

RISK ANALYSIS FOR LARGE POOLS OF LOANS
— ONLINE APPENDIX —

JUSTIN A. SIRIGNANO AND KAY GIESECKE

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 $(U^{1,v}, \dots, U^{N,v})$ which are indexed by $v = (v_0, \dots, v_T) \in \mathbb{R}^{d_V \times T+1}$:

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

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 $(U_t^{1,v}, \dots, U_t^{N,v}, Y^1, \dots, Y^N)$. Similarly, the loss given default for the n -th loan is $\ell_t^{n,v}(Y^n, v_t)$. The system (1) is simply the original system (2) given a realization v of V . In particular, $U_t^{1,V}, \dots, U_t^{N,V}$ has the same distribution as U_t^1, \dots, U_t^N . Define:

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

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 $\mu^{N,v} = (\mu_0^{N,v}, \mu_1^{N,v}, \dots, \mu_T^{N,v})$ for any $v = (v_0, \dots, v_T) \in \mathbb{R}^{d_V \times T+1}$, which in turn will yield the desired limiting laws for $\mu^N = (\mu_0^N, \mu_1^N, \dots, \mu_T^N)$.

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

Date: August 22, 2017.

random variable \bar{X} if:

$$(3) \quad \mathbb{E}[\Phi(X^N)] \rightarrow \mathbb{E}[\Phi(\bar{X})],$$

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 $\mathcal{S} : \mathcal{P}(\mathcal{U} \times \mathcal{Y}) \rightarrow \mathbb{R}$ be the collection of functions of the form $\Phi(\mu) = \phi_1(\langle f_1, \mu \rangle_E, \dots, \langle f_M, \mu \rangle_E)$ 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 [?] or [?]).

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*

$$(4) \quad \mathcal{M}_t^{1,N,v}(u) = \frac{1}{N} \sum_{n=1}^N \phi(Y^n)(\mathbf{1}_{U_t^{n,v}=u} - h_\theta(u, U_{t-1}^{n,v}, Y^n, v_{t-1}, H_{t-\tau:t-1}^{N,v})),$$

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:

$$(5) \quad \begin{aligned} \text{Var}[\mathcal{M}_t^{1,N,v}(u)] &= \mathbb{E}\left[\left(\frac{1}{N} \sum_{n=1}^N \phi(Y^n)(\mathbf{1}_{U_t^{n,v}=u} - h_\theta(u, U_{t-1}^{n,v}, Y^n, v_{t-1}, H_{t-\tau:t-1}^{N,v}))\right)^2\right] \\ &= \frac{1}{N^2} \sum_{n=1}^N \mathbb{E}\left[\left(\phi(Y^n)(\mathbf{1}_{U_t^{n,v}=u} - h_\theta(u, U_{t-1}^{n,v}, Y^n, v_{t-1}, H_{t-\tau:t-1}^{N,v}))\right)^2\right] \\ &\leq \frac{1}{N^2} \sum_{n=1}^N C \xrightarrow{N \rightarrow \infty} 0. \end{aligned}$$

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, $\mathcal{M}_t^{1,N,v}(u) \xrightarrow{p} 0$. \square

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

$$(6) \quad \left\langle \phi, \mu_t^{N,v}(u, dy) \right\rangle = \sum_{u' \in \mathcal{U}} \left\langle \phi(y) h_\theta(u, u', y, V_{t-1}, H_{t-\tau:t-1}^{N,v}), \mu_{t-1}^{N,v}(u', dy) \right\rangle + \mathcal{M}_t^{1,N,v}(u).$$

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 $V = v$:

$$(7) \quad \begin{aligned} \bar{\mu}_t^v(u, dy) &= \sum_{u' \in \mathcal{U}} h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v) \bar{\mu}_{t-1}^v(u', dy), \\ \bar{H}_t^v &= \langle f^H, \bar{\mu}_t^v \rangle_{\mathcal{U} \times \mathcal{Y}} \\ \bar{\mu}_0^v &= \bar{\mu}_0. \end{aligned}$$

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 :

$$(8) \quad \lim_{N \rightarrow \infty} \left\langle \phi, \mu_t^{N,v}(u, dy) \right\rangle = \sum_{u' \in \mathcal{U}} \left\langle \phi(y) h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v), \bar{\mu}_{t-1}^v(u', dy) \right\rangle = \langle \phi, \bar{\mu}_t^v(u, dy) \rangle.$$

The result (8) 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,

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

This last fact follows from:

$$(10) \quad \underbrace{\left\langle \phi(y) h_\theta(u, u', y, v_{t-1}, H_{t-\tau:t-1}^{N,v}), \mu_{t-1}^{N,v}(u', dy) \right\rangle - \left\langle \phi(y) h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v), \mu_{t-1}^{N,v}(u', dy) \right\rangle}_{(*)} = \left\langle \phi(y) [h_\theta(u, u', y, v_{t-1}, H_{t-\tau:t-1}^{N,v}) - h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v)], \mu_{t-1}^{N,v}(u', dy) \right\rangle.$$

The function h_θ is a continuous function, by assumption. Since \mathcal{Y} is compact and $H_{t-\tau:t-1}^{N,v}$ 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:

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

since $H_{t-\tau:t-1}^{N,v} \xrightarrow{P} \bar{H}_{t-\tau:t-1}^v$ and the desired result (9) holds.

By Lemma A.8, $\bar{\mu}_t^v$ from (5) is the unique measure which satisfies (8). Let $\Phi \in \mathcal{S}$ where \mathcal{S} is the convergence determining class specified earlier. Then, for any $\Phi \in \mathcal{S}$,

$$\lim_{N \rightarrow \infty} \mathbb{E}[\Phi(\mu_t^{N,v})] = \lim_{N \rightarrow \infty} \mathbb{E}[\phi_1(\langle f_1, \mu_t^{N,v} \rangle_B, \dots, \langle f_M, \mu_t^{N,v} \rangle_B)]$$

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

The above result follows from equation (8), the continuous mapping theorem, $\mathcal{P}(\mathcal{U} \times \mathcal{Y})$ being compact (and thus any continuous function of $\mu_t^N \in \mathcal{P}(\mathcal{U} \times \mathcal{Y})$ is bounded), and the dominated convergence theorem. Therefore, for any v , $\mu_t^{N,v} \xrightarrow{d} \bar{\mu}_t^v$ and, since $\bar{\mu}_t^v$ is deterministic, $\mu_t^{N,v}$ also converges in probability to $\bar{\mu}_t^v$. This also means that $\mu_{0:t}^{N,v} \xrightarrow{P} \bar{\mu}_{0:t}^v$ for any v .

Note that in Assumption 3.1, $\mu_0^{N,v}$ 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:

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

for any continuous bounded $f : \mathcal{P}(\mathcal{U} \times \mathbb{R}^{d_Y}) \rightarrow \mathbb{R}$. We have used above the tower property, dominated convergence, and previous convergence results for $\mu^{N,v}$. \square

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 [?]). 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:

$$(12) \qquad \qquad \qquad \mathbb{E}[\exp(i\alpha\phi(X^N))] \rightarrow \mathbb{E}[\exp(i\alpha\phi(\bar{X}))],$$

for every $\phi \in S$.¹ Of course, (12) is implied by:

$$(13) \qquad \qquad \qquad \phi(X^N) \xrightarrow{d} \phi(\bar{X}),$$

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

$$S(\mathbb{R}^{d_Y}) = \{\phi \in C^\infty(\mathbb{R}^{d_Y}) : \|\phi\|_{\alpha,\beta} < \infty \ \forall \ \alpha, \beta \in \mathbb{N}\},$$

¹This result was originally proven for both the space of tempered distributions S' and the space of distributions D' in [?]. The result was also independently proven by [?] (Proposition 3.3 and corresponding Supplementary Comments on pg. 247), [?], and [?]. Finally, such results have been extended to continuous-time processes taking values in S' and D' , see [?] and [?]. See [?] for a similar result for separable Banach spaces.

$$\|\phi\|_{\alpha,\beta} = \sup_{y \in \mathbb{R}^{d_Y}} |y^\alpha D^\beta \phi(y)|.$$

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.

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 (12) or (13) for each $\phi \in S$:

$$(14) \quad \langle \phi(y), \Xi^N(u, dy) \rangle \xrightarrow{d} \langle \phi(y), \bar{\Xi}(u, dy) \rangle,$$

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

Since the dual of the cartesian product of finitely many spaces is the cartesian product of the duals, the criterion (12) or (13) also of course covers the case of $\Xi^N \in \prod_{i=1}^{T \times |\mathcal{U}|} S'$, where it is sufficient to show (12) for $\phi \in \prod_{i=1}^{T \times |\mathcal{U}|} S$, or more simply convergence in distribution for the random vector $(\langle \phi_{t,u}, \Xi_t^N(u, dy) \rangle)_{t \in I, u \in \mathcal{U}}$ for every $(\phi_{t,u})_{t \in I, u \in \mathcal{U}} \in \prod_{i=1}^{T \times |\mathcal{U}|} S$.

