

Transition of α -mixing in Random Iterations with Applications in Queuing Theory*

Attila Lovas¹

¹*HUN-REN Alfréd Rényi Institute of Mathematics, Budapest, Hungary*

¹*Budapest University of Technology and Economics, Budapest, Hungary*

October 24, 2025

Abstract

Nonlinear time series models with exogenous regressors are essential in econometrics, queuing theory, and machine learning, though their statistical analysis remains incomplete. Key results, such as the law of large numbers and the functional central limit theorem, are known for weakly dependent variables. We demonstrate the transfer of mixing properties from the exogenous regressor to the response via coupling arguments. Additionally, we study Markov chains in random environments with drift and minorization conditions, even under non-stationary environments with favorable mixing properties, and apply this framework to single-server queuing models.

Introduction

It is very common in natural and social sciences that for describing the time evolution of certain quantity of interests, researchers build models incorporating input variables not influenced by other variables in the system and on which the output variable depends. Such explicative variables, especially in econometrics literature, are called exogeneous covariates. Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complete and separable metric spaces. The $\mathcal{X}$ -valued process $(X_t)_{t \in \mathbb{N}}$ represents the time series of interest and the $\mathcal{Y}$ -valued process $(Y_t)_{t \in \mathbb{Z}}$ denotes the exogeneous covariate. We postulate that $(X_t)_{t \in \mathbb{N}}$ satisfies the recursion

$$X_{t+1} = f(X_t, Y_t, \varepsilon_{t+1}), \quad (1)$$

where X_0 is a possibly random initial state, $f : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \rightarrow \mathcal{X}$ is a measurable function, and $\varepsilon_t \in \mathcal{Z}$, $t \in \mathbb{N}$ represents the noise entering to the system.

The exploration and analysis of non-linear autoregressive processes of this kind constitute a recent and actively developing area of research. In particular, there is a pronounced surge of interest within the fields of applied statistics and econometrics regarding the investigation of standard time series models that incorporate exogeneous regressors. Notable examples include a novel class of Poisson autoregressive models with exogeneous covariates (PARX) introduced by Agosto et al. [1] for modeling corporate defaults. Additionally, the recent research by Gorgi and Koopman [23] has provided valuable insights on observation-driven models involving beta autoregressive processes with exogeneous factors. Furthermore, the theory of non-linear autoregressive processes allows researchers for analyzing large-scale stochastic optimization algorithms, which play a pivotal role in machine learning applications, see [7, 35].

The statistical analysis of general non-linear time series models with exogenous covariates necessitates the law of large numbers (LLN), central limit theorem (CLT), and others. However, this

*This work was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. Additional support was provided by the National Research, Development and Innovation Office (NKFIH) under the Thematic Excellence Program 2021, National Research Subprogram “Artificial intelligence, large networks, data security: mathematical foundations and applications”, and by Grant K 143529.

framework is presently unavailable. Researchers have investigated these models under additional assumptions that facilitate their analysis. The ergodicity of iterations given by (1) has been extensively studied under the restrictive assumption that the data $(Y_t)_{t \in \mathbb{Z}}$ and the noise $(\varepsilon_t)_{t \in \mathbb{N}}$ are both i.i.d. and also independent of each other (See, [14], [30], and [48]). In this case, the process $(X_t)_{t \in \mathbb{N}}$ is a Markov chain, and this setting now can be considered to be textbook material. Moving beyond this simplifying yet unrealistic assumption, Debaly and Truquet established general results for getting stationarity, ergodicity and stochastic dependence properties for general nonlinear dynamics defined in terms of iterations of random maps [13]. Additionally, there are earlier contributions that consider more general schemes and investigate them without assuming independence. For instance, in the paper of Borovkov and Foss [6], Foss and Konstantopoulos [20] and also in the monograph of Borovkov [5] such processes are treated under the name "stochastically recursive sequences". Among the most recent results, we can mention the paper [24] by Györfi et al. that introduces a novel concept called strong stability and provides sufficient conditions for strong stability of iterations given by (1). Furthermore, new findings related to Langevin-type iterations with dependent noise and multitype branching processes were also established.

Assuming that the noise $(\varepsilon_t)_{t \in \mathbb{N}}$ is i.i.d. and independent of the regressor $\mathbf{Y} := (Y_t)_{t \in \mathbb{Z}}$, we have

$$\mathbb{P}(X_t \in B \mid (X_j)_{j < t}, \mathbf{Y}) = \int_{\mathcal{Z}} \mathbf{1}_{\{f(X_{t-1}, Y_{t-1}, z) \in B\}} \nu(dz), \quad t \geq 1, \quad (2)$$

