RISK ANALYSIS FOR LARGE POOLS OF LOANS

JUSTIN A. SIRIGNANO AND KAY GIESECKE

ABSTRACT. Financial institutions, government-sponsored enterprises, and asset-backed security investors are often exposed to delinquency and prepayment risk from large numbers of loans. Examples include mortgages, credit cards, auto, student, and business loans. Due to the size of the pools, the measurement and management of these exposures is computationally expensive. This paper develops and tests efficient numerical methods for the analysis of large pools of loans as well as asset-backed securities backed by such pools. For a broad class of dynamic, data-driven models of loan-level risk, we develop a law of large numbers and a central limit theorem that describe the behavior of pool-level risk. The asymptotics are then used to construct efficient Monte Carlo approximations 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 in practice. 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 profiles.

1. INTRODUCTION

Financial institutions, government sponsored enterprises such as 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. A major US bank might directly own a million mortgages. The GSEs are exposed to tens of millions of mortgages, either through directly owning the mortgages or providing credit guarantees against default for the mortgages. Major loan servicers 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. Mortgage-backed securities (MBS) typically have hundreds to tens of thousands of loans. Large pools of loans are also common for many other types of 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 tens of millions of credit card accounts. Over half of all consumer credit is eventually securitized, and each deal can consist of tens of thousands of credit cards. Due to their size, these loan pools and the securities they back are computationally challenging to analyze. A simulation of a typical pool might take many hours.

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-level risk. A loan might be in one of several states, such as 30 days past due, prepaid, or foreclosure. The conditional transition probabilities could come from a range of statistical or machine learning formulations such as generalized linear models (e.g., logistic regression), neural networks, decision trees, and support vector machines. The transition probability is allowed to 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, as well as the past behavior of the loans in the pool (through a "mean-field" term). These data-driven models are widely used in practice (see [33], [9], [19], [17], and [16]) and are fitted from historical loan performance data that are collected internally or acquired from data vendors.

For this important class of models, we prove a law of large numbers and a central limit theorem that describe pool-level risk. These limit theorems are then used to construct efficient Monte Carlo approximations

Date: December 22, 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 University of Illinois at Urbana-Champaign, Imperial College London, London-Paris Bachelier Workshop on Mathematical Finance, and the International Monetary Fund for comments. Special thanks are due to Kostas Spiliopoulos and Stefan Weber 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.

of the loss, prepayment, and cash flow distributions for a large loan pool. We also prove a uniform integrability result, which guarantees the convergence of our approximation for a large class of continuous functions of pool-level quantities such as loss, prepayment, and cashflow. 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, prepayment, and cash flow distributions across all time horizons at no extra computational cost. Given these distributions, risk analysis, pricing, and hedging of asset-backed securities such as MBSs backed by the pool is immediate.

We test the performance of our approximations on a detailed loan-level mortgage data set which includes over 25 million prime and subprime 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. The approximation's computational cost is often several orders of magnitude less than the cost of brute-force simulation of the actual pool. It has a similar level of accuracy, in the center as well as the tail of the distribution. 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. [46], [52], and [54] propose parallel computing approaches to MBS valuation. [20], [53], [43], [31] and [36] develop top-down models of mortgage pool behavior, and [2], [25], [15], [18] and others develop top-down models of corporate credit pools. The models are formulated without reference to the constituent loans and only model the pool in aggregate. The approximation that we develop for loan-level models is as tractable as a top-down model while taking full advantage of the loan-level feature data.

Limit theorems for credit pools have previously been proven in other model frameworks; see [7], [11], [12], [14], [23], [24], [42], [45], 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. Unlike earlier results, our limit theorems cover a wide range of statistical and machine learning models that are in widespread industry use. Our formulation is well-adapted to settings with high-dimensional loan-level feature data, where the features can be continuously valued. Earlier model frameworks for which limiting laws have been proven do not include loan-level data as a model input. Unlike the aforementioned papers, we allow for multiple types of events which may be mutually exclusive. In practice, many loans are subject to several types of events, such as delinquency, prepayment, and foreclosure. Finally, limiting laws in the above papers do not cover important applications where the function of interest is unbounded. This includes the expectation, variance, higher-order moments, and utility functions of the pool-level loss, prepayment, or cashflow. To address these applications, we prove uniform integrability, which allows our convergence results to cover a large class of functions of practical interest.

Discrete-time interacting particle models have been previously studied in the context of filtering; see [39], [40], and [10]. The setting they study is different than the one in this paper. They prove limiting laws for a system of particles where, given a sequence of observations, the particles transition according to particle filtering updates. 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.

1.2. **Structure of the paper.** The class of models we consider is described in Section 2. The limit theorems as well as uniform integrability for this class of models are presented in Section 3. These results are used to develop an efficient Monte Carlo approximation for large pools. Numerical methods to compute the approximation are described in Section 4. In Section 5, the approximation is tested on actual mortgage data. There are several technical appendices, one containing the proofs of our results.

2. Class of Loan-Level Models

We consider a broad family of dynamic models of loan-level risk in a pool at times $t \in I = \{0, 1, \dots, T\}$. 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_t^n takes values in a finite discrete space $\mathcal{U}$. Common states are current, 30, 60, and 90+ days delinquent, prepaid, charge-off, and foreclosure. Some states might be absorbing.

State transitions are influenced by several types of risk factors. 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 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 interest 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 state transitions. 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. In light of the mortgage meltdown, such a feedback mechanism has been supported by several recent empirical papers; see [1], [26], [28], [34], and [49].

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

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

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, we model the financial losses suffered by the loan originator, servicer, or investor upon certain state transitions. Given a state transition $U_{t-1}^n \neq \mathrm{d} \longrightarrow U_t^n = \mathrm{d}$, where $\mathrm{d} \in \mathcal{U}$ represents an absorbing "default" state such as 90+ days delinquent, charge-off, or foreclosure, the n-th loan suffers a loss $\ell_t^n(Y^n, V_t) \in [0, 1]$. Here, $\ell_t^n(Y^n, V_t)$ is itself a random variable conditional on Y^n and V_t, therefore allowing for idiosyncratic losses.[3] Conditional on V_t, the loan-level feature Y^n, and the n-th loan defaulting at time t, the loss given default is $\ell_t^n(Y^n, V_t)$ where $\mathbb{P}[\ell_t^n(y, v) \in A] = \nu_{t,y,v}(A)$. Conditional on V_t, Y^n, $Y^{n'}$, and both loans n, n' defaulting, $\ell_t^n(Y^n, V_t)$ and $\ell_t^{n'}(Y^{n'}, V_t)$ are independent. Since the default state is absorbing, losses for the same loan cannot occur twice. More generally, one could consider "costs" or losses associated with each of the state transitions. 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. Typical loss given default models include logit models or neural networks; an overview can be found in [35].

Many models commonly used to describe loan delinquency and prepayment fall under the model class (1): see [32], [3], [4], [5], [8], [47], [51], and many others. The model framework (1) is popular in practice (see [33], [9], [19], [17], and [16]) 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,

---

[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 $\mathcal{Y} \subseteq \mathbb{R}^{d_Y}$.

[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.

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.

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 assumptions required for the results in the paper require only mild assumptions on the form of the function h_θ; these assumptions are satisfied by many standard models. Some examples are presented below.

Example 2.1 (**Logistic regression 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* $u_k^1, \ldots, u_k^{N_k}$*. Then,*

$$
\begin{aligned}
h_\theta(u_k^n, u_k, y, v, H) &= \frac{\exp(\theta_{u_k^n, u_k} \cdot (y, v, H))}{1 + \sum_{n'=1}^{N_k - 1} \exp(\theta_{u_k^{n'}, u_k} \cdot (y, v, H))}, \quad n < N_k, \\
h_\theta(u_k^{N_k}, u_k, y, v, H) &= 1 - \sum_{n'=1}^{N_k - 1} h_\theta(u_k^{n'}, u_k, y, v, H),
\end{aligned}
$$

where $\theta = \{\theta_{u_k^n, u_k}\}_{k=1:K, n=1:N_k-1}$.

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* $u_k^1, \ldots, u_k^{N_k}$*. Then,*

$$
\begin{aligned}
h_\theta(u_k^n, u_k, y, v, H) &= \frac{\exp(\sigma_\theta(u_k^n, u_k, y, v, H))}{1 + \sum_{n'=1}^{N_k - 1} \exp(\sigma_\theta(u_k^{n'}, u_k, y, v, H))}, \quad n < N_k, \\
h_\theta(u_k^{N_k}, u_k, y, v, H) &= 1 - \sum_{n'=1}^{N_k - 1} h_\theta(u_k^{n'}, u_k, y, v, H).
\end{aligned}
$$

The activation function σ_θ *is given by*

$$
\sigma_\theta(u_k^n, u_k, y, v, H) = W_\theta^2 q(W_\theta^1 Q^n),
$$

where $Q^n = (u_k^n, u_k, y, v, H) \in \mathbb{R}^{d_Q \times 1}$*, and* $W_\theta^1 \in \mathbb{R}^{d_h \times d_Q}$ *and* $W_\theta^2 \in \mathbb{R}^{1 \times d_h}$ *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 [29] 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* $\{R_\theta^m\}_{m=1}^M$*, with the condition that these regions are disjoint and their union covers the entire space. The transition probabilities in region* m *are* $p_\theta^m(u, u', y, v, H)$*. The transition probability* h_θ *then is*

$$
h_\theta(u, u', y, v, H) = \sum_{m=1}^M p_\theta^m(u, u', y, v, H) \mathbf{1}_{(u', y, v, H) \in R_\theta^m}.
$$

The regions $\{R_\theta^m\}_{m=1}^M$ *are chosen by sequentially splitting the space via recursive binary splitting. For each region* m*, the transition probabilities* $p_\theta^m(u, u', y, v, H)$ *could be a logistic regression or, more simply, a constant transition matrix between the different states. See [29] for more details on decision trees.*

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 $(U_t^1, Y^1), \ldots, (U_t^N, Y^N)$, which is denoted by $\mu_t^N \in B = \mathcal{P}(\mathcal{U} \times \mathcal{Y})$, where $\mathcal{P}$ is the space of probability measures. Formally,

$$
\mu_t^N = \frac{1}{N} \sum_{n=1}^N \delta_{(U_t^n, Y^n)}, \tag{2}
$$

where δ is the Dirac measure. The empirical measure $\mu_t^N(u, A) = \int_A \mu_t^N(u, dy)$ 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 $A \subset \mathbb{R}^{d_Y}$. It completely encodes, up to permutations, the pool dynamics and features. For instance, the pool-level default rate can be expressed as $\int_{\mathcal{Y}} \mu_t^N(\mathrm{d}, dy)$. Representing prepayment by an absorbing state $\mathrm{p} \in \mathcal{U}$, the pool-level prepayment rate takes the form $\int_{\mathcal{Y}} \mu_t^N(\mathrm{p}, dy)$. The process $H^N \in \mathbb{R}^K$ specifying the mean-field term in (1) can also be expressed in terms of the empirical measure: $H_t^N = \langle f^H, \mu_t^N \rangle_{\mathcal{U}\times\mathcal{Y}}$. Here and below, we use the notation $\langle f, \nu \rangle_E = \int_E f(x)\nu(dx)$.[4] For notational convenience, let $\langle f(u,y), \nu(u,dy)\rangle = \langle f(u,\cdot), \nu(u,\cdot)\rangle_{\mathcal{Y}}$. The default loss rate L_t^N at time t is:

$$L_t^N = \frac{1}{N}\sum_{n=1}^{N} \ell_t^n(Y^n, V_t)(\mathbf{1}_{U_t^n=\mathrm{d}} - \mathbf{1}_{U_{t-1}^n=\mathrm{d}}). \tag{3}$$

The cumulative loss from default up until time t is simply $\sum_{s=1}^t L_s^N$. Let the loss process over all times be $L^N = (L_t^N)_{t\in I}$. Given the empirical measure and the loss from default of a pool, analysis of securities backed by the pool is immediate. For instance, the distribution of the cashflow from a pool of loans can be analyzed, an example of which is given below.

Example 2.4 (**Cashflow from a Loan Pool**)**.** *Let $\mathcal{U} = \{c, p, d\}$, where c represents a loan which is current on its payments, p represents prepayment, and d represents default. The total cashflow rate from a pool of loans at time t is:*

$$C_t = \langle r_{t,V_t}, \mu_t^N \rangle_{\mathcal{U}\times\mathcal{Y}} + \langle b_t, \mu_t^N(p,\cdot) - \mu_{t-1}^N(p,\cdot)\rangle + (1 - L_t^N).$$

The cashflow for a loan with feature y in state u at time t, when the common factor $V_t = v$, is $r_{t,v}(u,y)$. It is comprised of both interest and principal payments. The dependence on V_t allows one to capture a variable loan interest rate tied to a benchmark rate. The outstanding balance for a loan with feature y at time t is $b_t(y)$. Upon prepayment, the borrower will pay off the entire outstanding balance. The functions $r_{t,v}(u,y)$ and $b_t(y)$ are deterministic. The recovery from default at time t is $1 - L_t^N$. The total cashflow rate is just the summation of C_t over all times $t \in I$.

3. Limiting Laws and an Efficient Monte Carlo Approximation

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. We develop an efficient Monte Carlo approximation of these distributions for large pools. 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 $\bar{\mu}_0$ in $B = \mathcal{P}(\mathcal{U}\times\mathcal{Y})$, where $\bar{\mu}_0$ is deterministic, and that the feature space $\mathcal{Y}$ is compact. Also, h_θ is continuous in its third and fifth arguments and the function f^H is continuous and bounded in its second argument.*

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 $\bar{\mu}$ satisfies:*

$$\bar{\mu}_t(u, dy) = \sum_{u'\in\mathcal{U}} h_\theta(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1})\bar{\mu}_{t-1}(u', dy), \tag{4}$$

where $\bar{H}_t = \langle f^H, \bar{\mu}_t\rangle_{\mathcal{U}\times\mathcal{Y}}$ and $\bar{H}_{t:s} = (\bar{H}_t, \ldots, \bar{H}_s)$.

The assumptions required for Theorem 3.2 are relatively mild and are satisfied by standard models. 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 credit score, LTV ratio, interest rate, and original balance). The assumption that the function f^H is bounded and continuous still allows the mean-field term H^N to keep track of the fraction of the pool in any combination of states in $\mathcal{U}$ and categories in $\mathcal{Y}$. For example, H^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 (such term loan, credit line, etc.).

It is important to note that the LLN in Theorem 3.2 is dynamic and is also a random equation. Randomness enters through the path of the common factor V: the quantities $\bar{H}_t$ and $\bar{\mu}_t$ are deterministic functions of

[4]For instance, $\langle f, \nu\rangle_{\mathcal{U}\times\mathcal{Y}} = \sum_{u\in\mathcal{U}} \int_{\mathcal{Y}} f(u,y)\nu(u,dy)$ and $\langle f(u,\cdot), \nu(u,\cdot)\rangle_{\mathcal{Y}} = \int_{\mathcal{Y}} f(u,y)\nu(u,dy)$.

$V_{0:t-1} = (V_0, \ldots, V_{t-1})$. If the transition function h_θ depends on the past pool dynamics through $H^N_{t-\tau:t-1}$, the LLN is nonlinear.

We supplement the LLN with a CLT. Define the empirical fluctuation process $\Xi^N_t = \sqrt{N}(\mu^N_t - \bar{\mu}_t)$, which takes values in $W = \prod_{u=1}^{|\mathcal{U}|} S'(\mathbb{R}^{d_Y})$, where S' is the space of tempered distributions. An element in the space of tempered distributions is the finite-order (distributional) derivative of some continuous function with polynomial growth. The space of measures is a subset of the space of tempered distributions. Convergence of Ξ^N_t must be considered in a larger space than the space of measures since Ξ^N_t can take negative values and $\int_A \Xi^N_t(u, dy)$ is not bounded. Although Ξ^N_t is not an element of $\mathcal{P}(\mathcal{U} \times \mathbb{R}^{d_Y})$, it does exist in the larger space W. Moreover, its limit $\bar{\Xi}_t$ also exists in W. The CLT satisfies an equation linearized around the nonlinear dynamics of the LLN. Randomness enters both through V and a martingale term. Like the LLN, the CLT is also dynamic.

Assumption 3.3. *$\sqrt{N}(\mu^N_0 - \bar{\mu}_0)$ converges in distribution to $\bar{\Xi}_0$ in W. In addition, the function f^H is smooth in its second argument and h_θ is smooth in its third and fifth arguments.*

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:*

$$
\begin{aligned}
\bar{\Xi}_t(u, dy) \quad &= \quad \sum_{u' \in \mathcal{U}} h_\theta(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1}) \bar{\Xi}_{t-1}(u', dy) \\
&+ \quad \sum_{u' \in \mathcal{U}} \left(\frac{\partial}{\partial H} h_\theta(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1}) \cdot \bar{E}_{t-\tau:t-1} \right) \bar{\mu}_{t-1}(u', dy) + \bar{\mathcal{M}}_t(u, dy),
\end{aligned} \tag{5}
$$

where $E^N_t = \left\langle f^H, \Xi^N_t \right\rangle_{\mathcal{U} \times \mathcal{Y}}$, $\bar{E}_t = \left\langle f^H, \bar{\Xi}_t \right\rangle_{\mathcal{U} \times \mathcal{Y}}$, and $\bar{E}_{t:s} = (\bar{E}_t, \ldots, \bar{E}_s)$. 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 additional assumptions necessary for Theorem 3.4 are mild. The main additional assumption is that h_θ and f^H are smooth in the loan feature y. Typical models for h_θ such as logistic regression, neural networks, probit, generalized linear models with a smooth link function, kernel methods with smooth kernels (e.g., Gaussian kernel), and Gaussian process regression satisfy Assumptions 3.1 and 3.3. Due to their discontinuities, decision trees do not satisfy Assumptions 3.1 and 3.3. However, if the Y^n are i.i.d., Theorems 3.2 and 3.4 will hold if h_θ is piecewise smooth in its third argument and smooth in its fifth argument, which would cover decision trees. Although we prove convergence in S', whose dual is the space of Schwartz functions, smoothness of h_θ and f^H on $\mathcal{Y}$ suffices for Theorem 3.4 to hold since $\mathcal{Y}$ is compact. Finally, the restriction on f^H is not stringent; the mean-field term H^N can still keep track of the fraction of the pool in any combination of states in $\mathcal{U}$ and categories in $\mathcal{Y}$.