Define:

$$(15) \quad \Xi_t^{N,v} = \sqrt{N}(\mu_t^{N,v} - \bar{\mu}_t^v),$$

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

$$(16) \quad \begin{aligned} \bar{\Xi}_t^v(u, dy) &= \sum_{u' \in \mathcal{U}} h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v) \bar{\Xi}_{t-1}^v(u', dy) \\ &+ \sum_{u' \in \mathcal{U}} \left(\frac{\partial}{\partial H} h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v) \cdot \bar{E}_{t-\tau:t-1}^v \bar{\mu}_{t-1}^v(u', dy) + \bar{\mathcal{M}}_t^v(u, dy) \right), \\ \bar{\Xi}_0^v &= \bar{\Xi}_0. \end{aligned}$$

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

$$\begin{aligned} \text{Cov}[\bar{\mathcal{M}}_t^v(u_1, dy), \bar{\mathcal{M}}_t^v(u_2, dy)] &= - \sum_{u' \in \mathcal{U}} h_\theta(u_1, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v) h_\theta(u_2, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v) \bar{\mu}_{t-1}^v(u', dy), \\ \text{Var}[\bar{\mathcal{M}}_t^v(u, dy)] &= \sum_{u' \in \mathcal{U}} h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v) (1 - h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v)) \bar{\mu}_{t-1}^v(u', dy), \end{aligned}$$

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

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

$$(17) \quad \langle \phi, X \rangle \leq C_k \|\phi\|_k,$$

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

$$(18) \quad \|\phi\|_k = \max_{|\alpha|+|\beta| \leq k} \sup_{y \in \mathbb{R}^{d_Y}} |y^\alpha D^\beta \phi(y)|.$$

Then, we certainly have that:

$$(19) \quad \langle \phi, \Xi_t^N \rangle = \sqrt{N} (\langle \phi, \mu_t^{N,v} \rangle - \langle \phi, \bar{\mu}_t^v \rangle) \leq 2\sqrt{N} \sup_{y \in \mathcal{Y}} |\phi(y)| \leq 2\sqrt{N} \|\phi\|_0.$$

This completes the proof. □

Lemma A.3. For any $\phi \in S$, α , and v , the martingale difference array²

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

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

- (i) $\sup_N \mathbb{E}[(\max_{n \leq N} Z_t^{N,n,v}(\alpha))^2]$ is uniformly bounded,
- (ii) $\max_{n \leq N} |Z_t^{N,n,v}(\alpha)| \xrightarrow{P} 0$,
- (iii) For each v , $\sum_{n=1}^N (Z_t^{N,n,v}(\alpha))^2 \xrightarrow{P} \text{Var}[\sum_{u \in \mathcal{U}} \alpha_u \langle \phi, \mathcal{M}_t^v(u, dy) \rangle]$.

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 ,

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

²See [?] for details on martingale difference arrays.

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

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

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

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 (20) 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. \square

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

$$(21) \quad \mathcal{M}_t^{2,N,v}(u, dy) = \frac{1}{\sqrt{N}} \sum_{n=1}^N [\delta_{(U_t^{n,v}, Y^n)}(u, dy) - h_\theta(u, U_{t-1}^{n,v}, Y^n, v_{t-1}, H_{t-\tau:t-1}^{N,v}) \delta_{Y^n}(dy)].$$

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:

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

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

Proof. We first show that:

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

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

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

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

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

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

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

$$\begin{aligned} & \mathbb{E}[\exp(i\alpha_1 \langle \phi_1, \mathcal{M}_{t_1}^{2,N,v}(u, dy) \rangle) + i\alpha_2 \langle \phi_2, \mathcal{M}_{t_2}^{2,N,v}(u', dy) \rangle] \\ (22) \quad &= \mathbb{E}[\exp(i\alpha_1 \langle \phi_1, \mathcal{M}_{t_1}^{2,N,v}(u, dy) \rangle)] \mathbb{E}[\exp(i\alpha_2 \langle \phi_2, \mathcal{M}_{t_2}^{2,N,v}(u', dy) \rangle) | \mu_{0:t_2-1}^{N,v}]. \end{aligned}$$

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

$$\begin{aligned} & \mathbb{E}[\exp(i\alpha_2 \langle \phi_2, \mathcal{M}_{t_2}^{2,N,v}(u', dy) \rangle) | \mu_{0:t_2-1}^{N,v}] \\ &= \prod_{n=1}^N \mathbb{E}[e^{\frac{i\alpha_2}{\sqrt{N}} \phi_2(Y^n)(\mathbf{1}_{U_{t_2}^{n,v}=u'} - h_\theta(u', U_{t_2-1}^{n,v}, Y^n, v_{t_2-1}, H_{t_2-\tau:t_2-1}^{N,v}))} | H_{t_2-\tau:t_2-1}^{N,v}, U_{t_2-1}^{n,v}, Y^n] \\ &= \prod_{n=1}^N (1 - \frac{\alpha_2^2}{2N} \mathbb{E}[\phi_2(Y^n)^2 (\mathbf{1}_{U_{t_2}^{n,v}=u'} - h_\theta(u', U_{t_2-1}^{n,v}, Y^n, v_{t_2-1}, H_{t_2-\tau:t_2-1}^{N,v}))^2 | H_{t_2-\tau:t_2-1}^{N,v}, U_{t_2-1}^{n,v}, Y^n] + G_1^N), \\ &= \prod_{n=1}^N (1 - \frac{\alpha_2^2}{2N} \gamma(v, H_{t_2-\tau:t_2-1}^{N,v}, U_{t_2-1}^{n,v}, Y^n) + G_1^N), \end{aligned}$$

where we have used a Taylor expansion for the exponential, $|G_1^N| \leq \frac{C}{N^{3/2}}$, $\sup_{y \in \mathcal{Y}} 2|\phi_2(y)^3| < C < \infty$, and the quantity $\gamma(v, H_{t_2-\tau:t_2-1}^{N,v}, U_{t_2-1}^{n,v}, Y^n) = \mathbb{E}[\phi_2(Y^n)^2 (\mathbf{1}_{U_{t_2}^{n,v}=u'} - h_\theta(u', U_{t_2-1}^{n,v}, Y^n, v_{t_2-1}, H_{t_2-\tau:t_2-1}^{N,v}))^2 | H_{t_2-\tau:t_2-1}^{N,v}, U_{t_2-1}^{n,v}, Y^n]$.

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:

$$\begin{aligned} & \left| \prod_{n=1}^N (1 - \frac{\alpha_2^2}{2N} \gamma(v, H_{t_2-\tau:t_2-1}^{N,v}, U_{t_2-1}^{n,v}, Y^n) + G_1^N) - \prod_{n=1}^N (1 - \frac{1}{2N} \alpha_2^2 \gamma(v, H_{t_2-\tau:t_2-1}^{N,v}, U_{t_2-1}^{n,v}, Y^n)) \right| \\ (23) \quad & \leq \frac{C}{N^{1/2}} \xrightarrow{a.s.} 0. \end{aligned}$$

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

$$\begin{aligned} & \prod_{n=1}^N (1 - \frac{1}{2N} \alpha_2^2 \gamma(v, H_{t_2-\tau:t_2-1}^{N,v}, U_{t_2-1}^{n,v}, Y^n)) = \exp(\sum_{n=1}^N \log(1 - \frac{1}{2N} \alpha_2^2 \gamma(v, H_{t_2-\tau:t_2-1}^{N,v}, U_{t_2-1}^{n,v}, Y^n))) \\ &= \exp(\sum_{n=1}^N [-\frac{1}{2N} \alpha_2^2 \gamma(v, H_{t_2-\tau:t_2-1}^{N,v}, U_{t_2-1}^{n,v}, Y^n) + G_2^N]) \\ &= \exp(-\frac{1}{2} \alpha_2^2 \sum_{u'' \in \mathcal{U}} \langle \gamma(v, H_{t_2-\tau:t_2-1}^{N,v}, u'', y), \mu_{t_2-1}^{N,v}(u'', dy) \rangle + N G_2^N) \xrightarrow{p} \exp(-\frac{1}{2} \alpha_2^2 \text{Var}[\langle \phi_2, \bar{\mathcal{M}}_{t_2}^v \rangle]), \end{aligned}$$

where we have again used a Taylor expansion and $|G_2^N| \leq \frac{C}{N^2}$. 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 $\gamma(v, H_{t_2-\tau:t_2-1}^{N,v}, U_{t_2-1}^{n,v}, Y^n) = \phi_2(Y^n)^2 h_\theta(u, U_{t_2-1}^{n,v}, Y^n, v_{t_2-1}, \bar{H}_{t_2-\tau:t_2-1}^v)(1 - h_\theta(u, U_{t_2-1}^{n,v}, Y^n, v_{t_2-1}, \bar{H}_{t_2-\tau:t_2-1}^v))$, and the continuous mapping theorem.