where $\nu = \text{Law}(\varepsilon_0)$. Clearly, the process $(X_t)_{t \in \mathbb{N}}$ defines a time-inhomogeneous Markov chain conditionally on the exogenous process $(Y_t)_{t \in \mathbb{N}}$ being interpreted as random environment. This characterization leads us to term this process a Markov chain in a random environment (MCRE). This concept is proved to be a good compromise since, many interesting models can be treated as a MCRE. Furthermore, it is worth noting that the rich theory of general state Markov chains equips us with powerful analytical tools to study and understand these processes in-depth. Markov chains in random environments were first studied on countable state spaces in [11, 12, 44]. On general state spaces [31, 32, 47] investigated their ergodic properties under a rather stringent hypothesis: essentially, the Doeblin condition was assumed (see Chapter 16 of [42]). Such assumptions are acceptable on compact state spaces but they fail in most models evolving in $\mathbb{R}^d$. For non-compact state spaces the results of [48] apply (see also Chapter 3 of [5]) but the system dynamics is assumed to be strictly contracting, which, again, is too stringent for most applications. Markov chains in stationary random environments were first treated on non-compact state spaces under Lyapunov and "small set"-type conditions in [22] and [35]. The former paper was based on the control of the maximal process of the random environment but its techniques worked only assuming that the system dynamics is contractive with respect to a certain Lyapunov function, whatever the random environment is. In [35] this decreasing property is required only in an *averaged* sense. This result covers important model classes that none of the previous works could: queuing systems with non-independent service times (or inter-arrival times), linear systems that are stable in the average, and stochastic gradient Langevin dynamics when the data is merely stationary. In [49], under a notably weaker, yet in certain aspects, optimal form of the Lyapunov and the small set conditions, Truquet showed that for a given strongly stationary process $(Y_t)_{t \in \mathbb{N}}$, there exists a process $(X_t)_{t \in \mathbb{N}}$ satisfying the iteration in (1), and the distribution of the process $(X_t, Y_t)_{t \in \mathbb{N}}$ is unique. Additionally, if the process $(Y_t)_{t \in \mathbb{N}}$ is ergodic, then the process $(X_t, Y_t)_{t \in \mathbb{N}}$ is ergodic as well, hence the strong law of large numbers applies.

As far as we know, there are no known results regarding MCREs when the environment $(Y_t)_{t \in \mathbb{N}}$ is non-stationary. Furthermore, the sequence of iterates $(X_t)_{t \in \mathbb{N}}$ is typically non-stationary even in cases when the environment is stationary but the initial state $x_0 \in \mathcal{X}$ is independent of $\sigma(\{\varepsilon_t, Y_t \mid t \in \mathbb{N}\})$. Weak dependence assumptions offer a valuable approach to address this problem while allowing for long-range dependencies to be present. The recent work by Truquet [50] directed our attention to the fact that through arguments based on coupling inequalities, it can be established under general conditions that the mixing properties of the process $(Y_t)_{t \in \mathbb{N}}$ are inherited by the iterates $(X_t)_{t \in \mathbb{N}}$. Combining this idea with Corollary 2 from Herrndorf's paper [29], we were able to establish the functional central limit theorem for the stochastic gradient Langevin iteration in cases where the data stream is stationary and exhibits favorable mixing properties [36].

Rosenblatt introduced the alpha-mixing coefficient in 1965, defined the class of strongly mixing processes and proved the central limit theorem for strongly mixing stationary processes [46]. Over

the past decades, researchers have established numerous strong results for non-stationary mixing processes, including various versions of the law of large numbers and the central limit theorem. The main goal of this paper is to investigate the sequence of iterates $(X_t)_{t \in \mathbb{N}}$ through the transitions of mixing properties, leveraging these established results.

The paper is organized as follows: In the first section, we provide sufficient conditions for a recursion of the form (1) to inherit the mixing properties of the process $(Y_t)_{t \in \mathbb{N}}$. Leveraging these conditions along with existing results from the literature on strongly mixing sequences, we prove the strong and L^1 law of large numbers for suitable functionals of the process $(X_t)_{t \in \mathbb{N}}$. Furthermore, we also show the possibility of constructing confidence intervals.

The second section focuses on the investigation of MCREs under long-term contractivity and minorization conditions satisfied by models discussed in [35]. By establishing a coupling inequality and a moment estimate for such chains, the framework presented in Section 1 becomes directly applicable to these processes. Additionally, in this section, using the Cramér-Rao bound, we prove an inequality for variances of sums crucial for the functional central limit theorem by Merlevède and Peligrad [39]. To the best of our knowledge, this technique represents a novel contribution to the theory of MCREs. In the third and final section of the paper, we revisit single-server queuing models discussed in [35] and [34], and prove the functional central limit theorem for them.

Notations and conventions. Let $\mathbb{R}_+ := \{x \in \mathbb{R} : x \geq 0\}$ and $\mathbb{N}_+ := \{n \in \mathbb{N} : n \geq 1\}$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We denote by $\mathbb{E}[X]$ the expectation of a random variable X . For $1 \leq p < \infty$, L^p is used to denote the usual space of p -integrable real-valued random variables and $\|X\|_p$ stands for the L^p -norm of a random variable X .

In the sequel, we employ the convention that $\inf \emptyset = \infty$, $\sum_k^l = 0$ and $\prod_k^l = 1$ whenever $k, l \in \mathbb{Z}$, $k > l$. Lastly, $\langle \cdot | \cdot \rangle$ denotes the standard Euclidean inner product on finite dimensional vector spaces. For example, on $\mathbb{R}^d$, $\langle x | y \rangle = \sum_{i=1}^d x_i y_i$.

1 Transition of mixing properties

In this section, we study the transition of mixing properties of the covariate process to the response and its immediate consequences under minimal assumptions on the iteration (1). To this end, we first introduce the basic concepts that we use in our analysis. Several notions of mixing exist in the literature. The interested reader should consult the excellent survey by Bradley [8], for example. In our context α -mixing holds particular importance, therefore let us first recall the key concepts related to this type of mixing. We define the measure of dependence, denoted as $\alpha(\mathcal{G}, \mathcal{H})$, for any two sub- σ -algebras $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$, using the equation:

$$\alpha(\mathcal{G}, \mathcal{H}) = \sup_{G \in \mathcal{G}, H \in \mathcal{H}} |\mathbb{P}(G \cap H) - \mathbb{P}(G)\mathbb{P}(H)|. \quad (3)$$

