total number of loans initially in the pool is N. The process $U^n = (U^n_t)_{t\in I}$ prescribes the state of the n-th loan. The variable U^n_t takes values in a finite discrete space $\mathcal{U}$. A typical model would include states for when the loan is still outstanding, has been prepaid, or is in default (usually 90+ days late). The default and prepaid states are absorbing states. The outstanding, prepaid, and default states will be denoted by o, p, and d, respectively. Other possible states include 30 days late, 60 days late, etc.

We allow for idiosyncratic, systematic, and contagion factors to influence the dynamics of U^n. Each loan has an $\mathcal{F}_0$-measurable loan-level covariate vector $Y^n \in \mathcal{Y} \subseteq \mathbb{R}^{d_Y}$,[1] which can contain static variables such as

[1] Y^n includes both continuous and categorical variables. Categorical variables are encoded as a vector whose elements are each in $\{0, 1\}$. Of course, $\{0, 1\}$ is a subset of the real line, so $\mathcal{Y} \subseteq \mathbb{R}^{d_Y}$.

loan-to-value (LTV) ratio, credit score, geographic location, type of loan, and historical loan performance up until the initial time of interest $t = 0$ (for instance, how many days behind payment the loan is or whether it is in foreclosure). These loan-level factors are specific to each loan and are sources of idiosyncratic risk. We also consider systematic risk factors V which will have a common influence across many loans in the pool. The vector process $V = (V_t)_{t\in I}$, where $V_t \in \mathbb{R}^{d_V}$, might represent the behavior of local and national economic conditions such as the unemployment rate, housing prices, and mortgage rates.[2] The risk factor V is exogeneous in the sense that its dynamics are not affected by the states U^n nor the Y^n. Finally, we allow borrowers' past behavior to influence the future dynamics of the states. Define the "mean-field" process $H^N = (H_t^N)_{t\in I}$ as:

$$H_t^N = \frac{1}{N}\sum_{n=1}^{N} f^H(U_t^n, Y^n),$$

where $f^H = \left(f_1^H, \ldots, f_K^H\right)$ and $f_k^H : \mathcal{U} \times \mathbb{R}^{d_Y} \mapsto \mathbb{R}$. We let $H_{t:s}^N = (H_t^N, \ldots, H_s^N)$ denote the path of H^N between times t and $s \geq t$. A specific application for H^N would be to model a contagion effect for mortgages where past defaults of mortgages in areas geographically close to the n-th mortgage increase the likelihood of the n-th mortgage defaulting. The process H^N would then keep track of the number of defaults at each geographic location. (See Example 2.1 for a concrete formulation.) In light of the mortgage meltdown, such a feedback mechanism has been supported by several recent empirical papers; see [2], [25], [27], [31], and [46].

The dynamics of U^n are prescribed by the transition function:

$$\mathbb{P}[U_t^n = u|\mathcal{F}_{t-1}] = h_\theta(u, U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N), \quad t \in \{1, \ldots, T\}. \tag{1}$$

where $\tau \geq 1$ is some fixed integer and H_t^N is set to some predetermined constant (independent of N) for $t < 0$. Therefore, the dynamics of the states $U^1, \ldots, U^N$ can potentially depend upon the history of the mean field term H^N. Extension of the model (and convergence results) when h_θ depends upon the history of the common factor V is straightforward. Note that even conditional on the path of the common factor V, the dynamics of the loans $U^1, \ldots, U^N$ are not independent due to the mean-field term H^N. The function h_θ is specified by a parameter θ, which takes values in the compact Euclidean space Θ, and which must be estimated from data on loan performance.

Equation (1) gives the marginal probability for the transitions of the loans from their state at time $t-1$ to time t. Furthermore, we stipulate that conditional on $\mathcal{F}_{t-1}$, the states $U_t^1, \ldots, U_t^N$ are independent. This fully specifies the dynamics of the loans $n = 1, \ldots, N$. In addition, given a state transition $U_{t-1}^n \neq \mathrm{d} \longrightarrow U_t^n = \mathrm{d}$, the n-th loan suffers a loss $\ell_t^n(Y^n, V_t) \in [0, 1]$, where $\ell_t^n(Y^n, V_t)$ is itself a random variable conditional on Y^n and V_t, therefore allowing for idiosyncratic losses.[3] Conditional on V and the n-th loan defaulting at time t, the loss given default is $\ell_t^n(Y^n, V_t)$ where $\mathbb{P}[\ell_t^n(y, v) \in A] = \nu_{t,y,v}(A)$. Finally, conditional on V and both loans n, n' defaulting, the random variables $\ell_t^n(Y^n, V_t)$ and $\ell_t^{n'}(Y^{n'}, V_t)$ are independent. Moreover, since the default state is an absorbing state, losses for the same loan cannot occur twice. More generally, one could consider "costs" or losses associated with each of the states $u \in \mathcal{U}$. For example, there might be servicer costs associated with loans which are behind payment but have not defaulted yet. The results in this paper could be extended to this more general case.

Many models commonly used to describe loan delinquency and prepayment fall under the model class (1): see [1], [4], [5], [6], [10] [41], [44], [50], and many others. Some explicit examples of h_θ are provided at the end of this section. The model framework (1) is popular in practice because it allows for detailed modeling of loan-level dynamics, inclusion of high-dimensional loan-level data, flexible choice of transition functions from statistical and machine learning, and the ability to incorporate important characteristics of loan pools such as correlation between delinquency and prepayment rates for different types of loans, geographic diversity, and burnout. We describe below how the model class (1) can address these characteristics.

The model class (1) allows for correlation between the default and prepayment rates for different types of loans via the dependence on factors V and the mean-field term H^N. The mean-field term H^N can be used to model contagion effects where high default rates in one loan market spread to other loan markets. The

[2]The time t can also be included in the systematic factor vector V.

[3]Here we have implicitly assumed a unit notional for each loan. This paper's results can be extended to the case of loans with different notional sizes as long as the loans' notional sizes are bounded (which they are in practice). The notional size of the loan would be included in the Y^n variable.

dependence on V and H^N also produces correlation between the default and prepayment rates for a single loan type. During times of financial duress, the joint modeling of defaults and prepayments can become essential. For instance, during the financial crisis of 2007-09, even though interest rates were relatively low, mortgage prepayment levels were very low. Default rates will typically be higher when prepayment rates are lower, and vice versa.

The geographic composition of a pool of loans can be important. There may be different correlations between loans in different geographic locations; a loan from California will be less correlated with a loan from Kansas than with another loan from California. The loan-level feature Y^n can include the geographic location of the loan (countries, states, counties, or zip codes). The systematic factor V can include risk factors for each of these different geographic areas. Local contagion can be modeled via including the default rate in a loan's particular geographic location using the mean-field term H^N.

Burnout will naturally occur under model classes such as (1) due to the heterogeneity of a pool where different loans have different incentives or propensities to prepay. Such incentives (e.g., interest rates and prepayment penalties) can be included in the loan-level feature Y^n. The mortgage holders who have stronger incentives will prepay the first time interest rates fall, leaving the pool comprised of mortgages with lower incentives to prepay. Thus, the second time interest rates fall, the prepayment rate will be lower. The burnout effect is also often heightened by some mortgage holders paying closer attention to refinancing opportunities than others. The mortgage holders who pay close attention to refinancing opportunities will prepay the first time mortgage rates fall. This could be modeled by choosing an element of Y^n to be a frailty which represents how closely the mortgage holder pays attention to refinancing opportunities.

A typical choice for h_θ might be a generalized linear model (GLM), although many other choices are available. An example of a GLM is logistic regression. The results in the paper require only very mild assumptions on the form of the function h_θ; these assumptions are satisfied by many standard models such as generalized linear models and neural networks. Some examples are presented below.

Example 2.1 (**Logistic regression model for Mortgages).** *Consider a single zip code and let $\mathcal{U}$ be the state space $\{o, d, p\}$. Let Y^n be a vector containing the credit score, loan-to-value ratio, and earliest interest rate for the n-th loan and let V_t be the unemployment rate, housing price index, and mortgage rate in the zip code at time t. Finally, let $\mathcal{L}^N_{t-1}$ be the fraction of mortgages which defaulted in the $(t-1)$-th period in the zip code and specify $U^n_t = o$ for $t \le 0$. Note that $\mathcal{L}^N_t$ can be written as a function of $H^N_{t-2:t-1}$ since:*

$$\mathcal{L}^N_{t-1} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{1}_{U^n_{t-1}=d} - \mathbf{1}_{U^n_{t-2}=d}), \tag{2}$$

for $t \ge 2$ and $\mathcal{L}^N_t = 0$ for $t = 0, 1$. We choose logistic function for h_θ. Let the parameter vector $\theta = (\theta_d, \theta_p)$ and let $X^n_{t-1} = (1, Y^n, V_{t-1}, \mathcal{L}^N_{t-1})$. The constant 1 is included to provide baseline default and prepayment rates. Then, the dynamics for the n-th mortgage are

$$\begin{aligned}
h_\theta(d, o, Y^n, V_{t-1}, H^N_{t-2:t-1}) &= \frac{e^{\theta_d \cdot X^n_{t-1}}}{1 + e^{\theta_d \cdot X^n_{t-1}} + e^{\theta_p \cdot X^n_{t-1}}}, \\
h_\theta(p, o, Y^n, V_{t-1}, H^N_{t-2:t-1}) &= \frac{e^{\theta_p \cdot X^n_{t-1}}}{1 + e^{\theta_d \cdot X^n_{t-1}} + e^{\theta_p \cdot X^n_{t-1}}}, \\
h_\theta(o, o, Y^n, V_{t-1}, H^N_{t-2:t-1}) &= \frac{1}{1 + e^{\theta_d \cdot X^n_{t-1}} + e^{\theta_p \cdot X^n_{t-1}}}.
\end{aligned} \tag{3}$$

Of course, $h_\theta(o, d, y, v, H) = h_\theta(o, p, y, v, H) = h_\theta(p, d, y, v, H) = h_\theta(d, p, y, v, H) = 0$.

Here we considered a single zip code. This example could be extended to allow interaction between many zip codes; namely, the dynamics of mortgages in a particular zip code could depend upon the number of mortgages which have defaulted in neighboring zip codes.

Example 2.2 (**Neural network model).** *Let the elements of $\mathcal{U}$ be $u_1, \ldots, u_K$. If a loan is in state u_k, denote the states to which it can transition as $u^1_k, \ldots, u^{N_k}_k$. Then,*

$$h_\theta(u_n, u_k, Y^n, V_{t-1}, H^N_{t-1}) = \frac{e^{\sigma_\theta(u_n, u_k, Y^n, V_{t-1}, H^N_{t-1})}}{1 + \sum_{n'=1}^{N_k - 1} e^{\sigma_\theta(u^{n'}_k, u_k, Y^n, V_{t-1}, H^N_{t-1})}}, \quad n < N_k,$$

$$h_\theta(u_{N_k}, u_k, Y^n, V_{t-1}, H^N_{t-1}) \quad = \quad 1 - \sum_{n'=1}^{N_k - 1} h_\theta(u_k^{n'}, u_k, Y^n, V_{t-1}, H^N_{t-1}). \tag{4}$$

The function σ_θ is a neural network with a single hidden layer:

$$\sigma_\theta(u_n, u_k, Y^n, V_{t-1}, H^N_{t-1}) = W^2_\theta q(W^1_\theta Q^n), \tag{5}$$

where $Q^n = (u_n, u_k, Y^n, V_{t-1}, H^N_{t-1}) \in \mathbb{R}^{d_Q \times 1}$, and $W^1_\theta \in \mathbb{R}^{d_h \times d_Q}$ and $W^2_\theta \in \mathbb{R}^{1 \times d_h}$ are weights on the input Q^n and the output of the hidden layer q, respectively. The hidden layer $q : \mathbb{R}^{d_h \times 1} \mapsto \mathbb{R}^{d_h \times 1}$ is the function $q(x) = (q_1(x_1), q_2(x_2), \ldots, q_{d_h}(x_{d_h}))^\top$. Typical choices for $q_1, \ldots, q_{d_h}$ are sigmoid functions and d_h is the number of "neurons" in the hidden layer. See [28] for more details on neural networks.

Example 2.3 (**Decision tree model**). *Divide the space $\mathcal{U} \times \mathcal{Y} \times \mathbb{R}^{d_V} \times \mathbb{R}^K$ into the regions $\{R^m_\theta\}_{m=1}^M$, with the condition that the regions $\{R^m_\theta\}_{m=1}^M$ are disjoint and their union covers the entire space. The transition probability h_θ then is*

$$h_\theta(u, u', Y^n, V_{t-1}, H^N_{t-1}) = \sum_{m=1}^M p^m_\theta(u, u') \mathbf{1}_{(u', Y^n, V_{t-1}, H^N_{t-1}) \in R^m_\theta}, \tag{6}$$

where $\sum_u p^m_\theta(u, u') = 1$ for $m = 1, \ldots, M$. The regions $\{R^m_\theta\}_{m=1}^M$ are chosen by sequentially splitting the space via recursive binary splitting. See [28] for more details on decision trees.

2.2. **Pool dynamics.** For risk management and other applications, one is interested in the behavior of a pool of N loans under the model class (1). The behavior of the pool at time t can be described by the empirical measure of the variables $(U^1_t, Y^1), \ldots, (U^N_t, Y^N)$, which is denoted by $\mu^N_t \in B = \mathcal{P}(\mathcal{U} \times \mathbb{R}^{d_Y})$, where $\mathcal{P}$ is the space of probability measures. Formally,

$$\mu^N_t = \frac{1}{N} \sum_{n=1}^N \delta_{(U^n_t, Y^n)}, \tag{7}$$

where δ is the Dirac measure. The empirical measure $\mu^N_t(u, A) = \int_A \mu^N_t(u, dy)$ gives the fraction of the loans in pool at time t which are in state u and which have loan-level features in the set $A \subset \mathbb{R}^{d_Y}$. The empirical measure completely encodes, up to permutations, the pool dynamics and features. For instance, the default and prepayment rates for the pool can be expressed as $\int_{\mathbb{R}^{d_Y}} \mu^N_t(\mathrm{d}, dy)$ and $\int_{\mathbb{R}^{d_Y}} \mu^N_t(\mathrm{p}, dy)$, respectively. The vector process $H^N \in \mathbb{R}^K$ can also be expressed in terms of the empirical measure: $H^N_t = \sum_{u \in \mathcal{U}} \int_{\mathbb{R}^{d_Y}} f^H(u, y) \mu^N_t(u, dy)$. The loss from default L^N_t at time t is:

$$L^N_t = \frac{1}{N} \sum_{n=1}^N \ell^n_t(Y^n, V_t)(\mathbf{1}_{U^n_t = \mathrm{d}} - \mathbf{1}_{U^n_{t-1} = \mathrm{d}}). \tag{8}$$

The cumulative loss from default up until time t is simply $\sum_{s=1}^t L^N_s$. Let the loss process over all times be $L^N = \{L^N_0, \ldots, L^N_T\} \in [0, 1]^{T+1}$. Given the empirical measure and the loss from default of a pool, risk analysis for MBSs, CMOs, or ABSs backed by the pool is immediate.

For large pools of loans (i.e., large N), the computation of the distribution of the default rate, prepayment rate, and loss from default is computationally burdensome, especially for long time horizons. We address this problem in the remainder of the paper.

3. Limiting Laws and an Efficient Monte Carlo Approximation

Risk analysis for the large pools of loans common in practice is challenging. Due to their large size, brute-force simulation of entire pools is computationally expensive. We develop an efficient Monte Carlo approximation for the loss and prepayment levels in a pool. The approximation is based on a law of large numbers (LLN) and a central limit theorem (CLT) for the empirical measure μ^N and the loss process L^N. The proofs for the theorems below are given in Appendix A.

Assumption 3.1. *Suppose that μ_0^N converges in distribution to $\bar{\mu}_0$, where $\bar{\mu}_0$ is deterministic, and that the distribution of Y^n has compact support on $\mathbb{R}^{d_Y}$ (i.e., $\mathcal{Y}$ is compact). Also, h is continuous and the functions $f_1^H, \ldots, f_K^H$ are continuous and bounded.*[4] *Finally, $g(t,v,y) = \int_0^1 z\nu_{t,y,v}(dz)$ is continuous in y for each t, v.*

Note that since h is a transition probability function, it is automatically bounded.

Theorem 3.2. *Provided Assumption 3.1, the empirical measure μ^N converges in distribution to $\bar{\mu}$ in B^{T+1} as $N \longrightarrow \infty$, where $\bar{\mu}$ is the unique solution of the equation:*

$$\bar{\mu}_t(u, dy) = \sum_{u' \in \mathcal{U}} h_\theta(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1})\bar{\mu}_{t-1}(u', dy), \tag{9}$$

where $\bar{H}_t = \sum_{u \in \mathcal{U}} \int_{\mathbb{R}^{d_Y}} f^H(u,y)\bar{\mu}_t(u, dy)$ and $\bar{H}_{t-\tau:t-1} = (\bar{H}_{t-\tau}, \ldots, \bar{H}_{t-1})$. The loss process L^N converges in distribution to $\bar{L}$ in $[0,1]^{T+1}$ as $N \longrightarrow \infty$, where $\bar{L}$ satisfies the equation:

$$\bar{L}_t = \int_{\mathcal{Y}} \int_0^1 z\nu_{t,y,V_t}(dz)(\bar{\mu}_t(d, dy) - \bar{\mu}_{t-1}(d, dy)). \tag{10}$$

The assumptions required for Theorem 3.2 are relatively mild and are typically satisfied by models used in practice. For instance, logistic regression and neural networks both satisfy Assumption 3.1. The compactness assumption for the feature space $\mathcal{Y}$ is also satisfied by commonly used data features (such as FICO score, LTV ratio, interest rate, and original balance). The final condition on the conditionally expected loss simply states that the expected loss (conditional on V) must be continuous in the loan type $y \in \mathcal{Y}$.

It is important to note that the law of large numbers is dynamic and is also a random equation; randomness enters through the path of the common factor V. $\bar{H}_t$ and $\bar{\mu}_t$ are deterministic functions of the random variable $V_{0:t-1} = (V_0, \ldots, V_{t-1})$. The law of large numbers has a natural link with the original model (1). The function h_θ from (1) appears in the law of large numbers. If the model includes dependence on the past pool dynamics through $\bar{H}_{t-\tau:t-1}$, the law of large numbers is nonlinear. Furthermore, the law of large numbers for the loss $\bar{L}_t$ can be calculated immediately given the solution $\bar{\mu}_t$.

The law of large numbers can also be supplemented with a central limit theorem. Define the empirical fluctuation process $\Xi_t^N = \sqrt{N}(\mu_t^N - \bar{\mu}_t) \in W = S'(\mathcal{U} \times \mathbb{R}^{d_Y})$.[5] The central limit theorem satisfies an equation linearized around the nonlinear dynamics of the law of large numbers. Randomness enters both through V and a martingale term $\bar{\mathcal{M}}$. Like the law of large numbers, the central limit theorem is also dynamic. In addition, we will also prove a central limit theorem for the loss from default. To this end, define the loss fluctuations $\Lambda_t^N = \sqrt{N}(L_t^N - \bar{L}_t) \in \mathbb{R}$.