With this result in hand, we return to equation (22). Since $\mathbb{E}[\exp(i\alpha_2 \langle \phi_2, \mathcal{M}_{t_2}^{2,N,v}(u') \rangle) | \mu_{0:t_2-1}^{N,v}]$ converges to a constant and $\exp(i\alpha_1 \langle \phi_1, \mathcal{M}_{t_1}^{2,N,v}(u) \rangle) \xrightarrow{d} \exp(i\alpha_1 \langle \phi_1, \bar{\mathcal{M}}_{t_1}^v \rangle)$, Slutsky's theorem yields that $\exp(i\alpha_1 \langle \phi_1, \mathcal{M}_{t_1}^{2,N,v}(u, dy) \rangle) \mathbb{E}[\exp(i\alpha_2 \langle \phi_2, \mathcal{M}_{t_2}^{2,N,v}(u', dy) \rangle) | \mu_{0:t_2-1}^{N,v}]$ converges in distribution to the quantity $\exp(i\alpha_1 \langle \phi_1, \bar{\mathcal{M}}_{t_1}^v \rangle) \exp(-\frac{1}{2} \alpha_2^2 \text{Var}[\langle \phi_2, \bar{\mathcal{M}}_{t_2}^v \rangle])$. By the bounded convergence theorem, we then have that:

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

Consequently, $\langle \phi_1, \bar{\mathcal{M}}_{t_1}^v \rangle$ is independent of $\langle \phi_2, \bar{\mathcal{M}}_{t_2}^v \rangle$ 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 (12). \square

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

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

$$\begin{aligned} & \mathbb{E}[\exp(i \sum_{t' < t, u \in \mathcal{U}} \langle \phi_{1,u,t'}, \bar{\Xi}_{t'}^{N,v}(u, dy) \rangle + i \langle \phi_2, \mathcal{M}_t^{2,N,v}(u', dy) \rangle)] \\ (24) \quad &= \mathbb{E}[\exp(i \sum_{t' < t, u \in \mathcal{U}} \langle \phi_{1,u,t'}, \bar{\Xi}_{t'}^{N,v}(u, dy) \rangle) \mathbb{E}[\exp(i \langle \phi_2, \mathcal{M}_t^{2,N,v}(u', dy) \rangle) | \mu_{0:t-1}^{N,v}]]. \end{aligned}$$

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

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

$$(25) \quad \begin{aligned} \left\langle \phi(y), \Xi_t^{N,v}(u, dy) \right\rangle &= \sqrt{N} \sum_{u' \in \mathcal{U}} \left[\left\langle \phi(y), h_\theta(u, u', y, v_{t-1}, H_{t-\tau:t-1}^{N,v}) \mu_{t-1}^{N,v}(u', dy) \right\rangle \right. \\ &\quad \left. - \left\langle \phi(y), h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v) \bar{\mu}_{t-1}^v(u', dy) \right\rangle \right] + \left\langle \phi, \mathcal{M}_t^{2,N,v}(u) \right\rangle, \end{aligned}$$

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

$$(26) \quad \left\langle \phi, \mathcal{M}_t^{2,N,v}(u, dy) \right\rangle = \frac{1}{\sqrt{N}} \sum_{n=1}^N \phi(Y^n) (\mathbf{1}_{U_t^{n,v}=u} - h_\theta(u, U_{t-1}^{n,v}, Y^n, v_{t-1}, H_{t-\tau:t-1}^{N,v})).$$

Using a Taylor expansion, one can obtain

$$(27) \quad \begin{aligned} \left\langle \phi(y), \Xi_t^{N,v}(u, dy) \right\rangle &= \sum_{u' \in \mathcal{U}} \left\langle \phi(y) h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v), \Xi_{t-1}^N(u', dy) \right\rangle \\ &\quad + \sum_{u' \in \mathcal{U}} \left\langle \phi(y) h_{\theta,H}(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v) \cdot E_{t-\tau:t-1}^{N,v}, \mu_{t-1}^{N,v}(u', dy) \right\rangle \\ &\quad + \frac{1}{\sqrt{N}} \mathcal{R}^{N,v} + \left\langle \phi, \mathcal{M}_t^{2,N,v}(u, dy) \right\rangle, \end{aligned}$$

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

$$(28) \quad \begin{aligned} |\mathcal{R}^{N,v}| &= \left| \frac{1}{2} \left\langle \phi(y) (E_{t-\tau:t-1}^{N,v})^\top h_{HH}(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v) E_{t-\tau:t-1}^{N,v}, \mu_{t-1}^{N,v}(u', dy) \right\rangle \right| \\ &\leq \left| \frac{1}{2} \left\langle \sup_{y \in \mathcal{Y}} \phi(y) (E_{t-\tau:t-1}^{N,v})^\top h_{HH}(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v) E_{t-\tau:t-1}^{N,v}, \mu_{t-1}^{N,v}(u', dy) \right\rangle \right| \\ &\leq K \|E_{t-\tau:t-1}^{N,v}\|^2 \left\langle \sup_{y \in \mathcal{Y}} |\phi(y) h_{HH}(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v)|, \mu_{t-1}^{N,v}(u', dy) \right\rangle \\ &\leq \|E_{t-\tau:t-1}^{N,v}\|^2 \left\langle C, \mu_{t-1}^{N,v}(u', dy) \right\rangle = C \|E_{t-\tau:t-1}^{N,v}\|^2. \end{aligned}$$

Here, h_{HH} denotes the second partial derivative of h_θ with respect to its fifth argument. We have bounded $h_{HH}(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v)$ 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 $\|E_{t-\tau:t-1}^{N,v}\|^2 \xrightarrow{d} \|\bar{E}_{t-\tau:t-1}^v\|^2$. 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 $\Xi_{0:t-1}^{N,v}$ converges in distribution to $\bar{\Xi}_{0:t-1}^v$. 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 (27) also converges in distribution since $\mu_{t-1}^{N,v}$ converges in probability to $\bar{\mu}_{t-1}^v$ for any v (from the law of large numbers) and $E_{t-\tau:t-1}^{N,v} \xrightarrow{d} \bar{E}_{t-\tau:t-1}^v$ (since we assumed $\Xi_{0:t-1}^{N,v} \xrightarrow{d} \bar{\Xi}_{0:t-1}^v$ and f^H is continuous).

The third term converges in distribution due to Lemma A.4. Then, assuming $\Xi_{0:t-1}^{N,v}$ converges in distribution to $\bar{\Xi}_{0:t-1}^v$, $\langle \phi, \Xi_t^{N,v} \rangle \xrightarrow{d} \langle \phi, \bar{\Xi}_t^v \rangle$ for any $\phi \in S$ where $\langle \phi, \bar{\Xi}_t^v \rangle$ satisfies the evolution equation:

$$(29) \quad \begin{aligned} \langle \phi(y), \bar{\Xi}_t^v(u, dy) \rangle &= \sum_{u' \in \mathcal{U}} \langle \phi(y) h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}), \bar{\Xi}_{t-1}^v(u', dy) \rangle \\ &+ \sum_{u' \in \mathcal{U}} \left\langle \phi(y) \frac{\partial}{\partial H} h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v) \cdot \bar{E}_{t-\tau:t-1}^v, \bar{\mu}_{t-1}^v \right\rangle + \langle \phi, \bar{\mathcal{M}}_t^v(u, dy) \rangle, \quad \phi \in S. \end{aligned}$$

This is sufficient to show that the $\Xi_t^{N,v} \xrightarrow{d} \bar{\Xi}_t^v$, by the criterion (12). However, to make the stronger statement of the joint convergence $\Xi_{0:t}^{N,v} \xrightarrow{d} \bar{\Xi}_{0:t}^v$, we need an additional fact, namely that $(\Xi_{0:t-1}^{N,v}, \mathcal{M}_t^{2,N,v}) \xrightarrow{d} (\bar{\Xi}_{0:t}^v, \bar{\mathcal{M}}_t^v(u, dy))$. This has been proven in Lemma A.5, and therefore we have the desired result that $\Xi_{0:t}^{N,v} \xrightarrow{d} \bar{\Xi}_{0:t}^v$. By induction, we have that $\bar{\Xi}^{N,v} \xrightarrow{d} \bar{\Xi}^v$ for any v . Therefore, $\bar{\Xi}^N \xrightarrow{d} \bar{\Xi}$. \square

Proof of Corollary 3.7. Since $(\mu_t^{N,v}, L^{N,v}) \xrightarrow{p} (\bar{\mu}^v, \bar{L}^v)$ 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 $(v, \mu_t^{N,v}, L^{N,v}, \Xi^{N,v}, \Lambda^{N,v}) \xrightarrow{d} (v, \bar{\mu}^v, \bar{L}^v, \bar{\Xi}^v, \bar{\Lambda}^v)$. Since this holds for any v , the desired convergence follows. \square

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

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

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

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

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

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

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

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

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 . \square

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:

$$(30) \quad L_t^{N,v} = \left\langle \int_0^1 z \nu_{t,y,v_t}(dz) \mu_t^{N,v}(d, dy) - \mu_{t-1}^{N,v}(d, dy) \right\rangle + \mathcal{Z}_t^{N,v}.$$

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 :

$$(31) \quad \left\langle \int_0^1 z \nu_{t,y,v_t}(dz), \mu_t^{N,v}(d, dy) - \mu_{t-1}^{N,v}(d, dy) \right\rangle \xrightarrow{P} \left\langle \int_0^1 z \nu_{t,y,v_t}(dz), \bar{\mu}_t^v(d, dy) - \bar{\mu}_{t-1}^v(d, dy) \right\rangle,$$

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 (11), this implies that $(\mu^N, L^N) \xrightarrow{d} (\bar{\mu}, \bar{L})$.