Furthermore, considering an arbitrary sequence of random variables $(W_t)_{t \in \mathbb{Z}}$, we introduce the σ -algebras $\mathcal{F}_{t,s}^W := \sigma(W_k, t \leq k \leq s)$, where $-\infty \leq t \leq s \leq \infty$. Additionally, we define the dependence coefficients as follows:

$$\alpha_j^W(n) = \alpha(\mathcal{F}_{-\infty, j}^W, \mathcal{F}_{j+n, \infty}^W), \quad j \in \mathbb{Z}.$$

The mixing coefficient of W is $\alpha^W(n) = \sup_{j \in \mathbb{Z}} \alpha_j^W(n)$, $n \geq 1$ which is obviously non-increasing. Note that, for strictly stationary W , $\alpha_j^W(n)$ does not depend on j , and thus $\alpha^W(n) = \alpha_0^W(n)$. We classify W as strongly mixing if $\lim_{n \rightarrow \infty} \alpha^W(n) = 0$. Strongly mixing processes are significant because many fundamental theorems known for sequences of i.i.d. random variables can be appropriately extended to such processes. The primary importance of this lies in enabling the statistical analysis of heterogeneously distributed dependent data even when the exact underlying dynamics generating the time series are unknown. The available theoretical framework includes the strong law of large numbers by McLeish [38], the L^1 -law of large numbers by Hansen [28], and the Central Limit Theorem by White [52], among others. For better clarity, these results are collected in Appendix A.

Now, we turn to the analysis of the iteration (1). For better readability, we introduced the complete separable metric space $\mathcal{Z}$ in the introduction. However, we can always assume that

$\mathcal{Z} = [0, 1]$ without loss of generality. Indeed, since $\mathcal{Z}$ is a complete separable metric space, the Borel isomorphism theorem guarantees the existence of a bi-measurable bijection $\phi : \mathcal{Z} \rightarrow A$, where $A \subseteq [0, 1]$. Given an iteration of the form (1), we can always define an equivalent iteration:

$$X_{t+1} = \tilde{f}(X_t, Y_t, \tilde{\varepsilon}_{t+1}),$$

where $\tilde{\varepsilon}_t = \phi(\varepsilon_t) \in [0, 1]$, $t \in \mathbb{N}$, and

$$\tilde{f}(x, y, u) = f(x, y, \phi^{-1}(u)), \quad x \in \mathcal{X}, y \in \mathcal{Y}, u \in A.$$

Since this transformation does not alter the dynamics of the iteration, we conclude that it suffices to assume $\mathcal{Z} = [0, 1]$. Therefore, from this point onward, we adopt this assumption.

Let us introduce the notation $\tilde{Y}_n = (Y_n, \varepsilon_{n+1})$, $n \in \mathbb{N}$. Furthermore, for a fixed $x \in \mathcal{X}$, we define the process

$$Z_{s,t}^x = \begin{cases} x, & \text{if } t \leq s \\ f(Z_{s,t-1}^x, Y_{t-1}, \varepsilon_t), & \text{if } t > s. \end{cases} \quad (4)$$

Note that, for $s \in \mathbb{N}$ and $t \geq s$, $Z_{s,t}^{X_s} = X_t$.

Definition 1.1. We say that the iteration (1) satisfies the coupling condition if for some $x_0 \in \mathcal{X}$,

$$\sup_{j \in \mathbb{N}} \mathbb{P} \left(Z_{j,j+n}^{X_j} \neq Z_{j,j+n}^{x_0} \right) \rightarrow 0, \quad n \rightarrow \infty.$$

Bounding the mixing coefficient of a process is typically a non-trivial task in general. The following lemma presents an upper bound for the α -mixing coefficient of the iterates $(X_n)_{n \in \mathbb{N}}$ given $\alpha^{\tilde{Y}}$.

Lemma 1.2. Assume that X_0 is a random initial state independent of $\sigma(Y_n, \varepsilon_{n+1} \mid n \in \mathbb{N})$, moreover the iteration (1) satisfies the coupling condition with $x_0 \in \mathcal{X}$.

Then for $0 \leq m < n$, we have

$$\alpha^X(n) \leq \alpha^{\tilde{Y}}(m+1) + b(n-m),$$

where $b(n) = \sup_{j \in \mathbb{N}} \mathbb{P} \left(Z_{j,j+n}^{X_j} \neq Z_{j,j+n}^{x_0} \right)$.

Proof. Let $j \in \mathbb{N}$ and $n \geq 1$ be arbitrary. Suppose $A \in \mathcal{F}_{0,j}^X$ and $B \in \mathcal{F}_{j+n,\infty}^X$ are arbitrary events. Then by the definition of the generated σ -algebra, exist collections of Borel sets $(A_k)_{k \leq j}$, $(B_k)_{k \geq j+n} \subseteq \mathcal{B}(\mathcal{X})$ such that

$$A = (X_k \in A_k, 0 \leq k \leq j) \text{ and } B = (X_k \in B_k, k \geq j+n).$$

Let $0 \leq m < n$ be arbitrary, and introduce the event $\tilde{B} = \left(Z_{j+m,k}^{x_0} \in B_k, k \geq j+n \right)$. We can estimate

$$|P(A \cap B) - P(A)P(B)| = |\text{cov}(\mathbb{1}_A, \mathbb{1}_B)| \leq |\text{cov}(\mathbb{1}_A, \mathbb{1}_{\tilde{B}})| + |\text{cov}(\mathbb{1}_A, \mathbb{1}_B - \mathbb{1}_{\tilde{B}})|. \quad (5)$$