Assumption 3.3. *Assume that $\sqrt{N}(\mu_0^N - \bar{\mu}_0)$ converges in distribution to $\bar{\Xi}_0$. In addition, assume that $h_\theta(u, u', v, y, H)$ is twice-differentiable in H and its second derivative is bounded.*[6] *Finally, assume that $g_2(t,v,y) = \int_0^1 z^2\nu_{t,y,v}(dz)$ is continuous for each t, v.*

Theorem 3.4. *Provided Assumptions 3.1 and 3.3, Ξ^N converges in distribution to $\bar{\Xi}$ in W^{T+1} as $N \longrightarrow \infty$, where $\bar{\Xi}$ satisfies the equation:*

$$\begin{aligned} \bar{\Xi}_t(u, dy) &= \sum_{u' \in \mathcal{U}} h_\theta(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1})\bar{\Xi}_{t-1}(u', dy) \\ &+ \sum_{u' \in \mathcal{U}} \Big(\frac{\partial}{\partial H} h_\theta(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1}) \cdot \bar{E}_{t-\tau:t-1}\Big)\bar{\mu}_{t-1}(u', dy) + \bar{\mathcal{M}}_t(u, dy), \end{aligned} \tag{11}$$

[4]The assumption that the functions f_k^H are bounded and continuous still allows H_t^N to keep track of the fraction of the pool in any combination of performance states in $\mathcal{U}$ and categorical states in $\mathcal{Y}$. For example, H_t^N could track the fraction of the pool which has defaulted at each geographic location as well as the fraction of the pool which has defaulted for each loan product type (fixed rate, adjustable-rate, interest-only, etc.). The loss in the last time period in each of these categories would simply be $H_{t-1}^N - H_{t-2}^N$.

[5]S' is the space of tempered distributions.

[6]This is not a stringent assumption and is satisfied by a wide range of models used in practice. For instance, logistic regression and neural networks both satisfy this assumption.

where $E_t^N = \sum_{u\in\mathcal{U}} \int_{\mathbb{R}^{d_Y}} f^H(u,y)\Xi_t^N(u,dy)$, $\bar{E}_t = \sum_{u\in\mathcal{U}} \int_{\mathbb{R}^{d_Y}} f^H(u,y)\bar{\Xi}_t(u,dy)$, *and* $\bar{E}_{t-\tau:t-1} = (\bar{E}_{t-\tau},\ldots,\bar{E}_{t-1})$. *Given* V, $\bar{\mathcal{M}}(u,dy)$ *is a conditionally Gaussian process with zero mean and covariance:*

$$Cov\Big[\bar{\mathcal{M}}_t(u_1,dy),\bar{\mathcal{M}}_t(u_2,dy)\Big|V_{0:t-1}\Big] = -\sum_{u'\in\mathcal{U}} h_\theta(u_1,u',y,V_{t-1},\bar{H}_{t-\tau:t-1})h_\theta(u_2,u',y,V_{t-1},\bar{H}_{t-\tau:t-1})\bar{\mu}_{t-1}(u',dy),$$

$$Var\Big[\bar{\mathcal{M}}_t(u,dy)\Big|V_{0:t-1}\Big] = \sum_{u'\in\mathcal{U}} h_\theta(u,u',y,V_{t-1},\bar{H}_{t-\tau:t-1})(1-h_\theta(u,u',y,V_{t-1},\bar{H}_{t-\tau:t-1})\bar{\mu}_{t-1}(u',dy),$$

where $u_1\neq u_2$ *and* $V_{0:t} = (V_0,\ldots,V_t)$. *Given* V, $\bar{\mathcal{M}}_{t_2}$ *is independent of* $\bar{\mathcal{M}}_{t_1}$ *for* $t_2\neq t_1$ *and* $\bar{\mathcal{M}}_t(u,dy_1)$ *is independent of* $\bar{\mathcal{M}}_t(u',dy_2)$ *for* $y_1\neq y_2$. *The fluctuations for the loss* Λ^N *converge in distribution to* $\bar{\Lambda}$ *in* $\mathbb{R}^{T+1}$ *as* $N\to\infty$ *where* $\bar{\Lambda}$ *satisfies:*

$$\bar{\Lambda}_t = \sum_{u\in\mathcal{U}}\int_{\mathcal{Y}}\int_0^1 z\nu_{t,u,y,V_t}(dz)(\bar{\Xi}_t(d,dy)-\bar{\Xi}_{t-1}(d,dy)) + \bar{\mathcal{Z}}_t, \tag{12}$$

where, conditional on V, $\bar{\mathcal{Z}}_t$ *is a mean-zero Gaussian with variance:*

$$Var[\bar{\mathcal{Z}}_t|V_{0:t}] = \int_{\mathcal{Y}}[\int_0^1 z^2\nu_{t,y,V_t}(dz) - (\int_0^1 z\nu_{t,y,V_t}(dz))^2](\bar{\mu}_t(d,dy)-\bar{\mu}_{t-1}(d,dy)).$$

Given V, $\bar{\mathcal{Z}}_{t_1}$ *is independent of* $\bar{\mathcal{Z}}_{t_2}$ *for* $t_1\neq t_2$ *as well as* $\bar{\Xi}$.

The additional assumptions necessary for Theorem 3.4 are mild. The assumption of twice differentiability is satisfied by models commonly used in practice such as logistic regression and neural networks. The final condition on the conditionally expected loss simply states that the variance of the loss (conditional on V) must be continuous in the loan type $y\in\mathcal{Y}$.

Corollary 3.5. *Under the assumptions required for Theorems 3.2 and 3.4,* $(\mu^N, L^N, \Xi^N, \Lambda^N) \xrightarrow{d} (\bar{\mu},\bar{L},\bar{\Xi},\bar{\Lambda})$.

Using Corollary 3.5, the law of large numbers and central limit theorem can be combined to form an approximation for a finite pool of N loans:[7]

$$\begin{aligned} \mu^N &= \bar{\mu} + \frac{1}{\sqrt{N}}\Xi^N \overset{d}{\approx} \bar{\mu}^N = \bar{\mu} + \frac{1}{\sqrt{N}}\bar{\Xi}, \\ L_t^N &\overset{d}{\approx} \bar{L}_t^N = \bar{L}_t + \frac{1}{\sqrt{N}}\bar{\Lambda}_t. \end{aligned} \tag{13}$$

The LLN $\bar{\mu}$ is a first-order approximation of μ^N while the CLT $\bar{\Xi}$ provides a second-order correction. The large pool approximation (13) is conditionally Gaussian given V and can be utilized to simulate large pools of loans for the purposes of forecasting loan default conditions in different geographic areas, risk management for banks and other financial institutions which hold large quantities of loans, and pricing of MBSs, CMOs, and ABSs.[8]

Simulation of the approximation (13) proceeds as follows. First, many paths are simulated from the systematic process V. Conditional on each path of V, $\bar{\mu}$ can be calculated deterministically and the conditional covariance of $\bar{\Xi}$ can be evaluated in closed-form. (For details, see Section 4 below.) Since $\bar{\Xi}$ is conditionally Gaussian, this directly yields the conditional distribution of $\bar{\Xi}$. Then, the unconditional distribution for the approximation $\bar{\mu}^N$ can be found by averaging the conditional distributions across the paths of V since $\mathbb{P}[\bar{\mu}^N\in A] = \mathbb{E}[\mathbb{P}[\bar{\mu}^N\in A|V]]$. We highlight that the approximate loss $\bar{L}_t^N$ is simply a linear function of $(\bar{\mu},\bar{\Xi})$ plus some independent Gaussian noise, and therefore is also very straightforward to simulate. In effect, the approximated loss $\bar{L}_t^N$ can be simulated (at no extra cost) on top of the simulated paths of $(\bar{\mu},\bar{\Xi})$.

The ideas and results in this paper, which are developed for discrete time, can be extended to a continuous time framework where default and prepayment events are modeled in continuous time as counting processes with intensities governed by h_θ. The law of large numbers and central limit theorem for the continuous time framework satisfy a random ODE and an SDE, respectively. Their forms are very similar to the forms given above for the discrete-time model considered in this paper. In fact, the discrete-time limiting laws presented

[7] The addition of $\bar{\mu}$ and $\bar{\Xi}$ for the approximation is rigorous due to Corollary 3.5.

[8] In this application setting, the reference measure $\mathbb{P}$ would be a risk-neutral pricing measure.

here are the Euler discretization for the continuous time limiting laws. Some examples of continuous-time loan models which would be covered under such an extension are [14] and [32].

4. Simulating the Approximation

The numerical evaluation of $\bar{\mu}$ and $\bar{\Xi}$ conditional on each path of V requires discretizing $\mathbb{R}^{d_Y}$ into computational cells and then calculating $\bar{\mu}_t(u, dy)$ and $\bar{\Xi}_t(u, dy)$ at the center of each cell. Let the set of cells be $\mathcal{C} = \{c_1, \ldots, c_K\}$ where $\cup_{k=1}^K c_k = \mathcal{Y}$ with associated centers (i.e., grid points) $\mathcal{R} = \{y_1, \ldots, y_K\}$ where $y_k \in c_k$. The initial measures $\bar{\mu}_0$ would be calibrated to the data of the pool and the second-order correction $\bar{\Xi}_0$ would be set to zero. For each V, $\bar{\Xi}_t$'s distribution on the set of grid points can be computed in closed-form (since it is conditionally Gaussian). The approach is outlined below:

- Define $\bar{\mu}^\Psi$ and $\bar{\Xi}^\Psi$ to satisfy the LLN and CLT evolution equations, but with initial condition $\bar{\mu}^\Psi_{t=0} = \Psi(u, dy)$ where $\Psi(u, dy) = \sum_{k=1}^K \bar{\mu}_{t=0}(u, c_k)\delta_{y_k}$.[9]
- Simulate paths $V^1, \ldots, V^L$ from the random variable V. Conditional on a path V^l, let $\bar{\mu}^\Psi$ and $\bar{\Xi}^\Psi$ be $\bar{\mu}^{\Psi,l}$ and $\bar{\Xi}^{\Psi,l}$, respectively. For each V^l:
  - Given V, $\bar{\mu}$ is deterministic. Calculate $\bar{\mu}^{\Psi,l}_t(u, y_k)$ for all $y_k \in \mathcal{R}$ at the times $t = 1, \ldots, T$.
  - Given V, $\bar{\Xi}$ is a zero mean Gaussian; therefore, its conditional distribution is completely described by its conditional covariance. Then, provided $Cov[\bar{\Xi}_{t-1}(u, y_i), \bar{\Xi}_{t-1}(u', y_j)|V^l]$ and $Cov[\bar{\Xi}_{t-1}(u, y_i), \bar{E}_{0:t-1}|V^l]$ for all $y_i, y_j \in \mathcal{R}$ and $u, u' \in \mathcal{U}$, one can calculate the quantities $Cov[\bar{\Xi}_t(u, y_i), \bar{\Xi}_t(u', y_j)|V^l]$ and $Cov[\bar{\Xi}_t(u, y_i), \bar{E}_{0:t-1}|V^l]$ *in closed-form*. If there is no mean field term, it is only necessary to calculate at each time t the covariance $Cov[\bar{\Xi}_t(u, y_i), \bar{\Xi}_t(u', y_i)|V^l]$. Therefore, one can march forward through time, calculating the (closed-form) covariance of $\bar{\Xi}^{\Psi,l}_t$ across the grid points $y_1, \ldots, y_K$ at each time step.
  - For each V^l and any function f, a numerical approximation can be made:[10]

$$\begin{aligned}\left\langle f, \mu^{N,l}_t \right\rangle_{\mathcal{U}\times\mathcal{Y}} &\overset{d}{\approx} \left\langle f, \bar{\mu}^l_t \right\rangle_{\mathcal{U}\times\mathcal{Y}} + \frac{1}{\sqrt{N}} \left\langle f, \bar{\Xi}^l_t \right\rangle_{\mathcal{U}\times\mathcal{Y}} \\ &\approx \sum_{k=1}^K \sum_{u\in\mathcal{U}} f(u, y_k)[\bar{\mu}^{\Psi,l}_t(u, y_k) + \frac{1}{\sqrt{N}}\bar{\Xi}^{\Psi,l}_t(u, y_k)] \sim \mathcal{N}(m^l, \Sigma^l),\end{aligned}$$

    where $\mu^{N,l}_t$ is the empirical measure μ^N_t for the path V^l. Let $\phi(\cdot, m, \Sigma)$ is the density of Gaussian random variable with mean m and covariance Σ. Since $\bar{\mu}^{\Psi,l}$ is conditionally deterministic and $\bar{\Xi}^{\Psi,l}$ is conditionally Gaussian with covariance known in closed-form, m^l and Σ^l can again be calculated in closed-form.
- Collecting the conditional distributions for each V^l, the density of $\left\langle f, \mu^N_t \right\rangle_{\mathcal{U}\times\mathcal{Y}}$ can be directly computed:

$$p_\theta^{\langle f, \mu^N_t\rangle_{\mathcal{U}\times\mathcal{Y}}}(z) \overset{d}{\approx} \frac{1}{L}\sum_{l=1}^L \phi(z, m^l, \Sigma^l).$$

A convergence rate for the numerical error of the simulation scheme described above is presented in Appendix B. The convergence rate can guide the choice of the location of grid points, number of grid points, and number of Monte Carlo trials. Alternatively, $\bar{\Xi}_t$ could be directly simulated instead of finding its closed-form covariance conditional on each V^l. In this latter approach, one would simply simulate $\bar{\Xi}^\Psi_t$ for each V^l. Such simulation is straightforward since $\bar{\Xi}_t$ is conditionally Gaussian given V. As mentioned previously, given the simulation of the pool dynamics, risk analysis for an MBS, CMO, or ABS backed by the pool is immediate.

---

[9] This is the equivalent of solving the LLN and CLT if $\mathcal{Y}$ was finite-dimensional, i.e., at a finite set of grid points $y_1, \ldots, y_K$. A generalization allowing different sets of grid points for different states u is straightforward.

[10] Here we define $\langle f, \nu\rangle_E = \int_E f(x)\nu(dx)$. For instance, $\langle f, \nu\rangle_{\mathcal{U}\times\mathcal{Y}} = \sum_{u\in\mathcal{U}} \int_{\mathcal{Y}} f(u, y)\nu(u, dy)$. Also, for notational convenience, define $\langle f, \nu(u, dy)\rangle \equiv \langle f(u, \cdot), \nu(u, \cdot)\rangle_{\mathcal{Y}} = \int_{\mathcal{Y}} f(u, y)\nu(u, dy)$.

4.1. **Efficient Low-dimensional Approximation.** There is a potential drawback to the law of large numbers (9) and central limit theorem (11). If the loan-level feature space $\mathcal{Y}$ is very high-dimensional, the law of large numbers will be high-dimensional and computations using traditional *uniform grids* can become expensive due to the curse of dimensionality. The number of mesh points in a uniform grid will grow exponentially with the dimension d_Y of the loan-level feature space. Fortunately, for a reasonably large subclass of models, one can perform a transformation of the LLN in equation (9) which converts the law of large numbers into a low-dimensional problem. A similar transformation can be performed for the CLT equation. The transformation reduces the dimension d_Y of the LLN and CLT to a constant dimension d_W, *no matter how large the original dimension d_Y is*, paving the way for tractable computations.

Suppose that there is a function g_θ such that $h_\theta(u, u', y, v, H) = g_\theta(u, u', f(y), v, H)$ where $f : \mathcal{Y} \mapsto \mathbb{R}^{d_W}$. Then, h_θ is invariant under the coordinate transformation $w = f(y)$ and one can reduce the high-dimensional equation for $\bar{\mu}_t(u, dy)$ to the low-dimensional equation:

$$\bar{\mu}_t(u, dw) = \sum_{u' \in \mathcal{U}} g_\theta(u, u', w, V_{t-1}, \bar{H}_{t-\tau:t-1}) \bar{\mu}_{t-1}(u', dw), \quad w \in \mathbb{R}^{d_W}. \tag{14}$$

For instance, a function h_θ which satisfies this requirement and reduces to the low-dimensional representation in equation (14) is any generalized linear model (GLM), such as logistic regression. Generalized linear models are widely used in practice for default and prepayment modeling. Furthermore, it is to be emphasized that the transformation allows arbitrary complexity with respect to the input factors y and v; the sole restriction is that v can only interact with y through the function $f(y)$. For instance, one can always make a GLM arbitrarily nonlinear in the input space y by adding features which are nonlinear functions of the initial set of features. A common choice is to use basis functions; a very simple example might be polynomials while a more complicated choice might be wavelets. This can greatly increase the dimension d_Y. However, such an expansion of the feature space does not increase the dimension d_W of the low-dimensional LLN (14) and its computational cost remains the same. In fact, *the dimension d_W does not depend on how large the original dimension d_Y is.*

The initial distribution $\bar{\mu}_0(u, dw)$ can be calibrated directly from the data available for the pool of loans. Simply transform the features $Y = (Y^1, \ldots, Y^N)$ to $W = (W^1, \ldots, W^N)$ via the map f and then calculate the empirical distribution of W. Due to its broad applicability and computational tractability, the low-dimensional LLN (14) is of great practical use.

Similarly, one can use the exact same transformation to arrive at a low-dimensional CLT, which has similar computational tractability to the low-dimensional LLN:

$$\begin{aligned} \bar{\Xi}_t(u, dw) &= \sum_{u' \in \mathcal{U}} g_\theta(u, u', w, V_{t-1}, \bar{H}_{t-\tau:t-1}) \bar{\Xi}_{t-1}(u', dw) \\ &+ \sum_{u' \in \mathcal{U}} \big(\frac{\partial}{\partial H} g_\theta(u, u', w, V_{t-1}, \bar{H}_{t-\tau:t-1}) \cdot \bar{E}_{t-\tau:t-1}\big) \bar{\mu}_{t-1}(u', dw) + \bar{\mathcal{M}}_t(u, dw), \quad w \in \mathbb{R}^{d_W}. \end{aligned} \tag{15}$$

Given V, $\bar{\mathcal{M}}(u, dw)$ is a conditionally Gaussian process with zero mean and covariance:

$$\begin{aligned} \mathrm{Cov}\Big[\bar{\mathcal{M}}_t(u_1, dw), \bar{\mathcal{M}}_t(u_2, dw)\Big| V_{0:t-1}\Big] &= -\sum_{u' \in \mathcal{U}} g_\theta(u_1, u', w, V_{t-1}, \bar{H}_{t-\tau:t-1}) g_\theta(u_2, u', w, V_{t-1}, \bar{H}_{t-\tau:t-1}) \bar{\mu}_{t-1}(u', dw), \\ \mathrm{Var}\Big[\bar{\mathcal{M}}_t(u, dw)\Big| V_{0:t-1}\Big] &= \sum_{u' \in \mathcal{U}} g_\theta(u, u', w, V_{t-1}, \bar{H}_{t-\tau:t-1})(1 - g_\theta(u, u', w, V_{t-1}, \bar{H}_{t-\tau:t-1})) \bar{\mu}_{t-1}(u', dw), \end{aligned}$$

where $u_1 \neq u_2$. Some examples of the low-dimensional approximation are presented below.