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

$$(32) \quad \bar{\Lambda}_t^v = \left\langle \int_0^1 z \nu_{t,y,v_t}(dz), \bar{\Xi}_t^v(d, dy) - \bar{\Xi}_{t-1}^v(d, dy) \right\rangle + \bar{\mathcal{Z}}_t^v,$$

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

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

$\bar{\mathcal{Z}}_{t_1}^v$ 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:

$$(33) \quad \text{Var}[\bar{\mathcal{Z}}_t^v] = \text{Var}[\bar{\mathcal{Z}}_t^v] = \left\langle \int_0^1 z^2 \nu_{t,y,v_t}(dz) - \left(\int_0^1 z \nu_{t,y,v_t}(dz) \right)^2, \bar{\mu}_t^v(d, dy) - \bar{\mu}_{t-1}^v(d, dy) \right\rangle.$$

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

Proof. We have

$$(34) \quad \begin{aligned} & \mathbb{E}[\exp(i \sum_{t' \leq t, u \in \mathcal{U}} \langle \phi_{1,u,t'}, \bar{\Xi}_{t'}^{N,v}(u, dy) \rangle + i \sum_{t' < t} \alpha_{t'} \sqrt{N} Z_{t'}^{N,v} + i \alpha_t \sqrt{N} Z_t^{N,v})] \\ &= \mathbb{E}[\exp(i \sum_{t' \leq t, u \in \mathcal{U}} \langle \phi_{1,u,t'}, \bar{\Xi}_{t'}^{N,v}(u, dy) \rangle + i \sum_{t' < t} \alpha_{t'} \sqrt{N} Z_{t'}^{N,v})] \mathbb{E}[\exp(i \alpha_t \sqrt{N} Z_t^{N,v}) | \mu_{0:t}^N]. \end{aligned}$$

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

Define:

$$(35) \quad \Lambda_t^{N,v} = \sqrt{N}(L_t^{N,v} - \bar{L}_t^v).$$

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

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

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 , $(\Xi^N, \Lambda^N) \xrightarrow{d} (\bar{\Xi}, \bar{\Lambda})$. \square

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

Proof of Theorem 3.9. Note that the bound (11) implies

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

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

$$(36) \quad \mathbb{P} \left[\left| \sum_{t=1}^T \sum_{u \in \mathcal{U}} \langle \phi_{t,u}(y), \Xi_t^N(u, dy) \rangle + \sum_{t=1}^T c_t \Lambda_t^N \right| > \alpha \right] \leq C_1 \exp(-C_2 \alpha^2).$$

For notational convenience, define $X^N = \sum_{t=1}^T \sum_{u \in \mathcal{U}} \langle \phi_{t,u}(y), \Xi_t^N(u, dy) \rangle + \sum_{t=1}^T c_t \Lambda_t^N$. If the bound (36) holds and $K > C_0$, then:

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

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

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

Consequently, the bound (36), $\mathbb{P}\left[|X^N| > \alpha\right] \leq C_1 \exp(-C_2\alpha^2)$ for $\alpha > 0$, implies uniform integrability for $g(X^N)$. Next, recognize that the following bound implies the bound (36):

$$(37) \quad \mathbb{P}\left[|\langle \phi_{t,u}(y), \Xi_t^N(u, dy) \rangle + c_t \Lambda_t^N| > \alpha\right] \leq C_{1,t,u} \exp(-C_{2,t,u}\alpha^2).$$

The bound (37) implies the bound (36) since:

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

It is therefore sufficient to prove an exponentially decaying bound:

$$(38) \quad \mathbb{P}\left[|\langle \phi_{t,u}(y), \Xi_t^N(u, dy) \rangle + c_t \Lambda_t^N| > \alpha\right] \leq C_{1,t,u} \exp(-C_{2,t,u}\alpha^2),$$

for each t and u . We prove this result using induction. From equation (27):

$$(39) \quad \begin{aligned} \langle \phi(y), \Xi_t^N(u, dy) \rangle &= \sum_{u' \in \mathcal{U}} \langle \phi(y) h_\theta(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1}), \Xi_{t-1}^N(u', dy) \rangle \\ &+ \sum_{u' \in \mathcal{U}} \langle \phi(y) h_{\theta,H}(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1}) \cdot E_{t-\tau:t-1}^N \mu_{t-1}^N(u', dy) \rangle \\ &+ \frac{1}{\sqrt{N}} \mathcal{R}^{N,V} + \langle \phi, \mathcal{M}_t^{2,N,V}(u, dy) \rangle, \end{aligned}$$

where $|\mathcal{R}^{N,V}| \leq C \sup_{y \in \mathcal{Y}} |\phi(y)| \|E_{t-\tau:t-1}^N\|^2$ (see equation (28)) 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:

$$(40) \quad \mathbb{P}\left[\left(\sup_{y \in \mathcal{Y}} |\phi(y)|\right)^{-1} |\langle \phi(y), \Xi_{t'}^N(u, dy) \rangle| > \alpha \mid V\right] \leq K_{1,t',u} \exp(-K_{2,t',u}\alpha^2),$$

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 $\mathbb{P}\left[|\langle \phi(y), \Xi_{t'}^N(u, dy) \rangle| > \alpha \mid V\right] \leq C_{1,t',u} \exp(-C_{2,t',u}\alpha^2)$ 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 (39) satisfies:

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

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 (39).

$$\begin{aligned}
& \mathbb{P} \left[\left| \sum_{u' \in \mathcal{U}} \langle \phi(y) h_{\theta,H}(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1}) \cdot E_{t-\tau:t-1}^N, \mu_{t-1}^N(u', dy) \rangle \right| > \alpha \middle| V \right] \\
& \leq \mathbb{P} \left[\sup_{(y,v) \in \mathcal{Y} \times \mathcal{V}} \left| \sum_{u' \in \mathcal{U}} \phi(y) h_{\theta,H}(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1}) \right| \sum_{t'=t-\tau}^{t-1} |E_{t'}^N| > \alpha \middle| V \right] \\
& \leq \mathbb{P} \left[\sup_{y \in \mathcal{Y}} |\phi(y)| \sup_{(y,v) \in \mathcal{Y} \times \mathcal{V}} \left| \sum_{u' \in \mathcal{U}} h_{\theta,H}(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1}) \right| \sum_{t'=t-\tau}^{t-1} |E_{t'}^N| > \alpha \middle| V \right] \\
& \leq \mathbb{P} \left[\sup_{y \in \mathcal{Y}} |\phi(y)| K_0 \sum_{t'=t-\tau}^{t-1} |E_{t'}^N| > \alpha \middle| V \right] \leq C_{1,t,u}^2 \exp \left(- (\sup_{y \in \mathcal{Y}} |\phi(y)|) \right)^{-2} C_{2,t,u}^2 \alpha^2, \\
(42) \quad &
\end{aligned}$$

where $\sup_{(y,v) \in \mathcal{Y} \times \mathcal{V}} \left| \sum_{u' \in \mathcal{U}} h_{\theta,H}(u, u', y, V_{t-1}, \bar{H}_{t-\tau:t-1}) \right| \leq K_0$, 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 (39) can also be exponentially bounded by $C_{1,t,u}^3 \exp \left(- (\sup_{y \in \mathcal{Y}} |\phi(y)|) \right)^{-2} C_{2,t,u}^3 \alpha^2$. The fourth term on the RHS of (39) 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:

$$(43) \quad |S^{N,n} - S^{N,n-1}| = \frac{1}{\sqrt{N}} \left| \phi(Y^n) [\mathbf{1}_{U_t^n \in u} - h_\theta(u, U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N)] \right| \leq \frac{2}{\sqrt{N}} \sup_{y \in \mathcal{Y}} |\phi(y)|.$$

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

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

Combining the exponential bounds for the four terms on the RHS of (39) and applying the standard union bound, we then have a bound for the LHS of (39):

$$(44) \quad \mathbb{P}\left[\left|\langle \phi(y), \Xi_t^N(u, dy) \rangle\right| > \alpha \middle| V\right] \leq C_{1,t,u} \exp\left(-C_{2,t,u} \sup_{y \in \mathcal{Y}} |\phi(y)|^{-2}\alpha^2\right),$$

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

$$(45) \quad \begin{aligned} \mathbb{P}\left[\left(\sup_{y \in \mathcal{Y}} |\phi(y)|\right)^{-1} \left|\langle \phi(y), \Xi_t^N(u, dy) \rangle\right| > \alpha \middle| V\right] &= \mathbb{P}\left[\left|\langle \phi(y), \Xi_t^N(u, dy) \rangle\right| > \left(\sup_{y \in \mathcal{Y}} |\phi(y)|\right)\alpha \middle| V\right] \\ &= C_{1,t,u} \exp\left(-C_{2,t,u}\alpha^2\right) \\ &\equiv K_{1,t,u} \exp\left(-K_{2,t,u}\alpha^2\right). \end{aligned}$$

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

$$(46) \quad \Lambda_t^N = \sqrt{N}(L_t^N - \bar{L}_t) = \left\langle \int_0^1 z \nu_{t,y,V_t}(dz), \Xi_t^N(d, dy) - \Xi_{t-1}^N(d, dy) \right\rangle + \sqrt{N} Z_t^{N,V}.$$