Using that A is $\sigma(X_0) \vee \mathcal{F}_{0,j-1}^Y \vee \mathcal{F}_{1,j}^\varepsilon$ and $\tilde{B}$ is $\mathcal{F}_{j+m,\infty}^Y \vee \mathcal{F}_{j+m+1,\infty}^\varepsilon$ -measurable, and that X_0 is independent of $\sigma(Y_n, \varepsilon_{n+1} \mid n \in \mathbb{N})$, we have

$$\begin{aligned} |\text{cov}(\mathbb{1}_A, \mathbb{1}_{\tilde{B}})| &\leq \mathbb{E} \left| \mathbb{E} \left[\mathbb{1}_A (\mathbb{1}_{\tilde{B}} - \mathbb{P}(\tilde{B})) \mid X_0 \right] \right| \leq \alpha(\mathcal{F}_{0,j-1}^Y \vee \mathcal{F}_{1,j}^\varepsilon, \mathcal{F}_{j+m,\infty}^Y \vee \mathcal{F}_{j+m+1,\infty}^\varepsilon) \\ &= \alpha(\mathcal{F}_{0,j-1}^{\tilde{Y}}, \mathcal{F}_{j+m,\infty}^{\tilde{Y}}) \leq \alpha^{\tilde{Y}}(m+1). \end{aligned}$$

Observe that on the event $\{X_{j+n} = Z_{j+m,j+n}^{x_0}\}$, we have $\mathbb{1}_B = \mathbb{1}_{\tilde{B}}$, and thus the second term in (5) can be estimated as

$$\begin{aligned} |\text{Cov}(\mathbb{1}_A, \mathbb{1}_B - \mathbb{1}_{\tilde{B}})| &= |\mathbb{E}[(\mathbb{1}_A - \mathbb{P}(A))(\mathbb{1}_B - \mathbb{1}_{\tilde{B}})]| \leq \mathbb{P} \left(Z_{0,j+n}^{X_0} \neq Z_{j+m,j+n}^{x_0} \right) \\ &= \mathbb{P} \left(Z_{j+m,j+n}^{X_{j+m}} \neq Z_{j+m,j+n}^{x_0} \right) \leq b(n-m). \end{aligned}$$

Given that $A \in \mathcal{F}_{0,j}^X$ and $B \in \mathcal{F}_{j+n,\infty}^X$ were arbitrary, we have shown that

$$\alpha(\mathcal{F}_{0,j}^X, \mathcal{F}_{j+n,\infty}^X) \leq \alpha^{\tilde{Y}}(m+1) + b(n-m).$$

Since the upper bound is independent of j , taking the supremum over j , we arrive at the desired inequality. $\square$

Remark 1.3. *In the above Lemma, it is enough to prescribe that an appropriate version of X satisfy the coupling condition. The proof of the coupling condition for Markov chains in random environments (See Appendix C) hinges on this observation.*

Remark 1.4. *The recent work by Truquet [50] employs a similar inequality (equation (3) on page 3) to establish the transition of mixing in random iteration. However, that approach differs from ours in two significant aspects. Firstly, the analysis in [50] is limited to discrete-valued time series models with strictly stationary exogenous covariates. Secondly, unlike in Lemma 1.2, the upper estimate of the strong mixing coefficient in [50] incorporates the tail sum of non-coupling probabilities.*

Combining Lemma 1.2 with the theorems on strongly mixing processes presented in Appendix A, we obtain the following general umbrella theorem. This result equips us with essential tools for the statistical analysis of time series involving non-stationary exogenous covariates, including versions of the weak and strong law of large numbers and theoretical guarantees for constructing confidence intervals. However, this theorem does not cover all relevant aspects. For strongly mixing processes, there are further results concerning the distribution of extreme values [51], concentration inequalities [40], and the law of the iterated logarithm [45]. Leveraging Lemma 1.2, these results can be readily extended to random iterations driven by strongly mixing sequences.

Theorem 1.5. *Assume that X_0 is a random initial state independent of $\sigma(Y_n, \varepsilon_{n+1} \mid n \in \mathbb{N})$, moreover the iteration (1) satisfies the coupling condition with $x_0 \in \mathcal{X}$.*

Consider a measurable function $\Phi : \mathcal{X} \rightarrow \mathbb{R}$ such that $\mu_n = \mathbb{E}[\Phi(X_n)]$ exists and is finite for all $n \in \mathbb{N}$. Let $b(n)$, $n \in \mathbb{N}$ be as in Lemma 1.2, and define

$$S_n = \sum_{k=1}^n (\Phi(X_k) - \mu_k), \quad n \geq 1.$$

Then the following statements hold:

A) *If the process $(\tilde{Y}_n)_{n \in \mathbb{N}}$ is strongly mixing and $\sup_{n \in \mathbb{N}} \mathbb{E}|\Phi(X_n)|^p < \infty$ for some $p > 1$, then the L^1 law of large numbers holds:*

$$\frac{S_n}{n} \xrightarrow{L^1} 0, \quad \text{as } n \rightarrow \infty.$$

B) *Assume that there exist constants $c > 0$ and $r > 2$ such that*

$$\alpha^{\tilde{Y}}(n) + b(n) \leq c n^{-\frac{r}{r-2}}.$$

Moreover, suppose that for some $r/2 < p \leq r < \infty$, we have $\sup_{n \in \mathbb{N}} \mathbb{E}|\Phi(X_n)|^p < \infty$. Then,

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0, \quad \text{as } n \rightarrow \infty.$$