Theorems 3.2 and 3.4 can be extended to the loss from default (3). To this end, define the empirical loss fluctuations $\Lambda^N_t = \sqrt{N}(L^N_t - \bar{L}_t) \in \mathbb{R}$.

Assumption 3.5. *The first and second moments of the loss given default, $g_1(t, v, y) = \int_0^1 z \nu_{t,y,v}(dz)$ and $g_2(t, v, y) = \int_0^1 z^2 \nu_{t,y,v}(dz)$, are smooth in y for each t, v.*

Proposition 3.6. *Let Assumptions 3.1, 3.3, and 3.5 hold. The loss process L^N converges in distribution to $\bar{L}$ in $[0,1]^{T+1}$ as $N \longrightarrow \infty$, where $\bar{L}$ satisfies:*

$$
\bar{L}_t = \left\langle \int_0^1 z \nu_{t,y,V_t}(dz), \bar{\mu}_t(d, dy) - \bar{\mu}_{t-1}(d, dy) \right\rangle. \tag{6}
$$

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:

$$\bar{\Lambda}_t = \left\langle \int_0^1 z\nu_{t,y,V_t}(dz), \bar{\Xi}_t(d,dy) - \bar{\Xi}_{t-1}(d,dy) \right\rangle + \bar{\mathcal{Z}}_t, \tag{7}$$

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

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

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

Assumption 3.5 requires smoothness of the first and second moments of the loss given default. Standard models of the loss given default such as neural networks (see [35] for an overview) satisfy this assumption.

Corollary 3.7. *Under Assumptions 3.1, 3.3, and 3.5, $(V, \mu^N, L^N, \Xi^N, \Lambda^N)$ converges in distribution to $(V, \bar{\mu}, \bar{L}, \bar{\Xi}, \bar{\Lambda})$ in $(\mathbb{R}^{d_V})^{T+1} \times B^{T+1} \times \mathbb{R} \times W^{T+1} \times \mathbb{R}$ as $N \longrightarrow \infty$.*

Using Corollary 3.7, the law of large numbers and central limit theorem can be combined to form an approximation for a finite pool of N loans:

$$\begin{aligned} \mu^N = &\quad \bar{\mu} + \frac{1}{\sqrt{N}}\Xi^N \overset{d}{\approx} \bar{\mu}^N = \bar{\mu} + \frac{1}{\sqrt{N}}\bar{\Xi}, \\ L^N = &\quad \bar{L} + \frac{1}{\sqrt{N}}\Lambda^N \overset{d}{\approx} \bar{L}^N = \bar{L} + \frac{1}{\sqrt{N}}\bar{\Lambda}. \end{aligned} \tag{8}$$

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 approximations (8) are conditionally Gaussian given V and can be utilized to efficiently estimate pool-level quantities such as prepayment, default, loss, and cashflow rates for large pools of loans. Section 4 below explains the computation of the approximation.

The theorems 3.2 and 3.4 prove that the distribution of the finite pool and its fluctuations converge to the distribution of the LLN and CLT. However, this does not guarantee that the statistics, such as the moments, of the distribution of the finite pool and its fluctuations converge to the statistics of the LLN and CLT.[5] In particular, $X^N \xrightarrow{d} \bar{X}$ only implies $\mathbb{E}[f(X^N)] \to \mathbb{E}[f(\bar{X})]$ for f continuous and *bounded*. In order to prove $\mathbb{E}[f(X^N)] \to \mathbb{E}[f(\bar{X})]$ for continuous f, one needs to show $f(X^N)$ is uniformly integrable. Since $\mathcal{Y}$ is assumed to be compact, $\mathcal{P}(\mathcal{Y} \times \mathcal{U})$ is also compact. Consequently, the empirical measure μ^N lies in a compact space (the space of measures on a compact space is compact). The loss L^N also lies in the compact set $[0,1]$. $f(\mu^N, L^N)$ is therefore uniformly integrable for any continuous function f and $\mathbb{E}[f(\mu^N, L^N)] \to \mathbb{E}[f(\bar{\mu}, \bar{L})]$ for any continuous function f. The main challenge resides in showing that $f(\Xi^N, \Lambda^N)$ is also uniformly integrable. Unlike μ^N, Ξ^N and Λ^N are not restricted to compact sets.

For instance, convergence in distribution is not sufficient for the approximation (8) to hold for the second moment of the loss at time t:

$$\mathbb{E}[(L_t^N)^2] = \mathbb{E}[\bar{L}_t^2] + \frac{2}{\sqrt{N}}\mathbb{E}[\bar{L}_t\Lambda_t^N] + \frac{1}{N}\mathbb{E}[(\Lambda_t^N)^2]. \tag{9}$$

In order to justify the approximation $\mathbb{E}[(L_t^N)^2] \approx \mathbb{E}[(\bar{L}_t^N)^2]$, one needs the convergence results $\mathbb{E}[\bar{L}_t\Lambda_t^N] \to \mathbb{E}[\bar{L}_t\bar{\Lambda}_t]$ and $\mathbb{E}[(\Lambda_t^N)^2] \to \mathbb{E}[(\bar{\Lambda}_t)^2]$ as $N \to \infty$. Due to $|\bar{L}_t|$ being bounded, it is sufficient to show that Λ_t^N and $(\Lambda_t^N)^2$ are uniformly integrable.

Assumption 3.8. *For functions of the form $f(\Xi^N, \Lambda^N) = g(\sum_{t=1}^T \sum_{u \in \mathcal{U}} \langle \phi_{t,u}(y), \Xi_t^N(u,dy)\rangle + \sum_{t=1}^T c_t \Lambda_t^N)$ where $c_t \in \mathbb{R}$, g is continuous and polynomially bounded,[6] and the functions $\phi_{t,u}$ are continuous. In addition, $V_{0:T} = (V_0, V_1, \ldots, V_T)$ takes values in a compact set $\mathcal{V}$, and for any continuous ϕ, there are positive constants $K_{1,0,u}$ and $K_{2,0,u}$ such that*

$$\mathbb{P}\left[(\sup_{y \in \mathcal{Y}} |\phi(y)|)^{-1} \big| \langle \phi(y), \Xi_0^N(u,dy) \rangle \big| > \alpha \middle| V_{0:T} \right] \leq K_{1,0,u} \exp\big(-K_{2,0,u}\alpha^2\big). \tag{10}$$

[5] For instance, consider the random variable X^N where $X^N = 2^N$ with probability $\frac{1}{N}$ and is otherwise 0. $X^N \xrightarrow{p} 0$ but $\mathbb{E}[X^N] \to \infty$ as $N \to \infty$.

[6] I.e., there is some positive constant C_0 such that $|g(x)| \leq |x|^k$ for $|x| \geq C_0$ and $k \in \mathbb{N}$.

Theorem 3.9. *Under Assumptions 3.1, 3.3, 3.5, and 3.8, $f(\Xi^N, \Lambda^N)$ is uniformly integrable.*

The class of functions covered by Assumption 3.8 can represent any function g of the default, loss, prepayment, or cashflow rates where g does not have faster than polynomial growth in its tails. We are not aware of any applications with exponential growth in the tails, so this class of functions is quite general. A sufficient condition for the bound (10) is that the loan features Y^n are independent; then, by the Azuma-Hoeffding inequality, the bound is satisfied.

The ideas and results in this paper, which are developed for discrete time, can be extended to a continuous-time framework where events are governed by counting processes with intensities. 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 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 the mortgage models studied by [13] and [36].

4. Computing the Approximation

The approximation (8) can be computed by Monte Carlo simulation. 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. 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. We highlight that the approximate loss $\bar{L}^N$ 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 approximate loss $\bar{L}^N$ can be simulated (at no extra cost) on top of the simulated paths of $(\bar{\mu}, \bar{\Xi})$.

4.1. Algorithm.

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 $\bigcup_{k=1}^K c_k = \mathcal{Y}$ 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 $\Psi(u, dy) = \sum_{k=1}^K \bar{\mu}_{t=0}(u, c_k)\delta_{y_k}$.[7]
- 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:
  - Calculate $\bar{\mu}^{\Psi,l}_t(u, y_k)$ for all $y_k \in \mathcal{R}$ at times $t = 1, \ldots, T$.
  - Calculate the (closed-form) covariance of $\bar{\Xi}^{\Psi,l}_t$ across the grid points $y_1, \ldots, y_K$ at $t = 1, \ldots, T$.[8]
  - For each V^l and any function f, a numerical approximation can be made:

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

where $\mu_t^{N,l}$ is the empirical measure μ_t^N for the path V^l, $m^l = \sum_{k=1}^K \sum_{u\in\mathcal{U}} f(u, y_k)\bar{\mu}_t^{\Psi,l}(u, y_k)$, and $\Sigma^l = \frac{1}{N}\mathrm{Var}[\sum_{k=1}^K \sum_{u\in\mathcal{U}} f(u, y_k)\bar{\Xi}_t^{\Psi,l}(u, y_k)]$ (which can be calculated in closed-form).

---

[7] 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.

[8] 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*. Thus, one can march forward through time, calculating the (closed-form) covariance of $\bar{\Xi}_t^{\Psi,l}$ across the grid points $y_1, \ldots, y_K$ at each time $t = 1, \ldots, T$.

- Collecting the conditional distributions for each V^l, the density of $\left\langle f, \mu_t^N \right\rangle_{\mathcal{U}\times\mathcal{Y}}$ can be directly computed as $\frac{1}{L}\sum_{l=1}^{L}\phi(z, m^l, \Sigma^l)$ where $\phi(\cdot, m, \Sigma)$ is the density of Gaussian random variable with mean m and covariance Σ.

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 $\bar{\Xi}_t^{\Psi}$ for each V^l. Such simulation is straightforward since $\bar{\Xi}_t$ is conditionally Gaussian given V. The simulation scheme above has the advantage of lower variance per sample with the same bias.

The convergence for the numerical error of the simulation scheme is analyzed in Appendix B. We show that as the sizes of the computational cells tend to zero (i.e., the number of grid points tends to infinity), the error also converges to zero. Moreover, a specific convergence rate is provided which can guide the choice of the location of grid points, number of grid points, and number of Monte Carlo trials. Given a fixed computational budget, there is a tradeoff between more trials (reduction of variance) and more grid points (reduction of bias). In Appendix B, we derive a closed-form formula for the optimal allocation between the number of grid points and the number of trials provided a fixed budget.

The computational performance of the LLN and CLT can be further improved by the use of a non-uniform grid for discretizing $\mathbb{R}^{d_Y}$. Using a non-uniform grid, more points would be placed where $\bar{\mu}_0(u, dy)$ is large and less points would be placed where $\bar{\mu}_0(u, dy)$ is small. Appendix C provides an example of a particular non-uniform grid well-adapted to our problem. The proposed grid 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. Appendix D 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. Another choice, implemented in Section 5.7, is a non-uniform sparse grid using k-means clustering.

4.2. **Efficient Low-dimensional Approximation.** There is a potential drawback to the law of large numbers (4) and central limit theorem (28). 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 (4) which converts the LLN into a low-dimensional problem. 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:

$$\bar{\mu}_t(u, dw) \quad = \quad \sum_{u' \in \mathcal{U}} g_\theta(u, u', w, V_{t-1}, \bar{H}_{t-\tau:t-1})\bar{\mu}_{t-1}(u', dw), \qquad w \in \mathbb{R}^{d_W}. \tag{11}$$

The same coordinate transformation can be used to arrive at a low-dimensional CLT in terms of w. No matter how large the original dimension d_Y is, the transformation reduces the dimension d_Y of the LLN and CLT to a constant dimension d_W, paving the way for tractable computations.

An example of a function h_θ which satisfies the necessary requirement is any generalized linear model (GLM), such as logistic regression (see Example 2.1). GLMs 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 (11) and its computational cost remains the same. 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 their broad applicability and computational tractability, the low-dimensional LLN and CLT are very useful in practice.

The combination of the low-dimensional transformation and the non-uniform grid schemes mentioned in the previous section make simulating the pool loss and prepayment levels via the LLN and CLT very computationally tractable. 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 (8) for high-dimensions (without any transformation) via sparse non-uniform grids. We describe and numerically implement this latter approach in Section 5.7, which also strongly outperforms brute-force simulation of the actual pool.

5. 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 approximation has very high accuracy even for small pools in the hundreds of mortgages. Its computational cost is typically orders of magnitude lower than brute-force Monte Carlo simulation of the actual pool.

5.1. **Models and Data.** We test the performance of the approximation using several logistic regression and neural network models of the transition function h_θ in (1); see Example 2.1 and 2.2, respectively. We consider three states: current, prepayment, and default. The loss given default is normalized to one. We fit the model parameters θ to the data via maximum likelihood estimation; see Appendix E for details.[9] Our data is comprised of two loan-level data sets. The first is a subprime data set consisting of over 10 million mortgages during the time period $1995-2013$, obtained from the Trust Company of the West. The mortgages are spread across the entire US, covering over $36,000$ different zip codes. The feature 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. The second data set, obtained from Freddie Mac, contains agency mortgages over the time period $1999-2014$ and consists of 16 million mortgages. The feature data includes FICO score, first time homebuyer indicator, type of mortgage, maturity date, number of units, occupancy status, combined loan-to-value, original debt-to-income ratio, LTV ratio, initial interest rate, prepayment penalty mortgage 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.

5.2. **Performance of LLN and CLT.** First, we demonstrate the accuracy of the approximate distribution by comparing it with the actual distribution for the pool. The actual (or "true") distribution is found via brute-force Monte Carlo simulation. The pool is drawn at random from the subprime data set.

Here and in Sections 5.3 through 5.6, we use a logistic regression model for h_θ. The parameters θ are fitted using the entire subprime data set prior to January 1, 2012. 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. Unless otherwise stated, we do not include a mean field term H^N. A mean-field term is included in Section 5.4 where we study the accuracy of the approximation for the moments of the loss.

We perform $25,000$ Monte Carlo trials for both the brute-force simulation of the actual pool as well as the simulation of the approximation. 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.[10] We model both default and prepayment, and hence $d_W = 2$ for the efficient Monte Carlo approximation. Our implementation of the approximation uses non-uniform grids; see Appendix C.

[9] Parameter estimates are available upon request.

[10] 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.

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 LLN is very accurate. 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 $N = 500, 1,000, 2,500$, and $5,000$. 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. In the next section we assess the accuracy in the tails more thoroughly.

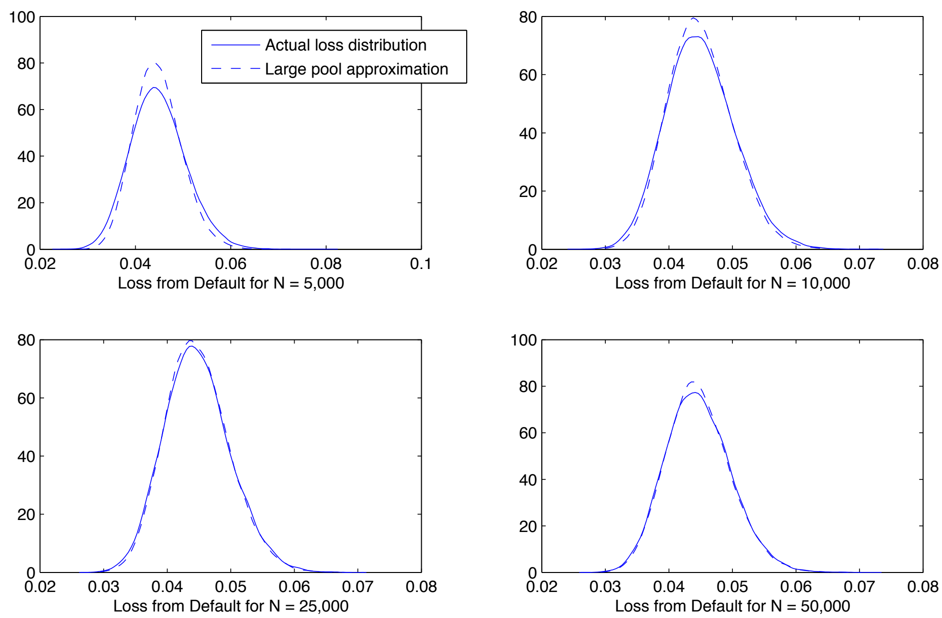

Figure 1. Comparison of actual loss distribution with LLN loss distribution (does not include CLT in approximation). Loss reported as fraction of pool which defaulted. The horizon is 12 months.

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

$$\frac{\text{Cost of LLN}}{\text{Cost of Simulation of Actual Pool}} = \frac{N_g}{N}, \tag{12}$$

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 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.

5.3. **Numerical Performance across Actual Deals.** The subprime data set covers over 6,000 deals.[11] As mentioned earlier, for each mortgage, a deal ID is available. One can therefore reconstruct the actual

[11] "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.