Let's first consider the first term on the RHS of (46). The function $\int_0^1 z \nu_{t,y,V_t}(dz)$ is continuous on y (by assumption) and $\sup_{y \in \mathcal{Y}} \int_0^1 z \nu_{t,y,V_t}(dz) \leq 1$ since ν is a measure on $[0, 1]$. Using the (now proven) bound $\mathbb{P}\left[\left|\langle \phi(y), \Xi_t^N(u, dy) \rangle\right| > \alpha \middle| V\right] \leq C_{1,t,u} \exp\left(-C_{2,t,u}\alpha^2\right)$ and the standard union bound, we have that:

$$(47) \quad \mathbb{P}\left[\left|\left\langle \int_0^1 z \nu_{t,y,V_t}(dz), \Xi_t^N(d, dy) - \Xi_{t-1}^N(d, dy) \right\rangle\right| \geq \alpha \middle| V\right] \leq A_{1,t,u} \exp\left(-A_{2,t,u}\alpha^2\right).$$

Next, let's consider the second term on the RHS of (46), $\sqrt{N} Z_t^{N,V}$. Define $Q^{N,n} = \frac{1}{\sqrt{N}} \left[\sum_{n'=1}^n \ell_t^{n'}(Y^{n'}, V_t) (\mathbf{1}_{U_t^{n'}=d} - \mathbf{1}_{U_{t-1}^{n'}=d}) - \sum_{n'=1}^n (\mathbf{1}_{U_t^{n'}=d} - \mathbf{1}_{U_{t-1}^{n'}=d}) \int_0^1 z \nu_{t,Y^{n'},V_t}(dz) \right]$. Conditional on \mathcal{F}_t , the sequence $Q^{N,1}, Q^{N,2}, \dots$

is a martingale. The sequence also has bounded differences since:

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

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:

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

Combining the bounds (47) and (49), we have by the standard union bound that:

$$(50) \quad \mathbb{P} \left[\left| c_t \Lambda_t^N \right| \geq \alpha \middle| V \right] \leq B_{1,t,u} \exp\left(-B_{2,t,u}\alpha^2\right).$$

Finally, again using the union bound, bounds (50) and (44) can be combined to prove the bound (38). \square

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 (7), which is the LLN (5) conditional on $V = v$. Suppose that $\bar{\mu}_{0,t-1}^v$ is unique; then, it can be proved by contradiction that $\bar{\mu}_t^v$ 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 (8), 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 [?] or [?]) 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 (7) 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. \square

Lemma A.9. *There exists a unique solution $\bar{\Xi} \in W^{T+1}$ to equation (16) in Theorem 3.4.*

Proof. We first show existence of $\bar{\Xi}^v$ in W^{T+1} for any v . Suppose $\bar{\Xi}_{t' < t}^v \in W^t$. We now seek to show that $\bar{\Xi}_t^v \in W$ or, equivalently, $\bar{\Xi}_t^v(u, dy) \in S'$ for each u . We recall (16):

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

Using the assumption that $\bar{\Xi}_{t' < t}^v \in W^t$, h_θ is smooth, and \mathcal{Y} is compact, it immediately follows via definition (17) 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 [?]) 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.³

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:

$$\begin{aligned} g_{\bar{\mathcal{M}}^v}(\phi) &= \mathbb{E} \left[\exp \left(i \langle \phi(y), \bar{\mathcal{M}}_t^v(u, y) \rangle \right) \right] \\ (5\ddagger) \quad &\exp \left(- \frac{1}{2} \sum_{u' \in \mathcal{U}} \langle \phi(y)^2, h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v) (1 - h_\theta(u, u', y, v_{t-1}, \bar{H}_{t-\tau:t-1}^v)) \bar{\mu}_{t-1}^v(u', dy) \rangle \right). \end{aligned}$$

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

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

Since $\eta(u', dy) \in S'$, $\langle \cdot, \eta(u', dy) \rangle : S \rightarrow \mathbb{R}$ 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, $\phi_m^2 \rightarrow \phi^2$ on \mathcal{Y} since the product operation $C_c^\infty(\mathbb{R}^{d_Y}) \times C_c^\infty(\mathbb{R}^{d_Y}) \rightarrow C_c^\infty(\mathbb{R}^{d_Y})$ 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, $\phi_m \rightarrow \phi$ implies that $\langle \phi_m(y)^2, \eta(u', dy) \rangle \rightarrow \langle \phi(y)^2, \eta(u', dy) \rangle$ for any v . This in turn implies that $\exp \left(- \frac{1}{2} \sum_{u' \in \mathcal{U}} \langle \phi_m(y)^2, \eta(u', dy) \rangle \right) \rightarrow \exp \left(- \frac{1}{2} \sum_{u' \in \mathcal{U}} \langle \phi(y)^2, \eta(u', dy) \rangle \right)$ for every V . The distribution $\eta(u', dy)$ is nonnegative and $\phi_m(y)^2 > 0$, so $\frac{1}{2} \sum_{u' \in \mathcal{U}} \langle \phi_m(y)^2, \eta(u', dy) \rangle \geq 0$ and $\exp \left(-$

³A function $f : E \rightarrow \mathbb{R}$ is pseudo-continuous if its restriction to any finite-dimensional subspace of E is continuous.

$\frac{1}{2} \sum_{u' \in \mathcal{U}} \langle \phi_m(y)^2, \eta(u', dy) \rangle \leq 1$. Therefore, $g_{\bar{\mathcal{M}}^v}(\phi_m) \rightarrow g_{\bar{\mathcal{M}}^v}(\phi)$ 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 $\phi_j \in S$, $j = 1, \dots, n$:

$$(52) \quad \sum_{j,l} z_j \bar{z}_l \exp \left(-\frac{1}{2} \sum_{u' \in \mathcal{U}} \langle (\phi_j - \phi_l)^2, \eta(u', dy) \rangle \right)$$

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:

$$(53) \quad \begin{aligned} \sum_{l,j} z_j \bar{z}_l \Sigma_{l,j} &= \sum_{l,j} z_j \bar{z}_l \sum_{u' \in \mathcal{U}} \langle \phi_j \phi_l, \eta(u', dy) \rangle = \sum_{u' \in \mathcal{U}} \left\langle \sum_{l,j} z_j \bar{z}_l \phi_j \phi_l, \eta(u', dy) \right\rangle \\ &= \sum_{u' \in \mathcal{U}} \left\langle \left| \sum_l z_l \phi_l \right|^2, \eta(u', dy) \right\rangle \geq 0. \end{aligned}$$

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

$$(54) \quad \sum_{l,j} z_j \bar{z}_l \mathbb{E}[\exp(i(e_l - e_j) \cdot Z)] \geq 0,$$

for $e_j \in \mathbb{R}^n$. Let $e_j = (\mathbf{1}_{j=1}, \mathbf{1}_{j=2}, \dots, \mathbf{1}_{j=n})$. Then, (54) is exactly the quantity (52), which proves Property (ii). Therefore, assuming $\bar{\Xi}_{t' < t}^v \in W^t$, $\bar{\mathcal{M}}_t^v(u, y) \in S'$. 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 be 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 (16), substituting ν into equation (16) yields that $\langle \phi, \nu \rangle_{\mathcal{U} \times \mathbb{R}^{d_Y}} = 0$ for every ϕ , which implies uniqueness.

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 $\bar{\mu}$:

$$(55) \quad \text{MSE} = \mathbb{E}[(\mathbb{E}[g(\langle f, \bar{\mu}_t \rangle_{\mathcal{U} \times \mathcal{Y}})] - \frac{1}{L} \sum_{l=1}^L g(m^l))^2]$$

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:

$$(56) \quad \begin{aligned} \zeta_t^{l,H}(u, y) &= \sum_{u' \in \mathcal{U}} h_\theta(u, u', y, V_{t-1}^l, H_{t-1}) \zeta_{t-1}^{l,H}(u', y), \\ \zeta_0^{l,H}(c, y) &= 1, \\ \zeta_0^{l,H}(u, y) &= 0, \quad u \neq c. \end{aligned}$$

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

$$|\zeta_t^{l, \bar{H}}(u, y_k) - \zeta_t^{l, H^\Psi}(u, y_k)| \leq |\zeta_t^{l, \bar{H}}(u, y_k) - \sum_{u' \in \mathcal{U}} h_\theta(u, u', y_k, V_{t-1}^l, \bar{H}_{t-1}) \zeta_{t-1}^{l, H^\Psi}(u', y_k)|$$

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

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 :

$$(58) \quad \max_{u, y_k} |\zeta_t^{l, \bar{H}}(u, y_k) - \zeta_t^{l, H^\Psi}(u, y_k)| \leq \max_{u, y_k} |\zeta_{t-1}^{l, \bar{H}}(u, y_k) - \zeta_{t-1}^{l, H^\Psi}(u, y_k)| + K_1 |\bar{H}_{t-1} - H_{t-1}^\Psi|.$$