C) *If for $r > 2$, $\sup_{n \in \mathbb{N}} \mathbb{E}|\Phi(X_n)|^r < \infty$ and*

$$\sum_{k=0}^{\infty} (k+1)^2 \left(\alpha^{\tilde{Y}}(k)^{\frac{r-2}{r+2}} + b(k)^{\frac{r-2}{r+2}} \right) < \infty,$$

then the distributions of $n^{-1/2}S_n$ and $\mathcal{N}(0, \text{Var}(n^{-1/2}S_n))$ are weakly approaching, that is for any bounded continuous function $g : \mathbb{R} \rightarrow \mathbb{R}$,

$$\mathbb{E} \left[g \left(n^{-1/2}S_n \right) \right] - \int_{\mathbb{R}} g \left(\text{Var}(n^{-1/2}S_n)^{1/2}t \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Furthermore, there exists $\sigma > 0$ such that for any $a > 0$, we have

$$\limsup_{n \rightarrow \infty} \mathbb{P}(n^{-1/2}|S_n| \geq a) \leq \int_{\mathbb{R}} \mathbb{1}_{[-a,a]^c}(\sigma t) \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

Proof. Define the process $W_n = \Phi(X_n) - \mu_n$, for $n \in \mathbb{N}$. By Lemma 1.2, we have

$$\alpha^W(n) \leq \alpha^X(n) \leq \alpha^{\tilde{Y}}(\lfloor n/2 \rfloor + 1) + b(n - \lfloor n/2 \rfloor). \quad (6)$$

From inequality (6), it follows immediately that if $(\tilde{Y}_n)_{n \in \mathbb{N}}$ is strongly mixing, then so is $(W_n)_{n \in \mathbb{N}}$. Thus, part A) follows directly from Theorem A.3 and Remark A.4.

For part B), again using (6) and applying Remark A.2, we see that the conditions of Theorem A.1 are satisfied, from which the result follows immediately.

Finally, under the assumptions of part C), we can verify that

$$\sum_{k=0}^{\infty} (k+1)^2 \alpha^W(k)^{\frac{r-2}{r+2}} < \infty,$$

and hence by Theorem A.5 and Corollary A.6, the distributions of $n^{-1/2}S_n$ and $\mathcal{N}(0, \text{Var}(n^{-1/2}S_n))$ are weakly approaching. Furthermore, Remark A.7 guarantees the existence of $\sigma > 0$ such that $\sigma_n := \text{Var}(n^{-1/2}S_n)^{1/2} < \sigma$ for all $n \geq 1$. Therefore, for any $a > 0$, we have

$$\begin{aligned} \mathbb{P}(n^{-1/2}|S_n| \geq a) &\leq \mathbb{P}(n^{-1/2}|S_n| \geq a) - \int_{\mathbb{R}} \mathbb{1}_{[-a,a]^c}(\sigma_n t) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \\ &\quad + \int_{\mathbb{R}} \mathbb{1}_{[-a,a]^c}(\sigma t) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt. \end{aligned}$$

Taking the upper limit as $n \rightarrow \infty$ yields

$$\limsup_{n \rightarrow \infty} \mathbb{P}(n^{-1/2}|S_n| \geq a) \leq \int_{\mathbb{R}} \mathbb{1}_{[-a,a]^c}(\sigma t) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

□

2 Markov chains in random environments

This section is devoted to study an important class of random iterations incorporating exogenous covariates, called Markov chains in random environments. For convenience, we adopt the parametric kernel formalism to set Lyapunov and “small set”-type conditions. Let us introduce

$$Q(y, x, B) = \int_{[0,1]} \mathbb{1}_{\{f(x,y,z) \in B\}} dz.$$

The function $Q : \mathcal{X} \times \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \rightarrow [0, 1]$ is a parametric probabilistic kernel, which means:

- i For each pair $(y, x) \in \mathcal{Y} \times \mathcal{X}$, the mapping $B \mapsto Q(y, x, B)$ defines a Borel probability measure on the Borel sigma-algebra $\mathcal{B}(\mathcal{X})$.
- ii For any choice of set $B \in \mathcal{B}(\mathcal{X})$, the mapping $(x, y) \mapsto Q(y, x, B)$ is a measurable function with respect to the product sigma-algebra $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y})$.

Definition 2.1. Let $P : \mathcal{X} \times \mathcal{B} \rightarrow [0, 1]$ be a probabilistic kernel. For a bounded measurable function $\phi : \mathcal{X} \rightarrow \mathbb{R}$, we define

$$[P\phi](x) = \int_{\mathcal{X}} \phi(z) P(x, dz), \quad x \in \mathcal{X}.$$

This definition makes sense for any non-negative measurable ϕ , too.

Consistently with Definition 2.1, for $y \in \mathcal{Y}$, we write $Q(y)\phi$ to denote the action of the kernel $Q(y, \cdot, \cdot)$ on ϕ . It is important to note that if $y_l, \dots, y_{k-1} \in \mathcal{Y}$ with $0 \leq l < k$, then the successive application of the kernels is interpreted in the order corresponding to the composition of conditional expectations:

$$[Q(y_{k-1}) \dots Q(y_l)\phi] = [Q(y_l)[\dots [Q(y_{k-1})\phi]]].$$

This convention will be important later in the proof of Lemma C.1.

We say that Q satisfies the drift (or Lyapunov) condition if there exists a measurable mapping $V : \mathcal{X} \rightarrow [0, \infty)$, which we call Lyapunov-function, and measurable functions $\gamma, K : \mathcal{Y} \rightarrow (0, \infty)$, such that for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$,

$$[Q(y)V](x) := \int_{\mathcal{X}} V(z) Q(y, x, dz) \leq \gamma(y)V(x) + K(y). \quad (7)$$

We may, and from now on, we will assume that $K(\cdot) \geq 1$ in the drift condition (7).