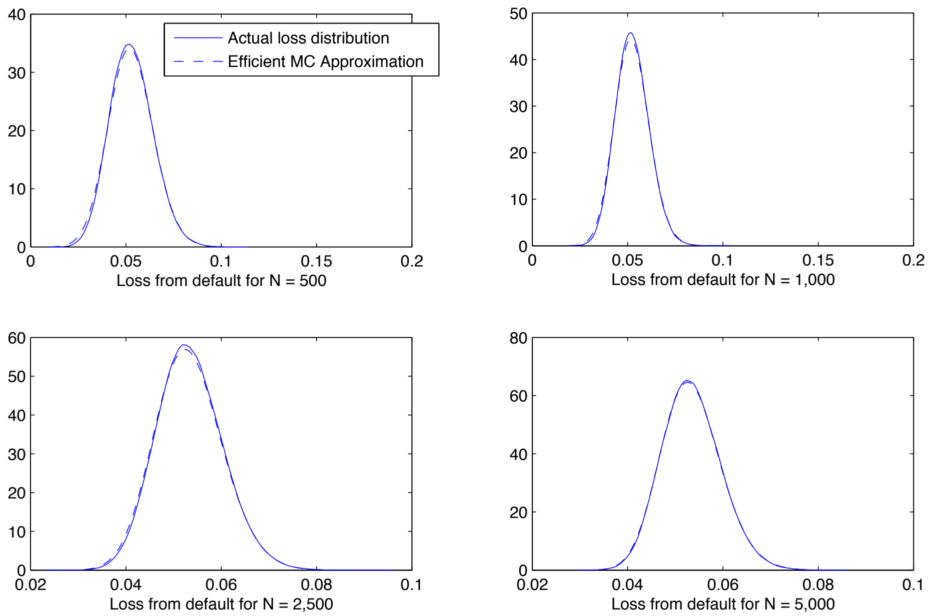

FIGURE 2. Comparison of actual distribution with approximate distribution (using both LLN and CLT). Loss reported as fraction of pool which defaulted. The horizon is 12 months.

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.

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.

Figure 3 compares the approximate distribution from the efficient Monte Carlo approximation with the true distribution (found via brute-force Monte Carlo simulation) 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 approximation, a set of deals is selected at random from the data set and the 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 4 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 0.22%. The average error just using the LLN (no

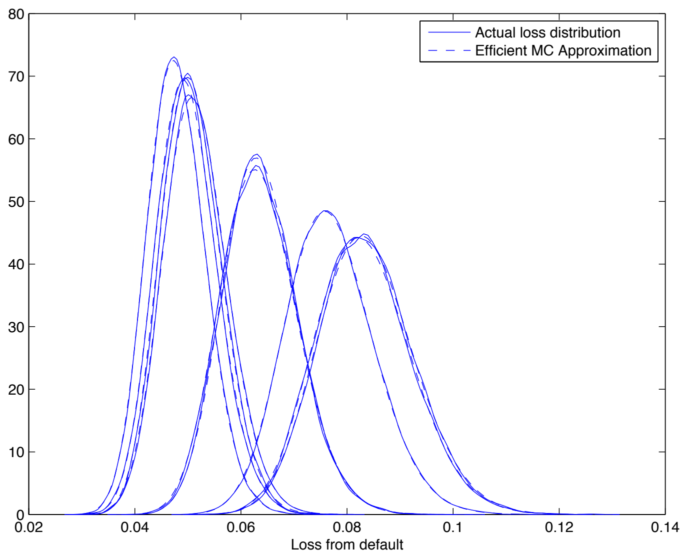

FIGURE 3. Comparison of actual distribution with approximate distribution for ten actual deals with $5,000 < N < 10,000$. Loss reported as fraction of pool which defaulted. The horizon is 12 months.

CLT) is 2.25%. For deals with more than $10,000$ mortgages, the average error for the approximation (LLN and CLT) is is 0.18%. The average error just using the LLN is 1.25%.

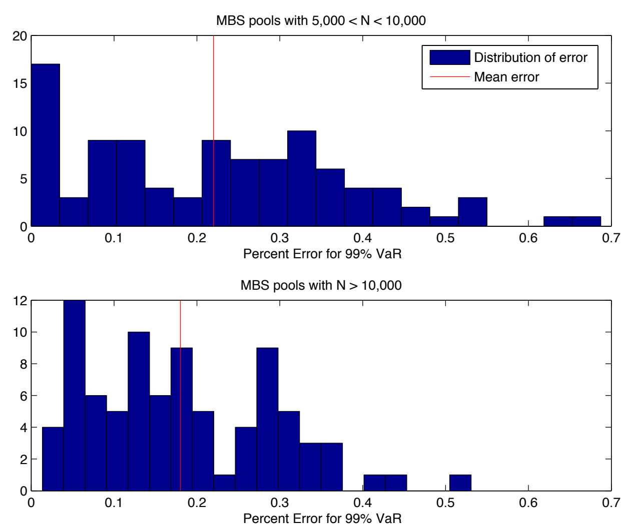

FIGURE 4. Distribution across deals of error for 99% VaR from the efficient Monte Carlo approximation. The time horizon is 12 months.

5.4. **Accuracy of Approximation for Moments.** We now study the accuracy of the Monte Carlo approximation for the moments for pools of subprime mortgages. In addition to the loan-level and common factors used in the previous sections, a mean-field term is also included. The mean-field term is the pool-level default rate from the previous month. Table 2 reports the percent error of the Monte Carlo approximation for the moments of the pool loss for different sizes N at a $T = 12$ month time horizon. The actual values for the moments are calculated via brute-force Monte Carlo simulation with 50,000 trials. 50,000 Monte Carlo trials are also used for the approximation of the moments. There is lower comparative accuracy for higher-order moments which represent the non-Gaussian characteristics of the finite pool. These non-Gaussian characteristics fade as N becomes large.

N	First Moment	Second Moment	Third Moment	Fourth Moment	Fifth Moment
500	.69	1.3	1.9	2.5	3.2
1,000	.44	.87	1.3	1.7	2.2
5,000	.09	.19	.28	.36	.45
10,000	.07	.14	.22	.30	.37

TABLE 2. Percent error for Monte Carlo approximation for moments of the loss.

5.5. **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, some 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 logistic regression model only including prepayment to the agency mortgage data. 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 statistical area (MSA); there are 430 metropolitan statistical areas in the data set. Therefore, in total, $d_Y = 452$.[12] We emphasize that, even though $d_Y = 452$, d_W only equals 1. The systematic factors are the national unemployment rate and the national mortgage rate, as above.

Figure 5 compares the approximation with the actual prepayment distribution for a randomly drawn pool of agency mortgages for a time horizon of 12 months. $50,000$ Monte Carlo simulations are used.

5.6. **Precomputation for Financial Institutions.** In Appendix D, we propose 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 this approach, 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 illustrate this approach using the parameter fits from Section 5.5. 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 5.5). $50,000$ 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.[13] This saves computational time since it allows one to simulate on a very sparse grid but still ultimately achieve a very accurate solution on a fine grid. Figure 6 is a histogram of the percent error over the 400 pools for the 99% VaR from the efficient

[12] The parameter estimates are available upon request.

[13] 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.

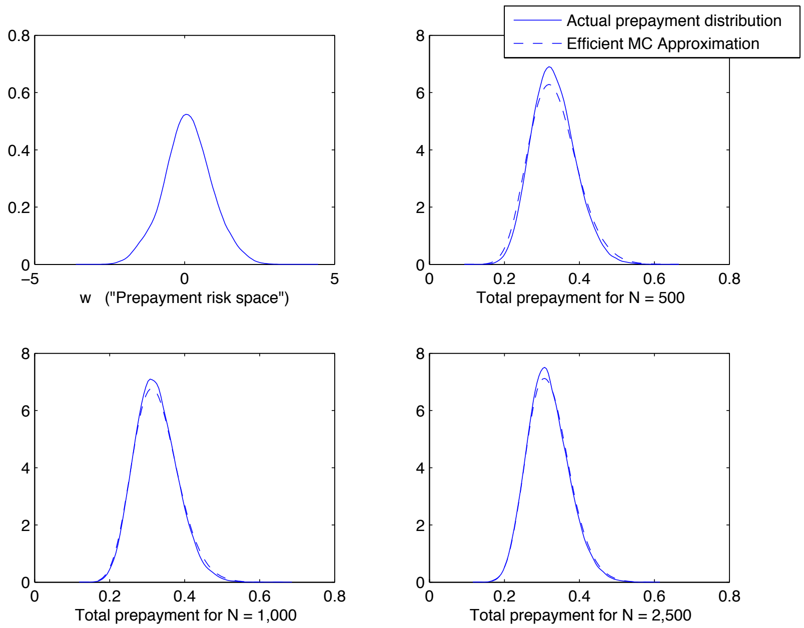

FIGURE 5. Top right plot and bottom plots compare the actual prepayment distribution with the efficient Monte Carlo approximation for $N = 500$, $N = 1,000$, and $N = 2,500$. The total prepayment is reported as the fraction of the pool which prepays. The time horizon is 12 months. The top left plot shows the distribution of the pool in the "prepayment risk space".

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.

5.7. **Efficient 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 V. Although that low-dimensional transformation still covers a wide range of models, for completeness, we now present some methods for the computation of the LLN and CLT 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 FICO score, LTV ratio, original balance, and 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 evaluation 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

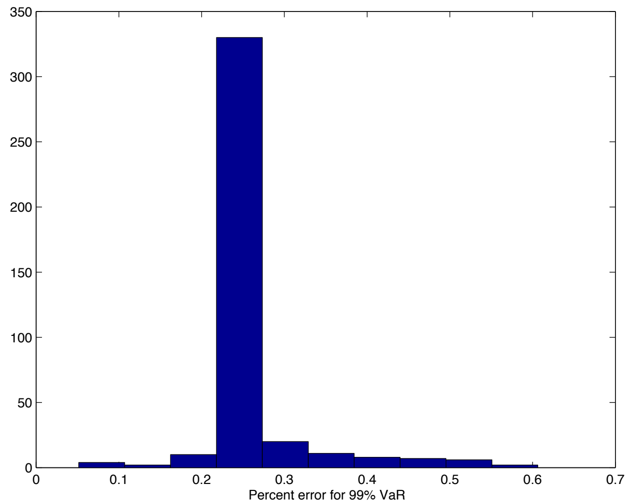

FIGURE 6. Distribution across deals of error for 99% VaR from the efficient Monte Carlo approximation using pre-simulation. The time horizon is 12 months.

third layer takes as inputs the output of the d_W neurons from the second layer as well as V. 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 7 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. The time horizon is 12 months.

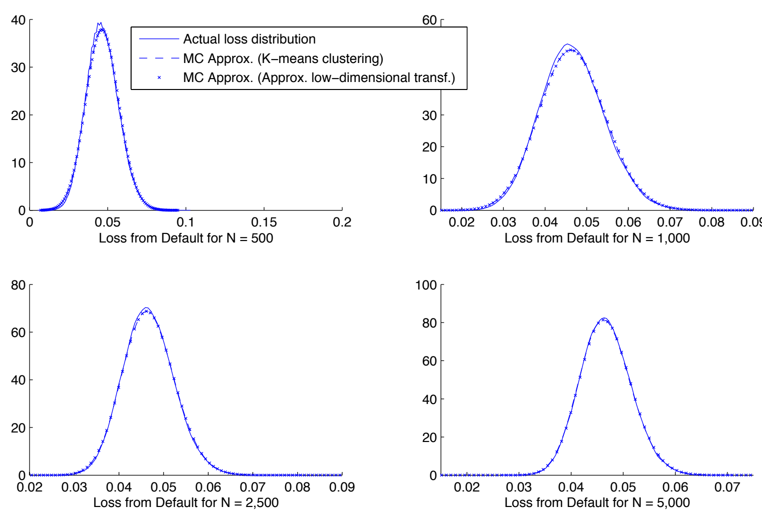

FIGURE 7. Comparison of actual loss distribution with Monte Carlo approximation using sparse grids when h_θ is a neural network.

6. 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. 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.

Appendix A. Proofs for LLN, CLT, and Uniform Integrability

This appendix proves the LLN, CLT, and uniform integrability. The LLN for the empirical measure is proven in A.1. The CLT for the empirical measure is proven in A.2. The LLN and CLT for the loss from default are proven in A.3. Uniform integrability is proven in A.4. Existence and uniqueness for the LLN and CLT equations for the empirical measure are proven in A.5.

The proofs of the limit theorems will make use of an auxiliary system of particles $(U^{1,v}, \ldots, U^{N,v})$ which are indexed by $v = (v_0, \ldots, v_T) \in \mathbb{R}^{d_V \times T+1}$:

$$\begin{aligned} \mathbb{P}[U_t^{n,v} = u | \mathcal{F}_{t-1}^v] &= h_\theta(u, U_{t-1}^{n,v}, Y^n, v_{t-1}, H_{t-\tau:t-1}^{N,v}), \quad t = 1, \ldots, T, \\ U_0^{n,v} &= U_0^n, \end{aligned} \tag{13}$$

where $H_t^{N,v} = \frac{1}{N}\sum_{n=1}^N f^H(U_t^{n,v}, Y^n)$, and $\mathcal{F}_t^v$ is the sigma-field generated by $(U_t^{1,v}, \ldots, U_t^{N,v}, Y^1, \ldots, Y^N)$. Similarly, the loss given default for the n-th loan is $\ell_t^{n,v}(Y^n, v_t)$. The system (13) is simply the original system (1) given a realization v of V. In particular, $U_t^{1,V}, \ldots, U_t^{N,V}$ has the same distribution as $U_t^1, \ldots, U_t^N$. Define:

$$\mu_t^{N,v} = \frac{1}{N}\sum_{n=1}^N \delta_{U_t^{n,v}, Y^n}. \tag{14}$$

The empirical measures $\mu_t^{N,V}$ and μ_t^N have the same distribution. In addition, $\mu_0^{N,v} = \mu_0^N$ for any v. We prove limiting laws for $\mu^{N,v} = (\mu_0^{N,v}, \mu_1^{N,v}, \ldots, \mu_T^{N,v})$ for any $v = (v_0, \ldots, v_T) \in \mathbb{R}^{d_V \times T+1}$, which in turn will yield the desired limiting laws for $\mu^N = (\mu_0^N, \mu_1^N, \ldots, \mu_T^N)$.

A.1. **Law of Large Numbers (Theorem 3.2).** 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:

$$\mathbb{E}[\Phi(X^N)] \to \mathbb{E}[\Phi(\bar{X})], \tag{15}$$

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 $\phi_1 \in C(\mathbb{R}^M), f_m \in C(E)$. (Recall that any function on a discrete metric space is continuous.) Then, $\mathcal{S}$ separates $\mathcal{P}(\mathcal{U} \times \mathcal{Y})$ and is a convergence determining class for $\mathcal{P}(\mathcal{P}(E))$ (see [48] or [23]).

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

We prove several facts, culminating in the law of large numbers $\mu^N \xrightarrow{d} \bar{\mu}$. The proof uses an induction argument and the convergence determining class $\mathcal{S}$.

Lemma A.1. *Define $\mathcal{M}^{1,N,v}(u)$ to be the martingale*

$$\mathcal{M}_t^{1,N,v}(u) = \frac{1}{N}\sum_{n=1}^{N}\phi(Y^n)(\mathbf{1}_{U_t^{n,v}=u} - h_\theta(u, U_{t-1}^{n,v}, Y^n, v_{t-1}, H_{t-\tau:t-1}^{N,v})), \tag{16}$$

where $\phi \in C(\mathbb{R}^{d_Y})$. Then, for each $u \in \mathcal{U}$ and any $v \in \mathbb{R}^{d_V \times T+1}$, we have $\mathcal{M}_t^{1,N,v}(u) \xrightarrow{p} 0$.

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

$$\begin{aligned}
\mathrm{Var}[\mathcal{M}_t^{1,N,v}(u)] &= \mathbb{E}[(\frac{1}{N}\sum_{n=1}^{N}\phi(Y^n)(\mathbf{1}_{U_t^{n,v}=u} - h_\theta(u, U_{t-1}^{n,v}, Y^n, v_{t-1}, H_{t-\tau:t-1}^{N,v}))])^2] \\
&= \frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[(\phi(Y^n)(\mathbf{1}_{U_t^{n,v}=u} - h_\theta(u, U_{t-1}^{n,v}, Y^n, v_{t-1}, H_{t-\tau:t-1}^{N,v})))^2] \\
&\le \frac{1}{N^2}\sum_{n=1}^{N} C \underset{N\to\infty}{\to} 0. \qquad (17)
\end{aligned}$$

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

Of course, if $\mathcal{M}_t^{1,N,v}(u) \xrightarrow{p} 0$ for each t, u, $\mathcal{M}^{1,N,v} = \{\mathcal{M}_t^{1,N,v}(u)\}_{u=1:|\mathcal{U}|,t=1:T} \xrightarrow{p} \{0\}_{u=1:|\mathcal{U}|,t=1:T}$.