Example 4.1 (One-dimensional Efficient Approximation). *Let $\mathcal{U}$ be $\{o, p\}$. Let Y^n be a vector containing loan-level features for the n-th loan and V_t be a vector of systematic factors at time t. For this example, h_θ does not depend upon the mean field term H^N. A logistic function is chosen for h_θ. Let the parameter vector $\theta = (\theta_p^Y, \theta_p^X)$ where $X_t = (1, V_t)$. Then, the dynamics for the n-th loan are*

$$\begin{aligned} h_\theta(p, o, Y^n, V_{t-1}, H^N_{t-\tau:t-1}) &= \frac{e^{\theta_p^X \cdot X_{t-1} + \theta_p^Y \cdot Y^n}}{1 + e^{\theta_p^X \cdot X_{t-1} + \theta_p^Y \cdot Y^n}}, \\ h_\theta(o, o, Y^n, V_{t-1}, H^N_{t-\tau:t-1}) &= 1 - h_\theta(p, o, Y^n, V_{t-1}, H^N_{t-\tau:t-1}). \end{aligned} \tag{16}$$

Let $w = \theta_p^Y \cdot y$. Then, both the low-dimensional LLN (14) and the low-dimensional CLT (15) only have one spatial dimension.

Example 4.2 (**Two-dimensional Efficient Approximation**). *Let $\mathcal{U}$ be $\{o, d, p\}$. Let Y^n be a vector containing loan-level features for the n-th loan and V_t be a vector of systematic factors at time t. Again, there is no dependence upon the mean field term H^N. We choose a logistic function for h_θ. Let the parameter vector $\theta = (\theta_d^Y, \theta_d^X, \theta_p^Y, \theta_p^X)$ where $X_t = (1, V_t)$. Then, the dynamics for the n-th loan are*

$$\begin{aligned}
h_\theta(d, o, Y^n, V_{t-1}, H^N_{t-\tau:t-1}) &= \frac{e^{\theta_d^X \cdot X_{t-1} + \theta_d^Y \cdot Y^n}}{1 + e^{\theta_d^X \cdot X_{t-1} + \theta_d^Y \cdot Y^n} + e^{\theta_p^X \cdot X_{t-1} + \theta_p^Y \cdot Y^n}}, \\
h_\theta(p, o, Y^n, V_{t-1}, H^N_{t-\tau:t-1}) &= \frac{e^{\theta^X \cdot X_{t-1} + \theta_p^Y \cdot Y^n}}{1 + e^{\theta_d^X \cdot X_{t-1} + \theta_d^Y \cdot Y^n} + e^{\theta_p^X \cdot X_{t-1} + \theta_p^Y \cdot Y^n}}, \\
h_\theta(o, o, Y^n, V_{t-1}, H^N_{t-\tau:t-1}) &= 1 - h_\theta(p, o, Y^n, V_{t-1}, H^N_{t-\tau:t-1}) - h_\theta(d, o, Y^n, V_{t-1}, H^N_{t-\tau:t-1}). \qquad (17)
\end{aligned}$$

Let $w = (w_1, w_2) = (\theta_d^Y \cdot y, \theta_p^Y \cdot y)$. The low-dimensional LLN (14) and the low-dimensional CLT (15) then have two spatial dimensions.

We emphasize that the low-dimensional approximations for these two examples only have one and two spatial dimensions, respectively, *no matter how large the number of loan-level features d_Y is*. Therefore, the computational expense of the approximation remains constant even as d_Y grows very large.

The approximation using the low-dimensional LLN and CLT is now very computationally tractable for simulating the loss and prepayment levels in a pool. The computational performance of the low-dimensional LLN and CLT can be further enhanced by the use of a non-uniform grid for discretizing $\mathbb{R}^{d_W}$. Using a non-uniform grid, more points would be placed where $\bar{\mu}_0(u, dw)$ is large and less points would be placed where $\bar{\mu}_0(u, dw)$ is small. Section 4.2 provides an example of a particular non-uniform grid well-adapted to our problem. The non-uniform grid proposed in Section 4.2 is highly accurate even with only a small number of grid points. A sparse grid can also be used in order to further decrease computational time. Section 4.3 describes this approach. Using a sparse grid, simulation is performed at only a few points and then the solution is evaluated on a finer grid via interpolation. A final advantage of the approximation is its reusability; one set of simulations on a pre-chosen grid can be used to calculate the distribution for many different pools by re-weighting the simulated CLT and LLN (see Section 4.3). The low-dimensional approximation combined with these computational methods provides *an efficient Monte Carlo approximation* for mortgage pools.

For models of h_θ which do not fall under the subclass of models for which the low-dimensional transformation is applicable, one can still compute the approximation (13) for high-dimensions (without any transformation) via sparse non-uniform grids. We describe and numerically implement this latter approach in Section 6.5, which also strongly outperforms brute-force simulation of the actual pool.

In many typical applications, the efficient Monte Carlo approximation will be several orders of magnitude faster than brute-force Monte Carlo simulation of the original pool with a similar level of accuracy. In fact, just the LLN (without the second-order correction provided by the CLT) is sufficiently accurate for many typically-sized pools of interest. Section 6 studies the numerical performance of the efficient Monte Carlo approximation using actual mortgage data.

4.2. **Non-uniform Grids.** To increase computational efficiency for the low-dimensional LLN and CLT, we recommend non-uniform grids. In a non-uniform grid, more points would be placed where $\bar{\mu}_0(u, dw)$ is large and less points would be placed where $\bar{\mu}_0(u, dw)$ is small. In the case where $h_\theta(y) = h_\theta(y) = g_\theta(w)$ is a logistic function, we propose the following non-uniform grid for $\mathbb{R}^{d_W}$:

- Divide $\mathbb{R}^{d_W}$ into K boxes, each with equal mass $\int_{\text{box } k} \bar{\nu}(dw) = 1/K$ where we take $\bar{\nu}(dw) = \bar{\mu}_0(\text{o}, dw)$. In one dimension, this can be done by finding the quantiles of the distribution $\bar{\nu}$. It is assumed that $\int_{\mathbb{R}^{d_W}} \bar{\mu}_0(\text{o}, dw) = 1$.
- In the k-th box, choose the grid point $w_k = \log K \int_{\text{box } k} e^w \bar{\nu}(dw)$.
- Evaluate the solution $\bar{\mu}_t(u, dw)$ at the grid points $w_1, \ldots, w_K$.

If g_θ is locally linear (at least within the k-th box) in e^w, the grid points y_k can make this scheme highly accurate. We demonstrate for one time-step to explain the choice of the points y_k. Define the function q_θ such that $q_\theta(u, e^w) = g_\theta(u, w)$. The exact mass within the k-th box at $t = 1$ is $\int_{\text{box } k} \bar{\mu}_1(u, dw) = \int_{\text{box } k} g_\theta(u, w) \bar{\nu}(dw) =$

$\int_{\text{box } k} q_\theta(u, e^w)\bar{\nu}(dw)$ where we have suppressed the other arguments of g for notational convenience. If g_θ is approximately locally linear in e^w (i.e., q is approximately linear) in the k-th box, one has that

$$\begin{aligned}\int_{\text{box } k} \bar{\mu}_1(u, dw) &= \int_{\text{box } k} q_\theta(u, e^w)\bar{\nu}(dw) \approx \frac{1}{K} q_\theta(u, K \int_{\text{box } k} e^w \bar{\nu}(dw)) \\ &= \frac{1}{K} g_\theta(u, w_k) = g_\theta(u, w_k) \int_{\text{box } k} \bar{\mu}_0(\mathrm{o}, dw).\end{aligned}$$

Then, if g is close to locally linear in e^w, the choice of the grid point w_k will lead to a very accurate solution for the total mass in the k-th box. In the end, the quantity of interest is the total mass in each state u (i.e., what fraction of loans are still alive, what fraction have defaulted, and what fraction have prepaid), so this is highly useful. One can simply sum up the mass in each box to find the total mass in state u. Although this scheme has been specifically tailored to the case where h_θ is a logistic function, generalizations can be made to other function choices.

4.3. **Pre-computation for Financial Institutions.** Even for the risk analysis of smaller, individual pools, the efficient Monte Carlo approximation can provide considerably faster computations. For instance, for a single pool of $1,000$ loans, although the approximation is accurate, it does not offer as large computational cost savings as for very large pools. However, a typical financial institution will deal with thousands of such pools. As mentioned earlier, a mortgage trading desk at a major bank will on a daily basis analyze thousands of MBSs and hundreds of CMOs.

Assuming there is no mean field dependence in equation (1), one can pre-simulate the LLN and CLT at a set of grid points $\mathcal{R} \in \mathbb{R}^{dw}$. This pre-simulation occurs only once. Then, one can find the distribution for the k-th pool by taking a weighted combination of the pre-simulated approximation $\bar{\mu}^N$ across the grid points $\mathcal{R}$, where the weights are chosen to match the k-th pool's loan-level feature distribution.

If the series of pools have sizes $N_1, \ldots, N_K$ with $N = N_1 + \cdots + N_K$, then the computational cost of the efficient Monte Carlo approximation compared with brute-force Monte Carlo simulation of the actual pool is N_g/N where $N_g = |\mathcal{R}|$ is the number of grid points. Furthermore, the method immediately yields the correlation between the different pools, which is essential for risk management purposes. Namely, for each path of the systematic factor V, we simultaneously have the default and prepayment behavior for all of the $1, \ldots, K$. The approach is summarized below:

- Pre-simulate the (finite-dimensional) LLN $\bar{\mu}^\Psi$ and CLT $\bar{\Xi}^\Psi$ on the grid $\mathcal{R} = \{w_1, \ldots, w_I\}$ with initial condition $\Psi(\mathrm{o}, dw) = \sum_{i=1}^I \delta_{w_i}$.
- For each pool $1, \ldots, K$: Find the k-th pool's distribution in the w-space and approximate it at the grid points $\mathcal{R}$; let z_i be the fraction at the i-th grid point. Then, the k-th pool's distribution is

$$\mu_t^{N_k}(u, w_i) = z_i \bar{\mu}_t^\Psi(u, w_i) + \sqrt{\frac{z_i}{N_k}} \bar{\Xi}^\Psi(u, w_i), \quad , i = 1, \ldots, I, \tag{18}$$

and zero otherwise. The method can be further improved by taking a sparse grid $\mathcal{R}$ in order to reduce the number of calculations and then, after the pre-simulation, interpolating on a finer grid. Due to the smoothness of $\bar{\mu}^N$ for typical functions h_θ, only a few grid points are usually needed in order to get an accurate interpolated solution. Using this approach, the efficient Monte Carlo approximation can be highly useful even for small pools of loans as long as the financial institution is dealing with many such pools in aggregate. The approach is implemented using actual mortgage data in Section 6.4.

5. Mortgage Data

We will use actual mortgage data to test the performance of the efficient Monte Carlo approximation. Our data is comprised of two loan-level data sets that are described below. These data sets will be used to statistically estimate the model (1) and then numerically test the performance of the efficient Monte Carlo approximation on real mortgage pools drawn from the data set.

5.1. **Subprime Mortgage Data.** The first is a subprime data set during the time period $1995 - 2013$, obtained from the Trust Company of the West. There are over 10 million mortgages in the data set. The mortgages are spread across the entire United States, covering over $36,000$ different zip codes. The data includes zip code, FICO score, loan-to-value (LTV) ratio, initial interest rate, initial balance, type of mortgage,

deal ID, and time of origination. Default, foreclosure, modification, real estate owned (REO) and prepayment events, if they occur, are also recorded in the data set on a monthly basis.

5.2. **Agency Mortgage Data.** The second data set, obtained from Freddie Mac, contains agency mortgages over the time period 1999−2014 and consists of 16 million mortgages. The data includes FICO score, first time homebuyer indicator, type of mortgage, maturity date, number of units, occupancy status, combined loan-to-value (CLTV), original debt-to-income (DTI) ratio, LTV ratio, initial interest rate, prepayment penalty mortgage (PPM) indicator, loan purpose (purchase, cash-out refinance, or no cash-out refinance), original loan term, and number of borrowers. The zip codes for each mortgage have been partially anonymized in this data set, but the metropolitan statistical area (MSA) is reported for each mortgage. There are roughly 430 different MSAs reported in the data set. Finally, events such as prepayment, foreclosure, 180+ days delinquent are also recorded on a monthly basis.

5.3. **Parameter Estimation.** Before one analyzes loan pools using the efficient Monte Carlo approximation, the parameter θ specifying the model (1) must be fitted to the loan data described above. More specifically, we wish to estimate θ given observations of $Y = (Y^1, \dots, Y^N)$ and $(U_t^1, \dots, U_t^N, V_t)_{t=1,\dots,T}$. Collectively, the observations of the states up to time T are $\mathsf{D}_{T,N} = (Z_1^N, \dots, Z_T^N)$ where $Z_t^N = (U_t^1, \dots, U_t^N)$. The log-likelihood function for $\mathsf{D}_{T,N}$ given V and Y is

$$\begin{aligned}\mathcal{L}(\theta) = \log \mathbb{P}_\theta(\mathsf{D}_{T,N} \,|\, V, Y) = \log \mathbb{P}_\theta[Z_1^N, \dots, Z_T^N | V, Y] = \log \prod_{t=1}^{T} \mathbb{P}_\theta[Z_t^N | Z_0^N, \dots, Z_{t-1}^N, V, Y] \\ = \log \prod_{t=1}^{T} \mathbb{P}_\theta[Z_t^N | Z_{t-1}^N, H_{t-\tau:t-1}^N, V, Y] = \log \prod_{t=1}^{T} \prod_{n=1}^{N} h_\theta(U_t^n, U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N) \\ = \sum_{t=1}^{T} \sum_{n=1}^{N} \log h_\theta(U_t^n, U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N). \end{aligned} \tag{19}$$

Note that we have used the conditional independence of $U_t^1, \dots, U_t^N$ with respect to $\mathcal{F}_{t-1}$ on the second line in equation (19). The maximum likelihood estimator $\hat{\theta} = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta)$. Stochastic gradient descent can be used to numerically optimize $\mathcal{L}(\theta)$. Typically, one will also choose a separate model for the systematic factors V with its own parameters. These parameters can be estimated separately from θ using standard methods; note that the likelihood for θ depends only on the observed values of V and is independent of V's exact form or parameterization since V is an exogenous process. The consistency and asymptotic normality for the maximum likelihood estimator $\hat{\theta}$ is studied in [41].

6. Computational Performance of the Efficient Monte Carlo Approximation

We compare the accuracy and computational cost of the efficient Monte Carlo approximation with brute-force Monte Carlo simulation using actual mortgage data. The efficient Monte Carlo approximation has very high accuracy even for small pools in the hundreds of mortgages. Mortgage-backed securities can range from a few thousand mortgages to hundreds of thousands of mortgages. Banks and other financial institutions often have credit exposure to hundreds of thousands or millions of mortgages. GSEs such as Fannie Mae and Freddie Mac have credit exposure to tens of millions of mortgages. The computational cost of the efficient Monte Carlo approximation is typically orders of magnitude lower than brute-force Monte Carlo simulation of the actual pool.

6.1. **Numerical Performance of LLN and CLT.** First, we demonstrate the accuracy of the law of large numbers by comparing the LLN distribution with the actual distribution for the pool. The actual (or "true") distribution is found via brute-force Monte Carlo simulation of the actual pool. The pool is drawn at random from the subprime data set, allowing us to assess the performance of the LLN with actual mortgage data.

A multinomial logistic regression model is used for h_θ and there is no dependence on the mean field term H^N (e.g., Example 4.2). The loan-level features Y^n used for both default and prepayment are FICO score, LTV, initial balance of the mortgage, and initial interest rate for the mortgage. The national unemployment rate is used as the systematic factor for default. Both the national unemployment rate and national mortgage rate are used as systematic factors for prepayments. We perform $25,000$ Monte Carlo simulations for both

the brute-force simulation of the actual pool as well as the simulation of the LLN. We simulate the pool for a one-year time horizon, with monthly discretization ($T = 12$). The mortgage rate and unemployment rate are simulated as (independent) discrete-time random walks with standard deviations fitted to their historical values prior to January 1, 2012.[11] The parameters θ are also fitted using the entire subprime data set prior to January 1, 2012. The numerical studies in Sections 6.1 and 6.2 are performed using these parameter fits.[12] We model both default and prepayment, and hence $d_W = 2$ for the efficient Monte Carlo approximation.

Figure 1 compares the actual distribution with the LLN distribution for the loss from default for pools of sizes $N = 5,000$, $10,000$, $25,000$ and $100,000$, respectively. The pools are drawn at random from the data set. The LLN is very accurate, especially in the right tail of the distribution. The right tail is essential for risk management purposes, such as calculating the value at risk (VaR). The LLN can be combined with the CLT to create a second-order accurate approximation. The approximation is accurate even for relatively small pools in the hundreds of mortgages. Figure 2 compares the approximate distribution (using both the LLN and CLT) with the actual distribution for pools with sizes $N = 500, 1000, 2500$, and 5000. Using the LLN alone can underestimate the tails of the distribution for small N. By including the CLT in addition to the LLN, one is able to accurately capture the tail of the distribution.

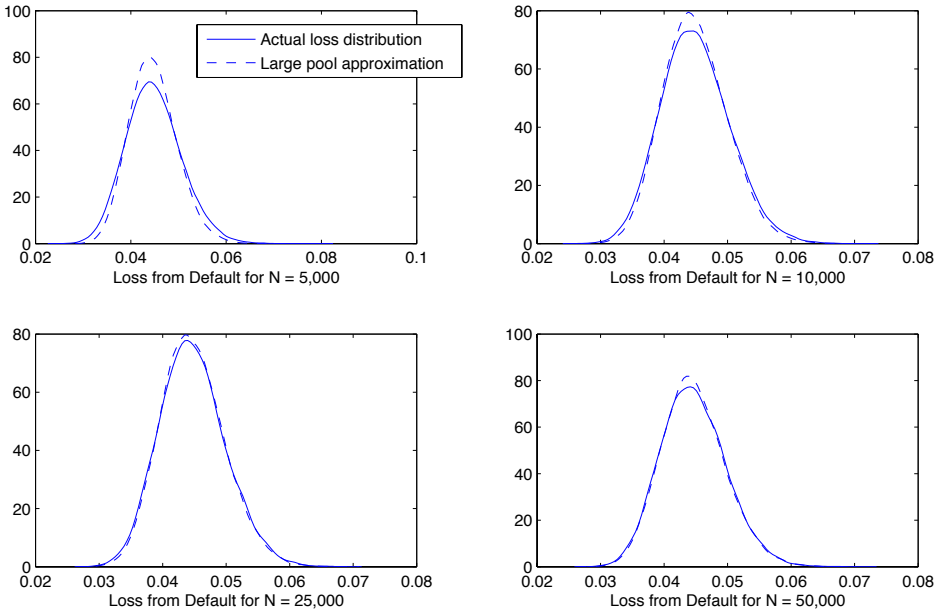

FIGURE 1. Comparison of actual loss distribution with LLN loss distribution (does not include CLT in approximation). Loss reported as fraction of pool which defaulted. The horizon is 12 months.

Computational costs are reported in Table 1 for different sized pools N. In general, a rough approximation of the ratio of the computational costs is

$$\frac{\text{Cost of LLN}}{\text{Cost of Simulation of Actual Pool}} = \frac{N_g}{N}, \tag{20}$$

where N_g is the number of grid points needed for the numerical solution of the LLN and CLT equations across the space $\mathbb{R}^{d_W}$ and N is the number of mortgages in the actual pool. For instance, if one only needs 25 grid points, the ratio of computational costs is $\frac{25}{N}$. For a million mortgages, that leads to a reduction in

[11] More complicated models for V may be worthwhile to implement in practice, which incorporate correlation between systematic factors as well as seasonal effects or trends.