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

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

Using a Taylor expansion:

$$\begin{aligned}
|\bar{H}_t - H_t^\Psi| & = \left| \left\langle f^H, \bar{\mu}_t^l \right\rangle_{\mathcal{U} \times \mathcal{Y}} - \sum_{u \in \mathcal{U}} \sum_{k=1}^K f_\theta^h(u, y_k) \bar{\mu}_t^{\Psi, l}(u, c_k) \right| \\
& \leq \sum_{u \in \mathcal{U}} \sum_{k=1}^K \int_{c_k} |f_\theta^h(u, y) \zeta_t^{l, \bar{H}}(u, y) - f_\theta^h(u, y_k) \zeta_t^{l, H^\Psi}(u, y_k)| \bar{\mu}_0(c, dy) \\
& \leq \sum_{u \in \mathcal{U}} \sum_{k=1}^K \int_{c_k} |f_\theta^h(u, y) \zeta_t^{l, \bar{H}}(u, y) - f_\theta^h(u, y_k) \zeta_t^{l, \bar{H}}(u, y_k)| \bar{\mu}_0(c, dy) \\
& + \sum_{u \in \mathcal{U}} \sum_{k=1}^K \int_{c_k} |f_\theta^h(u, y_k) \zeta_t^{l, \bar{H}}(u, y_k) - f_\theta^h(u, y_k) \zeta_t^{l, H^\Psi}(u, y_k)| \bar{\mu}_0(c, dy). \\
& \leq 2 \sum_{i=1}^{d_Y} r_{K, i} \sum_{u \in \mathcal{U}} \sup_{y \in \mathcal{Y}} \left| \frac{\partial}{\partial y_i} [f_\theta^h(u, y) \zeta_t^{l, \bar{H}}(u, y)] \right| + C_3 \max_{u, y_k} |\zeta_t^{l, \bar{H}}(u, y_k) - \zeta_t^{l, H^\Psi}(u, y_k)|
\end{aligned}$$

$$(59) \quad \equiv 2 \sum_{i=1}^{d_Y} C_{1,t,i}(X^l) r_{K,i} + C_3 \max_{u, y_k} |\zeta_t^{l, \bar{H}}(u, y_k) - \zeta_t^{l, H^\Psi}(u, y_k)|$$

Returning to equation (58), we now have:

$$\begin{aligned}
\max_{u, y_k} |\zeta_t^{l, \bar{H}}(u, y_k) - \zeta_t^{l, H^\Psi}(u, y_k)| &\leq \max_{u, y_k} |\zeta_{t-1}^{l, \bar{H}}(u, y_k) - \zeta_{t-1}^{l, H^\Psi}(u, y_k)| \\
&+ 2K_1 \sum_{i=1}^{d_Y} C_{1,t-1,i}(V^l) r_{K,i} + C_3 K_1 \max_{u, y_k} |\zeta_{t-1}^{l, \bar{H}}(u, y_k) - \zeta_{t-1}^{l, H^\Psi}(u, y_k)| \\
&\leq 2K_1 \sum_{i=1}^{d_Y} C_{1,t-1,i}(V^l) r_{K,i} + (1 + C_3 K_1) \sum_{t'=0}^{t-1} \max_{u, y_k} |\zeta_{t'}^{l, \bar{H}}(u, y_k) - \zeta_{t'}^{l, H^\Psi}(u, y_k)| \\
&\leq 2K_1 \sum_{i=1}^{d_Y} C_{1,t-1,i}(V^l) r_{K,i} + 2K_1 (1 + C_3 K_1) \sum_{t'=0}^{t-1} e^{(t-t'-1)(1+C_3 K_1)} \sum_{i=1}^{d_Y} C_{1,t',i}(V^l) r_{K,i} \\
(60) \quad &\equiv \sum_{i=1}^{d_Y} C_{2,t,i}(V^l) r_{K,i}
\end{aligned}$$

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

$$(61) \quad |g(\langle f, \bar{\mu}_t^l \rangle_{\mathcal{U} \times \mathcal{Y}}) - g(\langle f, \bar{\mu}_t^{l, \Psi} \rangle_{\mathcal{U} \times \mathcal{Y}})| \leq \sum_{i=1}^{d_Y} C_{3,t,i}(V^l) r_{K,i}.$$

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:

$$\begin{aligned}
\mathbb{E} \left[\left(\mathbb{E} g(\langle f, \bar{\mu}_t \rangle_{\mathcal{U} \times \mathcal{Y}}) - \frac{1}{L} \sum_{l=1}^L g(m^l) \right)^2 \right] &\leq 2 \mathbb{E} \left[\left(\mathbb{E} g(\langle f, \bar{\mu}_t \rangle_{\mathcal{U} \times \mathcal{Y}}) - \frac{1}{L} \sum_{l=1}^L g(\langle f, \bar{\mu}_t^l \rangle_{\mathcal{U} \times \mathcal{Y}}) \right)^2 \right] \\
+ (62) \left[\left(\frac{1}{L} \sum_{l=1}^L g(\langle f, \bar{\mu}_t^l \rangle_{\mathcal{U} \times \mathcal{Y}}) - \frac{1}{L} \sum_{l=1}^L g(m^l) \right)^2 \right] &\leq \underbrace{\frac{2}{L} \text{Var} \left[g(\langle f, \bar{\mu}_t \rangle_{\mathcal{U} \times \mathcal{Y}}) \right]}_{\text{variance}} + \underbrace{C_5 d_Y \sum_{i=1}^{d_Y} \mathbb{E} \left[C_{3,t,i}(V)^2 \right]}_{\text{bias}} r_{K,i}^2
\end{aligned}$$

The assumption that \mathcal{V} is compact was made in order that $C_{3,t,i}(V)^2 < C_3^2$ and thus $C_{3,t,i}(V)^2$ would be integrable. This assumption can be relaxed to simply requiring that $C_{3,t,i}(V)^2$ be integrable. As a consequence of equation (62), the mean-squared error of the numerical approximation converges to

zero as $L \rightarrow \infty$ and $\max_i r_{K,i} \rightarrow 0$. Equation (62) 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 numbers $\bar{\mu}_t$ 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 (62) 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 5.6 for a better alternative) with a total computational budget B , the budget equation is:

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

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

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

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

APPENDIX C. CONVERGENCE RATE OF THE APPROXIMATION

The error bound (62) shows the rate of convergence of the numerical scheme to the law of large numbers. We now build on this result and derive explicit error bounds for the numerical scheme and the finite system expectation $\mathbb{E}g(\langle f, \mu_t^N \rangle_{\mathcal{U} \times \mathcal{Y}})$. That is, we show the rate of convergence of the numerical scheme to $\mathbb{E}g(\langle f, \mu_t^N \rangle_{\mathcal{U} \times \mathcal{Y}})$ as a function of the size of the system N , the number of Monte Carlo samples L , and the size of the cells $R_{K,i}$. This new result accounts for the quadrature error of the numerical scheme to solve the law of large numbers, Monte Carlo error from the numerical scheme, and the approximation error of $\bar{\mu}$ for $\bar{\mu}^N$. Specifically, we seek a convergence rate for the mean-squared error $\mathbb{E}[(\mathbb{E}g(\langle f, \mu_t^N \rangle_{\mathcal{U} \times \mathcal{Y}}) - \frac{1}{L} \sum_{l=1}^L g(m^l))^2]$.

Proving such a convergence rate for weak convergence limits of dynamical systems is typically difficult and requires significant effort beyond simply proving the convergence to the law of large numbers and central limit theorem. To prove the convergence rate, we use the uniform integrability results from Theorem 3.9 and Section A.4. As previously mentioned, such uniform integrability results are uncommon and usually challenging to prove (we are unaware of other instances where this has been proven for mean-field systems). The uniform integrability result allows us to show the rate of convergence. For the remainder of this section B, Assumption 3.8 of Theorem 3.9 is also in force.

The mean-squared error can be decomposed into the mean-squared error of the numerical scheme for the law of large numbers, which was previously derived in (62), and the bias of the law of large numbers for the finite system.

$$\begin{aligned}
 & \mathbb{E} \left[\left(\mathbb{E}g(\langle f, \mu_t^N \rangle_{\mathcal{U} \times \mathcal{Y}}) - \frac{1}{L} \sum_{l=1}^L g(m^l) \right)^2 \right] \\
 (63) \quad & \leq 2 \left(\mathbb{E}g(\langle f, \mu_t^N \rangle_{\mathcal{U} \times \mathcal{Y}}) - \mathbb{E}g(\langle f, \bar{\mu}_t \rangle_{\mathcal{U} \times \mathcal{Y}}) \right)^2 + 2 \mathbb{E} \left[\left(\mathbb{E}g(\langle f, \bar{\mu}_t \rangle_{\mathcal{U} \times \mathcal{Y}}) - \frac{1}{L} \sum_{l=1}^L g(m^l) \right)^2 \right]
 \end{aligned}$$

The second term in (63) has already been analyzed in (62). The first term is the bias of the law of large numbers for the finite system. In order to bound the first term, we first use a Taylor expansion to show that:

$$(64) \quad g(\langle f, \mu_t^N \rangle_{\mathcal{U} \times \mathcal{Y}}) - g(\langle f, \bar{\mu}_t \rangle_{\mathcal{U} \times \mathcal{Y}}) = \frac{1}{\sqrt{N}} g'(Z) \langle f, \Xi_t^N \rangle_{\mathcal{U} \times \mathcal{Y}},$$

where $Z \in [\langle f, \mu_t^N \rangle, \langle f, \bar{\mu}_t \rangle]$. Z takes values in a compact set. This implies that:

$$\left| \mathbb{E}g(\langle f, \mu_t^N \rangle_{\mathcal{U} \times \mathcal{Y}}) - \mathbb{E}g(\langle f, \bar{\mu}_t \rangle_{\mathcal{U} \times \mathcal{Y}}) \right| \leq \frac{1}{\sqrt{N}} \mathbb{E} \left[\left| g'(Z) \langle f, \Xi_t^N \rangle_{\mathcal{U} \times \mathcal{Y}} \right| \right].$$

$$(65) \quad \leq \frac{C_z}{\sqrt{N}} \mathbb{E} \left[\left| \langle f, \Xi_t^N \rangle_{\mathcal{U} \times \mathcal{Y}} \right| \right]$$