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

$$\left\langle \phi, \mu_t^{N,v}(u, dy)\right\rangle = \sum_{u' \in \mathcal{U}}\left\langle \phi(y)h_\theta(u, u', y, V_{t-1}, H_{t-\tau:t-1}^{N,v}), \mu_{t-1}^{N,v}(u', dy)\right\rangle + \mathcal{M}_t^{1,N,v}(u). \tag{18}$$

Define $\bar{\mu}_t^v = \mathbb{E}[\bar{\mu}_t|V = v]$ and $\bar{H}_t^v = \mathbb{E}[\bar{H}_t|V = v]$. Note that $\bar{\mu}_t^V = \bar{\mu}_t$ and $\bar{H}_t^V = \bar{H}_t$ since $\bar{\mu}_t^v$ and $\bar{H}_t^v$ are simply $\bar{\mu}_t$ and $\bar{H}_t$ conditional on $V = v$:

$$\begin{aligned}
\bar{\mu}_t^v(u, dy) &= \sum_{u' \in \mathcal{U}} h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v)\bar{\mu}_{t-1}^v(u', dy), \\
\bar{H}_t^v &= \left\langle f^H, \bar{\mu}_t^v\right\rangle_{\mathcal{U}\times\mathcal{Y}} \qquad (19) \\
\bar{\mu}_0^v &= \bar{\mu}_0.
\end{aligned}$$

We will now use an induction argument to prove the law of large numbers for $\mu^{N,v} = (\mu_0^{N,v}, \dots, \mu_T^{N,v})$. Assuming $\mu_{0:t-1}^{N,v} \xrightarrow{p} \bar{\mu}_{0:t-1}^v$, we have that for any v:

$$\lim_{N\to\infty}\left\langle \phi, \mu_t^{N,v}(u, dy)\right\rangle = \sum_{u' \in \mathcal{U}}\left\langle \phi(y)h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v), \bar{\mu}_{t-1}^v(u', dy)\right\rangle = \left\langle \phi, \bar{\mu}_t^v(u, dy)\right\rangle. \tag{20}$$

The result (20) is a consequence of the following facts. First, $\mathcal{M}^{1,N,v} \xrightarrow{p} 0$ by Lemma A.1. Secondly, since $\mu_{0:t-1}^{N,v} \xrightarrow{p} \bar{\mu}_{0:t-1}^v$, $H_{t-\tau:t-1}^{N,v} \xrightarrow{p} \bar{H}_{t-\tau:t-1}^v$ (f^H is continuous by assumption). Thirdly,

$$\left\langle \phi(y)h_\theta(u, u', y, v_{t-1}, H_{t-\tau:t-1}^{N,v}), \mu_{t-1}^{N,v}(u', dy)\right\rangle \xrightarrow{p} \left\langle \phi(y)h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v), \bar{\mu}_{t-1}^v(u', dy)\right\rangle. \tag{21}$$

This last fact follows from:

$$\begin{aligned}
&\left\langle \phi(y)h_\theta(u, u', y, v_{t-t}, H_{t-\tau:t-1}^{N,v}), \mu_{t-1}^{N,v}(u', dy)\right\rangle = \left\langle \phi(y)h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v), \mu_{t-1}^{N,v}(u', dy)\right\rangle \\
(22)\quad + &\underbrace{\left\langle \phi(y)[h_\theta(u, u', y, v_{t-1}, H_{t-\tau:t-1}^{N,v}) - h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v)], \mu_{t-1}^{N,v}(u', dy)\right\rangle}_{(*)}.
\end{aligned}$$

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

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

since $H^{N,v}_{t-\tau:t-1} \xrightarrow{p} \bar{H}^{v}_{t-\tau:t-1}$ and the desired result (21) holds.

By Lemma A.8, $\bar{\mu}^v_t$ from (4) is the unique measure which satisfies (20). Let $\Phi \in \mathcal{S}$ where $\mathcal{S}$ is the convergence determining class specified earlier. Then, for any $\Phi \in \mathcal{S}$,

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

The above result follows from equation (20), the continuous mapping theorem, $\mathcal{P}(\mathcal{U}\times\mathcal{Y})$ being compact (and thus any continuous function of $\mu^N_t \in \mathcal{P}(\mathcal{U}\times\mathcal{Y})$ is bounded), and the dominated convergence theorem. Therefore, for any v, $\mu^{N,v}_t \xrightarrow{d} \bar{\mu}^v_t$ and, since $\bar{\mu}^v_t$ is deterministic, $\mu^{N,v}_t$ also converges in probability to $\bar{\mu}^v_t$. This also means that $\mu^{N,v}_{0:t} \xrightarrow{p} \bar{\mu}^v_{0:t}$ for any v.

Note that in Assumption 3.1, $\mu^{N,v}_0$ converges in distribution to $\bar{\mu}_0$, where $\bar{\mu}_0$ is deterministic. By Assumption 3.1 and induction, it follows that $\mu^{N,v}$ converges in probability to $\bar{\mu}^v$ for any v. Since the convergence in probability holds for any v, we certainly have that μ^N converges in distribution to $\bar{\mu}$, since:

$$\begin{aligned}
\lim_{N\to\infty} \mathbb{E}[f(\mu^N)] &= \lim_{N\to\infty} \mathbb{E}[f(\mu^{N,V})] = \lim_{N\to\infty} \mathbb{E}[\mathbb{E}[f(\mu^{N,V})|V]] \\
&= \mathbb{E}[\lim_{N\to\infty} \mathbb{E}[f(\mu^{N,V})|V]] = \mathbb{E}[\mathbb{E}[f(\bar{\mu}^V)|V]] \\
&= \mathbb{E}[f(\bar{\mu}^V)] = \mathbb{E}[f(\bar{\mu})], \qquad (23)
\end{aligned}$$

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 $\mu^{N,v}$. □

A.2. **Central Limit Theorem (Theorem 3.4).** Let S be the space of Schwartz functions, and let its topological dual $S' = S'(\mathbb{R}^{d_Y})$ be the space of tempered distributions on $\mathbb{R}^{d_Y}$ (equipped with the weak topology). S is a nuclear Fréchet space and therefore reflexive (see [44]). A random variable $X^N \in S'$ itself has a probability measure $P^N \in \mathcal{P}(S')$. The measure P^N for the random variable X^N weakly converges to the measure P of a random variable $\bar{X}$ if:

$$\mathbb{E}[\exp\left(i\alpha\phi(X^N)\right)] \to \mathbb{E}[\exp\left(i\alpha\phi(\bar{X})\right)], \qquad (24)$$

for every $\phi \in S$.[14] Of course, (24) is implied by:

$$\phi(X^N) \xrightarrow{d} \phi(\bar{X}), \qquad (25)$$

for every $\phi \in S$. S, the space of Schwartz functions, is defined as:

$$\begin{aligned}
S(\mathbb{R}^{d_Y}) &= \{\phi \in C^\infty(\mathbb{R}^{d_Y}) : ||\phi||_{\alpha,\beta} < \infty \ \ \forall \ \ \alpha,\beta \in \mathbb{N}\}, \\
||\phi||_{\alpha,\beta} &= \sup_{y\in\mathbb{R}^{d_Y}} |y^\alpha D^\beta \phi(y)|.
\end{aligned}$$

The requirement that ϕ be rapidly decreasing for large y does not have an impact on our setting since $\mathcal{Y}$ is assumed to be compact.

[14]This result was originally proven for both the space of tempered distributions S' and the space of distributions D' in [21]. The result was also independently proven by [50] (Proposition 3.3 and corresponding Supplementary Comments on pg. 247), [41], and [27]. Finally, such results have been extended to continuous-time processes taking values in S' and D', see [38] and [22]. See [30] for a similar result for separable Banach spaces.

We wish to show that $\Xi_t^N(u,dy) \in S'(\mathbb{R}^{d_Y})$ for each u. To prove convergence in distribution of $\Xi_t^N(u,dy)$ for each u, one needs to show (24) or (25) for each $\phi \in S$:

$$\left\langle \phi(y), \Xi^N(u,dy) \right\rangle \xrightarrow{d} \left\langle \phi(y), \bar{\Xi}(u,dy) \right\rangle, \tag{26}$$

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 μ_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 (24) or (25) also of course covers the case of $\Xi^N \in \prod_{i=1}^{T\times|\mathcal{U}|} S'$, where it is sufficient to show (24) 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 $\left(\phi_{t,u}\right)_{t\in I, u\in\mathcal{U}} \in \prod_{i=1}^{T\times|\mathcal{U}|} S$.

Define:

$$\Xi_t^{N,v} = \sqrt{N}(\mu_t^{N,v} - \bar{\mu}_t^v), \tag{27}$$

and note that $\Xi_t^N = \Xi_t^{N,V}$. We also define $\bar{\Xi}_t^v$:

$$\begin{aligned}
\bar{\Xi}_t^v(u,dy) &= \sum_{u'\in\mathcal{U}} h_\theta(u,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1})\bar{\Xi}^v_{t-1}(u',dy) \\
&+ \sum_{u'\in\mathcal{U}} \big(\frac{\partial}{\partial H} h_\theta(u,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1}) \cdot \bar{E}^v_{t-\tau:t-1}\big)\bar{\mu}^v_{t-1}(u',dy) + \bar{\mathcal{M}}^v_t(u,dy), \\
\bar{\Xi}_0^v &= \bar{\Xi}_0.
\end{aligned} \tag{28}$$

where $E_t^{N,v} = \left\langle f^H, \Xi_t^{N,v}\right\rangle_{\mathcal{U}\times\mathcal{Y}}$, $\bar{E}_t^v = \left\langle f^H, \bar{\Xi}_t^v\right\rangle_{\mathcal{U}\times\mathcal{Y}}$, and $\bar{E}^v_{t-\tau:t-1} = (\bar{E}^v_{t-\tau}, \ldots, \bar{E}^v_{t-1})$. $\bar{\mathcal{M}}^v(u,dy)$ is a Gaussian process with zero mean and covariance:

$$\mathrm{Cov}\Big[\bar{\mathcal{M}}^v_t(u_1,dy), \bar{\mathcal{M}}^v_t(u_2,dy)\Big] = -\sum_{u'\in\mathcal{U}} h_\theta(u_1,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1})h_\theta(u_2,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1})\bar{\mu}^v_{t-1}(u',dy),$$

$$\mathrm{Var}\Big[\bar{\mathcal{M}}^v_t(u,dy)\Big] = \sum_{u'\in\mathcal{U}} h_\theta(u,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1})(1 - h_\theta(u,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1})\bar{\mu}^v_{t-1}(u',dy),$$

where $u_1 \neq u_2$. $\bar{\mathcal{M}}^v_{t_2}$ is independent of $\bar{\mathcal{M}}^v_{t_1}$ for $t_2 \neq t_1$ and $\bar{\mathcal{M}}^v_t(u,dy_1)$ is independent of $\bar{\mathcal{M}}^v_t(u',dy_2)$ for $y_1 \neq y_2$. $\bar{\Xi}^v_t$ is $\bar{\Xi}_t$ conditional on $V = v$; $\bar{\Xi}^V_t = \bar{\Xi}_t$.

Lemma A.2. *For every u and N, $\Xi_t^{N,v}(u,dy) \in S'(\mathbb{R}^{d_Y})$ for any v.*

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

$$\langle \phi, X\rangle \leq C_k||\phi||_k, \tag{29}$$

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

$$||\phi||_k = \max_{|\alpha|+|\beta|\leq k} \sup_{y\in\mathbb{R}^{d_Y}} |y^\alpha D^\beta \phi(y)|. \tag{30}$$

Then, we certainly have that:

$$\left\langle \phi, \Xi_t^N\right\rangle = \sqrt{N}(\left\langle \phi, \mu_t^{N,v}\right\rangle - \langle \phi, \bar{\mu}_t^v\rangle) \leq 2\sqrt{N}\sup_{y\in\mathcal{Y}}|\phi(y)| \leq 2\sqrt{N}||\phi||_0. \tag{31}$$

This completes the proof. □

Lemma A.3. *For any $\phi \in S$, α, and v, the martingale difference array*[15]

$$Z_t^{N,n,v}(\alpha) = \frac{1}{\sqrt{N}}\sum_{u\in\mathcal{U}} \alpha_u[\phi(Y^n)(\mathbf{1}_{U_t^{n,v}=u} - h_\theta(u, U_{t-1}^{n,v}, Y^n, v_{t-1}, H^{N,v}_{t-\tau:t-1})]$$

satisfies the following properties for every $\alpha \in \mathbb{R}^{|\mathcal{U}|}$:

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

[15] See [37] for details on martingale difference arrays.

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

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

Proof. First, recognize that:

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

This certainly implies properties (i) and (ii). Finally, for each v,

$$\begin{aligned}
\sum_{n=1}^{N}(Z_t^{N,n,v}(\alpha))^2 \;&=\; \frac{1}{N}\sum_{n=1}^{N}[\sum_{u\in\mathcal{U}}\alpha_u\phi_u(Y^n)(\mathbf{1}_{U_t^{n,v}=u} - h_\theta(u,U_{t-1}^{n,v},Y^n,v_{t-1},H_{t-\tau:t-1}^{N,v})]^2\\
&=\; \frac{1}{N}\sum_{n=1}^{N}\sum_{u',u\in\mathcal{U}}\alpha_u\alpha_{u'}\phi_{u'}(Y^n)\phi_u(Y^n)[\mathbf{1}_{U_t^{n,v}=u}\mathbf{1}_{U_t^{n,v}=u'}\\
&+\; h_\theta(u,U_{t-1}^{n,v},Y^n,v_{t-1},H_{t-\tau:t-1}^{N,v})h_\theta(u',U_{t-1}^{n,v},Y^n,v_{t-1},H_{t-\tau:t-1}^{N,v})\\
&-\; \mathbf{1}_{U_t^{n,v}=u'}h_\theta(u,U_{t-1}^{n,v},Y^n,v_{t-1},H_{t-\tau:t-1}^{N,v}) - \mathbf{1}_{U_t^{n,v}=u}h_\theta(u',U_{t-1}^{n,v},Y^n,v_{t-1},H_{t-\tau:t-1}^{N,v})]\\
&=\; \frac{1}{N}\sum_{n=1}^{N}\sum_{u',u\in\mathcal{U}}\alpha_u\alpha_{u'}\phi_{u'}(Y^n)\phi_u(Y^n)[\mathbf{1}_{u=u'}h_\theta(u,U_{t-1}^{n,v},Y^n,v_{t-1},H_{t-\tau:t-1}^{N,v})\\
&-\; h_\theta(u,U_{t-1}^{n,v},Y^n,v_{t-1},H_{t-\tau:t-1}^{N,v})h_\theta(u',U_{t-1}^{n,v},Y^n,v_{t-1},H_{t-\tau:t-1}^{N,v})] + \mathcal{M}_t^{3,N,v}\\
&=\; \sum_{u',u\in\mathcal{U}}\alpha_u\alpha_{u'}\sum_{u''\in\mathcal{U}}[\mathbf{1}_{u=u'}\left\langle\phi_{u'}(y)\phi_u(y)h_\theta(u,u'',y,v_{t-1},H_{t-\tau:t-1}^{N,v}),\mu_{t-1}^{N,v}\right\rangle\\
&-\; \left\langle\phi_{u'}(y)\phi_u(y)h_\theta(u,u'',y,v_{t-1},H_{t-\tau:t-1}^{N,v})h_\theta(u',u'',y,v_{t-1},H_{t-\tau:t-1}^{N,v}),\mu_{t-1}^{N,v}\right\rangle] + \mathcal{M}_t^{3,N,v}\\
&\xrightarrow{p}\; \sum_{u',u\in\mathcal{U}}\alpha_u\alpha_{u'}\sum_{u''\in\mathcal{U}}[\mathbf{1}_{u=u'}\left\langle\phi_{u'}(y)\phi_u(y)h_\theta(u,u'',y,v_{t-1},\bar{H}_{t-\tau:t-1}^{v}),\bar{\mu}_{t-1}^{v}\right\rangle\\
&-\; \left\langle\phi_{u'}(y)\phi_u(y)h_\theta(u,u'',y,v_{t-1},\bar{H}_{t-\tau:t-1}^{v})h_\theta(u',u'',y,v_{t-1},\bar{H}_{t-\tau:t-1}^{v}),\bar{\mu}_{t-1}^{v}\right\rangle].
\end{aligned}\tag{32}$$

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

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

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

Using the exact same argument as was employed in the proof of Theorem 3.2, the final line in equation (32) follows from $\mu^{N,v} \xrightarrow{p} \bar{\mu}^v$ for any v and the uniform continuity of h_θ 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 $\bar{\mathcal{M}}_t^v$, whose distribution is given in Lemma A.4. □

Lemma A.4. *Let* $\mathcal{M}_t^{2,N,v}$ *be:*

$$\mathcal{M}_t^{2,N,v}(u,dy) = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}[\delta_{(U_t^{n,v},Y^n)}(u,dy) - h_\theta(u,U_{t-1}^{n,v},Y^n,v_{t-1},H_{t-\tau:t-1}^{N,v})\delta_{Y^n}(dy)].\tag{33}$$

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

$$Cov\Big[\left\langle\phi_1,\bar{\mathcal{M}}_t^v(u_1,dy)\right\rangle,\left\langle\phi_2,\bar{\mathcal{M}}_t^v(u_2,dy)\right\rangle\Big] = Cov\Big[\left\langle\phi_1,\bar{\mathcal{M}}_t^v(u_1,dy)\right\rangle,\left\langle\phi_2,\bar{\mathcal{M}}_t^v(u_2,dy)\right\rangle\Big]$$

$$= -\sum_{u' \in \mathcal{U}} \left\langle \phi_1(y)\phi_2(y) h_\theta(u_1, u', y, v_{t-1}, \bar{H}^v_{t-\tau:t-1}) h_\theta(u_2, u', y, v_{t-1}, \bar{H}^v_{t-\tau:t-1}), \bar{\mu}^v_{t-1} \right\rangle,$$

$$Var\Big[\left\langle \phi, \bar{\mathcal{M}}^v_t(u, dy) \right\rangle\Big] = Var\Big[\left\langle \phi, \bar{\mathcal{M}}^v_t(u, dy) \right\rangle\Big]$$
$$= \sum_{u' \in \mathcal{U}} \left\langle \phi(y)^2 h_\theta(u, u', y, v_{t-1}, \bar{H}^v_{t-1})(1 - h_\theta(u, u', y, v_{t-1}, \bar{H}^v_{t-\tau:t-1}), \bar{\mu}^v_{t-1} \right\rangle,$$

where $u_1 \neq u_2$. *Furthermore,* $\left\langle \phi, \bar{\mathcal{M}}^v_{t_1} \right\rangle$ *is independent of* $\left\langle \phi, \bar{\mathcal{M}}^v_{t_2} \right\rangle$ *for* $t_1 \neq t_2$.