[12] Parameter estimates are available from the authors upon request.

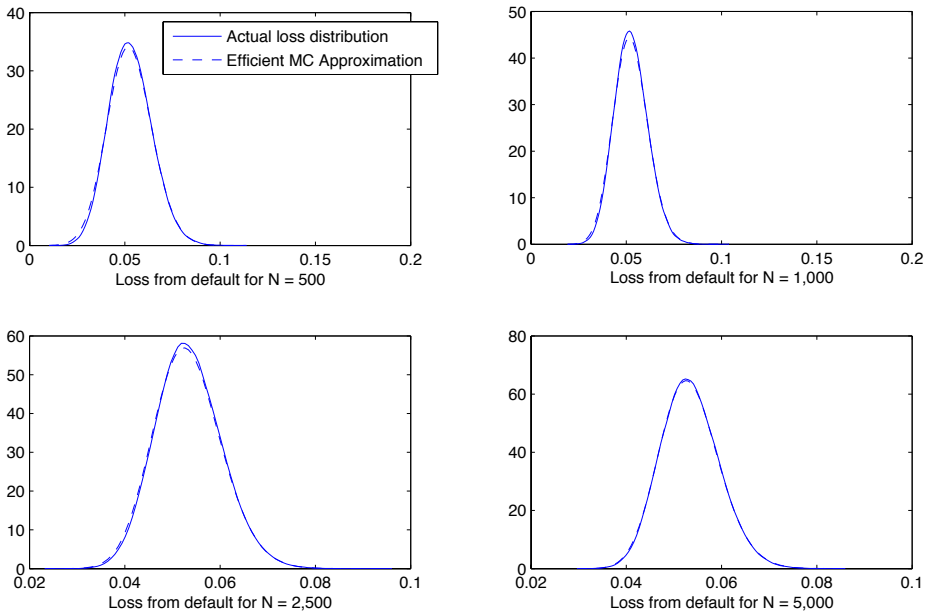

FIGURE 2. Comparison of actual distribution with approximate distribution (using both LLN and CLT). Loss reported as fraction of pool which defaulted. The horizon is 12 months.

computational time of well over 4 orders in magnitude. Computational times listed in Table 1 for brute-force simulation of the pool, simulation of the LLN, and simulation of the full approximation (LLN combined with the CLT) are for $d_W = 2$ (model includes both prepayment and default). These computational times are for a twelve-month time horizon.

N	Time for Brute-force Simulation	Time for LLN	Time for LLN and CLT
1,000	44.34	1.03	2.67
5,000	153.78	1.03	2.67
10,000	273.39	1.03	2.67
25,000	608.94	1.03	2.67
100,000	2,847.43	1.03	2.67
1,000,000	28,563.68	1.03	2.67

TABLE 1. Comparison of computational times (seconds) for efficient Monte Carlo approximation and brute-force Monte Carlo simulation of the pool.

6.2. **Numerical Performance across Actual Deals.** The subprime data set covers over 6,000 deals.[13] As mentioned earlier, for each mortgage, a deal ID is available. One can therefore reconstruct the actual pools for deals. We will perform some numerical tests to study the efficient Monte Carlo approximation's accuracy across the wide diversity of deals in the data universe.

Recall that we transformed the loan-level features y into the two-dimensional vector $w = (w_1, w_2) = (\theta_{\mathrm{d}}^Y \cdot y, \theta_{\mathrm{p}}^Y \cdot y)$. One can think of w_1 as the level of default risk. The distribution of a pool over the w_1 space

[13] "Deal" refers to the securitization of a pool of mortgages into an MBS. An MBS' structure may vary widely: it could be anything from a pass-through to a collateralized mortgage obligation (CMO). A CMO has a tranched structure, with different payment rules for the different tranches.

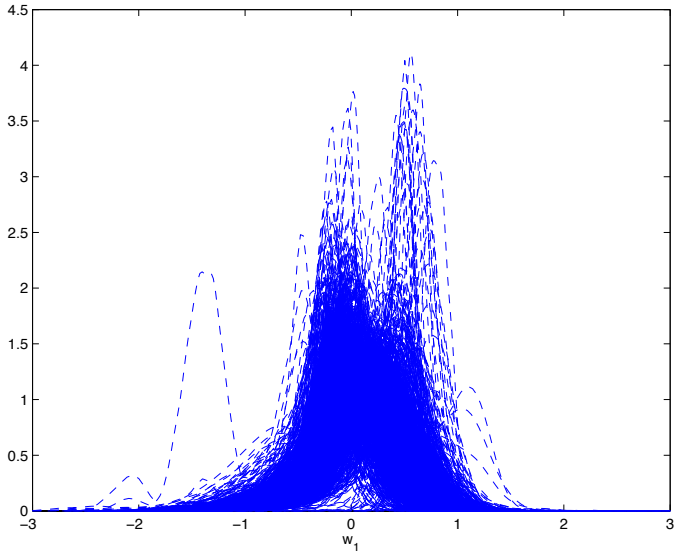

FIGURE 3. Comparison of risk for the different deals in the subprime data set.

indicates the distribution of risk for the pool. Figure 3 compares the distribution of risk for the deals in the subprime data set. One can see that some deals are very risky, while others are less risky.

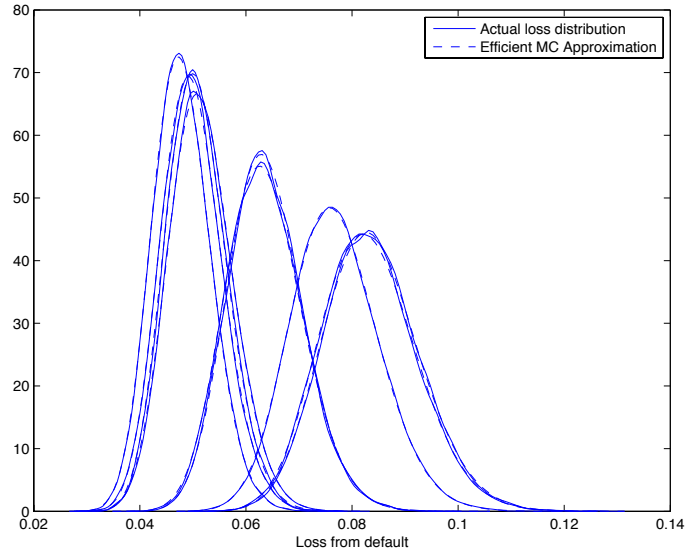

FIGURE 4. Comparison of actual distribution with approximate distribution for ten actual deals with $5,000 < N < 10,000$. Loss reported as fraction of pool which defaulted. The horizon is 12 months.

Figure 4 compares the approximate distribution from the efficient Monte Carlo approximation with the true distribution (found via brute-force Monte Carlo approximation) at a time horizon of 12 months for ten actual deals in the data set. The deals are chosen at random and each contains between $5,000$ and $10,000$

mortgages. It is interesting that the default rate can vary considerably between deals as a consequence of the quality of the underlying mortgages, demonstrating how important it is for a model to consider the loan-level characteristics of the mortgages in the pool.

To further assess the accuracy of the efficient Monte Carlo approximation, a set of deals is selected at random from the data set and the efficient Monte Carlo approximation's 99% value at risk (VaR) is compared with the true 99% value at risk. In total, we look at 185 deals and report the average error of just the LLN by itself as well as the average error for the full efficient Monte Carlo approximation (LLN and CLT combined). In addition, Figure 5 shows the distribution of the efficient Monte Carlo approximation's error across the set of deals. $50,000$ Monte Carlo simulations are performed and the time horizon is again twelve months. For deals with $5,000 - 10,000$ mortgages, the average error for the approximation (LLN and CLT) is .22%. The average error just using the LLN (no CLT) is 2.25%. For deals with more than $10,000$ mortgages, the average error for the approximation (LLN and CLT) is is .18%. The average error just using the LLN is 1.25%.

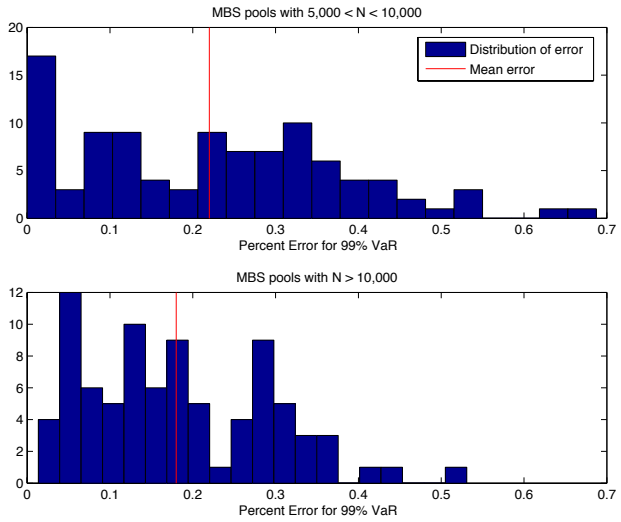

FIGURE 5. Distribution across deals of error for 99% VaR from the efficient Monte Carlo approximation. The time horizon is 12 months.

6.3. **One-dimensional Efficient Monte Carlo Approximation.** So far, we have focused on the case where $d_W = 2$. In this case, one models both default and prepayment. However, as mentioned previously, many types of loans only have default risk and do not have prepayment risk. In addition, for very high quality mortgage pools, default risk is small and it may be reasonable to consider only prepayment risk. Finally, for agency mortgage pools, the GSEs insure against any default losses, so prepayment and default can be treated as the same event. To demonstrate the one-dimensional approach, we fit a model only including prepayment to the agency mortgage data.[14]

For prepayment, we consider a number of loan-level features: FICO score, whether a first time homebuyer, the number of units, occupancy status (owner occupied, investment property, or second home), combined loan-to-value, loan-to-value, initial interest rate for the mortgage, debt-to-income ratio, whether there is a prepayment penalty in the mortgage contract, property type (condo, leasehold, PUD, manufactured housing, $1-4$ fee simple, Co-op), loan purpose (purchase, cash-out refinance, or no cash-out refinance), and number of borrowers. These features amount to 22 dimensions in the feature space. We also include the metropolitan

[14]The agency mortgage data set is particularly challenging to fit. It has over 500 million rows of data; therefore, one cannot read the data into memory all at once. We read chunks of the data one by one. In order to remove any bias from the ordering of the data, we randomly shuffle the data from the original files into a new set of files. Then, we use stochastic gradient descent to fit the model.

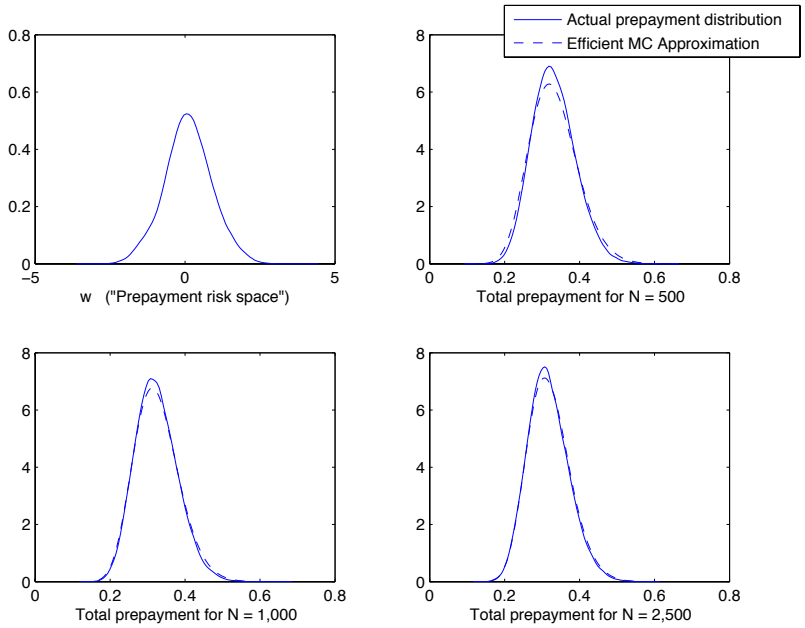

FIGURE 6. Top right plot and bottom plots compare the actual prepayment distribution with the efficient Monte Carlo approximation for $N = 500$, $N = 1,000$, and $N = 2,500$. The total prepayment is reported as the fraction of the pool which prepays. The time horizon is 12 months. The top left plot shows the distribution of the pool in the "prepayment risk space".

statistical area (MSA); there are 430 metropolitan statistical areas in the data set. Therefore, in total, $d_Y = 452$.[15] We emphasize that, even though $d_Y = 452$, d_W only equals 1. Figure 6 compares the efficient Monte Carlo approximation with the actual prepayment distribution for a randomly drawn pool of agency mortgages for a time horizon of 12 months.

6.4. **Precomputation for Large Financial Institutions.** In Section 4.3, we proposed to pre-simulate $\bar{\mu}^N$ on a pre-chosen grid and then use this one set of simulations on the single grid in order to find the distribution for many different mortgage-backed securities. Using the methodology proposed in Section 4.3, the efficient Monte Carlo approximation can provide great computational cost savings even for very small mortgage-backed securities, as long as a financial institution is dealing with many of these small mortgage-backed securities in aggregate.

We now implement this approach using the parameter fits from Section 6.3. 400 pools, each of size $N = 2,500$, are drawn at random from the agency mortgage data set. Each of these pools is simulated $50,000$ times using brute-force Monte Carlo simulation. Using the efficient Monte Carlo approximation, we also pre-simulate on a pre-chosen grid (as described in Section 6.3). $50,000$ Monte Carlo paths are also used for the efficient Monte Carlo simulation on this grid. The pre-chosen grid only has 20 grid points, placed at uniform intervals. The distribution of $\bar{\mu}^N$ is smooth in w, so one can numerically interpolate from the sparse grid points to get a finer solution. We use piecewise cubic spline interpolation.[16] This saves computational time since it allows one to simulate on a very sparse grid but still ultimately achieve a very accurate solution

[15]The parameter estimates are available upon request.

[16]An interesting question is whether one is guaranteed that the interpolated solution at any point y converges to the correct value as the number of Monte Carlo samples increases and the distance between grid points decreases. If h is continuously differentiable and X takes values in a compact set, such convergence holds (by the application of a Taylor expansion, boundedness of continuous function on a compact metric space, and the dominated convergence theorem). More general cases may require additional technical conditions.

on a fine grid. Figure 7 is a histogram of the percent error over the 400 pools for the 99% VaR from the efficient Monte Carlo approximation using pre-simulation. The time horizon is 12 months. The efficient Monte Carlo approximation is very accurate; the average percent error across the 400 pools for the 99% VaR from the approximation is only 0.25%. Brute-force simulation of the 400 pools takes $36,905.86$ seconds while the efficient Monte Carlo approximation of the 400 pools takes 4.72 seconds. The approximation provides cost savings of nearly 4 orders of magnitude versus brute-force simulation.

6.5. **Numerical Evaluation of Large Pool Approximation without Low-dimensional Transformation.** The previous numerical solutions of the law of large numbers and central limit theorem rely upon an *exact* low-dimensional transformation, which requires a restriction on the interaction between the factors Y^n and X. Although that low-dimensional transformation still covers a wide range of models, for completeness, we now present some methods for the computation of the law of large numbers and central limit theorem for the full class of models where a high-dimensional feature space $\mathcal{Y}$ may make a traditional Cartesian grid infeasible.

We let h_θ be a function of two neural networks, one for defaults and one for prepayments (see Example 2.2). Both neural networks have a single hidden layer of five neurons. The loan-level features are $\mathcal{Y} =$ (FICO score, LTV ratio, original balance, initial interest rate). The neural network for defaults also takes as an input the national unemployment rate while the neural network for prepayments takes both the national unemployment rate and the national mortgage rate as inputs. The model is trained on the subprime mortgage data set.

We implement two methods for the evaulation of the LLN and CLT for this case where the low-dimensional transformation is not applicable. In the first method, we cluster the pool into K clusters using k-means clustering on $Y^1, \ldots, Y^N$. The K centroids are chosen as the grid points. This is a sparse non-uniform grid in a high-dimensional space. The second method is an *approximate* low-dimensional transformation via an additional multi-layer neural network $\tilde{h}_\theta(u, u', y, V)$. The multi-layer neural network has three layers: the first (multi-neuron) layer takes as an input y, the second layer has d_W neurons (where $d_W \ll d_Y$), and the third layer takes as inputs the output of the d_W neurons from the second layer as well as X. This multi-layer neural network is trained to match the output of h_θ using the Levenberg-Marquardt algorithm. Note that $\tilde{h}_\theta(u, u', y, V)$ satisfies the exact low-dimensional transformation and $y \in \mathbb{R}^{d_Y}$ can be transformed into $w \in \mathbb{R}^{d_W}$. In the case we numerically implement, we use $d_W = 1$ and the w-space is discretized using k-means clustering. Figure compares the LLN and CLT with the actual distribution. $100,000$ Monte Carlo trials were performed and $K = 50$ clusters were used for the sparse grids. Figure 8 compares the Monte Carlo approximation using these sparse grid methods with the actual loss distribution.

7. Conclusion

This paper develops an efficient Monte Carlo approximation for a general class of loan-by-loan default and prepayment models for pools of loans. The approximation is based upon a law of large numbers and a central limit theorem for the pool default and prepayment processes. We extensively test our approach on actual mortgage data. The approximation is highly accurate even for relatively small pools (as little as 500 loans). In practice, pools commonly range from a few thousand to hundreds of thousands of loans. Brute-force simulation for large pools can be computationally expensive; our approximation can save several orders of magnitude in computational time. The efficient Monte Carlo approximation accounts for the full loan-level dynamics, taking advantage of the detailed loan-level information typically available (such as credit score, loan-to-value ratio, initial interest rate, and type of loan). A key feature of our approximation is that its computational cost is constant no matter the dimension of the loan-level data. This feature is desirable since loan-level data can be high-dimensional.

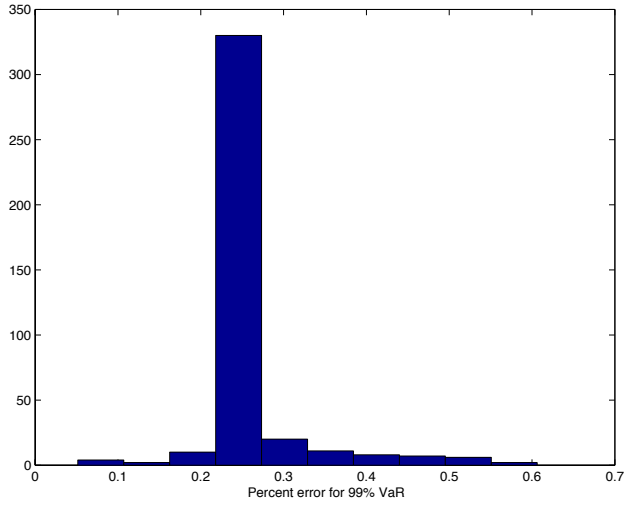

FIGURE 7. Distribution across deals of error for 99% VaR from the efficient Monte Carlo approximation using pre-simulation. The time horizon is 12 months.