The second lines follows from g being continuously differentiable and Z lying in a compact set, which gives that $\sup_z |g'(z)| \leq C_z < \infty$. Due to the bound (36) proven in Appendix A.4 for Theorem 3.9 on uniform integrability, it now follows that:

$$(66) \quad \left| \mathbb{E}g(\langle f, \mu_t^N \rangle_{\mathcal{U} \times \mathcal{Y}}) - \mathbb{E}g(\langle f, \bar{\mu}_t \rangle_{\mathcal{U} \times \mathcal{Y}}) \right| \leq \frac{C}{\sqrt{N}}.$$

We therefore have the mean-squared error of the numerical scheme for the finite system expected value:

$$(67) \quad \mathbb{E} \left[\left(\mathbb{E}g(\langle f, \mu_t^N \rangle_{\mathcal{U} \times \mathcal{Y}}) - \frac{1}{L} \sum_{l=1}^L g(m^l) \right)^2 \right] \\ \leq \underbrace{\frac{C_6}{L} \text{Var}[g(\langle f, \bar{\mu}_t \rangle_{\mathcal{U} \times \mathcal{Y}})]}_{\text{variance}} + \underbrace{C_7 d_Y \sum_{i=1}^{d_Y} \mathbb{E}[C_{3,t,i}(V)^2] r_{K,i}^2}_{\text{bias from numerical quadrature}} + \underbrace{\frac{C_8}{N}}_{\text{bias from approximation of finite system}}.$$

The first term is the error from the Monte Carlo variance and can be reduced by increasing the number of Monte Carlo samples L . The second term is the numerical quadrature error and can be reduced by decreasing the cell size $r_{K,i}$. The third term is the approximation error for the finite system and decreases as the size of the system gets large. Numerical experiments in Section 5 suggest that the approximation bias for the law of large numbers is very small even for N in the low thousands. The constant C_8 does not depend upon the number of Monte Carlo samples L nor the cell size $r_{K,i}$.

Traditional Monte Carlo simulation of the system is unbiased and has mean-squared error which decreases as $\frac{C_{MC}}{L}$. However, each simulation of the system requires simulating N individual loans. That is, NL paths must be simulated in order to get L samples from $g(\langle f, \mu_t^N \rangle_{\mathcal{U} \times \mathcal{Y}})$. The cost of the approximation depends upon the number of grid points required; as can be seen from (67), this typically increases as d_Y increases. Therefore a rough comparison of the computational costs of the approximation versus traditional Monte Carlo is the number of grid points versus the size of the system N ; for instance, see equation (15) and the corresponding discussion.

APPENDIX D. 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}^{dw} :

- Divide \mathbb{R}^{d_w} into K boxes, each with equal mass $\int_{\text{box } k} \bar{\nu}(dw) = 1/K$ where we take $\bar{\nu}(dw) = \bar{\mu}_0(c, dw)$. In one dimension, this can be done by finding the quantiles of the distribution $\bar{\nu}$. It is assumed that $\int_{\mathbb{R}^{d_w}} \bar{\mu}_0(c, dw) = 1$.
- 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, \dots, w_K .

If g_θ is locally linear (at least within the k -th box) in e^w , the grid points w_k can make this scheme highly accurate. We demonstrate for one time-step to explain the choice of the points w_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

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

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 E. PRE-COMPUTATION FOR FINANCIAL INSTITUTIONS

Even for the risk analysis of smaller, individual pools, the 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 (2), 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, \dots, N_K with $N = N_1 + \dots + N_K$, then the computational cost of the 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 $1, \dots, K$. The approach is summarized below:

- Pre-simulate the (finite-dimensional) LLN $\bar{\mu}^\Psi$ and CLT $\bar{\Xi}^\Psi$ on the grid $\mathcal{R} = \{w_1, \dots, w_I\}$ with initial condition $\Psi(c, dw) = \sum_{i=1}^I \delta_{w_i}$.
- For each pool $1, \dots, K$: 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

$$(68) \quad \mu_t^{N_k}(u, w_i) = z_i \bar{\mu}_t^\Psi(u, w_i) + \sqrt{\frac{z_i}{N_k}} \bar{\Xi}^\Psi(u, w_i), \quad , i = 1, \dots, I,$$

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 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 5.5.

APPENDIX F. PARAMETER ESTIMATION

The parameter θ specifying the model (2) 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 $D_{T,N} = (Z_1^N, \dots, Z_T^N)$ where $Z_t^N = (U_t^1, \dots, U_t^N)$. The log-likelihood function for $D_{T,N}$ given V and Y is

$$(69) \quad \begin{aligned} \mathcal{L}(\theta) &= \log \mathbb{P}_\theta(D_{T,N} | V, Y) = \log \mathbb{P}_\theta[Z_1^N, \dots, Z_T^N | V, Y] = \log \prod_{t=1}^T \mathbb{P}_\theta[Z_t^N | Z_0^N, \dots, Z_{t-1}^N, V, Y] \\ &= \log \prod_{t=1}^T \mathbb{P}_\theta[Z_t^N | Z_{t-1}^N, H_{t-\tau:t-1}^N, V, Y] = \log \prod_{t=1}^T \prod_{n=1}^N h_\theta(U_t^n, U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N) \\ &= \sum_{t=1}^T \sum_{n=1}^N \log h_\theta(U_t^n, U_{t-1}^n, Y^n, V_{t-1}, H_{t-\tau:t-1}^N). \end{aligned}$$

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 (69). The maximum likelihood estimator $\hat{\theta} = \arg \max_{\theta \in \Theta} \mathcal{L}(\theta)$.⁴ The log-likelihood can be maximized using stochastic gradient descent; see [?] for details on implementation. Traditional gradient descent may be computationally costly due to the large amount of data and memory constraints.

Parameter fits for the subprime and agency mortgage models are reported in Tables 1 and 2, respectively. Here, the transition function h_θ is a (multinomial) logistic function, see Example 2.1. For the subprime loans, this is a competing risks model which jointly models transitions to prepayment and default. For the agency loans, only transitions to prepayment are modeled.

Factor	Default	Standard Error	Prepayment	Standard Error
Constant	-5.906	4.0×10^{-3}	-4.363	2.9×10^{-3}
National unemployment rate	.7593	3.5×10^{-3}	-.9782	7.4×10^{-4}
National mortgage rate	NA	NA	-.1517	2.7×10^{-3}
LTV ratio	.3072	3.2×10^{-3}	-.0657	9.5×10^{-4}
Initial Balance	.1001	3.4×10^{-3}	.0656	9.4×10^{-4}
Initial interest rate	.2859	3.4×10^{-3}	.0503	1.1×10^{-3}
FICO score	.0706	3.7×10^{-3}	-.0402	1.1×10^{-3}

TABLE 1. Parameter fits for default and prepayment model for subprime mortgage data set.

APPENDIX G. ADDITIONAL NUMERICAL STUDIES

G.1. Monte Carlo Error. The “actual distribution” is calculated using traditional Monte Carlo simulation with L samples and therefore itself has some Monte Carlo error. We show in Table 3 that the Monte Carlo error is relatively small and therefore it is reasonable to compare our approximation against a Monte Carlo simulation with a finite number of samples.

G.2. Additional Figures. Additional figures comparing the Monte Carlo approximation and the actual distribution are included below. 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 $100,000$,

⁴One typically also chooses a separate model for the behavior of the systematic factors V with its own parameters. These parameters can be estimated separately from θ ; 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.

Factor	Parameter
Constant	-4.476
National unemployment rate	-0.1502
National mortgage rate	-0.4464
FICO score	0.1787
First time homebuyer	-.0392
Number of units	-0.0282
Occupancy status: owned	0.0992
Occupancy status: investment property	-0.1272
Occupancy status: second home	-0.0063
Combined loan-to-value ratio	-0.0046
Loan-to-value ratio	-0.0296
Initial interest rate	0.7689
Prepayment penalty flag	0.0108
Property type: condo	0.0683
Property type: leasehold	.0422
Property type: PUD	0.1698
Property type: manufactured housing	-.0583
Property type: 1-4 fee simple	.1884
Property type: Co-op	.0464
Loan purpose: purchase	0.0531
Property type: cash-out refinance	-0.0541
Property type: no cash-out refinance	-0.0051
Number of borrowers: 1 borrower	-0.2330
Number of borrowers: more than 1 borrower	-0.0703
Debt-to-income	-0.0201

TABLE 2. Parameter fits for prepayment model for agency mortgage data set.