Proof. We first show that:

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

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

$$\sum_{u \in \mathcal{U}} \alpha_u \left\langle \phi_u, \mathcal{M}^{2,N,v}_t(u, dy) \right\rangle = \frac{1}{\sqrt{N}} \sum_{u \in \mathcal{U}} \alpha_u [\phi_u(Y^n)(\mathbf{1}_{U^{n,v}_t = u} - h_\theta(u, U^{n,v}_{t-1}, Y^n, v_{t-1}, H^{N,v}_{t-\tau:t-1})] = Z^{N,n,v}_t(\alpha).$$

Lemma A.3 shows that the conditions of the martingale central limit theorem of [37] hold. The martingale central limit theorem of [37], in conjunction with the Cramer-Wold theorem, gives:

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

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

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

$$\begin{aligned} & \mathbb{E}[\exp\big(i\alpha_1 \left\langle \phi, \mathcal{M}^{2,N,v}_{t_1}(u, dy) \right\rangle + i\alpha_2 \left\langle \phi, \mathcal{M}^{2,N,v}_{t_2}(u', dy) \right\rangle] \\ = \; & \mathbb{E}\big[\exp\big(i\alpha_1 \left\langle \phi_1, \mathcal{M}^{2,N,v}_{t_1}(u, dy) \right\rangle\big) \mathbb{E}[\exp\big(i\alpha_2 \left\langle \phi_2, \mathcal{M}^{2,N,v}_{t_2}(u', dy) \right\rangle\big) | \mu^{N,v}_{0:t_2-1}]\big]. \end{aligned} \tag{34}$$

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

$$\begin{aligned} & \mathbb{E}[\exp\big(i\alpha_2 \left\langle \phi_2, \mathcal{M}^{2,N,v}_{t_2}(u', dy) \right\rangle\big) | \mu^{N,v}_{0:t_2-1}] \\ = & \prod_{n=1}^N \mathbb{E}[e^{\frac{i\alpha_2}{\sqrt{N}} \phi_2(Y^n)(\mathbf{1}_{U^{n,v}_{t_2} = u'} - h_\theta(u', U^{n,v}_{t_2-1}, Y^n, v_{t_2-1}, H^{N,v}_{t_2-\tau:t_2-1}))} | H^{N,v}_{t_2-\tau:t_2-1}, U^{n,v}_{t_2-1}, Y^n] \\ = & \prod_{n=1}^N (1 - \frac{\alpha_2^2}{2N} \mathbb{E}[\phi_2(Y^n)^2 (\mathbf{1}_{U^{n,v}_{t_2} = u'} - h_\theta(u', U^{n,v}_{t_2-1}, Y^n, v_{t_2-1}, H^{N,v}_{t_2-\tau:t_2-1}))^2 | H^{N,v}_{t_2-\tau:t_2-1}, U^{n,v}_{t_2-1}, Y^n] + \frac{C}{N^{3/2}}), \\ = & \prod_{n=1}^N (1 - \frac{\alpha_2^2}{2N} \gamma(v, H^{N,v}_{t_2-\tau:t_2-1}, U^{n,v}_{t_2-1}, Y^n) + \frac{C}{N^{3/2}},) \end{aligned}$$

where we have used a Taylor expansion for the exponential, $\sup_{y \in \mathcal{Y}} 2\phi_2(y)^3 < C < \infty$, and the quantity $\gamma(v, H^{N,v}_{t_2-\tau:t_2-1}, U^{n,v}_{t_2-1}, Y^n) = \mathbb{E}[\phi_2(Y^n)^2 (\mathbf{1}_{U^{n,v}_{t_2} = u'} - h_\theta(u', U^{n,v}_{t_2-1}, Y^n, v_{t_2-1}, H^{N,v}_{t_2-\tau:t_2-1}))^2 | H^{N,v}_{t_2-\tau:t_2-1}, U^{n,v}_{t_2-1}, Y^n]$. 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:

$$\begin{aligned} | \; & \prod_{n=1}^N (1 - \frac{\alpha_2^2}{2N} \gamma(v, H^{N,v}_{t_2-\tau:t_2-1}, U^{n,v}_{t_2-1}, Y^n) + \frac{C}{N^{3/2}}) - \prod_{n=1}^N (1 - \frac{1}{2N} \alpha_2^2 \gamma(v, H^{N,v}_{t_2-\tau:t_2-1}, U^{n,v}_{t_2-1}, Y^n))| \\ < \; & \frac{C}{N^{1/2}} \xrightarrow{a.s.} 0. \end{aligned} \tag{35}$$

Then, all the remains to be shown is that $\prod_{n=1}^N (1 - \frac{1}{2N} \alpha_2^2 \gamma(v, H^{N,v}_{t_2-\tau:t_2-1}, U^{n,v}_{t_2-1}, Y^n))$ converges in probability to $\exp(-\frac{\alpha_2^2}{2\text{Var}[\left\langle \phi_2, \bar{\mathcal{M}}^v_{t_2}(u, dy) \right\rangle]})$ as $N \to \infty$. We have

$$\prod_{n=1}^N (1 - \frac{1}{2N} \alpha_2^2 \gamma(v, H^{N,v}_{t_2-\tau:t_2-1}, U^{n,v}_{t_2-1}, Y^n)) = \exp(\sum_{n=1}^N \log(1 - \frac{1}{2N} \alpha_2^2 \gamma(v, H^{N,v}_{t_2-\tau:t_2-1}, U^{n,v}_{t_2-1}, Y^n)))$$

$$= \exp(\sum_{n=1}^{N}[-\frac{1}{2N}\alpha_2^2\gamma(V, H^N_{t_2-\tau:t_2-1}, U^n_{t_2-1}, Y^n) + \frac{C}{N^2}])$$

$$= \exp(-\frac{1}{2}\alpha_2^2\sum_{u'\in\mathcal{U}}\left\langle \gamma(v, H^{N,v}_{t_2-\tau:t_2-1}, u', y), \mu^{N,v}_{t_2-1}(u', dy)\right\rangle + \frac{C}{N}) \xrightarrow{p} \exp(-\frac{1}{2}\alpha_2^2\mathrm{Var}[\langle\phi_2, \bar{\mathcal{M}}^v_{t_2}\rangle])$$

The last line follows from the previous convergence result $\mu^{N,v} \xrightarrow{p} \bar{\mu}^v$ (see Proof of Theorem 3.2), the fact that $\gamma(v, H^{N,v}_{t_2-\tau:t_2-1}, U^{n,v}_{t_2-1}, Y^n) = \phi_2(Y^n)^2 h_\theta(u, U^{n,v}_{t_2-1}, Y^n, v_{t_2-1}, \bar{H}^v_{t_2-\tau:t_2-1})(1-h_\theta(u, U^{n,v}_{t_2-1}, Y^n, v_{t_2-1}, \bar{H}^v_{t_2-\tau:t_2-1}))$, and the continuous mapping theorem.

With this result in hand, we return to equation (34). Since $\mathbb{E}[\exp\left(i\alpha_2\left\langle\phi_2, \mathcal{M}^{2,N,v}_{t_2}(u')\right\rangle\right)|\mu^{N,v}_{0:t_2-1}]$ converges to a constant and $\exp\left(i\alpha_1\left\langle\phi_1, \mathcal{M}^{2,N,v}_{t_1}(u)\right\rangle\right) \xrightarrow{d} \exp\left(i\alpha_1\left\langle\phi_1, \bar{\mathcal{M}}^v_{t_1}\right\rangle\right)$, Slutsky's theorem yields that $\exp\left(i\alpha_1\left\langle\phi_1, \mathcal{M}^{2,N,v}_{t_1}(u, dy)\right\rangle\right)\mathbb{E}[\exp\left(i\alpha_2\left\langle\phi_2, \mathcal{M}^{2,N,v}_{t_2}(u', dy)\right\rangle\right)|\mu^{N,v}_{0:t_2-1}]$ converges in distribution to the quantity $\exp\left(i\alpha_1\left\langle\phi_1, \bar{\mathcal{M}}^v_{t_1}\right\rangle\right)\exp(-\frac{1}{2}\alpha_2^2\mathrm{Var}[\langle\phi_2, \bar{\mathcal{M}}^v_{t_2}\rangle])$. By the bounded convergence theorem, we then have that:

$$\mathbb{E}[\exp\left(i\alpha_1\left\langle\phi, \mathcal{M}^{2,N,v}_{t_1}(u, dy)\right\rangle + i\alpha_2\left\langle\phi, \mathcal{M}^{2,N,v}_{t_2}(u', dy)\right\rangle\right)]$$
$$\to \exp(-\frac{1}{2}\alpha_1^2\mathrm{Var}[\langle\phi_1, \bar{\mathcal{M}}^v_{t_1}\rangle])\exp(-\frac{1}{2}\alpha_2^2\mathrm{Var}[\langle\phi_2, \bar{\mathcal{M}}^v_{t_2}\rangle])$$

Consequently, $\left\langle\phi_1, \bar{\mathcal{M}}^v_{t_1}\right\rangle$ is independent of $\left\langle\phi_2, \bar{\mathcal{M}}^v_{t_2}\right\rangle$ for $t_1 \neq t_2$.

Since these results hold for any choice of $\phi \in S$, we also have the stronger result that $\mathcal{M}^{2,N,v} \xrightarrow{d} \bar{\mathcal{M}}^v$, using the criterion (24). □

Lemma A.5. *Suppose that* $\Xi^{N,v}_{0:t-1} \xrightarrow{d} \bar{\Xi}^v_{0:t-1}$. *Then,* $(\Xi^{N,v}_{0:t-1}, \mathcal{M}^{2,N,v}_t) \xrightarrow{d} (\bar{\Xi}^v_{0:t-1}, \bar{\mathcal{M}}^v_t(u, dy))$.

Proof. The proof for this lemma is the same as in Lemma A.4. Namely, we show that $\bar{\Xi}^v_{0:t-1}$ and $\bar{\mathcal{M}}^v_t(u, dy)$ are independent. For $t' < t$,

$$\mathbb{E}[\exp\left(i\sum_{t'<t, u\in\mathcal{U}}\left\langle\phi_{1,u,t'}, \Xi^{N,v}_{t'}(u, dy)\right\rangle + i\left\langle\phi_2, \mathcal{M}^{2,N,v}_t(u', dy)\right\rangle\right)]$$
$$= \mathbb{E}\big[\exp\left(i\sum_{t'<t, u\in\mathcal{U}}\left\langle\phi_{1,u,t'}, \Xi^{N,v}_{t'}(u, dy)\right\rangle\right)\mathbb{E}[\exp\left(i\left\langle\phi_2, \mathcal{M}^{2,N,v}_t(u', dy)\right\rangle\right)|\mu^{N,v}_{0:t-1}]\big]. \tag{36}$$

Now, using the exact same steps as in Lemma A.4, the desired result can be obtained. □

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

$$\left\langle\phi(y), \Xi^{N,v}_t(u, dy)\right\rangle = \sqrt{N}\sum_{u'\in\mathcal{U}}\Bigg[\left\langle\phi(y), h_\theta(u, u', y, v_{t-1}, H^{N,v}_{t-\tau:t-1})\mu^{N,v}_{t-1}(u', dy)\right\rangle$$
$$- \left\langle\phi(y), h_\theta(u, u', y, v_{t-1}, \bar{H}^v_{t-\tau:t-1})\bar{\mu}^v_{t-1}(u', dy)\right\rangle\Bigg] + \left\langle\phi, \mathcal{M}^{2,N,v}_t(u)\right\rangle, \tag{37}$$

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

$$\left\langle\phi, \mathcal{M}^{2,N,v}_t(u, dy)\right\rangle = \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\phi(Y^n)(\mathbf{1}_{U^{n,v}_t=u} - h_\theta(u, U^{n,v}_{t-1}, Y^n, v_{t-1}, H^{N,v}_{t-\tau:t-1}). \tag{38}$$

Using a Taylor expansion, one can obtain

$$\left\langle\phi(y), \Xi^{N,v}_t(u, dy)\right\rangle = \sum_{u'\in\mathcal{U}}\left\langle\phi(y)h_\theta(u, u', y, v_{t-1}, \bar{H}^v_{t-\tau:t-1}), \Xi^N_{t-1}(u', dy)\right\rangle$$
$$+ \sum_{u'\in\mathcal{U}}\left\langle\phi(y)h_{\theta,H}(u, u', y, v_{t-1}, \bar{H}^v_{t-\tau:t-1})\cdot E^{N,v}_{t-\tau:t-1}, \mu^{N,v}_{t-1}(u', dy)\right\rangle$$
$$+ \frac{1}{\sqrt{N}}\mathcal{R}^{N,v} + \left\langle\phi, \mathcal{M}^{2,N,v}_t(u, dy)\right\rangle, \tag{39}$$

where $\mathcal{R}^{N,v}$ is the Taylor remainder term:

$$
\begin{aligned}
\left|\mathcal{R}^{N,v}\right| &= \Big|\frac{1}{2}\left\langle \phi(y)(E^{N,v}_{t-\tau:t-1})^\top h_{HH}(u,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1})E^{N,v}_{t-\tau:t-1}, \mu^{N,v}_{t-1}(u',dy)\right\rangle\Big| \\
&\leq \Big|\frac{1}{2}\left\langle \sup_{y\in\mathcal{Y}} \phi(y)(E^{N,v}_{t-\tau:t-1})^\top h_{HH}(u,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1})E^{N,v}_{t-\tau:t-1}, \mu^{N,v}_{t-1}(u',dy)\right\rangle\Big| \\
&\leq K||E^{N,v}_{t-\tau:t-1}||^2 \left\langle \sup_{y\in\mathcal{Y}} |\phi(y)h_{HH}(u,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1})|, \mu^{N,v}_{t-1}(u',dy)\right\rangle \\
&\leq ||E^{N,v}_{t-\tau:t-1}||^2 \left\langle C, \mu^{N,v}_{t-1}(u',dy)\right\rangle = C||E^{N,v}_{t-\tau:t-1}||^2. \qquad (40)
\end{aligned}
$$

Here, h_{HH} denotes the second partial derivative of h_θ with respect to its fifth argument. We have bounded $h_{HH}(u,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1})$ due to the continuous differentiability of h_θ and the compactness of $\mathcal{Y}$. Similarly, $\phi(y)$ is bounded due to the continuity of ϕ and the compactness of $\mathcal{Y}$. Since $||\cdot||^2$ is a continuous function, we have by the continuous mapping theorem that $||E^{N,v}_{t-\tau:t-1}||^2 \xrightarrow{d} ||\bar{E}^v_{t-\tau:t-1}||^2$. Consequently, $\frac{1}{\sqrt{N}}\mathcal{R}^{N,v} \xrightarrow{d} 0$.