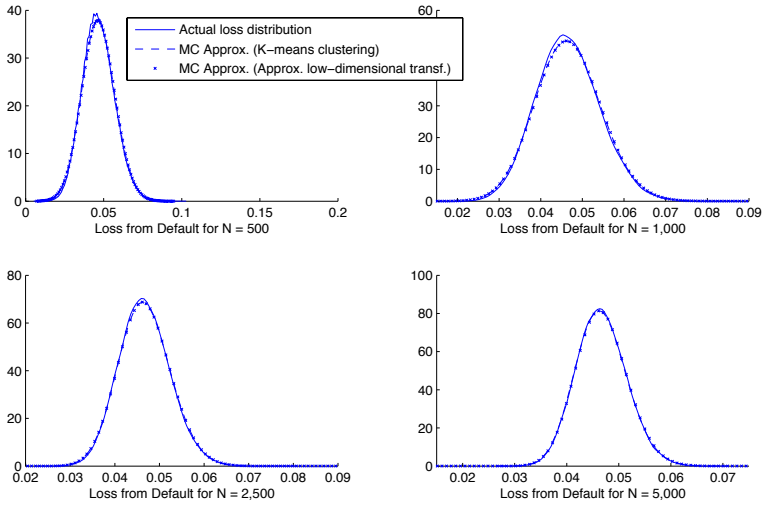

FIGURE 8. Comparison of actual loss distribution with Monte Carlo approximation using sparse grids when h_θ is a neural network.

APPENDIX A. PROOFS

A.1. **Law of Large Numbers.** Let $\mathcal{P}(E)$ be the space of measures on a complete separable space E. The topology for $\mathcal{P}(E)$ is the topology of weak convergence, which is metrized by the Prokhorov metric. A random variable $X^N \in \mathcal{P}(E)$ itself has a probability measure $P^N \in \mathcal{P}(\mathcal{P}(E))$. The measure P^N for the random variable X^N weakly converges to the measure P of a random variable $\bar{X}$ if:

$$\mathbb{E}[\Phi(X^N)] \to \mathbb{E}[\Phi(\bar{X})], \tag{21}$$

for every $\Phi \in \mathcal{S}$ where $\mathcal{S}$ is the class of functions which separates $\mathcal{P}(E)$. In our case, $E = \mathcal{U} \times \mathcal{Y}$. We let $\mathcal{S} : \mathcal{P}(\mathcal{U} \times \mathcal{Y}) \longrightarrow \mathbb{R}$ be the collection of functions of the form $\Phi(\mu) = \phi_1(\langle f_1, \mu \rangle_E, \ldots, \langle f_M, \mu \rangle_E)$ for

$\phi_1 \in C(\mathbb{R}^M), f_m \in C(E)$. Then, $\mathcal{S}$ separates $\mathcal{P}(\mathcal{U}\times\mathcal{Y})$ and is a convergence determining class for $\mathcal{P}(\mathcal{P}(E))$ (see [45] or [21]).

Recall that we have defined $\langle f,\nu\rangle_E = \int_E f(x)\nu(dx)$. For instance, if $E=\mathcal{U}\times\mathcal{Y}$, $\langle f,\nu\rangle_{\mathcal{U}\times\mathcal{Y}} = \sum_{u\in\mathcal{U}}\int_{\mathcal{Y}} f(u,y)\nu(u,dy)$. Also, for notational convenience, define $\langle f,\nu(u,dy)\rangle \equiv \langle f(u,\cdot),\nu(u,\cdot)\rangle_{\mathcal{Y}} = \int_{\mathcal{Y}} f(u,y)\nu(u,dy)$.

We prove several facts, culminating in the law of large numbers $\mu^N \xrightarrow{d} \bar{\mu}$. In Lemma A.1, we prove that the solution to the law of large numbers equation (3.2) is unique. The proof for the actual convergence $\mu^N \xrightarrow{d} \bar{\mu}$ follows, which makes use of an induction argument and the convergence determining class $\mathcal{S}$.

Lemma A.1. *The solution $\bar{\mu}$ to the law of large numbers equation (3.2) is the unique solution for equation (27) in B^{T+1}, where $B=\mathcal{P}(\mathcal{U}\times\mathbb{R}^{d_Y})$.*

Proof. Note that equation (3.2) satisfies equation (27) for every ϕ. Suppose that $\bar{\mu}_{t'<t}$ is unique; $\bar{\mu}_t$ must then be unique as well. Suppose there are different solutions $\bar{\mu}_t^1$ and $\bar{\mu}_t^2$, and let $\nu=\bar{\mu}_t^1-\bar{\mu}_t^2$. From equation (27), this implies that $\langle\phi,\nu\rangle_B = 0$ for every bounded, continuous ϕ. Since $\phi\in C_b(\mathcal{U}\times\mathbb{R}^{d_Y})$ is separating (see [8] or [45]) for $\mathcal{P}(\mathcal{U}\times\mathbb{R}^{d_Y})$, $\bar{\mu}_t^1=\bar{\mu}_t^2$. This is a contradiction and therefore $\bar{\mu}_t$ is unique. Since $\bar{\mu}_0$ is unique (by Assumption 3.1), $\bar{\mu}$ is unique by induction. □

Lemma A.2. *Define $\mathcal{M}_t^{1,N}$ to be the martingale*

$$\mathcal{M}_t^{1,N}(u) = \frac{1}{N}\sum_{n=1}^N \phi(Y^n)(\mathbf{1}_{U_t^n=u} - h(u,U_{t-1}^n,Y^n,V_{t-1},H_{t-\tau:t-1}^N)). \tag{22}$$

Then, for each V, $\mathcal{M}_t^{1,N}\xrightarrow{p}0$.

Proof. The variance of $\mathcal{M}_t^{1,N}$ converges to zero:

$$\begin{aligned}
\mathrm{Var}[\mathcal{M}_t^{1,N}|V] &= \mathrm{Var}[\mathcal{M}_t^{1,N}|V_{0:t-1}] = \mathrm{Var}[\frac{1}{N}\sum_{n=1}^N \phi(Y^n)(\mathbf{1}_{U_t^n=u} - h(u,U_{t-1}^n,Y^n,V_{t-1},H_{t-\tau:t-1}^N))|V_{0:t-1}] \\
&= \mathbb{E}[(\frac{1}{N}\sum_{n=1}^N \phi(Y^n)(\mathbf{1}_{U_t^n=u} - h(u,U_{t-1}^n,Y^n,V_{t-1},H_{t-\tau:t-1}^N))])^2|V_{0:t-1}] \\
&= \frac{1}{N^2}\sum_{n=1}^N \mathbb{E}[(\phi(Y^n)(\mathbf{1}_{U_t^n=u} - h(u,U_{t-1}^n,Y^n,V_{t-1},H_{t-\tau:t-1}^N)))^2|V_{0:t-1}] \le \frac{1}{N^2}\sum_{n=1}^N C \underset{N\to\infty}{\to} 0. \qquad (23)
\end{aligned}$$

The second and third equalities use the tower property and the independence of the processes U_t^n conditional on $\mathcal{F}_{t-1}$. The fourth equality uses the fact that $\phi(Y^n)(\mathbf{1}_{U_t^n=u} - h(u,U_{t-1}^n,Y^n,V_{t-1},H_{t-\tau:t-1}^N))$ is bounded. This follows from the facts that ϕ is a continuous function, $Y^n\in\mathcal{Y}$ where $\mathcal{Y}$ is compact, and h is a probability transition function (and so is bounded). Recall that a continuous function on a compact space is bounded. By Chebyshev's inequality, $\mathcal{M}_t^{1,N}\xrightarrow{p}0$. □

Lemma A.3. *Define $\mathcal{Z}_t^N$ to be the martingale*

$$\mathcal{Z}_t^N = \frac{1}{N}\sum_{n=1}^N \ell_t^n(Y^n,V_t)(\mathbf{1}_{U_t^n=d} - \mathbf{1}_{U_{t-1}^n=d}) - \int_{\mathcal{Y}}\int_0^1 z\nu_{t,y,V_t}(dz)(\mu_t^N(d,dy)-\mu_t^N(d,dy)), \tag{24}$$

where d is the default state. Then, for each V, $\mathcal{Z}_t^N\xrightarrow{p}0$.

Proof. First, rewrite $\mathcal{Z}_t^N$ as:

$$\mathcal{Z}_t^N = \frac{1}{N}\sum_{n=1}^N \ell_t^n(Y^n,V_t)(\mathbf{1}_{U_t^n=d} - \mathbf{1}_{U_{t-1}^n=d}) - \frac{1}{N}\sum_{n=1}^N(\mathbf{1}_{U_t^n=d} - \mathbf{1}_{U_{t-1}^n=d})\int_0^1 z\nu_{t,Y^n,V_t}(dz). \tag{25}$$

Similar to Lemma A.2, we will show that the variance of $\mathcal{Z}_t^N$ tends to zero as $N\to\infty$.

$$\mathrm{Var}[\mathcal{Z}_t^N|V] = \mathrm{Var}[\mathcal{Z}_t^N|V_{0:t}] = \mathbb{E}[(\frac{1}{N}\sum_{n=1}^N \ell_t^n(Y^n,V_t)(\mathbf{1}_{U_t^n=d} - \mathbf{1}_{U_{t-1}^n=d}) - \frac{1}{N}\sum_{n=1}^N(\mathbf{1}_{U_t^n=d} - \mathbf{1}_{U_{t-1}^n=d})\int_0^1 z\nu_{t,Y^n,V_t}(dz))^2|V_{0:t}]$$

$$
\begin{aligned}
&= \mathbb{E}[\mathbb{E}[(\frac{1}{N}\sum_{n=1}^{N}(\mathbf{1}_{U_t^n=d} - \mathbf{1}_{U_{t-1}^n=d})(\ell_t^n(Y^n, V_t) \\
&- \int_0^1 z\nu_{t,Y^n,V_t}(dz)))^2 | U_t^1, \ldots, U_t^N, U_{t-1}^1, \ldots, U_{t-1}^N, Y^1, \ldots, Y^N, V_t] | V_{0:t}] \\
&\leq \mathbb{E}[\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[(\ell_t^n(Y^n, V_t) - \int_0^1 z\nu_{t,Y^n,V_t}(dz))^2 | U_t^1, \ldots, U_t^N, Y^1, \ldots, Y^N, V_t] | V_{0:t}] \leq \frac{C}{N}.
\end{aligned}
$$

We have again used the tower property, the conditional independence of the ℓ^n given $(U_t^1, \ldots, U_t^N, Y^1, \ldots, Y^N, V_t)$, and the fact that the loss is bounded (i.e., $\ell^n \in [0,1]$). By Chebyshev's inequality, $\mathcal{Z}_t^N \xrightarrow{p} 0$. □

Proof of Theorem 3.2. For each $u \in \mathcal{U}$ and any $\phi \in C_b(\mathbb{R}^{d_Y})$, we have that

$$
\langle \phi, \mu_t^N(u, dy) \rangle = \sum_{u' \in \mathcal{U}} \langle \phi(y) h(u, u', y, V_{t-1}, H_{t-\tau:t-1}^N), \mu_{t-1}^N(u', dy) \rangle + \mathcal{M}_t^{1,N}(u). \tag{26}
$$

We will now use an induction argument to prove the law of large numbers. Assuming $\mu_{0:t-1}^N \xrightarrow{p} \bar{\mu}_{0:t-1}$ for each V, we have that for each V:

$$
\lim_{N\to\infty} \langle \phi, \mu_t^N(u, dy) \rangle = \sum_{u' \in \mathcal{U}} \langle \phi(y) h(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1}), \bar{\mu}_{t-1}(u', dy) \rangle = \langle \phi, \bar{\mu}_t(u, dy) \rangle . \tag{27}
$$

The result (27) is a consequence of the following facts. First, $\mathcal{M}^{1,N} \xrightarrow{p} 0$ Lemma A.2). Secondly, since $\mu_{0:t-1}^N \xrightarrow{p} \bar{\mu}_{0:t-1}$ for each V, $H_{t-\tau:t-1}^N \xrightarrow{p} \bar{H}_{t-\tau:t-1}$ for each V (f^H is continuous by assumption). Thirdly, for each V,

$$
\langle \phi(y) h(u, u', y, V_{t-1}, H_{t-\tau:t-1}^N), \mu_{t-1}^N(u', dy) \rangle \xrightarrow{p} \langle \phi(y) h(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1}), \bar{\mu}_{t-1}(u', dy) \rangle . \tag{28}
$$

This last fact follows from:

$$
\begin{aligned}
&\langle \phi(y) h(u, u', y, V_{t-t}, H_{t-\tau:t-1}^N), \mu_{t-1}^N(u', dy) \rangle = \langle \phi(y) h(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1}), \mu_{t-1}^N(u', dy) \rangle \\
+ \; &\underbrace{\langle \phi(y)[h(u, u', y, V_{t-1}, H_{t-\tau:t-1}^N) - h(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1})], \mu_{t-1}^N(u', dy) \rangle}_{(2)}
\end{aligned} \tag{29}
$$

The function h is a continuous function, by assumption. Since $\mathcal{Y}$ is compact and $H_{t-\tau:t-1}^N$ is bounded (due to the assumption that f^H is continuous and $\mathcal{U} \times \mathcal{Y}$ is compact), h is in fact uniformly continuous on the space that its arguments live on (a continuous function a compact space is uniformly continuous). Therefore:

$$
|(2)| \leq \sup_{y \in \mathcal{Y}} |\phi(y)| \sup_{y \in \mathcal{Y}, u \in \mathcal{U}, u' \in \mathcal{U}} |h(u, u', y, V_{t-1}, H_{t-\tau:t-1}^N) - h(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1})| \xrightarrow{p} 0,
$$

since $H_{t-\tau:t-1}^N \xrightarrow{p} \bar{H}_{t-\tau:t-1}$ and the desired result (28) holds.

By Lemma A.1, $\bar{\mu}_t$ from (3.2) is the unique measure which satisfies (27). Let $\Phi \in \mathcal{S}$ where $\mathcal{S}$ is the convergence determining class specified earlier. Then, for any $\Phi \in \mathcal{S}$,

$$
\begin{aligned}
\lim_{N\to\infty} \mathbb{E}[\Phi(\mu_t^N) | V] &= \lim_{N\to\infty} \mathbb{E}[\phi_1(\langle f_1, \mu_t^N \rangle_B, \ldots, \langle f_M, \mu_t^N \rangle_B) | V] \\
&= \lim_{N\to\infty} \mathbb{E}[\phi_1(\sum_{u \in \mathcal{U}} \langle f_1(u, y), \mu_t^N(u, dy) \rangle, \ldots, \sum_{u \in \mathcal{U}} \langle f_M(u, y), \mu_t^N(u, dy) \rangle) | V] \\
&= \mathbb{E}[\Phi(\bar{\mu}_t) | V].
\end{aligned}
$$

The above result follows from equation (27), the continuous mapping theorem, $\mathcal{P}(\mathcal{U} \times \mathcal{Y})$ being compact (and thus any continuous function of $\mu_t^N \in \mathcal{P}(\mathcal{U} \times \mathcal{Y})$ is bounded), and the dominated convergence theorem. Therefore, for each V, $\mu_t^N \xrightarrow{d} \bar{\mu}_t$ and, since $\bar{\mu}_t$ is deterministic given V, μ_t^N also converges in probability to $\bar{\mu}_t$ for each V. This also means that $\mu_{0:t}^N \xrightarrow{p} \bar{\mu}_{0:t}$ for each V.

Note that in Assumption 3.1, μ_0^N converges in distribution to $\bar{\mu}_0$, where $\bar{\mu}_0$ is deterministic. By Assumption 3.1 and induction, it follows that μ^N converges in probability to $\bar{\mu}$ for each V. Since the convergence in

probability holds for every V, we certainly have that μ^N converges in distribution to $\bar{\mu}$, since:

$$\begin{aligned}\lim_{N\to\infty}\mathbb{E}[f(\mu^N)] &= \lim_{N\to\infty}\mathbb{E}[\mathbb{E}[f(\mu^N)|V]] \\ &= \mathbb{E}[\lim_{N\to\infty}\mathbb{E}[f(\mu^N)|V]] = \mathbb{E}[\mathbb{E}[f(\bar{\mu})|V]] \\ &= \mathbb{E}[f(\bar{\mu})], \end{aligned} \tag{30}$$

for any continuous bounded $f : \mathcal{P}(\mathcal{U}\times\mathbb{R}^{d_Y}) \to \mathbb{R}$. We have used above the tower property, dominated convergence, and previous convergence results for μ^N conditional upon a path V.

Finally, we now show that $L^N \xrightarrow{d} \bar{L}$. First, recognize that:

$$L_t^N = \int_{\mathcal{Y}}\int_0^1 z\nu_{t,y,V_t}(dz)(\mu_t^N(d,dy) - \mu_{t-1}^N(d,dy)) + \mathcal{Z}_t^N, \tag{31}$$

where $d = \{\text{default}\}$. For each V, $\mathcal{Z}_t^N \xrightarrow{p} 0$ by Lemma A.3. In addition, since $g(t,v,y) = \int_0^1 z\nu_{t,y,v}(dz)$ is continuous and bounded for each y, v:

$$\int_{\mathcal{Y}}\int_0^1 z\nu_{t,y,V_t}(dz)(\mu_t^N(d,dy) - \mu_{t-1}^N(d,dy)) \xrightarrow{p} \int_{\mathcal{Y}}\int_0^1 z\nu_{t,y,V_t}(dz)(\bar{\mu}_t(d,dy) - \bar{\mu}_{t-1}(d,dy)), \tag{32}$$

for each V. This proves $L_t^N \xrightarrow{p} \bar{L}_t$ for each t, V, which (like before) implies that $L^N \xrightarrow{p} \bar{L}$ for each V. By similar reasoning as in (30), this implies that $(\mu^N, L^N) \xrightarrow{d} (\bar{\mu}, \bar{L})$.

A.2. **Central Limit Theorem.** Let E' be a Banach space (equipped with the weak topology) with a topological dual E. A random variable $X^N \in E'$ itself has a probability measure $P^N \in \mathcal{P}(E')$. The measure P^N for the random variable X^N weakly converges to the measure P of a random variable $\bar{X}$ if:

$$\mathbb{E}[\exp\left(i\alpha\phi(X^N)\right)] \to \mathbb{E}[\exp\left(i\alpha\phi(\bar{X})\right)], \tag{33}$$

for every $\phi \in E$.[17] Of course, (33) is implied by:

$$\phi(X^N) \xrightarrow{d} \phi(\bar{X}), \tag{34}$$

for every $\phi \in E$.

For instance, for each u, $\Xi_t^N(u,dy) \in S'(\mathbb{R}^{d_Y})$. S' is the space of tempered distributions and its dual is S, the space Schwartz functions on $\mathbb{R}^{d_Y}$. To prove convergence in distribution of $\Xi_t^N(u,dy)$ for each u, one needs to show (33) or (35) for each $\phi \in S$:

$$\left\langle \phi(y), X^N(u,dy)\right\rangle \xrightarrow{d} \left\langle \phi(y), \bar{X}(u,dy)\right\rangle, \tag{35}$$

where, as previously defined, $\langle f, \nu(u,dy)\rangle = \int_{\mathcal{Y}} f(y)\nu(u,dy)$. Note that since the $Y^n \in \mathcal{Y} \subset \mathbb{R}^{d_Y}$, $\left\langle f, \Xi_t^N(u,dy)\right\rangle = \int_{\mathcal{Y}} f(y)\Xi_t^N(u,dy) = \int_{\mathbb{R}^{d_Y}} f(y)\Xi_t^N(u,dy)$ (and similarly for $\mu_t^N, \bar{\mu}$, and $\bar{\Xi}$).