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

L	Bias	Standard Deviation
25,000	3.25×10^{-5}	1.34×10^{-4}
50,000	2.95×10^{-5}	1.00×10^{-4}
100,000	4.5×10^{-5}	7.67×10^{-5}
250,000	4.40×10^{-5}	5.19×10^{-5}

TABLE 3. 2 million samples of a pool of size 10,000 (drawn from the agency loan dataset) are generated. The 99% VaR of the loss from default for the pool are calculated using bootstrap samples of size L . 100 bootstrap samples are generated for each size of L . The bias and standard deviation for the 99% VaR using L Monte Carlo samples is estimated by comparing the bootstrap samples against the 99% VaR using the full 2 million samples.

LLN alone can underestimate the tails of the distribution for small N . By including the CLT in addition to the LLN, one is able to accurately capture the tail of the distribution.

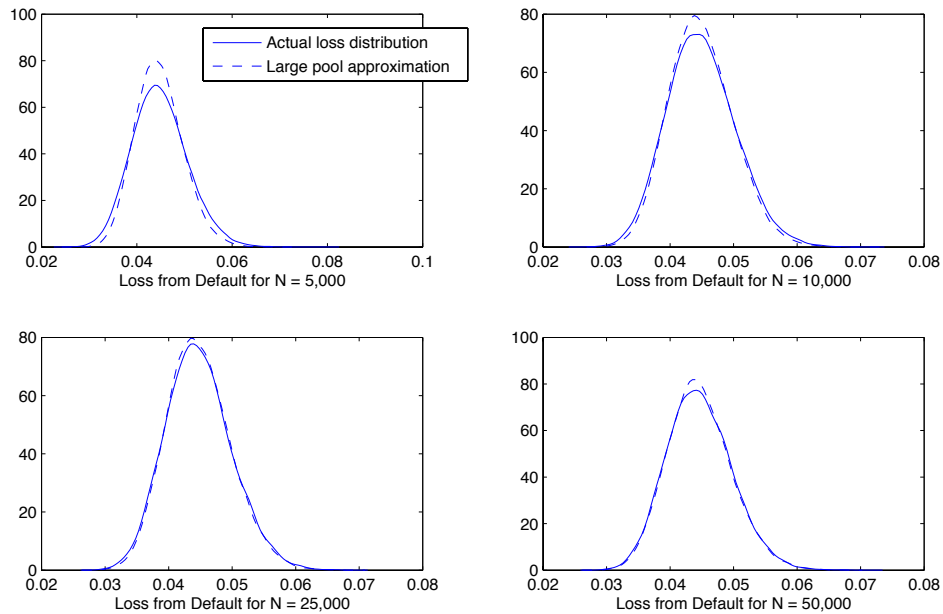


FIGURE 1. Comparison of actual loss distribution with LLN loss distribution (does not include CLT in approximation). Loss reported as fraction of pool which defaulted. The horizon is 12 months.

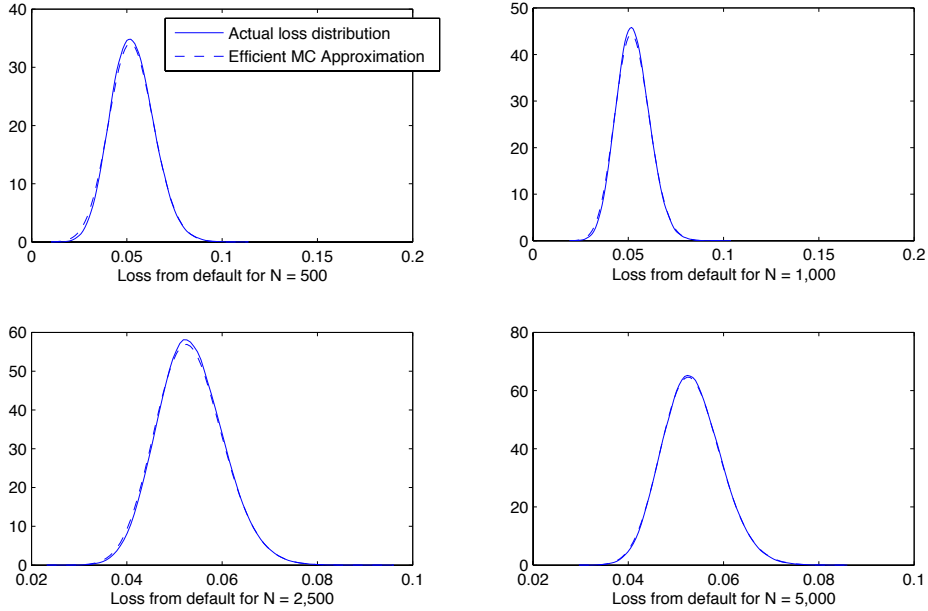


FIGURE 2. Comparison of actual distribution with approximate distribution (using both LLN and CLT). Loss reported as fraction of pool which defaulted. The horizon is 12 months.

G.3. Sparse Grids. When the low-dimensional application is not applicable, different methods must be used to efficiently evaluate the LLN and CLT. One approach is to cluster the pool into K clusters using k-means clustering on Y^1, \dots, Y^N . The K centroids are chosen as the grid points. This is a sparse non-uniform grid in a high-dimensional space. Another approach is to perform a *approximate* low-dimensional transformation via an additional multi-layer neural network $\tilde{h}_\theta(u, u', y, V)$. The multi-layer neural network has three layers: the first (multi-neuron) layer takes as an input y , the second layer has d_W neurons (where $d_W \ll d_Y$), and the third layer takes as inputs the output of the d_W neurons from the second layer as well as 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 4 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.

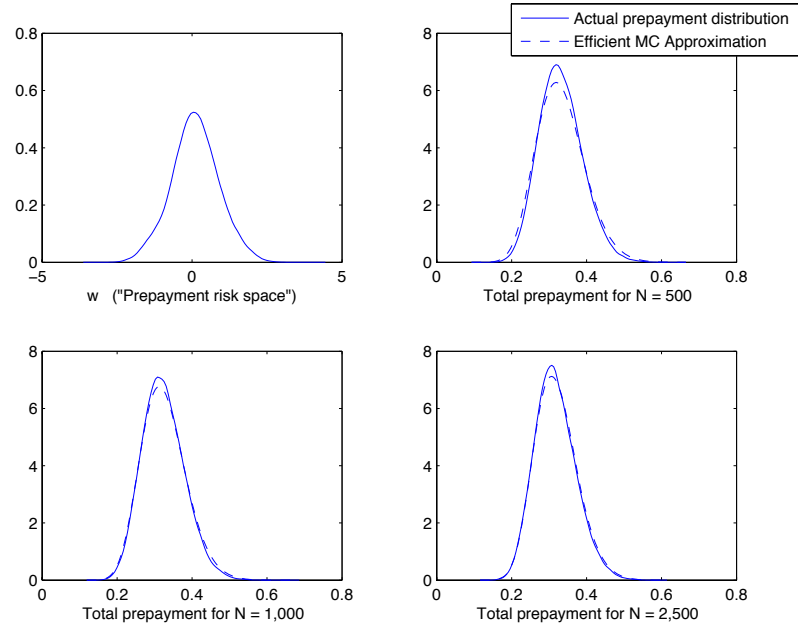


FIGURE 3. Top right plot and bottom plots compare the actual prepayment distribution with the approximation for $N = 500$, $N = 1,000$, and $N = 2,500$. The total prepayment is reported as the fraction of the pool which prepays. The time horizon is 12 months. The top left plot shows the distribution of the pool in the “prepayment risk space”.

DEPARTMENT OF INDUSTRIAL AND ENTERPRISE SYSTEMS ENGINEERING, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, URBANA, IL 61801

E-mail address: jasirign@illinois.edu

DEPARTMENT OF MANAGEMENT SCIENCE AND ENGINEERING, STANFORD UNIVERSITY, STANFORD, CA 94305

E-mail address: giesecke@stanford.edu

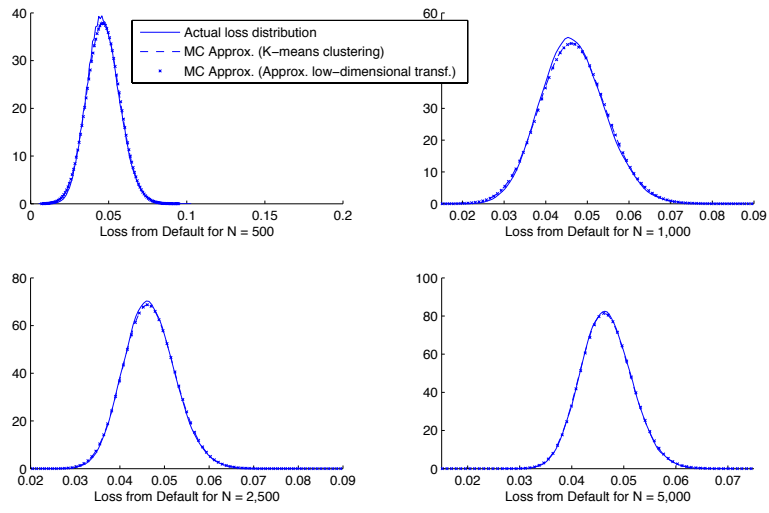


FIGURE 4. Comparison of actual loss distribution with Monte Carlo approximation using sparse grids when h_θ is a neural network.