The parametric kernel obeys the minorization condition with $R > 0$, if there exists a probability kernel $\kappa_R : \mathcal{Y} \times \mathcal{B}(\mathcal{X}) \rightarrow [0, 1]$ and a measurable function $\beta : [0, \infty) \times \mathcal{Y} \rightarrow [0, 1]$ such that for all $(y, x, A) \in \mathcal{Y} \times \mathcal{V}^{-1}([0, R]) \times \mathcal{B}(\mathcal{X})$,

$$Q(y, x, A) \geq (1 - \beta(R, y))\kappa_R(y, A). \quad (8)$$

The minorization condition stipulates the existence of “small sets”. Therefore, it is also referred to as a “small set”-type condition.

If γ, K are independent of y and $\gamma < 1$ then (7) is the standard drift condition for geometrically ergodic Markov chains, see Chapter 15 of [42]. Ergodic properties of Markov chains in stationary random environments was studied by Lovas and Rásonyi in [35] when $\gamma(y) \geq 1$ may well occur but the environment satisfies the following long-term contractivity condition:

$$\limsup_{n \rightarrow \infty} \mathbb{E}^{1/n} \left(K(Y_0) \prod_{k=1}^n \gamma(Y_k) \right) < 1. \quad (9)$$

Under the assumption that $\mathbb{E}[\log(\gamma(Y_0))_+] + \mathbb{E}[\log(K(Y_0))_+] < \infty$ and

$$\limsup_{n \rightarrow \infty} \prod_{k=1}^n \gamma(Y_{-k})^{1/n} < 1, \quad \mathbb{P} - \text{a.s.}, \quad (10)$$

which is notably weaker than (9), in [49] Truquet proved that there exists a stationary process $((Y_t, X_t^*))_{t \in \mathbb{Z}}$ satisfying the iteration (1), and the distribution of this process is unique. If, in addition, the environment $(Y_t)_{t \in \mathbb{Z}}$ is ergodic, the process $((Y_t, X_t^*))_{t \in \mathbb{Z}}$ is also ergodic. As a result, the strong law of large numbers holds for any measurable function $\Psi : \mathcal{Y} \times \mathcal{X} \rightarrow \mathbb{R}$ with $\mathbb{E}(|\Psi(Y_0, X_0^*)|) < \infty$ i.e.

$$\frac{1}{n} \sum_{k=1}^n \Psi(Y_k, X_k^*) \rightarrow \mathbb{E}(\Psi(Y_0, X_0^*)), \quad \text{as } n \rightarrow \infty, \quad \mathbb{P} - \text{a.s.}$$

In this case, the condition (10) boils down to $\mathbb{E}[\log(\gamma(Y_0))] < 0$. Truquet has also shown that if the iteration (1) is initialized with deterministic $x_0 \in \mathcal{X}$, denoted as $(X_n^{x_0})_{n \in \mathbb{N}}$, then $\text{Law}(X_n^{x_0})$ converges to $\text{Law}(X_0^*)$ in total variation as $n \rightarrow \infty$, where $\text{Law}(X_0^*)$ here denotes the marginal distribution of the stationary solution $((Y_t, X_t^*))_{t \in \mathbb{Z}}$. On the other hand, contrarily to [35], Truquet did not provide a rate estimate.

Independent and identically distributed sequences of random variables $(Y_n)_{n \in \mathbb{N}}$ satisfy the long-term contractivity condition (9) if $\mathbb{E}(\gamma(Y_0)) < 1$. Naturally, the question arises whether the inequality (9) still holds when $\mathbb{E}(\gamma(Y_0)) < 1$ and the sequence $\alpha^Y(n)$, $n \in \mathbb{N}$, tends to zero rapidly enough. In his Master’s thesis on the stability of general state Markov chains [19], Dániel Felsmann provided an example of a strongly stationary stochastic process $(Y_n)_{n \in \mathbb{Z}}$ and a function $\gamma : \mathcal{Y} \rightarrow (0, \infty)$ such that $\mathbb{E}(\gamma(Y_0)) < 1$, yet $\lim_{n \rightarrow \infty} \mathbb{E}(\prod_{k=1}^n \gamma(Y_k))$, i.e., the long-term contractivity

condition in (9) is not satisfied. Since the thesis is available only in Hungarian, the example in question is presented in Appendix B.

In the forthcoming discussion, we refrain from assuming stationarity for the environment $(Y_n)_{n \in \mathbb{N}}$. Instead, we regard it purely as a sequence of weakly dependent random variables. Consequently, the anticipation of the existence of limiting distributions, as demonstrated in [35], [36], or [49], is not viable. Instead, we employ the methodology delineated in Section 1 to establish the L^1 -law of large numbers and the functional central limit theorem.

We impose the following additional assumptions on the environment. In absence of stationarity, it is required that the long-term contractivity condition (9) holds uniformly along trajectories. With the second condition, essentially, we stipulate that the minorization coefficient $\beta : [0, \infty) \times \mathcal{Y} \rightarrow [0, 1]$ appearing in (8) can be substituted with a constant on appropriately chosen level sets of the Lyapunov function. This latter criterion is satisfied in all applications discussed in [35].