Since the dual of the cartesian product of finitely many spaces is the cartesian product of the duals, the criterion (33) or (35) also of course covers the case of $\Xi^N \in E = \prod_{i=1}^{T\times|\mathcal{U}|} S'$, where it is sufficient to show (33) for $\phi \in \prod_{i=1}^{T\times|\mathcal{U}|} S$, or more simply convergence in distribution for the random vector $\left(\left\langle \phi_{t,u}, \Xi_t^N(u,dy)\right\rangle\right)_{t\in I, u\in\mathcal{U}}$ for every $\left(\phi_{t,u}\right)_{t\in I,u\in\mathcal{U}} \in \prod_{i=1}^{T\times|\mathcal{U}|} S$.

Lemma A.4. *For every u and N, $\Xi_t^N(u,dy) \in S'(\mathbb{R}^{d_Y})$.*

[17] This result appears to have been developed independently by several authors. See [47] (Proposition 3.3 and corresponding Supplementary Comments on pg. 247), [29], and [38]. Similar results for nuclear Frechet spaces, albeit in French, can be found in [19]. A similar result for the space of tempered distributions (and which describes the results in [19]) can be found in [26]. Finally, such results have been extended to continuous-time processes taking values in nuclear Frechet spaces, see [34] and [20].

Proof. $X \in S'$ if for any $\phi \in S$ there exists a k, C_k such that

$$\langle \phi, X \rangle \leq C_k ||\phi||_k, \tag{36}$$

where the norm $||\cdot||_k$ is defined as:

$$||\phi||_k = \max_{|\alpha|+|\beta| \leq k} \sup_{y \in \mathbb{R}^{d_Y}} |y^\alpha D^\beta \phi(y)|. \tag{37}$$

Then, we certainly have that:

$$\left\langle \phi, \Xi_t^N \right\rangle = \sqrt{N}(\left\langle \phi, \mu_t^N \right\rangle - \langle \phi, \bar{\mu} \rangle) \leq 2\sqrt{N} \sup_{y \in \mathcal{Y}} |\phi(y)| \leq 2\sqrt{N}||\phi||_0. \tag{38}$$

□

Lemma A.5. *There is a unique solution to equation (11) in Theorem 3.4.*

Proof. Suppose there are two solutions $\bar{\Xi}_t^1$ and $\bar{\Xi}_t^2$; let their difference be $\nu = \bar{\Xi}_t^1 - \bar{\Xi}_t^2$. Substituting into equation (11) yields that $\langle \phi, \nu \rangle_{\mathcal{U} \times \mathbb{R}^{d_Y}} = 0$ for every ϕ, which implies uniqueness. □

Lemma A.6. *The martingale difference array* $Z_t^{N,n}(\alpha) = \frac{1}{\sqrt{N}} \sum_{u \in \mathcal{U}} \alpha_u [\phi(Y^n)(\mathbf{1}_{U_t^n = u} - h(u, U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N)]$ *satisfies:*[18]

(i) $\sup_N \mathbb{E}[(\max_{n \leq N} Z_t^{N,n})^2]$ *is uniformly bounded,*

(ii) $\max_{n \leq N} |Z_t^{N,n}| \xrightarrow{p} 0$,

(iii) For each V, $\sum_{n=1}^N Z_{N,n}^2 \xrightarrow{p} Var[\sum_{u \in \mathcal{U}} \alpha_u \left\langle \phi, \bar{\mathcal{M}}_t(u, dy) \right\rangle | V]$,

for every $\alpha \in \mathbb{R}^{|\mathcal{U}|}$.

Proof. First, recognize that:

$$|Z_t^{N,n}(\alpha)| \leq \frac{2}{\sqrt{N}} \sum_{u \in \mathcal{U}} |\alpha_u| \phi_u(Y^n) \leq \frac{2}{\sqrt{N}} \sum_{u \in \mathcal{U}} |\alpha_u| \sup_{y \in \mathcal{Y}} \phi_u(y) \leq \frac{C}{\sqrt{N}}.$$

This certainly implies properties 1 and 2. Finally, for each V,

$$\begin{aligned}
\sum_{n=1}^N (Z_t^{N,n})^2 &= \frac{1}{N} \sum_{n=1}^N [\sum_{u \in \mathcal{U}} \alpha_u \phi_u(Y^n)(\mathbf{1}_{U_t^n = u} - h(u, U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N)]^2 \\
&= \frac{1}{N} \sum_{n=1}^N \sum_{u', u \in \mathcal{U}} \alpha_u \alpha_{u'} \phi_{u'}(Y^n) \phi_u(Y^n) [\mathbf{1}_{U_t^n = u} \mathbf{1}_{U_t^n = u'} \\
&+ h(u, U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N) h(u', U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N) \\
&- \mathbf{1}_{U_t^n = u'} h(u, U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N) - \mathbf{1}_{U_t^n = u} h(u', U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N)] \\
&= \frac{1}{N} \sum_{n=1}^N \sum_{u', u \in \mathcal{U}} \alpha_u \alpha_{u'} \phi_{u'}(Y^n) \phi_u(Y^n) [\mathbf{1}_{u = u'} h(u, U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N) \\
&- h(u, U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N) h(u', U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N)] + \mathcal{M}_t^{N,3} \\
&= \sum_{u', u \in \mathcal{U}} \alpha_u \alpha_{u'} \sum_{u'' \in \mathcal{U}} [\mathbf{1}_{u = u'} \left\langle \phi_{u'}(y) \phi_u(y) h(u, u'', y, V_{t-1}, H_{t-\tau:t-1}^N), \mu_{t-1}^N \right\rangle \\
&- \left\langle \phi_{u'}(y) \phi_u(y) h(u, u'', y, V_{t-1}, H_{t-\tau:t-1}^N) h(u', u'', y, V_{t-1}, H_{t-\tau:t-1}^N), \mu_{t-1}^N \right\rangle] + \mathcal{M}_t^{N,3} \\
&\xrightarrow{p} \sum_{u', u \in \mathcal{U}} \alpha_u \alpha_{u'} \sum_{u'' \in \mathcal{U}} [\mathbf{1}_{u = u'} \left\langle \phi_{u'}(y) \phi_u(y) h(u, u'', y, V_{t-1}, \bar{H}_{t-\tau:t-1}), \bar{\mu}_{t-1} \right\rangle \\
&- \left\langle \phi_{u'}(y) \phi_u(y) h(u, u'', y, V_{t-1}, \bar{H}_{t-\tau:t-1}) h(u', u'', y, V_{t-1}, \bar{H}_{t-\tau:t-1}), \bar{\mu}_{t-1} \right\rangle]. \qquad (39)
\end{aligned}$$

[18] See [33] for details on martingale difference arrays.

$\mathcal{M}_t^{N,3}$ is the remainder term:

$$\begin{aligned}\mathcal{M}_t^{N,3} &= \frac{1}{N}\sum_{n=1}^{N}\sum_{u',u\in\mathcal{U}}\alpha_u\alpha_{u'}\phi_{u'}(Y^n)\phi_u(Y^n)[\mathbf{1}_{U_t^n=u}\mathbf{1}_{U_t^n=u'} - \mathbf{1}_{u=u'}h(u,U_{t-1}^n,Y^n,V_{t-1},H_{t-\tau:t-1}^N)\\ &- \mathbf{1}_{U_t^n=u'}h(u,U_{t-1}^n,Y^n,V_{t-1},H_{t-\tau:t-1}^N) - \mathbf{1}_{U_t^n=u}h(u',U_{t-1}^n,Y^n,V_{t-1},H_{t-\tau:t-1}^N)\\ &+ 2h(u,U_{t-1}^n,Y^n,V_{t-1},H_{t-\tau:t-1}^N)h(u',U_{t-1}^n,Y^n,V_{t-1},H_{t-\tau:t-1}^N)].\end{aligned}$$

This is of the same form as $\mathcal{M}_t^{N,1}$, and the exact same procedure as used in Lemma A.2 can be applied here to show that $\mathcal{M}_t^{N,3} \xrightarrow{p} 0$

Using the exact same argument as was employed in the proof of Theorem 3.2, the final line in equation (39) follows from $\mu^N \xrightarrow{p} \bar{\mu}$ for each V and the uniform continuity of h on the compact set $\mathcal{Y}$ (a continuous function on a compact set is uniformly continuous). Finally, note that the last line is exactly the covariance of $\bar{\mathcal{M}}_t$ (conditional on V), whose distribution is given in Lemma A.7.

□

Lemma A.7. *Let $\mathcal{M}_t^{2,N}$ be:*

$$\mathcal{M}_t^{2,N}(u,dy) = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}[\delta_{(U_t^n,Y^n)}(u,dy) - h(u,U_{t-1}^n,Y^n,V_{t-1},H_{t-\tau:t-1}^N)\delta_{Y^n}(dy)]. \tag{40}$$

Then, $\mathcal{M}_t^{2,N} \xrightarrow{d} \bar{\mathcal{M}}_t$ where, conditional on V, $\bar{\mathcal{M}}_t$ is a mean-zero Gaussian with covariance:

$$\begin{aligned}&Cov\Big[\left\langle\phi_1,\bar{\mathcal{M}}_t(u_1,dy)\right\rangle,\left\langle\phi_2,\bar{\mathcal{M}}_t(u_2,dy)\right\rangle\Big|V] = Cov\Big[\left\langle\phi_1,\bar{\mathcal{M}}_t(u_1,dy)\right\rangle,\left\langle\phi_2,\bar{\mathcal{M}}_t(u_2,dy)\right\rangle\Big|V_{0:t-1}]\\ &= -\sum_{u'\in\mathcal{U}}\left\langle\phi_1(y)\phi_2(y)h(u_1,u',y,V_{t-1},\bar{H}_{t-\tau:t-1})h(u_2,u',y,V_{t-1},\bar{H}_{t-\tau:t-1}),\bar{\mu}_{t-1}\right\rangle,\\ &Var\Big[\left\langle\phi,\bar{\mathcal{M}}_t(u,dy)\right\rangle\Big|V] = Var\Big[\left\langle\phi,\bar{\mathcal{M}}_t(u,dy)\right\rangle\Big|V_{0:t-1}]\\ &= \sum_{u'\in\mathcal{U}}\left\langle\phi(y)^2h(u,u',y,V_{t-1},\bar{H}_{t-1})(1-h(u,u',y,V_{t-1},\bar{H}_{t-\tau:t-1}),\bar{\mu}_{t-1}\right\rangle,\end{aligned}$$

where $u_1 \neq u_2$. Furthermore, conditional on each V, $\left\langle\phi,\bar{\mathcal{M}}_{t_1}\right\rangle$ is independent of $\left\langle\phi,\bar{\mathcal{M}}_{t_2}\right\rangle$ for $t_1 \neq t_2$.

Proof. We first show that, conditional on V:

$$(\Big\langle\phi_1,\mathcal{M}_t^{2,N}(u_1,dy)\Big\rangle,\ldots,\Big\langle\phi_{|\mathcal{U}|},\mathcal{M}_t^{2,N}(u_{|\mathcal{U}|},dy)\Big\rangle) \xrightarrow{d} (\left\langle\phi_1,\bar{\mathcal{M}}_t(u_1,dy)\right\rangle,\ldots,\left\langle\phi_{|\mathcal{U}|},\bar{\mathcal{M}}_t(u_{|\mathcal{U}|},dy)\right\rangle),$$

for any $\phi_1,\ldots,\phi_{|\mathcal{U}|} \in S$. Note that:

$$\sum_{u\in\mathcal{U}}\alpha_u\Big\langle\phi_u,\mathcal{M}_t^{2,N}(u,dy)\Big\rangle = \frac{1}{\sqrt{N}}\sum_{u\in\mathcal{U}}\alpha_u[\phi_u(Y^n)(\mathbf{1}_{U_t^n=u} - h(u,U_{t-1}^n,Y^n,V_{t-1},H_{t-\tau:t-1}^N)] = Z_t^{N,n}(\alpha).$$

By Lemma A.6, the martingale central limit theorem of [33], and the Cramer-Wold theorem:

$$(\Big\langle\phi_1,\mathcal{M}_t^{2,N}(u_1,dy)\Big\rangle,\ldots,\Big\langle\phi_{|\mathcal{U}|},\mathcal{M}_t^{2,N}(u_{|\mathcal{U}|},dy)\Big\rangle) \xrightarrow{d} (\left\langle\phi_1,\bar{\mathcal{M}}_t(u_1,dy)\right\rangle,\ldots,\left\langle\phi_{|\mathcal{U}|},\bar{\mathcal{M}}_t(u_{|\mathcal{U}|},dy)\right\rangle),$$

for any $\phi_1,\ldots,\phi_{|\mathcal{U}|} \in S$.

Furthermore, conditional on each V, $\left\langle\phi_1,\bar{\mathcal{M}}_{t_1}\right\rangle$ is independent of $\left\langle\phi_2,\bar{\mathcal{M}}_{t_2}\right\rangle$ for $t_1 \neq t_2$. To see this, assume $t_1 < t_2$, and first apply the tower property to write:

$$\begin{aligned}&\mathbb{E}[\exp\big(i\alpha_1\Big\langle\phi,\mathcal{M}_{t_1}^{2,N}(u,dy)\Big\rangle + i\alpha_2\Big\langle\phi,\mathcal{M}_{t_2}^{2,N}(u',dy)\Big\rangle\big)|V]\\ &= \mathbb{E}\big[\exp\big(i\alpha_1\Big\langle\phi_1,\mathcal{M}_{t_1}^{2,N}(u,dy)\Big\rangle\big)\mathbb{E}[\exp\big(i\alpha_2\Big\langle\phi_2,\mathcal{M}_{t_2}^{2,N}(u',dy)\Big\rangle\big)|\mu_{0:t_2-1}^N,V]|V\big].\end{aligned} \tag{41}$$

Next, we can express the inner expectation as a product of exponentials:

$$\mathbb{E}[\exp\big(i\alpha_2\Big\langle\phi_2,\mathcal{M}_{t_2}^{2,N}(u',dy)\Big\rangle\big)|\mu_{0:t_2-1}^N,V]$$

$$
\begin{aligned}
&= \prod_{n=1}^{N} \mathbb{E}[e^{\frac{i\alpha_2}{\sqrt{N}}\phi_2(Y^n)(\mathbf{1}_{U^n_{t_2}=u'} - h(u', U^n_{t_2-1}, Y^n, V_{t_2-1}, H^N_{t_2-\tau:t_2-1}))} | V, H^N_{t_2-\tau:t_2-1}, U^n_{t_2-1}, Y^n] \\
&= \prod_{n=1}^{N} (1 - \frac{\alpha_2^2}{2N}\mathbb{E}[\phi_2(Y^n)^2(\mathbf{1}_{U^n_{t_2}=u'} - h(u', U^n_{t_2-1}, Y^n, V_{t_2-1}, H^N_{t_2-\tau:t_2-1}))^2 | V, H^N_{t_2-\tau:t_2-1}, U^n_{t_2-1}, Y^n] + \frac{C}{N^{3/2}}), \\
&= \prod_{n=1}^{N} (1 - \frac{\alpha_2^2}{2N}\gamma(V, H^N_{t_2-\tau:t_2-1}, U^n_{t_2-1}, Y^n) + \frac{C}{N^{3/2}},)
\end{aligned}
$$

where we have used a Taylor expansion for the exponential, $\sup_{y\in\mathcal{Y}} 2\phi_2(y)^3 < C < \infty$, and $\gamma(V, H^N_{t_2-\tau:t_2-1}, U^n_{t_2-1}, Y^n) \equiv \mathbb{E}[\phi_2(Y^n)^2(\mathbf{1}_{U^n_{t_2}=u'} - h(u', U^n_{t_2-1}, Y^n, V_{t_2-1}, H^N_{t_2-\tau:t_2-1}))^2 | V, H^N_{t_2-\tau:t_2-1}, U^n_{t_2-1}, Y^n]$. Using the standard inequality $|\prod_{n=1}^{N} w_n - \prod_{n=1}^{N} z_n| \leq \sum_{n=1}^{N} |w_n - z_n|$ for $|w_n|, |z_n| \leq 1$, we have that:

$$
\begin{aligned}
&| \quad \prod_{n=1}^{N} (1 - \frac{\alpha_2^2}{2N}\gamma(V, H^N_{t_2-\tau:t_2-1}, U^n_{t_2-1}, Y^n) + \frac{C}{N^{3/2}}) - \prod_{n=1}^{N} (1 - \frac{1}{2N}\alpha_2^2\gamma(V, H^N_{t_2-\tau:t_2-1}, U^n_{t_2-1}, Y^n))| \\
(42) \quad &< \quad \frac{C}{N^{1/2}} \overset{a.s.}{\to} 0.
\end{aligned}
$$

Then, all the remains to be shown is that $\prod_{n=1}^{N} (1 - \frac{1}{2N}\alpha_2^2\gamma(V, H^N_{t_2-\tau:t_2-1}, U^n_{t_2-1}, Y^n)) \overset{p}{\to} \exp(-\frac{\alpha_2^2}{2\text{Var}[\langle \phi_2, \bar{\mathcal{M}}_{t_2}(u, dy)\rangle]})$ as $N \to \infty$.

$$
\begin{aligned}
&\prod_{n=1}^{N} (1 - \frac{1}{2N}\alpha_2^2\gamma(V, H^N_{t_2-\tau:t_2-1}, U^n_{t_2-1}, Y^n)) = \exp(\sum_{n=1}^{N} \log(1 - \frac{1}{2N}\alpha_2^2\gamma(V, H^N_{t_2-\tau:t_2-1}, U^n_{t_2-1}, Y^n))) \\
= \quad &\exp(\sum_{n=1}^{N} [-\frac{1}{2N}\alpha_2^2\gamma(V, H^N_{t_2-\tau:t_2-1}, U^n_{t_2-1}, Y^n) + \frac{C}{N^2}]) \\
= \quad &\exp(-\frac{1}{2}\alpha_2^2 \sum_{u'\in\mathcal{U}} \langle \gamma(V, H^N_{t_2-\tau:t_2-1}, u', y), \mu^N_{t_2-1}(u', dy)\rangle + \frac{C}{N}) \overset{p}{\to} \exp(-\frac{1}{2}\alpha_2^2 \text{Var}[\langle \phi_2, \bar{\mathcal{M}}_{t_2}\rangle])
\end{aligned}
$$

(43)

The last line follows from the previous convergence result $\mu^N \overset{p}{\to} \bar{\mu}$ (see Proof of Theorem 3.2), $\gamma(V, H^N_{t_2-\tau:t_2-1}, U^n_{t_2-1}, Y^n) = \phi_2(Y^n)^2 h(u, U^n_{t_2-1}, Y^n, V_{t_2-1}, \bar{H}_{t_2-\tau:t_2-1})(1 - h(u, U^n_{t_2-1}, Y^n, V_{t_2-1}, \bar{H}_{t_2-\tau:t_2-1}))$, and the continuous mapping theorem. With this result in hand, we return to equation (46). Since $\mathbb{E}[\exp\left(i\alpha_2 \left\langle \phi_2, \mathcal{M}^{2,N}_{t_2}(u')\right\rangle\right) | \mu^N_{0:t_2-1}, V]$ converges to a constant and $\exp\left(i\alpha_1 \left\langle \phi_1, \mathcal{M}^{2,N}_{t_1}(u)\right\rangle\right) \overset{d}{\to} \exp\left(i\alpha_1 \left\langle \phi_1, \bar{\mathcal{M}}_{t_1}\right\rangle\right)$, Slutsky's theorem yields $\exp\left(i\alpha_1 \left\langle \phi_1, \mathcal{M}^{2,N}_{t_1}(u, dy)\right\rangle\right)\mathbb{E}[\exp\left(i\alpha_2 \left\langle \phi_2, \mathcal{M}^{2,N}_{t_2}(u', dy)\right\rangle\right) | \mu^N_{0:t_2-1}, V] \overset{d}{\to} \exp\left(i\alpha_1 \left\langle \phi_1, \bar{\mathcal{M}}_{t_1}\right\rangle\right) \exp(-\frac{1}{2}\alpha_2^2 \text{Var}[\langle \phi_2, \bar{\mathcal{M}}_{t_2}\rangle])$. By the bounded convergence theorem, we then have that:

$$
\begin{aligned}
&\mathbb{E}[\exp\left(i\alpha_1 \left\langle \phi, \mathcal{M}^{2,N}_{t_1}(u, dy)\right\rangle + i\alpha_2 \left\langle \phi, \mathcal{M}^{2,N}_{t_2}(u', dy)\right\rangle\right) | V] \\
\to \quad &\exp(-\frac{1}{2}\alpha_1^2 \text{Var}[\langle \phi_1, \bar{\mathcal{M}}_{t_1}\rangle]) \exp(-\frac{1}{2}\alpha_2^2 \text{Var}[\langle \phi_2, \bar{\mathcal{M}}_{t_2}\rangle])
\end{aligned}
$$

Consequently, conditional on each V, $\langle \phi_1, \bar{\mathcal{M}}_{t_1}\rangle$ is independent of $\langle \phi_2, \bar{\mathcal{M}}_{t_2}\rangle$ for $t_1 \neq t_2$.