We will now use an induction argument to prove the desired result. Assume that $\Xi^{N,v}_{0:t-1}$ converges in distribution to $\bar{\Xi}^v_{0:t-1}$. Since convergence in distribution to a constant implies convergence in probability to a constant, $\frac{1}{\sqrt{N}}\mathcal{R}^{N,v}$ converges in probability to zero. The second term in equation (39) also converges in distribution since $\mu^{N,v}_{t-1}$ converges in probability to $\bar{\mu}^v_{t-1}$ for any v (from the law of large numbers) and $E^{N,v}_{t-\tau:t} \xrightarrow{d} \bar{E}^v_{t-\tau:t-1}$ (since we assumed $\Xi^{N,v}_{0:t-1} \xrightarrow{d} \bar{\Xi}^v_{0:t-1}$ and f^H is continuous). The third term converges in distribution due to Lemma A.4. Then, assuming $\Xi^{N,v}_{0:t-1}$ converges in distribution to $\bar{\Xi}^v_{0:t-1}$, $\left\langle \phi, \Xi^{N,v}_t\right\rangle \xrightarrow{d} \left\langle \phi, \bar{\Xi}^v_t\right\rangle$ for any $\phi \in S$ where $\left\langle \phi, \bar{\Xi}^v_t\right\rangle$ satisfies the evolution equation:

$$
\begin{aligned}
\left\langle \phi(y), \bar{\Xi}^v_t(u,dy)\right\rangle &= \sum_{u'\in\mathcal{U}} \left\langle \phi(y)h_\theta(u,u',y,v_{t-1},\bar{H}_{t-\tau:t-1}), \bar{\Xi}^v_{t-1}(u',dy)\right\rangle \\
&+ \sum_{u'\in\mathcal{U}} \left\langle \phi(y)\frac{\partial}{\partial H}h_\theta(u,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1})\cdot \bar{E}^v_{t-\tau:t-1}, \bar{\mu}^v_{t-1}\right\rangle + \left\langle \phi, \bar{\mathcal{M}}^v_t(u,dy)\right\rangle, \quad \phi\in S. \qquad (41)
\end{aligned}
$$

This is sufficient to show that the $\Xi^{N,v}_t \xrightarrow{d} \bar{\Xi}^v_t$, by the criterion (24). However, to make the stronger statement of the joint convergence $\Xi^{N,v}_{0:t} \xrightarrow{d} \bar{\Xi}^v_{0:t}$, we need an additional fact, namely that $(\Xi^{N,v}_{0:t-1}, \mathcal{M}^{2,N,v}_t) \xrightarrow{d} (\bar{\Xi}^v_{0:t}, \bar{\mathcal{M}}^v_t(u,dy))$. This has been proven in Lemma A.5, and therefore we have the desired result that $\Xi^{N,v}_{0:t} \xrightarrow{d} \bar{\Xi}^v_{0:t}$. By induction, we have that $\bar{\Xi}^{N,v} \xrightarrow{d} \bar{\Xi}^v$ for any v. Therefore, $\bar{\Xi}^N \xrightarrow{d} \bar{\Xi}$. ☐

Proof of Corollary 3.7. Since $(\mu^{N,v}_t, L^{N,v}) \xrightarrow{p} (\bar{\mu}^v, \bar{L}^v)$ where $(\bar{\mu}^v, \bar{L}^v)$ are constants and $(\Xi^{N,v}, \Lambda^{N,v}) \xrightarrow{d} (\bar{\Xi}^v, \bar{\Lambda}^v)$, we have that $(v, \mu^{N,v}_t, L^{N,v}, \Xi^{N,v}, \Lambda^{N,v}) \xrightarrow{d} (v, \bar{\mu}^v, \bar{L}^v, \bar{\Xi}^v, \bar{\Lambda}^v)$. Since this holds for any v, the desired convergence follows. ☐

A.3. **Convergence for Loss From Default (Proposition 3.6).**

Lemma A.6. *Define $\mathcal{Z}^{N,v}$ to be the martingale*

$$
\mathcal{Z}^{N,v}_t = \frac{1}{N}\sum_{n=1}^N \ell^n_t(Y^n, v_t)(\mathbf{1}_{U^{n,v}_t = d} - \mathbf{1}_{U^{n,v}_{t-1}=d}) - \left\langle \int_0^1 z\nu_{t,y,v_t}(dz), \mu^{N,v}_t(d,dy) - \mu^{N,v}_t(d,dy)\right\rangle,
$$

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

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

$$
\mathcal{Z}^{N,v}_t = \frac{1}{N}\sum_{n=1}^N \ell^n_t(Y^n, v_t)(\mathbf{1}_{U^{n,v}_t = \mathrm{d}} - \mathbf{1}_{U^{n,v}_{t-1}=\mathrm{d}}) - \frac{1}{N}\sum_{n=1}^N (\mathbf{1}_{U^{n,v}_t = \mathrm{d}} - \mathbf{1}_{U^{n,v}_{t-1}=\mathrm{d}})\int_0^1 z\nu_{t,Y^n,v_t}(dz).
$$

Similar to Lemma A.1, we will show that the variance of $\mathcal{Z}_t^{N,v}$ tends to zero as $N \to \infty$.

$$\begin{aligned}
\mathrm{Var}[\mathcal{Z}_t^{N,v}] &= \mathbb{E}[\mathbb{E}[(\frac{1}{N}\sum_{n=1}^{N}(\mathbf{1}_{U_t^{n,v}=\mathrm{d}} - \mathbf{1}_{U_{t-1}^{n,v}=\mathrm{d}})(\ell_t^n(Y^n, v_t) \\
&- \int_0^1 z\nu_{t,Y^n,v_t}(dz)))^2 | U_t^{1,v}, \dots, U_t^{N,v}, U_{t-1}^{1,v}, \dots, U_{t-1}^{N,v}, Y^1, \dots, Y^N]] \\
&\le \mathbb{E}[\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}[(\ell_t^n(Y^n, V_t) - \int_0^1 z\nu_{t,Y^n,V_t}(dz))^2 | U_t^{1,v}, \dots, U_t^{N,v}, Y^1, \dots, Y^N]] \le \frac{C}{N}.
\end{aligned}$$

We have again used the tower property, the independence of the $\ell^n(Y^n, v_t)$ given $(U_t^{1,v}, \dots, U_t^{N,v}, Y^1, \dots, Y^N)$, and the fact that the loss given default is bounded. By Chebyshev's inequality, $\mathcal{Z}_t^{N,v} \xrightarrow{p} 0$ for any v. ☐

Using Lemma A.6, we can show that $L^N \xrightarrow{d} \bar{L}$. Define $\bar{L}^v = \mathbb{E}[\bar{L} | V = v]$ and note that $\bar{L}^V = \bar{L}$. We also define:

$$L_t^{N,v} = \left\langle \int_0^1 z\nu_{t,y,v_t}(dz)\mu_t^{N,v}(\mathrm{d}, dy) - \mu_{t-1}^{N,v}(\mathrm{d}, dy) \right\rangle + \mathcal{Z}_t^{N,v}. \tag{42}$$

For any v, $\mathcal{Z}_t^{N,v} \xrightarrow{p} 0$ by Lemma A.6. Note that $L_t^{N,V} = L_t^N$. Since $g_1(t, v, y) = \int_0^1 z\nu_{t,y,v}(dz)$ is continuous and bounded for each y, v:

$$\left\langle \int_0^1 z\nu_{t,y,v_t}(dz), \mu_t^{N,v}(\mathrm{d}, dy) - \mu_{t-1}^{N,v}(\mathrm{d}, dy) \right\rangle \xrightarrow{p} \left\langle \int_0^1 z\nu_{t,y,v_t}(dz), \bar{\mu}_t^v(\mathrm{d}, dy) - \bar{\mu}_{t-1}^v(\mathrm{d}, dy) \right\rangle, \tag{43}$$

for any v. This proves $L_t^{N,v} \xrightarrow{p} \bar{L}_t^v$ for each t and any v, which (like before) implies that $(L_0^{N,v}, \dots, L_T^{N,v}) \xrightarrow{p} (\bar{L}_0^v, \dots, \bar{L}_T^v)$ for any v. By similar reasoning as in (23), this implies that $(\mu^N, L^N) \xrightarrow{d} (\bar{\mu}, \bar{L})$.

We now turn to proving the CLT for the loss. Define $\bar{\Lambda}^v$:

$$\bar{\Lambda}_t^v = \left\langle \int_0^1 z\nu_{t,y,v_t}(dz), \bar{\Xi}_t^v(\mathrm{d}, dy) - \bar{\Xi}_{t-1}^v(\mathrm{d}, dy) \right\rangle + \bar{\mathcal{Z}}_t^v, \tag{44}$$

where $\bar{\mathcal{Z}}_t^v$ is a mean-zero Gaussian with variance:

$$\mathrm{Var}[\bar{\mathcal{Z}}_t^v] = \left\langle [\int_0^1 z^2\nu_{t,y,v_t}(dz) - (\int_0^1 z\nu_{t,y,v_t}(dz))^2], \bar{\mu}_t^v(\mathrm{d}, dy) - \bar{\mu}_{t-1}^v(\mathrm{d}, dy) \right\rangle.$$

$\bar{\mathcal{Z}}_{t_1}^v$ is independent of $\bar{\mathcal{Z}}_{t_2}^v$ for $t_1 \neq t_2$ as well as $\bar{\Xi}^v$.

Lemma A.7. *Suppose that $(\Xi_{0:t}^{N,v}, \sqrt{N}\mathcal{Z}_{0:t-1}^{N,v}) \xrightarrow{d} (\bar{\Xi}_{0:t}^v, \bar{\mathcal{Z}}_{0:t-1}^v)$. Then, $(\{\Xi_{0:t}^{N,v}\}, \sqrt{N}\mathcal{Z}_{0:t}^{N,v}) \xrightarrow{d} (\bar{\Xi}_{0:t}^v, \bar{\mathcal{Z}}_{0:t}^v)$. $\bar{\mathcal{Z}}_t^v$ is a mean zero Gaussian with variance:*

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

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

Proof. We have

$$\begin{aligned}
&\mathbb{E}[\exp\big(i\sum_{t' \le t, u \in \mathcal{U}}\left\langle \phi_{1,u,t'}, \Xi_{t'}^{N,v}(u, dy)\right\rangle + i\sum_{t'<t}\alpha_{t'}\sqrt{N}Z_{t'}^{N,v} + i\alpha_t\sqrt{N}Z_t^{N,v})] \\
&= \mathbb{E}\big[\exp\big(i\sum_{t' \le t, u \in \mathcal{U}}\left\langle \phi_{1,u,t'}, \Xi_{t'}^{N,v}(u, dy)\right\rangle + i\sum_{t'<t}\alpha_{t'}\sqrt{N}Z_{t'}^{N,v}\big)\mathbb{E}[\exp\big(i\alpha_t\sqrt{N}Z_t^{N,v}\big)) | \mu_{0:t}^N]\big].
\end{aligned} \tag{46}$$

The result follows via the exact same procedure as used in Lemma A.4. ☐

Define:

$$\Lambda_t^{N,v} = \sqrt{N}(L_t^{N,v} - \bar{L}_t^v). \tag{47}$$

Using Lemma A.7, we have that $(\Xi^{N,v}, \Lambda^{N,v}) \xrightarrow{d} (\bar{\Xi}^v, \bar{\Lambda}^v)$ since:

$$\begin{aligned}\Lambda_t^{N,v} &= \Big\langle \int_0^1 z\nu_{t,y,v_t}(dz), \Xi_t^{N,v}(\mathrm{d},dy) - \Xi_{t-1}^{N,v}(\mathrm{d},dy) \Big\rangle + \sqrt{N}\mathcal{Z}_t^{N,v} \\ &\xrightarrow{d} \Big\langle \int_0^1 z\nu_{t,y,v_t}(dz), \bar{\Xi}_t^{v}(\mathrm{d},dy) - \bar{\Xi}_{t-1}^{v}(\mathrm{d},dy) \Big\rangle + \bar{\mathcal{Z}}_t^{v} = \bar{\Lambda}_t^v.\end{aligned}$$

This result follows since $\bar{\Xi}^{N,v} \xrightarrow{d} \bar{\Xi}^v$, the assumption of continuity in y for $\int_0^1 z\nu_{t,y,v_t}(dz)$ for any t, v, and Lemma A.7. Since the results holds for any v, $(\Xi^N, \Lambda^N) \xrightarrow{d} (\bar{\Xi}, \bar{\Lambda})$. ☐

A.4. **Uniform Integrability of Fluctuations (Theorem 3.9).**

Proof of Theorem 3.9. Note that the bound (10) implies

$$\mathbb{P}\Big[(\sup_{y\in\mathcal{Y}} |\phi(y)|)^{-1} \big| \left\langle \phi(y), \Xi_0^N(u,dy) \right\rangle \big| > \alpha\Big] \le K_{1,0,u} \exp\big(-K_{2,0,u}\alpha^2\big).$$

Consider the following bound for $\alpha > 0$:

$$\mathbb{P}\Big[\big|\sum_{t=1}^T \sum_{u\in\mathcal{U}} \left\langle \phi_{t,u}(y), \Xi_t^N(u,dy) \right\rangle + \sum_{t=1}^T c_t\Lambda_t^N)\big| > \alpha\Big] \le C_1 \exp\big(-C_2\alpha^2\big). \tag{48}$$

For notational convenience, define $X^N = \sum_{t=1}^T \sum_{u\in\mathcal{U}} \left\langle \phi_{t,u}(y), \Xi_t^N(u,dy) \right\rangle + \sum_{t=1}^T c_t\Lambda_t^N$. If the bound (48) holds and $K > C_0$, then:

$$\begin{aligned}\mathbb{E}\Big[\big|g(X^N)\big|\mathbf{1}_{|g(X^N)|>K}\Big] &= \int_K^\infty \mathbb{P}\Big[\big|g(X^N)\big| \ge \alpha\Big] d\alpha \\ &\le \int_K^\infty \mathbb{P}\Big[\big|X^N\big|^k \ge \alpha\Big] d\alpha = \int_K^\infty \mathbb{P}\Big[\big|X^N\big| \ge \alpha^{1/k}\Big] d\alpha \\ &\le \int_K^\infty C_1 \exp\big(-C_2\alpha^{2/k}\big) d\alpha \\ &= C_1 \int_K^\infty \exp\big(-C_2 x\big)\frac{x^{k-1}}{k} dx \le C_1 \int_0^\infty \exp\big(-C_2 x\big)\frac{x^{k-1}}{k} dx \\ &= \frac{C_1}{k} C_2^k \Gamma(k),\end{aligned}$$

where Γ is the gamma function. Then, for any $\epsilon > 0$, there is a $K > 0$ such that:

$$\mathbb{E}\Big[\big|g(X^N)\big|\mathbf{1}_{|X^N|>K}\Big] < \epsilon.$$

Consequently, the bound (48), $\mathbb{P}\Big[\big|X^N\big| > \alpha\Big] \le C_1 \exp\big(-C_2\alpha^2\big)$ for $\alpha > 0$, implies uniform integrability for $g(X^N)$. Next, recognize that the following bound implies the bound (48):

$$\mathbb{P}\Big[\big| \left\langle \phi_{t,u}(y), \Xi_t^N(u,dy) \right\rangle + c_t\Lambda_t^N \big| > \alpha\Big] \le C_{1,t,u} \exp\big(-C_{2,t,u}\alpha^2\big). \tag{49}$$

The bound (49) implies the bound (48) since:

$$\begin{aligned}&\mathbb{P}\Big[\big|\sum_{t=1}^T \sum_{u\in\mathcal{U}} \left\langle \phi_{t,u}(y), \Xi_t^N(u,dy) \right\rangle + \sum_{t=1}^T c_t\Lambda_t^N\big| > \alpha\Big] \\ \le\ &\mathbb{P}\Big[\sum_t \sum_u \big| \left\langle \phi_{t,u}(y), \Xi_t^N(u,dy) \right\rangle + c_t\Lambda_t^N \big| > \alpha\Big] \\ \le\ &\mathbb{P}\Big[\bigcup_{t,u} \{\big| \left\langle \phi_{t,u}(y), \Xi_t^N(u,dy) \right\rangle + c_t\Lambda_t^N \big| > \frac{\alpha}{|\mathcal{U}|T}\}\Big]\end{aligned}$$

$$
\begin{aligned}
&= \sum_{t,u} \mathbb{P}\Big[\{ | \langle \phi_{t,u}(y), \Xi_t^N(u,dy) \rangle + c_t \Lambda_t^N | > \frac{\alpha}{|\mathcal{U}| T} \}\Big] \\
&\leq \sum_{t,u} \mathbb{P}\Big[| \langle \phi_{t,u}(y), \Xi_t^N(u,dy) \rangle + c_t \Lambda_t^N | > \frac{\alpha}{|\mathcal{U}| T} \Big] \\
&\leq \sum_{t,u} C_{1,t,u} \exp\big(- C_{2,t,u} \alpha^2 \big) \\
&\leq |T||\mathcal{U}| \max_{t,u} C_{1,t,u} \exp\big(- \min_{t,u} C_{2,t,u} \alpha^2 \big) = C_1 \exp\big(- C_2 \alpha^2 \big).
\end{aligned}
$$

It is therefore sufficient to prove an exponentially decaying bound:

$$
\mathbb{P}\Big[| \langle \phi_{t,u}(y), \Xi_t^N(u,dy) \rangle + c_t \Lambda_t^N | > \alpha \Big] \leq C_{1,t,u} \exp\big(- C_{2,t,u} \alpha^2 \big), \tag{50}
$$

for each t and u. We prove this result using induction. From equation (39):

$$
\begin{aligned}
\langle \phi(y), \Xi_t^N(u,dy) \rangle &= \sum_{u' \in \mathcal{U}} \langle \phi(y) h_\theta(u,u',y,V_{t-1},\bar{H}_{t-\tau:t-1}), \Xi_{t-1}^N(u',dy) \rangle \\
&+ \sum_{u' \in \mathcal{U}} \langle \phi(y) h_{\theta,H}(u,u',y,V_{t-1},\bar{H}_{t-\tau:t-1}) \cdot E_{t-\tau:t-1}^N, \mu_{t-1}^N(u',dy) \rangle \\
&+ \frac{1}{\sqrt{N}} \mathcal{R}^{N,V} + \Big\langle \phi, \mathcal{M}_t^{2,N,V}(u,dy) \Big\rangle,
\end{aligned} \tag{51}
$$

where $\big|\mathcal{R}^{N,V}\big| \leq C \sup_{y \in \mathcal{Y}} |\phi(y)| ||E_{t-\tau:t-1}^N||^2$ (see equation (40)) and $h_{\theta,H}$ is the gradient of h_θ with respect to H. $\mathcal{M}_t^{2,N,V}(u,dy)$ is defined in Lemma A.4. First, suppose that:

$$
\mathbb{P}\Big[(\sup_{y \in \mathcal{Y}} |\phi(y)|)^{-1} | \langle \phi(y), \Xi_{t'}^N(u,dy) \rangle | > \alpha \Big| V \Big] \leq K_{1,t',u} \exp\big(- K_{2,t',u} \alpha^2 \big), \tag{52}
$$

for any continuous ϕ and all $t' \leq t-1$. Here and below, we use the notation $V = V_{0:T}$. This is equivalent to $\mathbb{P}\Big[| \langle \phi(y), \Xi_{t'}^N(u,dy) \rangle | > \alpha \Big| V \Big] \leq C_{1,t',u} \exp\big(- C_{2,t',u} \alpha^2 \big)$ where $C_{2,t',u} = K_{2,t',u} (\sup_{y \in \mathcal{Y}} |\phi(y)|)^{-2}$ and $K_{1,t',u} = C_{1,t',u}$. Since ϕh_θ is continuous on a compact space, we also have that the first term of the RHS of (51) satisfies:

$$
\begin{aligned}
&\mathbb{P}\Big[\Big| \sum_{u' \in \mathcal{U}} \langle \phi(y) h_\theta(u,u',y,V_{t-1},\bar{H}_{t-\tau:t-1}), \Xi_{t-1}^N(u',dy) \rangle \Big| > \alpha \Big| V \Big] \\
\leq& \sum_{u' \in \mathcal{U}} \mathbb{P}\Big[\Big| \langle \phi(y) h_\theta(u,u',y,V_{t-1},\bar{H}_{t-\tau:t-1}), \Xi_{t-1}^N(u',dy) \rangle \Big| > \frac{\alpha}{|\mathcal{U}|} \Big| V \Big] \\
\leq& \sum_{u' \in \mathcal{U}} K_{1,t-1,u'} \exp\big(- K_{2,t-1,u'} (\sup_{y \in \mathcal{Y}} |\phi(y) h(u,u',y,V_{t-1},\bar{H}_{t-\tau:t-1})|)^{-2} \frac{\alpha^2}{|\mathcal{U}|^2} \big), \\
\leq& |\mathcal{U}| \max_{u' \in \mathcal{U}} K_{1,t-1,u'} \exp\big(- (\sup_{y \in \mathcal{Y}} |\phi(y)|)^{-2} \min_{u' \in \mathcal{U}} K_{2,t-1,u'} \frac{\alpha^2}{|\mathcal{U}|^2} \big) \\
\leq& C_{1,t,u}^1 \exp\big(- (\sup_{y \in \mathcal{Y}} |\phi(y)|)^{-2} C_{2,t,u}^1 \alpha^2 \big)
\end{aligned} \tag{53}
$$

where the second-to-last inequality uses the fact that $0 \leq h_\theta \leq 1$. Note that the final coefficients do not depend upon V nor ϕ.