Since these results hold for any choice of $\phi \in S$, we also have the stronger result that $\mathcal{M}^{2,N} \overset{d}{\to} \bar{\mathcal{M}}$, using the criterion (33).

□

Lemma A.8. *For each V, suppose that $\Xi^N_{0:t-1} \overset{d}{\to} \bar{\Xi}_{0:t-1}$. Then, for each V, $(\Xi^N_{0:t-1}, \mathcal{M}^{2,N}_t) \overset{d}{\to} (\bar{\Xi}_{0:t-1}, \bar{\mathcal{M}}_t(u, dy))$.*

Proof. For convenient reference, the form of $\bar{\Xi}_t$ is given in Theorem 3.4. The proof for this lemma is the same as in Lemma A.7. For $t' < t$,

$$\begin{aligned} & \mathbb{E}[\exp\big(i \sum_{t'<t, u\in\mathcal{U}} \left\langle \phi_{1,u,t'}, \Xi^N_{t'}(u,dy)\right\rangle + i\left\langle \phi_2, \mathcal{M}^{2,N}_t(u',dy)\right\rangle\big)|V] \\ = \; & \mathbb{E}\big[\exp\big(i \sum_{t'<t, u\in\mathcal{U}} \left\langle \phi_{1,u,t'}, \Xi^N_{t'}(u,dy)\right\rangle\big)\mathbb{E}[\exp\big(i\left\langle \phi_2, \mathcal{M}^{2,N}_t(u',dy)\right\rangle\big)|\mu^N_{0:t-1}, V]|V\big]. \end{aligned} \tag{44}$$

Now, using the exact same steps as in Lemma A.7, the desired result can be obtained.

□

Lemma A.9. *For each V, suppose that $(\Xi^N_{0:t}, \sqrt{N}\mathcal{Z}^N_{0:t-1}) \xrightarrow{d} (\bar{\Xi}_{0:t}, \bar{\mathcal{Z}}_{0:t-1})$. Then, for each V, $(\{\Xi^N_{0:t}\}, \sqrt{N}\mathcal{Z}^N_{0:t}) \xrightarrow{d} (\bar{\Xi}_{0:t}, \bar{\mathcal{Z}}_{0:t})$. Given V, $\bar{\mathcal{Z}}_t$ is a mean zero Gaussian with variance:*

$$Var[\bar{\mathcal{Z}}_t|V] = Var[\bar{\mathcal{Z}}_t|V_{0:t}] = \left\langle \int_0^1 z^2 \nu_{t,y,V_t}(dz) - (\int_0^1 z\nu_{t,y,V_t}(dz))^2, \bar{\mu}_t(d,dy) - \bar{\mu}_{t-1}(d,dy)\right\rangle. \tag{45}$$

Given V, $\bar{\mathcal{Z}}_t$ is independent of $\bar{\mathcal{Z}}_{0:t-1}$ and $\bar{\Xi}_{0:t}$.

Proof.

$$\begin{aligned} & \mathbb{E}[\exp\big(i \sum_{t'\le t, u\in\mathcal{U}} \left\langle \phi_{1,u,t'}, \Xi^N_{t'}(u,dy)\right\rangle + i\sum_{t'<t} \alpha_{t'}\sqrt{N}Z^N_{t'} + i\alpha_t\sqrt{N}Z^N_t)|V] \\ = \; & \mathbb{E}\big[\exp\big(i \sum_{t'\le t, u\in\mathcal{U}} \left\langle \phi_{1,u,t'}, \Xi^N_{t'}(u,dy)\right\rangle + i\sum_{t'<t} \alpha_{t'}\sqrt{N}Z^N_{t'}\big)\mathbb{E}[\exp\big(i\alpha_t\sqrt{N}Z^N_t\big))|\mu^N_{0:t}, V]|V\big]. \end{aligned} \tag{46}$$

The result follows via the exact same procedure as used in Lemma A.7. □

Proof of Theorem 3.4. For each $u \in \mathcal{U}$ and any $\phi \in S$, we have that

$$\begin{aligned} \left\langle \phi(y), \Xi^N_t(u,dy)\right\rangle = & \sum_{u'\in\mathcal{U}} \Bigg[\left\langle \phi(y), h(u,u',y,V_{t-1}, H^N_{t-\tau:t-1})\mu^N_{t-1}(u',dy)\right\rangle \\ & - \left\langle \phi(y), h(u,u',y,V_{t-1}, H^N_{t-\tau:t-1})\bar{\mu}_{t-1}(u',dy)\right\rangle \Bigg] + \left\langle \phi, \mathcal{M}^{2,N}_t(u)\right\rangle, \end{aligned} \tag{47}$$

where $\mathcal{M}^{2,N}_t$ satisfies:

$$\left\langle \phi, \mathcal{M}^{2,N}_t(u,dy)\right\rangle = \frac{1}{\sqrt{N}} \sum_{n=1}^N \phi(Y^n)(\mathbf{1}_{U^n_t = u} - h(u, U^n_{t-1}, Y^n, V_{t-1}, H^N_{t-\tau:t-1}). \tag{48}$$

Using a Taylor expansion, one can obtain

$$\begin{aligned} \left\langle \phi(y), \Xi^N_t(u,dy)\right\rangle = & \sum_{u'\in\mathcal{U}} \left\langle \phi(y)h(u,u',y,V_{t-1}, \bar{H}_{t-\tau:t-1}), \Xi^N_{t-1}(u',dy)\right\rangle \\ & + \sum_{u'\in\mathcal{U}} \left\langle \phi(y)h_H(u,u',y,V_{t-1}, \bar{H}_{t-\tau:t-1})\cdot E^N_{t-\tau:t-1}, \mu^N_{t-1}(u',dy)\right\rangle + \frac{1}{\sqrt{N}}\mathcal{R}^N + \left\langle \phi, \mathcal{M}^{2,N}_t(u,dy)\right\rangle, \end{aligned} \tag{49}$$

where $\mathcal{R}^N$ is the Taylor remainder term which can be bound using Assumption 3.3 (continuous differentiability of h and its arguments live on a compact space):

$$\mathcal{R}^N \le C||E^N_{t'<t}||^2. \tag{50}$$

We will now use an induction argument to prove the desired result. Conditional on each V, assume that $\Xi^N_{0:t-1}$ converges in distribution to $\bar{\Xi}_{0:t-1}$. By the continuous mapping theorem (and the fact that convergence in distribution to a constant implies convergence in probability to a constant), we have that $\frac{1}{\sqrt{N}}\mathcal{R}^N$ converges in probability to zero. The second term in equation (49) also converges in distribution since μ^N_{t-1} converges in probability to $\bar{\mu}_{t-1}$ for each V (from the law of large numbers) and $E^N_{t-\tau:t} \xrightarrow{d} \bar{E}_{t-\tau:t-1}$ (since we assumed $\Xi^N_{0:t-1} \xrightarrow{d} \bar{\Xi}_{0:t-1}$ and f^H is continuous). The third term converges in distribution due to Lemma A.7. Then,

for each V, assuming $\Xi^N_{0:t-1}$ converges in distribution to $\bar{\Xi}_{0:t-1}$, $\langle \phi, \Xi^N_t \rangle \xrightarrow{d} \langle \phi, \bar{\Xi}_t \rangle$ for any $\phi \in S$ where $\langle \phi, \bar{\Xi}_t \rangle$ satisfies the evolution equation:

$$\begin{aligned} \langle \phi(y), \bar{\Xi}_t(u, dy) \rangle &= \sum_{u' \in \mathcal{U}} \langle \phi(y) h(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1}), \bar{\Xi}_{t-1}(u', dy) \rangle \\ &+ \sum_{u' \in \mathcal{U}} \langle \phi(y) h_H(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1}) \cdot \bar{E}_{t-\tau:t-1}, \bar{\mu}_{t-1} \rangle + \langle \phi, \bar{\mathcal{M}}_t(u, dy) \rangle, \quad \phi \in S. \end{aligned} \tag{51}$$

This is sufficient to show that the $\Xi^N_t \xrightarrow{d} \bar{\Xi}_t$, by the criterion (33). However, to make the stronger statement of the joint convergence $\Xi^N_{0:t} \xrightarrow{d} \bar{\Xi}_{0:t}$, we need an additional fact, namely that $(\Xi^N_{0:t-1}, \mathcal{M}^{2,N}_t) \xrightarrow{d} (\bar{\Xi}_{0:t}, \bar{\mathcal{M}}_t(u, dy))$. This has been proven in Lemma A.8, and therefore we have the desired result that $\Xi^N_{0:t} \xrightarrow{d} \bar{\Xi}_{0:t}$. By induction, we have that $\bar{\Xi}^N \xrightarrow{d} \bar{\Xi}$ for each V. Therefore, $\bar{\Xi}^N \xrightarrow{d} \bar{\Xi}$. Finally, using Lemma A.9, we have that $(\Xi^N, \Lambda^N) \xrightarrow{d} (\bar{\Xi}, \bar{\Lambda})$ since:

$$\begin{aligned} \Lambda^N_t = \sqrt{N}(L^N_t - \bar{L}_t) &= \int_{\mathcal{Y}} \int_0^1 z \nu_{t,y,V_t}(dz)(\Xi^N_t(d, dy) - \Xi^N_{t-1}(d, dy)) + \sqrt{N} \mathcal{Z}^N_t \\ &\xrightarrow{d} \int_{\mathcal{Y}} \int_0^1 z \nu_{t,y,V_t}(dz)(\bar{\Xi}_t(d, dy) - \bar{\Xi}_{t-1}(d, dy)) + \bar{\mathcal{Z}}_t = \bar{\Lambda}_t. \end{aligned}$$

This result follows since $\bar{\Xi}^N \xrightarrow{d} \bar{\Xi}$, the assumption of continuity in y for $\int_0^1 z \nu_{t,y,V_t}(dz)$ for each t, V, and Lemma A.9. □

Proof of Corollary 3.5. Since for each V, $(\mu^N_t, L^N) \xrightarrow{p} (\bar{\mu}, \bar{L})$ where $(\bar{\mu}, \bar{L})$ are constants (conditional on V) and $(\Xi^N, \Lambda^N) \xrightarrow{d} (\bar{\Xi}, \bar{\Lambda})$, we have the result by Slutsky's theorem and dominated convergence. □

Appendix B. Convergence Rate of a Quadrature Scheme for the Evaluation of the LLN

In Section 4, a quadrature scheme was proposed in order to simulate the Monte Carlo approximation $\bar{\mu}^N$ (which is a linear combination of a law of large numbers $\bar{\mu}$ and a central limit theorem $\bar{\Xi}$). Under some technical conditions, we show a convergence rate for that quadrature scheme for the law of large numbers. The convergence rate for the simulation scheme for the central limit theorem can be proven in a similar fashion. The results in this section can be used as a practical guideline to determine the number of grid points for the simulation scheme in Section 4.

Let the maximum "radius" of the computational cells be:

$$r_{K,i} = \frac{1}{2} \max_{k=1,\ldots,K} \max_{y_i, y_i' \in c_k} |y_i - y_i'|,$$

where y_i is the i-th element of the vector $y \in \mathcal{Y} \subset \mathbb{R}^{d_Y}$. As in Section 4, for a sample V^l of the process X:

$$\langle f, \bar{\mu}^l_t \rangle_{\mathcal{U} \times \mathcal{Y}} \approx \sum_{k=1}^K f(u, y_k) \bar{\mu}^{\Psi,l}_t(u, y_k) \equiv m^l,$$

where $\bar{\mu}^{\Psi,l}_t$ is the law of large numbers under the quadrature scheme, $\bar{\mu}^l_t$ is the law of large numbers conditional on the path V^l, and we define $H^\Psi = \sum_{u \in \mathcal{U}} \int_{\mathcal{Y}} f^H(u, y) \bar{\mu}^\Psi(u, dy)$. For notational convenience, we have suppressed the Monte Carlo sample notation "l" for H^Ψ and $\bar{H}$. Under certain conditions, we will prove a convergence rate for the mean-squared error (MSE) of the simulation scheme given in Section 4 for the law of large numbers $\bar{\mu}$:

$$\text{MSE} = \mathbb{E}[(\mathbb{E}[g(\langle f, \bar{\mu}_t \rangle_{\mathcal{U} \times \mathcal{Y}})] - \frac{1}{L} \sum_{l=1}^L g(m^l))^2] \tag{52}$$

The convergence rate will give insight into how to optimally design a grid for the simulation scheme described in Section 4.

Assumption B.1. *$\bar{\mu}_0(u, dy) = 0$ if $u \neq o$ (i.e., all of the pool is initially in the {outstanding} state), $V_t \in \mathcal{V} \subset \mathbb{R}^{d_V}$ where $\mathcal{V}$ is compact, $\mathcal{Y}$ is compact, g, f, f^H are continuously differentiable, and h is twice continuously differentiable.*

Let $\zeta_t^{l,H}(u, y)$ satisfy:

$$\begin{aligned}
\zeta_t^{l,H}(u,y) &= \sum_{u' \in \mathcal{U}} h_\theta(u, u', y, V_{t-1}^l, H_{t-1}) \zeta_{t-1}^{l,H}(u', y), \\
\zeta_0^{l,H}(\mathrm{o}, y) &= 1, \\
\zeta_0^{l,H}(u, y) &= 0, \quad u \neq \mathrm{o}. \qquad (53)
\end{aligned}$$

Then, $\bar{\mu}_t^l(u, dy) = \zeta_t^{l,\bar{H}}(u, y)\bar{\mu}_0(\mathrm{o}, dy)$ and $\bar{\mu}_t^{l,\Psi}(u, y_k) = \zeta_t^{l,H^\Psi}(u, y_k)\bar{\mu}_0(\mathrm{o}, c_k)$. Since h is continuously differentiable, $\zeta_t^{l,H}(u, y)$ is continuously differentiable in y for each u.

$$\begin{aligned}
|\zeta_t^{l,\bar{H}}(u,y_k) - \zeta_t^{l,H^\Psi}(u,y_k)| &\leq |\zeta_t^{l,\bar{H}}(u,y_k) - \sum_{u' \in \mathcal{U}} h_\theta(u,u',y_k,V_{t-1}^l,\bar{H}_{t-1})\zeta_{t-1}^{l,H^\Psi}(u',y_k)| \\
&+ |\sum_{u' \in \mathcal{U}} h_\theta(u,u',y_k,V_{t-1}^l,\bar{H}_{t-1})\zeta_{t-1}^{l,H^\Psi}(u',y_k) - \zeta_t^{l,H^\Psi}(u,y_k)| \\
&\leq \sum_{u' \in \mathcal{U}} h_\theta(u,u',y_k,V_{t-1}^l,\bar{H}_{t-1})|\zeta_{t-1}^{l,\bar{H}}(u',y_k) - \zeta_{t-1}^{l,H^\Psi}(u',y_k)| \\
&+ \sum_{u' \in \mathcal{U}} |\frac{\partial h_\theta}{\partial H}(u,u',y_k,V_{t-1}^l,h^*)||\zeta_{t-1}^{l,H^\Psi}(u',y_k)||H_{t-1}^\Psi - \bar{H}_{t-1}| \\
&\leq \sum_{u' \in \mathcal{U}} h_\theta(u,u',y_k,V_{t-1}^l,\bar{H}_{t-1})|\zeta_{t-1}^{l,\bar{H}}(u',y_k) - \zeta_{t-1}^{l,H^\Psi}(u',y_k)| \\
&+ \sum_{u' \in \mathcal{U}} |\frac{\partial h_\theta}{\partial H}(u,u',y_k,V_{t-1}^l,h^*)||\bar{H}_{t-1} - H_{t-1}^\Psi|, \\
&\leq \sum_{u' \in \mathcal{U}} h_\theta(u,u',y_k,V_{t-1}^l,\bar{H}_{t-1})|\zeta_{t-1}^{l,\bar{H}}(u',y_k) - \zeta_{t-1}^{l,H^\Psi}(u',y_k)| + K_1|\bar{H}_{t-1} - H_{t-1}^\Psi| \\
&\leq \max_{u,y_k} |\zeta_{t-1}^{l,\bar{H}}(u,y_k) - \zeta_{t-1}^{l,H^\Psi}(u,y_k)| + K_1|\bar{H}_{t-1} - H_{t-1}^\Psi| \qquad (54)
\end{aligned}$$

where we have bounded $\frac{\partial h_\theta}{\partial H}$ using the compactness of the space its arguments live on and its continuity. In addition, we have used the fact that $h(\cdot, u', \cdot)$ is a probability kernel and therefore sums to one. Taking the maximum over the $u \in \mathcal{U}$ and the grid points y_k:

$$\max_{u,y_k} |\zeta_t^{l,\bar{H}}(u,y_k) - \zeta_t^{l,H^\Psi}(u,y_k)| \leq \max_{u,y_k} |\zeta_{t-1}^{l,\bar{H}}(u,y_k) - \zeta_{t-1}^{l,H^\Psi}(u,y_k)| + K_1|\bar{H}_{t-1} - H_{t-1}^\Psi|. \qquad (55)$$

Next, we find a bound for $|\bar{H}_t - H_t^\Psi|$ in terms of $|\zeta_t^{l,\bar{H}}(u, y_k) - \zeta_t^{l,H^\Psi}(u, y_k)|$.

$$H_t^\Psi = \sum_{u \in \mathcal{U}} \sum_{k=1}^K f^H(u, y_k)\bar{\mu}_t^{\Psi,l}(u, c_k) = \sum_{k=1}^K \sum_{u \in \mathcal{U}} f(u, y_k)\zeta_t^{l,H^\Psi}(u, y_k)\bar{\mu}_0(\mathrm{o}, c_k).$$

Using a Taylor expansion:

$$\begin{aligned}
|\bar{H}_t - H_t^\Psi| &= |\left\langle f^H, \bar{\mu}_t^l \right\rangle_{\mathcal{U}\times\mathcal{Y}} - \sum_{u \in \mathcal{U}} \sum_{k=1}^K f^H(u, y_k)\bar{\mu}_t^{\Psi,l}(u, c_k)| \\
&\leq \sum_{u \in \mathcal{U}} \sum_{k=1}^K \int_{c_k} |f^H(u,y)\zeta_t^{l,\bar{H}}(u,y) - f^H(u,y_k)\zeta_t^{l,H^\Psi}(u,y_k)|\bar{\mu}_0(\mathrm{o}, dy) \\
&\leq \sum_{u \in \mathcal{U}} \sum_{k=1}^K \int_{c_k} |f^H(u,y)\zeta_t^{l,\bar{H}}(u,y) - f^H(u,y_k)\zeta_t^{l,\bar{H}}(u,y_k)|\bar{\mu}_0(\mathrm{o}, dy)
\end{aligned}$$

$$
\begin{aligned}
&+ \quad \sum_{u\in\mathcal{U}}\sum_{k=1}^{K}\int_{c_k} |f^H(u,y_k)\zeta_t^{l,\bar{H}}(u,y_k) - f^H(u,y_k)\zeta_t^{l,H^\Psi}(u,y_k)|\bar{\mu}_0(\mathrm{o},dy).\\
&\leq \quad 2\sum_{i=1}^{d_Y} r_{K,i}\sum_{u\in\mathcal{U}}\sup_{y\in\mathcal{Y}}|\frac{\partial}{\partial y_i}[f^H(u,y)\zeta_t^{l,\bar{H}}(u,y)]| + C_3\max_{u,y_k}|\zeta_t^{l,\bar{H}}(u,y_k) - \zeta_t^{l,H^\Psi}(u,y_k)|\\
&\equiv \quad 2\sum_{i=1}^{d_Y} C_{1,t,i}(X^l)r_{K,i} + C_3\max_{u,y_k}|\zeta_t^{l,\bar{H}}(u,y_k) - \zeta_t^{l,H^\Psi}(u,y_k)| \qquad (56)
\end{aligned}
$$

Returning to equation (55), we now have:

$$
\begin{aligned}
\max_{u,y_k}|\zeta_t^{l,\bar{H}}(u,y_k) - \zeta_t^{l,H^\Psi}(u,y_k)| &\leq \max_{u,y_k}|\zeta_{t-1}^{l,\bar{H}}(u,y_k) - \zeta_{t-1}^{l,H^\Psi}(u,y_k)|\\
&+ \quad 2K_1\sum_{i=1}^{d_Y} C_{1,t-1,i}(V^l)r_{K,i} + C_3K_1\max_{u,y_k}|\zeta_{t-1}^{l,\bar{H}}(u,y_k) - \zeta_{t-1}^{l,H^\Psi}(u,y_k)|\\
&\leq \quad 2K_1\sum_{i=1}^{d_Y} C_{1,t-1,i}(V^l)r_{K,i} + (1+C_3K_1)\sum_{t=0}^{t-1}\max_{u,y_k}|\zeta_{t'}^{l,\bar{H}}(u,y_k) - \zeta_{t'}^{l,H^\Psi}(u,y_k)|\\
&\leq \quad 2K_1\sum_{i=1}^{d_Y} C_{1,t-1,i}(V^l)r_{K,i} + 2K_1(1+C_3K_1)\sum_{t'=0}^{t-1} e^{(t-t'-1)(1+C_3K_1)}\sum_{i=1}^{d_Y} C_{1,t',i}(V^l)r_{K,i},\\
&\equiv \quad \sum_{i=1}^{d_Y} C_{2,t,i}(V^l)r_{K,i} \qquad (57)
\end{aligned}
$$

where we have used Gronwall's lemma for the last inequality. We note that numerical scheme converges as the size of the cells $r_{K,1},\ldots,r_{K,d_Y}\to 0$. The error depends upon the magnitude of the derivative of the law of large numbers is with respect to each dimension i and how fine the grid is along the dimension i. The sensitivity of the error to the magnitude of the derivative of the law of large numbers with respect to the dimension i is captured in the term $C_{2,t,i}(V^l)$. By using a Taylor expansion in the exact same manner as shown previously, this of course implies that:

$$
|g(\langle f,\bar{\mu}_t^l\rangle_{\mathcal{U}\times\mathcal{Y}}) - g(\left\langle f,\bar{\mu}_t^{l,\Psi}\right\rangle_{\mathcal{U}\times\mathcal{Y}})| \leq \sum_{i=1}^{d_Y} C_{3,t,i}(V^l)r_{K,i}. \qquad (58)
$$

We also note that $C_{3,t,i}(v) < C_4 < \infty$ since it is a continuous function on a compact set (due to assumptions that $\mathcal{V}$ is compact and h is twice differentiable).

We now find the desired convergence rate:

$$
\begin{aligned}
\mathbb{E} \quad &[(\mathbb{E}[g(\langle f,\bar{\mu}_t\rangle_{\mathcal{U}\times\mathcal{Y}})] - \frac{1}{L}\sum_{l=1}^{L} g(m^l))^2] \leq \mathbb{E}[(\mathbb{E}[g(\langle f,\bar{\mu}_t\rangle_{\mathcal{U}\times\mathcal{Y}})] - \frac{1}{L}\sum_{l=1}^{L} g(\langle f,\bar{\mu}_t^l\rangle_{\mathcal{U}\times\mathcal{Y}}))^2]\\
+ \quad &\mathbb{E}[(\frac{1}{L}\sum_{l=1}^{L} g(\langle f,\bar{\mu}_t^l\rangle_{\mathcal{U}\times\mathcal{Y}}) - \frac{1}{L}\sum_{l=1}^{L} g(m^l))^2] \leq \underbrace{\frac{1}{L}\mathrm{Var}[g(\langle f,\bar{\mu}_t\rangle_{\mathcal{U}\times\mathcal{Y}})]}_{variance} + \underbrace{C_5 d_Y\sum_{i=1}^{d_Y}\mathbb{E}[C_{3,t,i}(V)^2]r_{K,i}^2}_{bias}
\end{aligned}
$$

The assumption that $\mathcal{V}$ is compact was made in order that $C_{3,t,i}(V)^2 < C_3^2$ and thus $C_{3,t,i}(V)^2$ would be integrable. This assumption can be relaxed to simply requiring that $C_{3,t,i}(V)^2$ be integrable. As a consequence of equation (59), the mean-squared error of the numerical approximation converges to zero as $L\to\infty$ and $\max_i r_{K,i}\to 0$. Equation (59) also provides insight into the factors driving the numerical error of the simulation scheme. The mean-squared error is composed of a variance and a bias term. The variance term is the variance of a sample without numerical error; i.e., the variance term would remain even if one could produce samples with no numerical error. The bias term is a consequence of the error produced by the quadrature scheme. It is the sum of the maximum length of the cells along a particular dimension multiplied by the *average* of the squared partial derivative of a function of the law of large numbers $\bar{\mu}_t$ with respect

to that dimension (i.e., $\mathbb{E}[C_{3,t,i}(V)^2]$). Therefore, it is desirable to have a finer grid with respect to the dimensions along which the partial derivatives of the law of large nubmers $\bar{\mu}_t$ are most rapidly changing. Overall, one can reduce the MSE by either choosing a finer grid with respect to a particular dimension (thus reducing the bias) or by generating more Monte Carlo samples (thus reducing the variance).

The convergence rate (59) suggests an optimal allocation of a computational budget. Given a fixed computational cost (i.e., maximum allowed computational time), one must choose the optimal number of Monte Carlo trials L and the cell radii $r_{K,1}, \ldots, r_{K,K}$ in order to minimize the mean-squared error. For instance, assuming a rectangular grid (which is not the best approach in higher-dimensions; see Section 6.5 for a better alternative) with a total computational budget B, the budget equation is:

$$L \prod_{i=1}^{d_Y} \frac{1}{r_{K,i}} = \frac{B}{C_6},$$

where the constant C_6 involves the cost of each simulation as well as the size of the space $\mathcal{Y}$ which one is discretizing over. For notational convenience, define $C_{\text{variance}} = \text{Var}[g(\langle f, \bar{\mu}_t \rangle)]$ and $C_{\text{bias},i} = C_5 d_Y \mathbb{E}[C_{3,t,i}(V)^2]$. The constant $C_{\text{bias},i}$ is larger for dimensions i along which the solution varies more rapidly. The optimal choices for L and $r_{K,1}, \ldots, r_{K,K}$ satisfy a system of hyperbolic equations which can be solved explicitly. For instance, in two dimensions ($d_Y = 2$), the optimal choices are:

$$\begin{aligned}
r_{K,1} &= (C_{\text{variance}} C_6)^{2/10} \left(\frac{1}{2BC_{\text{bias},1}}\right)^{3/10} \left(\frac{1}{2BC_{\text{bias},2}}\right)^{-1/10}, \\
r_{K,2} &= (C_{\text{variance}} C_6)^{2/10} \left(\frac{1}{2BC_{\text{bias},2}}\right)^{3/10} \left(\frac{1}{2BC_{\text{bias},1}}\right)^{-1/10}, \\
L &= B^{3/5} C_6^{-4/5} C_{\text{variance}}^{4/10} \left(\frac{1}{2C_{\text{bias},1}}\right)^{1/5} \left(\frac{1}{2BC_{\text{bias},2}}\right)^{1/5}.
\end{aligned}$$

As expected, the optimal number of Monte Carlo trials L increases with the variance. Similarly, the fineness of the grid in dimension i decreases the larger the solution's derivative is with respect to that dimension. □

References

[1] A. Lo A. Khandani, A. Kim. Consumer credit-risk models via machine-learning algorithms. *Journal of Banking and Finance*, 34(11):2767–2787, 2010.

[2] B. Ambrose and C. Capone. Modeling the conditional probability of foreclosure in the context of single-family mortgage default resolutions. *Real Estate Economics*, 26(3):391–429, 1998.

[3] Matthias Arnsdorf and Igor Halperin. BSLP: markovian bivariate spread-loss model for portfolio credit derivatives. *Journal of Computational Finance*, 12:77–100, 2008.

[4] B. Baesens. Neural network survival analysis for personal loan data. *Journal of the Operational Research Society*, 56(9):1089–1098, 2005.

[5] J. Banasik, J. Crook, and L. Thomas. Not if but when will borrowers default. *Journal of Operational Research Society*, pages 1185–1190, 1999.

[6] J. Bastos. Forecasting bank loans loss-given-default. *Journal of Banking and Finance*, 34(10):2510–2517, 2010.

[7] M. Benaim and J. Boudec. A class of mean field interaction models for computer and communication systems. *Performance Evaluation*, 65(11):823–838, 2008.

[8] P. Billingsley. *Convergence of Probability Measures*. John Wiley and Sons, 2008.

[9] N. Bush, B. M. Hambly, H. Haworth, L. Jin, and C. Reisinger. Stochastic evolution equations in portfolio credit modelling. *SIAM Journal of Financial Mathematics*, 2(1):627–664, 2011.

[10] D. Capozza, D. Kazarian, and T. Thomson. Mortgage default in local markets. *Real Estate Economics*, 25(4):631–655, 1997.

[11] D. Crisan, P. Del Moral, and T. Lyons. Discrete filtering using branching and interacting particle systems. Technical report, Laboratoire de Statistique et Probabilities, Universite de Toulouse, 1998.

[12] J. Cvitanic, J. Ma, and J. Zhang. The law of large numbers for self-exciting correlated defaults. *Stochastic Processes and their Applications*, 122(8):2781–2810, 2012.

[13] P. Dai Pra, W.J. Runggaldier, E. Sartori, and M. Tolotti. Large portfolio losses: A dynamic contagion model. *The Annals of Applied Probability*, 19(1):347–394, 2009.

[14] Y. Deng, J. Quigley, and R. Van Order. Mortgage terminations, heterogeneity, and the exercise of mortgage options. *Econometrica*, 68(2):275–307, 2000.

[15] N. Diener, R. Jarrow, and P. Protter. Relating top-down with bottom-up approaches in the evaluation of abs with large collateral pools. *International Journal of Theoretical and Applied Finance*, 15(2), 2012.

[16] Xiaowei Ding, Kay Giesecke, and Pascal Tomecek. Time-changed birth processes and multi-name credit derivatives. *Operations Research*, 57(4):990–1005, 2009.

[17] E. Errais, K. Giesecke, and L. Goldberg. Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1:642–665, 2010.

[18] J. Fermanian. A top-down approach for MBS, ABS and CDO of ABS: a consistent way to manage prepayment, default and interest rate risks. *Journal of Real Estate Finance and Economics*, 46(3), February 2013.

[19] X. Fernique. Processus lineaires, processus generalises. *Ann. Inst. Fourier*, 17(1):1–92, 1967.

[20] Jean-Pierre Fouque. La convergence en loi pour les processus a valeurs dans un espace nucleaire. *Annales de l'I.H.P.*, 20:225–245, 1984.

[21] K. Giesecke, K. Spiliopoulos, R.B. Sowers, and J.A. Sirignano. Large portfolio asymptotics for loss from default. *Mathematical Finance*, 2014, in press.

[22] K. Giesecke and S. Weber. Credit contagion and aggregate losses. *Journal of Economic Dynamics and Control*, 30(5):741–767, 2006.

[23] Kay Giesecke, Lisa Goldberg, and Xiaowei Ding. A top-down approach to multi-name credit. *Operations Research*, 59(2):283–300, 2011.

[24] D. Gomes, J. Mohr, and R. Souza. Discrete time, finite state space mean field games. *Journal de Mathematiques Pures et Appliquees*, 93(3):308–328, 2010.

[25] R. Goodstein, P. Hanouna, C. Ramirez, and C. Stahel. Contagion effects in strategic mortgage defaults. GMU Working Paper in Economics No. 13-07, 2011.

[26] K. Harada and H. Saigo. The space of tempered distributions as a k-space. arXiv preprint arXiv:1009.1429, 2010.

[27] J. Harding, Eric Rosenblatt, and V. Yao. The contagion effect of foreclosed properties. *Journal of Urban Economics*, 66:164–178, July 2009.

[28] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.

[29] J. Hoffmann-Jorgensen and G. Pisier. The law of large numbers and the central limit theorem in banach spaces. *The Annals of Probability*, pages 587–599, 1976.

[30] P. Kang and S. Zenios. Complete prepayment models for mortgage-backed securities. *Management Science*, 38(11):1665–1685, 1992.

[31] Z. Lin, E. Rosenblatt, and V. Yao. Spillover effects of foreclosures on neighborhood property values. *Journal of Real Estate Finance and Economics*, 38(4):387–407, May 2009.

[32] J. Mattey and N. Wallace. Housing-price cycles and prepayment rates of us mortgage pools. *The Journal of Real Estate Finance and Economics*, 23(2):161–184, 2001.

[33] D. McLeish. Dependent central limit theorems and invariance principles. *The Annals of Probability*, pages 620–628, 1974.

[34] I. Mitoma. Tightess of probabilities on c([0,1]; y') and d([0,1]; y'). *The Annals of Probability*, pages 989–999, 1983.

[35] P. Del Moral. Measure-valued processes and interacting particle systems: Application to nonlinear filtering problems. *Annals of Applied Probability*, 1998.

[36] P. Del Moral and A. Guionnet. Central limit theorem for nonlinear filtering and interacting particle systems. *Annals of Applied Probability*, pages 275–297, 1999.

[37] M.Schreckenberg, A. Schadschneider, K. Nageland, and N. Ito. Discrete stochastic models for traffic flow. *Physical Review E*, 51(4), 1995.

[38] D. Mushtari. Levy type criteria for weak convergence of probabilities in frechet spaces. *Theory of Probability and its Applications*, 24(3):587–592, 1980.

[39] P. Dai Pra and M. Tolotti. Heterogeneous credit portfolios and the dynamics of the aggregate losses. *Stochastic Processes and their Applications*, 119(9):2913–2944, 2009.

[40] S. Richard and R. Roll. Prepayments on fixed-rate mortgage-backed securities. *Journal of Portfolio Management*, 15(3):73–82, 1989.

[41] J. Sirignano and K. Giesecke. Correlated default and prepayment modeling for mortgages. Working Paper, April 2015.

[42] K. Spiliopoulos, J.A. Sirignano, and K. Giesecke. Fluctuation analysis for the loss from default. *Stochastic Processes and their Applications*, 124:2322–2362, 2014.

[43] H. Stein, A. Belikoff, K. Levin, and X. Tian. Analysis of mortgage-backed securities: before and after the credit crisis. *Credit Risk Frontiers: Subprime Crisis, Pricing and Hedging, CVA, MBS, Ratings, and Liquidity*, pages 345–394, 2007.

[44] M. Stepanova and L. Thomas. Survival analysis methods for personal loan data. *Operations Research*, 50(2):277–289, 2002.

[45] E. Stewart and T. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley and Sons Inc., 1986.

[46] C. Towe and C. Lawley. The contagion effect on neighboring foreclosures on own foreclosures. Working paper, University of Maryland and University of Manitoba, 2010.

[47] N. Vakhania, V. Tarieladze, and S. Chobanyan. *Probability Distributions on Banach Spaces*, volume 14. Springer Sciences and Business Media, 1987.

[48] G. Weintraub, C. Benkard, and B. Van Roy. Oblivious equilibrium: A mean field approximation for large-scale dynamic games. *NIPS*, 2005.

[49] G. Weintraub, C. Benkard, and B. Van Roy. Markov perfect industry dynamics with many firms. *Econometrica*, 76(6):1375–1411, 2008.

[50] S. Westgaard and N. Van der Wijst. Default probabilities in a corporate bank portfolio: a logistic model approach. *European Journal of Operational Research*, 135(2):338–349, 2001.

[51] T. Williams. Distributed calculations on fixed-income securities. In *Proceedings of the 2nd Workshop on High Performance Computational Finance*. Portland, OR, November 2009.
[52] S. Wu, L. Jiang, and J. Liang. Intensity-based models for pricing mortgage-backed securities with repayment risk under a cir process. *International Journal of Theoretical and Applied Finance*, 15(3), 2012.
[53] S. Zenios. Parallel monte carlo simulation of mortgage-backed securities. *Financial Optimization*, page 325, 1996.

Department of Management Science and Engineering, Stanford University, Stanford, CA 94305
E-mail address: jasirign@stanford.edu

Department of Management Science and Engineering, Stanford University, Stanford, CA 94305
E-mail address: giesecke@stanford.edu