Recall that $E_{t'}^N = \langle f^H, \Xi_{t'}^N \rangle$ where f^H is continuous. Then, $E_{t'}^N$ has an exponential bound, which allows one to exponentially bound the second term of the RHS of (51).

$$
\mathbb{P}\Big[\Big| \sum_{u' \in \mathcal{U}} \langle \phi(y) h_{\theta,H}(u,u',y,V_{t-1},\bar{H}_{t-\tau:t-1}) \cdot E_{t-\tau:t-1}^N, \mu_{t-1}^N(u',dy) \rangle \Big| > \alpha \Big| V \Big]
$$

$$
\begin{aligned}
&\le \mathbb{P}\Big[\sup_{(y,v)\in\mathcal{Y}\times\mathcal{V}} |\sum_{u'\in\mathcal{U}} \phi(y) h_{\theta,H}(u,u',y,V_{t-1},\bar{H}_{t-\tau:t-1})| \sum_{t'=t-\tau}^{t-1} |E_{t'}^N| > \alpha \Big| V\Big] \\
&\le \mathbb{P}\Big[\sup_{y\in\mathcal{Y}} |\phi(y)| \sup_{(y,v)\in\mathcal{Y}\times\mathcal{V}} |\sum_{u'\in\mathcal{U}} h_{\theta,H}(u,u',y,V_{t-1},\bar{H}_{t-\tau:t-1})| \sum_{t'=t-\tau}^{t-1} |E_{t'}^N| > \alpha \Big| V\Big] \\
&\le \mathbb{P}\Big[\sup_{y\in\mathcal{Y}} |\phi(y)| K_0 \sum_{t'=t-\tau}^{t-1} |E_{t'}^N| > \alpha \Big| V\Big] \le C_{1,t,u}^2 \exp\big(-(\sup_{y\in\mathcal{Y}}|\phi(y)|)^{-2} C_{2,t,u}^2 \alpha^2\big),
\end{aligned} \tag{54}
$$

where $\sup_{(y,v)\in\mathcal{Y}\times\mathcal{V}} |\sum_{u'\in\mathcal{U}} h_{\theta,H}(u,u',y,V_{t-1},\bar{H}_{t-\tau:t-1}) \le K_0$, a constant that does not depend upon V nor ϕ, since $\mathcal{Y}\times\mathcal{V}$ is compact and h is smooth. Similarly, the third term on the RHS of (51) can also be exponentially bounded by $C_{1,t,u}^3 \exp\big(-(\sup_{y\in\mathcal{Y}}|\phi(y)|)^{-2} C_{2,t,u}^3 \alpha^2\big)$. The fourth term on the RHS of (51) can be exponentially bounded via the Azuma-Hoeffding inequality. Define $S^{N,n} = \frac{1}{\sqrt{N}} \sum_{n'=1}^{n} \phi(Y^{n'})[\mathbf{1}_{U_t^{n'}\in u} - h_\theta(u, U_{t-1}^{n'}, Y^{n'}, V_{t-1}, H_{t-\tau:t-1}^N)]$ for $n = 1,\dots,N$. Conditional on $\mathcal{F}_{t-1}$, the sequence $S^{N,1}, S^{N,2},\dots$ is a martingale and has bounded differences:

$$
|S^{N,n} - S^{N,n-1}| = \frac{1}{\sqrt{N}} \big|\phi(Y^n)[\mathbf{1}_{U_t^n\in u} - h_\theta(u, U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N)]\big| \le \frac{2}{\sqrt{N}} \sup_{y\in\mathcal{Y}} |\phi(y)|. \tag{55}
$$

Therefore, conditional on $\mathcal{F}_{t-1}$, $S^{N,n}$ satisfies the Azuma-Hoeffding inequality. Note that $S^{N,N} = \Big\langle \phi(y), \mathcal{M}_t^{2,N,V}(u,dy)\Big\rangle$. Then, by the Azuma-Hoeffding inequality:

$$
\begin{aligned}
\mathbb{P}\Big[| \Big\langle \phi(y), \mathcal{M}_t^{2,N,V}(u,dy)\Big\rangle > \alpha| \Big| V\Big] &= \mathbb{E}\Big[\mathbb{P}\Big[| \Big\langle \phi(y), \mathcal{M}_t^{2,N,V}(u,dy)\Big\rangle \Big| > \alpha \big| \mathcal{F}_{t-1}\Big]\Big| V\Big] \\
&\le 2\mathbb{E}\Big[\exp(-\frac{1}{4}(\sup_{y\in\mathcal{Y}}|\phi(y)|^2)^{-2}\alpha^2)\Big| V\Big] \\
&= 2\exp(-\frac{1}{8}(\sup_{y\in\mathcal{Y}}|\phi(y)|)^{-2}\alpha^2).
\end{aligned}
$$

Combining the exponential bounds for the four terms on the RHS of (51) and applying the standard union bound, we then have a bound for the LHS of (51):

$$
\mathbb{P}\Big[| \langle \phi(y), \Xi_t^N(u,dy)\rangle | > \alpha \Big| V\Big] \le C_{1,t,u} \exp\big(-C_{2,t,u} \sup_{y\in\mathcal{Y}}|\phi(y)|)^{-2}\alpha^2\big), \tag{56}
$$

where $C_{1,t,u}$ and $C_{2,t,u}$ do not depend upon V nor ϕ. This of course implies $\mathbb{P}\Big[| \langle \phi(y), \Xi_t^N(u,dy)\rangle | > \alpha\Big] \le C_{1,t,u} \exp\big(-C_{2,t,u} \sup_{y\in\mathcal{Y}}|\phi(y)|)^{-2}\alpha^2\big)$. It also proves the inductive step:

$$
\begin{aligned}
\mathbb{P}\Big[(\sup_{y\in\mathcal{Y}}|\phi(y)|)^{-1} \big| \langle \phi(y), \Xi_t^N(u,dy)\rangle \big| > \alpha \Big| V\Big] &= \mathbb{P}\Big[| \langle \phi(y), \Xi_t^N(u,dy)\rangle | > (\sup_{y\in\mathcal{Y}}|\phi(y)|)\alpha \Big| V\Big] \\
&= C_{1,t,u} \exp\big(-C_{2,t,u}\alpha^2\big) \\
&\equiv K_{1,t,u} \exp\big(-K_{2,t,u}\alpha^2\big).
\end{aligned} \tag{57}
$$

Since $\mathbb{P}\Big[(\sup_{y\in\mathcal{Y}}|\phi(y)|)^{-1} \big| \langle \phi(y), \Xi_0^N(u,dy)\rangle \big| > \alpha \Big| V\Big] \le K_{1,0,u} \exp\big(-K_{2,0,u}\alpha^2\big)$ by assumption, this completes the induction proof to show that $\mathbb{P}\Big[| \langle \phi_{t,u}(y), \Xi_t^N(u,dy)\rangle | > \alpha\Big] \le C_{1,t,u} \exp\big(-C_{2,t,u}\alpha^2\big)$ for each t and u. The final step requires to proving a similar bound for $c_t \Lambda_t^N$. Recall that:

$$
\Lambda_t^N = \sqrt{N}(L_t^N - \bar{L}_t) \quad = \quad \Big\langle \int_0^1 z \nu_{t,y,V_t}(dz), \Xi_t^N(d,dy) - \Xi_{t-1}^N(d,dy) \Big\rangle + \sqrt{N} \mathcal{Z}_t^{N,V}. \tag{58}
$$

Let's first consider the first term on the RHS of (58). The function $\int_0^1 z\nu_{t,y,V_t}(dz)$ is continuous on y (by assumption) and $\sup_{y\in\mathcal{Y}} \int_0^1 z\nu_{t,y,V_t}(dz) \le 1$ since ν is a measure on $[0,1]$. Using the (now proven) bound

$\mathbb{P}\Big[| \left\langle \phi(y), \Xi_t^N(u,dy) \right\rangle | > \alpha \Big| V \Big] \leq C_{1,t,u} \exp\big(- C_{2,t,u} \alpha^2 \big)$ and the standard union bound, we have that:

$$\mathbb{P}\Big[\Big| \Big\langle \int_0^1 z \nu_{t,y,V_t}(dz), \Xi_t^N(d,dy) - \Xi_{t-1}^N(d,dy) \Big\rangle \Big| \geq \alpha \Big| V \Big] \leq A_{1,t,u} \exp\big(- A_{2,t,u}\alpha^2 \big). \tag{59}$$

Next, let's consider the second term on the RHS of (58), $\sqrt{N}\mathcal{Z}_t^{N,V}$. Define $Q^{N,n} = \frac{1}{\sqrt{N}} \big[\sum_{n'=1}^n \ell_t^{n'}(Y^{n'}, V_t)(\mathbf{1}_{U_t^{n'}=d} - \mathbf{1}_{U_{t-1}^{n'}=d}) - \sum_{n'=1}^n (\mathbf{1}_{U_t^{n'}=d} - \mathbf{1}_{U_{t-1}^{n'}=d}) \int_0^1 z\nu_{t,Y^{n'},V_t}(dz) \big]$. Conditional on $\mathcal{F}_t$, the sequence $Q^{N,1}, Q^{N,2}, \ldots$ is a martingale. The sequence also has bounded differences since:

$$\begin{aligned} |Q^{N,n} - Q^{N,n-1}| &= \Big| \frac{1}{\sqrt{N}} [\ell_t^n(Y^n, V_t)(\mathbf{1}_{U_t^{n'}=d} - \mathbf{1}_{U_{t-1}^n=d}) - (\mathbf{1}_{U_t^n=d} - \mathbf{1}_{U_{t-1}^n=d}) \int_0^1 z\nu_{t,Y^{n'},V_t}(dz)] \Big| \\ &\leq \frac{2}{\sqrt{N}}. \end{aligned} \tag{60}$$

Therefore, conditional on $\mathcal{F}_t$, $Q^{N,n}$ satisfies the Azuma-Hoeffding inequality. Note that $Q^{N,N} = \sqrt{N}\mathcal{Z}_t^{N,V}$. Then, by the Azuma-Hoeffding inequality:

$$\begin{aligned} \mathbb{P}\Big[|\sqrt{N}\mathcal{Z}_t^{N,V}| > \alpha \Big| V \Big] &= \mathbb{E}\Big[\mathbb{P}\Big[|\sqrt{N}\mathcal{Z}_t^{N,V}| > \alpha \Big| \mathcal{F}_t \Big] \Big| V \Big] \\ &\leq \mathbb{E}\Big[2\exp(-\frac{1}{8}\alpha^2) \Big| V \Big] \\ &= 2\exp(-\frac{1}{8}\alpha^2). \end{aligned} \tag{61}$$

Combining the bounds (59) and (61), we have by the standard union bound that:

$$\mathbb{P}\Big[|c_t \Lambda_t^N| \geq \alpha \Big| V \Big] \leq B_{1,t,u} \exp\big(- B_{2,t,u}\alpha^2 \big). \tag{62}$$

Finally, again using the union bound, bounds (62) and (56) can be combined to prove the bound (50). ☐

A.5. **Existence and Uniqueness for LLN and CLT.**

Lemma A.8. *The solution $\bar{\mu} \in B^{T+1}$ to the law of large numbers equation in Theorem 3.2 is unique, where $B = \mathcal{P}(\mathcal{U} \times \mathbb{R}^{d_Y})$.*

Proof. Recall that $\bar{\mu}^v$ is the solution to equation (19), which is the LLN (4) conditional on $V = v$. Suppose that $\bar{\mu}^v_{0:t-1}$ is unique; then, it can be proved by contradiction that $\bar{\mu}^v_t$ must be unique as well. Suppose there are different solutions $\bar{\mu}_t^{1,v}$ and $\bar{\mu}_t^{2,v}$, and let $\nu^v = \bar{\mu}_t^{1,v} - \bar{\mu}_t^{2,v}$. From equation (20), this implies that $\langle \phi, \nu^v \rangle_B = 0$ for every bounded, continuous ϕ. Since $\phi \in C_b(\mathcal{U} \times \mathbb{R}^{d_Y})$ is separating (see [6] or [48]) for $\mathcal{P}(\mathcal{U} \times \mathbb{R}^{d_Y})$, $\bar{\mu}_t^{1,v} = \bar{\mu}_t^{2,v}$. This is a contradiction and therefore $\bar{\mu}_t^v$ is unique. Since $\bar{\mu}_0$ is unique (by Assumption 3.1), $\bar{\mu}^v$ is unique by induction.

Of course, existence of a solution in B^{T+1} for (19) follows from the fact that h_θ is a probability transition function and induction. Assume $\bar{\mu}_t^v \in B$. Since h_θ is a probability transition function, $\bar{\mu}_{t+1}^v \in B$ as well. By assumption, $\bar{\mu}_0 \in B$ and therefore we have existence of a solution $\bar{\mu}^v \in B^{T+1}$ by induction.

Finally, since $\bar{\mu} = \bar{\mu}^V$, it immediately follows that $\bar{\mu}$ exists and is unique. ☐

Lemma A.9. *There exists a unique solution $\bar{\Xi} \in W^{T+1}$ to equation (28) in Theorem 3.4.*

Proof. We first show existence of $\bar{\Xi}^v$ in W^{T+1} for any v. Suppose $\bar{\Xi}^v_{t'<t} \in W^t$. We now seek to show that $\bar{\Xi}^v_t \in W$ or, equivalently, $\bar{\Xi}^v_t(u,dy) \in S'$ for each u. We recall (28):

$$\begin{aligned} \bar{\Xi}_t^v(u,dy) &= \sum_{u' \in \mathcal{U}} h_\theta(u,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1}) \bar{\Xi}^v_{t-1}(u',dy) \\ &+ \sum_{u' \in \mathcal{U}} \big(\frac{\partial}{\partial H} h_\theta(u,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1}) \cdot \bar{E}^v_{t-\tau:t-1} \big) \bar{\mu}^v_{t-1}(u',dy) + \bar{\mathcal{M}}^v_t(u,dy). \end{aligned}$$

Using the assumption that $\bar{\Xi}^v_{t'<t} \in W^t$, h_θ is smooth, and $\mathcal{Y}$ is compact, it immediately follows via defintion (29) that the first two terms of are in S'.

It remains to verify that $\bar{\mathcal{M}}_t^v(u,y) \in S'$ for each u. For this, we use the Bochner-Minlos theorem (see [50]) for S' which states that a necessary and sufficient condition for the existence of a random variable X in $S'(\mathbb{R}^{d_Y})$ with a characteristic functional $g(\phi) = \mathbb{E}[\exp(i\langle \phi, X\rangle)]$ is:

(i) $g(0) = 1$

(ii) g is positive definite in the sense that $\sum_{j,l=1}^n z_j \bar{z}_l g(\phi_j - \phi_l) \geq 0$ for any $z_j \in \mathbb{C}$, $\phi_j \in S$.

(iii) g is pseudo-continuous.[16]

Since $\bar{\mathcal{M}}_t^v(u,y)$ is Gaussian where its covariance is known in closed-form (see Lemma A.4), the characteristic functional for $\bar{\mathcal{M}}_t^v(u,y)$ is:

$$
\begin{aligned}
g_{\bar{\mathcal{M}}^v}(\phi) &= \mathbb{E}\bigg[\exp\bigg(i\left\langle \phi(y), \bar{\mathcal{M}}_t^v(u,y)\right\rangle\bigg)\bigg] \\
&= \exp\bigg(-\frac{1}{2}\sum_{u'\in\mathcal{U}}\left\langle \phi(y)^2, h_\theta(u,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1})(1-h_\theta(u,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1})\bar{\mu}^v_{t-1}(u',dy)\right\rangle\bigg). \qquad (63)
\end{aligned}
$$

Property (i) is trivially satisfied. To show Property (iii), first note that

$$
h_\theta(u,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1})(1-h_\theta(u,u',y,v_{t-1},\bar{H}^v_{t-\tau:t-1})\bar{\mu}^v_{t-1}(u',dy) \equiv \eta(u',dy) \in S'.
$$

Since $\eta(u',dy) \in S'$, $\langle \cdot, \eta(u',dy)\rangle : S \to \mathbb{R}$ is a continuous linear operator. Since S is metrizable, it suffices to check sequential continuity in order to prove Property (iii). Recall that $\eta(u',dy)$ has compact support on the compact set $\mathcal{Y}$. Therefore, $\phi_m^2 \to \phi^2$ on $\mathcal{Y}$ since the product operation $C_c^\infty(\mathbb{R}^{d_Y}) \times C_c^\infty(\mathbb{R}^{d_Y}) \to C_c^\infty(\mathbb{R}^{d_Y})$ is continuous where C_c^∞ is the space of smooth functions with compact support. Since $\langle \cdot, \eta(u',dy)\rangle$ is a continuous linear operator, $\phi_m \to \phi$ implies that $\left\langle \phi_m(y)^2, \eta(u',dy)\right\rangle \to \left\langle \phi(y)^2, \eta(u',dy)\right\rangle$ for any v. This in turn implies that $\exp\bigg(-\frac{1}{2}\sum_{u'\in\mathcal{U}}\left\langle \phi_m(y)^2, \eta(u',dy)\right\rangle\bigg) \to \exp\bigg(-\frac{1}{2}\sum_{u'\in\mathcal{U}}\left\langle \phi(y)^2, \eta(u',dy)\right\rangle\bigg)$ for every V. The distribution $\eta(u',dy)$ is nonnegative and $\phi_m(y)^2 > 0$, so $\frac{1}{2}\sum_{u'\in\mathcal{U}}\left\langle \phi_m(y)^2, \eta(u',dy)\right\rangle \geq 0$ and $\exp\bigg(-\frac{1}{2}\sum_{u'\in\mathcal{U}}\left\langle \phi_m(y)^2, \eta(u',dy)\right\rangle\bigg) \leq 1$. Therefore, $g_{\bar{\mathcal{M}}^v}(\phi_m) \to g_{\bar{\mathcal{M}}^v}(\phi)$ by the dominated convergence theorem, which proves Property (iii).

It remains to show Property (ii). We need to show that the following quantity is nonnegative for any $z_j \in \mathbb{C}$ and $\phi_j \in S$, $j = 1, \ldots, n$:

$$
\sum_{j,l}^n z_j \bar{z}_l \exp\bigg(-\frac{1}{2}\sum_{u'\in\mathcal{U}}\left\langle (\phi_j - \phi_l)^2, \eta(u',dy)\right\rangle\bigg) \qquad (64)
$$

Let the matrix Σ have elements $\Sigma_{l,j} = \sum_{u'\in\mathcal{U}}\langle \phi_j\phi_l, \eta(u',dy)\rangle$. The matrix Σ is positive semi-definite:

$$
\begin{aligned}
\sum_{l,j}^n z_j\bar{z}_l\Sigma_{l,j} &= \sum_{l,j}^n z_j\bar{z}_l \sum_{u'\in\mathcal{U}}\langle \phi_j\phi_l, \eta(u',dy)\rangle = \sum_{u'\in\mathcal{U}}\bigg\langle \sum_{l,j}^n z_j\bar{z}_l\phi_j\phi_l, \eta(u',dy)\bigg\rangle \\
&= \sum_{u'\in\mathcal{U}}\bigg\langle \Big|\sum_{l}^n z_l\phi_l\Big|^2, \eta(u',dy)\bigg\rangle \geq 0. \qquad (65)
\end{aligned}
$$

The last inequality comes from $\eta(u',dy) \geq 0$. Consider the mean-zero Gaussian random variables $Z = (Z_1, \ldots, Z_n)$ with covariance matrix $\Sigma_{l,j}$. Z's characteristic functional is positive-definite, meaning that:

$$
\sum_{l,j}^n z_j\bar{z}_l\mathbb{E}[\exp\bigg(i(e_l - e_j)\cdot Z\bigg)] \geq 0, \qquad (66)
$$

for $e_j \in \mathbb{R}^n$. Let $e_j = (\mathbf{1}_{j=1}, \mathbf{1}_{j=2}, \ldots, \mathbf{1}_{j=n})$. Then, (66) is exactly the quantity (64), which proves Property (ii). Therefore, assuming $\bar{\Xi}^v_{t'<t} \in W^t$, $\bar{\mathcal{M}}_t^v(u,y) \in S'$. By induction and $\bar{\Xi}_0^v \in W$, $\bar{\Xi}_t^v(u,dy) \in S'$ for each u.

The uniqueness of the solution can proven by supposing there are two solutions $\bar{\Xi}_t^{1,v}$ and $\bar{\Xi}_t^{2,v}$. Let their difference be $\nu = \bar{\Xi}_t^1 - \bar{\Xi}_t^2$. Due to the linearity of (28), substituting ν into equation (28) yields that $\langle \phi, \nu\rangle_{\mathcal{U}\times\mathbb{R}^{d_Y}} = 0$ for every ϕ, which implies uniqueness.

[16] A function $f : E \to \mathbb{R}$ is pseudo-continuous if its restriction to any finite-dimensional subspace of E is continuous.

Finally, since $\bar{\Xi} = \bar{\Xi}^V$, it immediately follows that $\bar{\Xi}$ exists and is unique. $\square$

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:

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

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:

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

where $\bar{\mu}_t^{\Psi,l}$ is the law of large numbers under the quadrature scheme, $\bar{\mu}_t^l$ is the law of large numbers conditional on the path V^l, and we define $H^\Psi = \sum_{u\in\mathcal{U}} \int_{\mathcal{Y}} f_\theta^h(u,y) \bar{\mu}^\Psi(u, dy)$. 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 $\bar{\mu}$:

$$\text{MSE} = \mathbb{E}[(\mathbb{E}[g(\langle f, \bar{\mu}_t \rangle_{\mathcal{U}\times\mathcal{Y}})] - \frac{1}{L}\sum_{l=1}^{L} g(m^l))^2] \tag{67}$$

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 $u \neq c$ (i.e., all of the pool is initially in the $\{current\}$ 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_θ is twice continuously differentiable.*

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

$$\begin{aligned}
\zeta_t^{l,H}(u, y) &= \sum_{u'\in\mathcal{U}} h_\theta(u, u', y, V_{t-1}^l, H_{t-1}) \zeta_{t-1}^{l,H}(u', y), \\
\zeta_0^{l,H}(\mathrm{c}, y) &= 1, \\
\zeta_0^{l,H}(u, y) &= 0, \quad u \neq \mathrm{c}.
\end{aligned} \tag{68}$$

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

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

$$
\begin{aligned}
&+ \quad \sum_{u'\in\mathcal{U}} |\frac{\partial h_\theta}{\partial H}(u,u',y_k,V^l_{t-1},h^*)||\bar{H}_{t-1}-H^\Psi_{t-1}|,\\
&\leq \quad \sum_{u'\in\mathcal{U}} h_\theta(u,u',y_k,V^l_{t-1},\bar{H}_{t-1})|\zeta^{l,\bar{H}}_{t-1}(u',y_k)-\zeta^{l,H^\Psi}_{t-1}(u',y_k)|+K_1|\bar{H}_{t-1}-H^\Psi_{t-1}|\\
&\leq \quad \max_{u,y_k}|\zeta^{l,\bar{H}}_{t-1}(u,y_k)-\zeta^{l,H^\Psi}_{t-1}(u,y_k)|+K_1|\bar{H}_{t-1}-H^\Psi_{t-1}| \qquad (69)
\end{aligned}
$$

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:

$$
\max_{u,y_k}|\zeta^{l,\bar{H}}_{t}(u,y_k)-\zeta^{l,H^\Psi}_{t}(u,y_k)| \leq \max_{u,y_k}|\zeta^{l,\bar{H}}_{t-1}(u,y_k)-\zeta^{l,H^\Psi}_{t-1}(u,y_k)|+K_1|\bar{H}_{t-1}-H^\Psi_{t-1}|. \qquad (70)
$$

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

$$
H^\Psi_t=\sum_{u\in\mathcal{U}}\sum_{k=1}^{K} f^h_\theta(u,y_k)\bar{\mu}^{\Psi,l}_t(u,c_k)=\sum_{k=1}^{K}\sum_{u\in\mathcal{U}} f(u,y_k)\zeta^{l,H^\Psi}_{t}(u,y_k)\bar{\mu}_0(\mathrm{c},c_k).
$$

Using a Taylor expansion:

$$
\begin{aligned}
|\bar{H}_t-H^\Psi_t| \quad &= \quad |\left\langle f^H,\bar{\mu}^l_t\right\rangle_{\mathcal{U}\times\mathcal{Y}}-\sum_{u\in\mathcal{U}}\sum_{k=1}^{K} f^h_\theta(u,y_k)\bar{\mu}^{\Psi,l}_t(u,c_k)|\\
&\leq \quad \sum_{u\in\mathcal{U}}\sum_{k=1}^{K}\int_{c_k}|f^h_\theta(u,y)\zeta^{l,\bar{H}}_t(u,y)-f^h_\theta(u,y_k)\zeta^{l,H^\Psi}_t(u,y_k)|\bar{\mu}_0(\mathrm{c},dy)\\
&\leq \quad \sum_{u\in\mathcal{U}}\sum_{k=1}^{K}\int_{c_k}|f^h_\theta(u,y)\zeta^{l,\bar{H}}_t(u,y)-f^h_\theta(u,y_k)\zeta^{l,\bar{H}}_t(u,y_k)|\bar{\mu}_0(\mathrm{c},dy)\\
&+ \quad \sum_{u\in\mathcal{U}}\sum_{k=1}^{K}\int_{c_k}|f^h_\theta(u,y_k)\zeta^{l,\bar{H}}_t(u,y_k)-f^h_\theta(u,y_k)\zeta^{l,H^\Psi}_t(u,y_k)|\bar{\mu}_0(\mathrm{c},dy).\\
&\leq \quad 2\sum_{i=1}^{d_Y} r_{K,i}\sum_{u\in\mathcal{U}}\sup_{y\in\mathcal{Y}}|\frac{\partial}{\partial y_i}[f^h_\theta(u,y)\zeta^{l,\bar{H}}_t(u,y)]|+C_3\max_{u,y_k}|\zeta^{l,\bar{H}}_t(u,y_k)-\zeta^{l,H^\Psi}_t(u,y_k)|\\
&\equiv \quad 2\sum_{i=1}^{d_Y} C_{1,t,i}(X^l)r_{K,i}+C_3\max_{u,y_k}|\zeta^{l,\bar{H}}_t(u,y_k)-\zeta^{l,H^\Psi}_t(u,y_k)| \qquad (71)
\end{aligned}
$$

Returning to equation (70), we now have:

$$
\begin{aligned}
\max_{u,y_k}|\zeta^{l,\bar{H}}_{t}(u,y_k)-\zeta^{l,H^\Psi}_{t}(u,y_k)| \quad &\leq \quad \max_{u,y_k}|\zeta^{l,\bar{H}}_{t-1}(u,y_k)-\zeta^{l,H^\Psi}_{t-1}(u,y_k)|\\
&+ \quad 2K_1\sum_{i=1}^{d_Y} C_{1,t-1,i}(V^l)r_{K,i}+C_3K_1\max_{u,y_k}|\zeta^{l,\bar{H}}_{t-1}(u,y_k)-\zeta^{l,H^\Psi}_{t-1}(u,y_k)|\\
&\leq \quad 2K_1\sum_{i=1}^{d_Y} C_{1,t-1,i}(V^l)r_{K,i}+(1+C_3K_1)\sum_{t=0}^{t-1}\max_{u,y_k}|\zeta^{l,\bar{H}}_{t'}(u,y_k)-\zeta^{l,H^\Psi}_{t'}(u,y_k)|\\
&\leq \quad 2K_1\sum_{i=1}^{d_Y} C_{1,t-1,i}(V^l)r_{K,i}+2K_1(1+C_3K_1)\sum_{t'=0}^{t-1}e^{(t-t'-1)(1+C_3K_1)}\sum_{i=1}^{d_Y}C_{1,t',i}(V^l)r_{K,i},\\
&\equiv \quad \sum_{i=1}^{d_Y} C_{2,t,i}(V^l)r_{K,i} \qquad (72)
\end{aligned}
$$

where we have used Gronwall's lemma for the last inequality. We note that numerical scheme converges as the size of the cells $r_{K,1}, \dots, r_{K,d_Y} \to 0$. 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:

$$|g(\langle f, \bar{\mu}_t^l \rangle_{\mathcal{U}\times\mathcal{Y}}) - g(\left\langle f, \bar{\mu}_t^{l,\Psi} \right\rangle_{\mathcal{U}\times\mathcal{Y}})| \le \sum_{i=1}^{d_Y} C_{3,t,i}(V^l) r_{K,i}. \tag{73}$$

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:

$$\begin{aligned}
\mathbb{E} \quad & [(\mathbb{E}[g(\langle f, \bar{\mu}_t \rangle_{\mathcal{U}\times\mathcal{Y}})] - \frac{1}{L}\sum_{l=1}^{L} g(m^l))^2] \le \mathbb{E}[(\mathbb{E}[g(\langle f, \bar{\mu}_t \rangle_{\mathcal{U}\times\mathcal{Y}})] - \frac{1}{L}\sum_{l=1}^{L} g(\langle f, \bar{\mu}_t^l \rangle_{\mathcal{U}\times\mathcal{Y}}))^2] \\
+ \quad & \mathbb{E}[(\frac{1}{L}\sum_{l=1}^{L} g(\langle f, \bar{\mu}_t^l \rangle_{\mathcal{U}\times\mathcal{Y}}) - \frac{1}{L}\sum_{l=1}^{L} g(m^l))^2] \le \underbrace{\frac{1}{L}\mathrm{Var}[g(\langle f, \bar{\mu}_t \rangle_{\mathcal{U}\times\mathcal{Y}})]}_{variance} + \underbrace{C_5 d_Y \sum_{i=1}^{d_Y} \mathbb{E}[C_{3,t,i}(V)^2] r_{K,i}^2}_{bias}
\end{aligned} \tag{74}$$

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 (74), the mean-squared error of the numerical approximation converges to zero as $L \to \infty$ and $\max_i r_{K,i} \to 0$. Equation (74) 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 (74) 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 $r_{K,1}, \dots, r_{K,K}$ 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 5.7 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 $C_{\text{variance}} = \mathrm{Var}[g(\langle f, \bar{\mu}_t \rangle)]$ 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 $r_{K,1}, \dots, r_{K,K}$ satisfy a system of hyperbolic equations which can be solved explicitly. For instance, in two dimensions ($d_Y = 2$), the optimal choices are:

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

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.

Appendix C. 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 $h_\theta(y) = g_\theta(w)$ 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 $\bar{\nu}(dw) = \bar{\mu}_0(\mathrm{c}, dw)$. In one dimension, this can be done by finding the quantiles of the distribution $\bar{\nu}$. It is assumed that $\int_{\mathbb{R}^{d_W}} \bar{\mu}_0(\mathrm{c}, dw) = 1$.
- 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

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

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.

Appendix D. 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 $\mathcal{R} \in \mathbb{R}^{d_W}$. 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 $\Psi(\mathrm{c}, dw) = \sum_{i=1}^I \delta_{w_i}$.
- 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

$$\mu_t^{N_k}(u, w_i) = z_i \bar{\mu}_t^\Psi(u, w_i) + \sqrt{\frac{z_i}{N_k}} \bar{\Xi}^\Psi(u, w_i), \quad , i = 1, \ldots, I, \tag{75}$$

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 5.6.

Appendix E. Parameter Estimation

The parameter θ specifying the model (1) can be estimated by the method of maximum likelihood. We are 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

$$\begin{aligned}
\mathcal{L}(\theta) = \log \mathbb{P}_\theta(\mathsf{D}_{T,N} \,|\, V, Y) &= \log \mathbb{P}_\theta[Z_1^N, \dots, Z_T^N | V, Y] = \log \prod_{t=1}^{T} \mathbb{P}_\theta[Z_t^N | Z_0^N, \dots, Z_{t-1}^N, V, Y] \\
&= \log \prod_{t=1}^{T} \mathbb{P}_\theta[Z_t^N | Z_{t-1}^N, H_{t-\tau:t-1}^N, V, Y] = \log \prod_{t=1}^{T} \prod_{n=1}^{N} h_\theta(U_t^n, U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N) \\
&= \sum_{t=1}^{T} \sum_{n=1}^{N} \log h_\theta(U_t^n, U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N). \qquad (76)
\end{aligned}$$

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 (76). 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.

References

[1] 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.

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

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

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

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

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

[7] 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.

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

[9] CoreLogic. Riskmodel. Technical report, 2014.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

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

[16] R. Dunsky, X. Zhou, M. Kane, M. Chow, C. Hu, and A. Varrieur. FHFA mortgage analytics platform. Technical report, Federal Housing Finance Agency, 2014.

[17] R. Elul, S. Chomsisengphet, D. Glennon, and R. Hunt. What "triggers" mortgage default? Technical report, Research Department, Federal Reserve Bank of Philadelphia, 2010.

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

[19] Experian. Home-equity indicators with new credit data methods for improved mortgage risk analytics. Technical report, Experian, 2014.

[20] 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.

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

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

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

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

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

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

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

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

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

[30] 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.

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

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

[33] Lending Club. Private Communication, 2015.

[34] 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.

[35] G. Loterman, I. Brown, D. Martens, C. Mues, and B. Baesens. Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting*, 28(1):161–170, 2012.

[36] 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.

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

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

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

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

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

[42] 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.

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

[44] H. Schaefer. *Topological Vector Spaces*. Springer, 2 edition, 1999.

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

[46] 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.

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

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

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

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

[51] 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.
[52] T. Williams. Distributed calculations on fixed-income securities. In *Proceedings of the 2nd Workshop on High Performance Computational Finance*. Portland, OR, November 2009.
[53] 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.
[54] 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