Assumption 2.2. *We assume that the parametric kernel $Q : \mathcal{Y} \times \mathcal{X} \times \mathcal{B}(\mathcal{X}) \rightarrow [0, 1]$ satisfies the drift condition (7) with $\gamma, K : \mathcal{Y} \rightarrow (0, \infty)$ such that*

$$\mathbb{E} \left(K(Y_j) \prod_{k=1}^n \gamma(Y_{k+j}) \right) < \infty, \quad j \in \mathbb{N}, \quad n \geq 1.$$

Furthermore the following two conditions hold:

$$A) \quad \bar{\gamma} := \limsup_{n \rightarrow \infty} \sup_{j \geq -1} \mathbb{E}^{1/n} \left(K(Y_j) \prod_{k=1}^n \gamma(Y_{k+j}) \right) < 1, \text{ where we define } K(Y_{-1}) := 1.$$

B) For some $0 < r < 1/\bar{\gamma} - 1$, the parametric kernel Q obeys the minorization condition with

$$R(y) = \frac{2K(y)}{r\gamma(y)}, \text{ and } \bar{\beta} := \sup_{y \in \mathcal{Y}} \beta(R(y), y) < 1.$$

Lemma 2.3. *Let $(X_n)_{n \in \mathbb{N}}$ be a Markov chain in a random environment $(Y_n)_{n \in \mathbb{N}}$ with parametric kernel $Q : \mathcal{Y} \times \mathcal{X} \times \mathcal{B}(\mathcal{X}) \rightarrow [0, 1]$ satisfying part A) of Assumption 2.2. Additionally, let X_0 be a random initial state independent of $\sigma(Y_n, \varepsilon_{n+1} \mid n \in \mathbb{N})$ and such that $\mathbb{E}(V(X_0)) < \infty$.*

Then for any measurable function $\Phi : \mathcal{X} \rightarrow \mathbb{R}$ satisfying

$$|\Phi(x)|^r \leq C(1 + V(x)), \quad x \in \mathcal{X}.$$

for some constants $C > 0$ and $r \in \mathbb{R}$, it follows that

$$\sup_{n \in \mathbb{N}} \mathbb{E}(|\Phi(X_n)|^r) < \infty.$$

Proof. Using Lemma C.1 from Appendix C, we can derive the following estimate:

$$\mathbb{E}([Q(Y_{n-1}) \dots Q(Y_0)V](X_0)) \leq \mathbb{E}(V(X_0)) \mathbb{E} \left(\prod_{k=0}^{n-1} \gamma(Y_k) \right) + \sum_{k=0}^{n-1} \mathbb{E} \left(K(Y_k) \prod_{j=k+1}^{n-1} \gamma(Y_j) \right).$$

Given Assumption 2.2, the Cauchy criterion applies to the sum in the second term. Consequently,

$$\sup_{n \in \mathbb{N}} \mathbb{E}([Q(Y_{n-1}) \dots Q(Y_0)V](X_0)) < \infty,$$

which further implies

$$\sup_{n \in \mathbb{N}} \mathbb{E}(|\Phi(X_n)|^r) \leq C(1 + \sup_{n \in \mathbb{N}} \mathbb{E}(V(X_n))) = C(1 + \sup_{n \in \mathbb{N}} \mathbb{E}([Q(Y_{n-1}) \dots Q(Y_0)V](X_0))) < \infty.$$

□

The following lemma ensures that under the drift and minorization conditions specified in Definition 2.2, the coupling condition holds for a suitable version of the chain.

Lemma 2.4. *Let $(X_n)_{n \in \mathbb{N}}$ be a Markov chain in a random environment $(Y_n)_{n \in \mathbb{N}}$ with random initial state X_0 independent of $\sigma(Y_n, \varepsilon_{n+1} \mid n \in \mathbb{N})$ such that $\mathbb{E}(V(X_0)) < \infty$. Then under Assumption 2.2, the chain $(X_n)_{n \in \mathbb{N}}$ admits a random iteration representation of the form (1) with appropriate $f : \mathcal{X} \times \mathcal{Y} \times [0, 1] \rightarrow \mathcal{X}$ which satisfies the coupling condition. More precisely, there exist $c_1, c_2 > 0$ constants depending only on $\bar{\gamma}, \bar{\beta}$ and r such that*

$$\sup_{j \in \mathbb{N}} \mathbb{P} \left(Z_{j,j+n}^{X_j} \neq Z_{j,j+n}^x \right) \leq c_1 (V(x) + \mathbb{E}(V(X_0)) + 1) e^{-c_2 n^{1/2}}, \quad n \geq N$$

holds for $x \in \mathcal{X}$ and appropriate $N > 0$.

Proof. The proof follows a similar argument as in [36]. For detailed steps, see Appendix C. $\square$

With the above Lemma and the results obtained for general random iterations in Section 1 in hand, we are now ready to prove the following theorem.

Theorem 2.5. *Let $(X_n)_{n \in \mathbb{N}}$ be a Markov chain in a random environment $(Y_n)_{n \in \mathbb{N}}$ satisfying Assumption 2.2, with X_0 a random initial state independent of $\sigma(Y_n, \varepsilon_{n+1} \mid n \in \mathbb{N})$, such that $\mathbb{E}[V(X_0)] < \infty$.*

Consider a measurable function $\Phi : \mathcal{X} \rightarrow \mathbb{R}$ satisfying

$$|\Phi(x)|^p \leq C(1 + V(x)), \quad x \in \mathcal{X}, \quad (11)$$

for some constants $C > 0$ and $p > 1$, and define

$$S_n := \sum_{k=1}^n (\Phi(X_k) - \mathbb{E}[\Phi(X_k)]), \quad n \in \mathbb{N}.$$

Under the above conditions, the following statements hold:

A) If the environment process $(Y_n)_{n \in \mathbb{N}}$ is strongly mixing, then the L^1 law of large numbers holds:

$$\frac{S_n}{n} \xrightarrow{L^1} 0, \quad \text{as } n \rightarrow \infty.$$

B) Suppose that for some constants $c > 0$ and $r > 2$, the mixing coefficients satisfy $\alpha^Y(n) \leq c n^{-\frac{r}{r-2}}$, $n \in \mathbb{N}$, and (11) holds with an exponent p such that $r/2 < p \leq r < \infty$. Then,

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0, \quad \text{as } n \rightarrow \infty.$$

C) If (11) holds with $p > 2$ and the mixing coefficients satisfy

$$\sum_{n \in \mathbb{N}} (n+1)^2 \alpha^Y(n)^{\frac{p-2}{p+2}} < \infty, \quad (12)$$

then the distribution of $S_n/\sqrt{n}$ and $\mathcal{N}(0, \text{Var}(n^{-1/2}S_n))$ are weakly approaching. Moreover, there exists $\sigma > 0$ such that for any $a > 0$, we have

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left(n^{-1/2} |S_n| \geq a \right) \leq \int_{\mathbb{R}} \mathbb{1}_{[-a,a]^c}(\sigma t) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

Proof. For any $n, j \in \mathbb{N}$, the sigma-algebras $\mathcal{F}_{0,j}^Y \vee \mathcal{F}_{n+j,\infty}^Y$ and $\mathcal{F}_{1,j}^\varepsilon \vee \mathcal{F}_{n+j,\infty}^\varepsilon$ are independent. Hence, by [9, Lemma 8 on page 13], we obtain

$$\begin{aligned} \alpha \left(\mathcal{F}_{0,j}^{\tilde{Y}}, \mathcal{F}_{n+j,\infty}^{\tilde{Y}} \right) &= \alpha \left(\mathcal{F}_{0,j}^Y \vee \mathcal{F}_{1,j}^\varepsilon, \mathcal{F}_{n+j,\infty}^Y \vee \mathcal{F}_{n+j,\infty}^\varepsilon \right) \\ &\leq \alpha \left(\mathcal{F}_{0,j}^Y, \mathcal{F}_{n+j,\infty}^Y \right) + \alpha \left(\mathcal{F}_{1,j}^\varepsilon, \mathcal{F}_{n+j,\infty}^\varepsilon \right), \end{aligned}$$

where $\tilde{Y}_n = (Y_n, \varepsilon_{n+1})$ for $n \in \mathbb{N}$.

On the other hand, since $\mathcal{F}_{1,j}^\varepsilon$ and $\mathcal{F}_{n+j,\infty}^\varepsilon$ are also independent, we have $\alpha(\mathcal{F}_{1,j}^\varepsilon, \mathcal{F}_{n+j,\infty}^\varepsilon) = 0$. By the definition of the dependence coefficient (3), the reverse inequality holds trivially, and therefore

$$\alpha(\mathcal{F}_{0,j}^{\tilde{Y}}, \mathcal{F}_{n+j,\infty}^{\tilde{Y}}) = \alpha(\mathcal{F}_{0,j}^Y, \mathcal{F}_{n+j,\infty}^Y), \quad n, j \in \mathbb{N}.$$

This implies that $\alpha^{\tilde{Y}}(n) = \alpha^Y(n)$ for all $n \in \mathbb{N}$.

Furthermore, by Lemma 2.3, the bound $\sup_{n \in \mathbb{N}} \mathbb{E}|\Phi(X_n)|^p < \infty$ holds for the same $p > 1$ as in (11).

Finally, by Lemma 2.4, the chain $(X_n)_{n \in \mathbb{N}}$ admits a random iteration representation satisfying the coupling condition with $b(n) = O(e^{-cn^{1/2}})$ for some $c > 0$.

Therefore, all conditions of Theorem 1.5 are satisfied, which completes the proof. $\square$

Following Lindvall [33], Györfi and Morvai introduced a stronger concept of stability known as *forward coupling* [25]. This notion of stability has proven useful in the study of queuing systems. Specifically, under mild ergodicity assumptions, it was found that waiting times in single-server queuing systems operating in a subcritical regime are forward coupled with a stationary and ergodic sequence (See in Section 3).

Definition 2.6. *We say that the sequence $(W_n)_{n \in \mathbb{N}}$ is forward coupled with the sequence $(W'_n)_{n \in \mathbb{N}}$ if there exists an almost surely finite random time τ such that*

$$W_n = W'_n$$

for $n > \tau$.

The next theorem revisits the case of a stationary environment, previously studied by Gerencsér [21], Rásonyi [22], Lovas [35], Truquet [49], and others, and establishes novel results.

Theorem 2.7. *Let $(Y_n)_{n \in \mathbb{Z}}$ be a strongly stationary process, and assume that Assumption 2.2 holds. Furthermore, let X_0 be a random initial state independent of $\sigma(Y_m, \varepsilon_{n+1} \mid m \in \mathbb{Z}, n \in \mathbb{N})$. Consider the stationary process $(Y_n, X_n^*)_{n \in \mathbb{Z}}$ satisfying the iteration (1). Then, appropriate versions of $(X_n)_{n \in \mathbb{N}}$ and $(X_n^*)_{n \in \mathbb{N}}$ are forward coupled. Moreover, the tail probability of the random coupling time τ satisfies the estimate*

$$\mathbb{P}(\tau > n) \leq c_1 (1 + \mathbb{E}(V(X_0))) e^{-c_2 n^{1/2}}$$

with constants $c_1, c_2 > 0$ depending only on $\bar{\beta}, \bar{\gamma}$ and r .

Proof. The proof follows similar lines as the proof of Lemma 2.4. For readability, it is provided in Appendix C. $\square$

The following important corollary of the above theorem provides an explicit and tractable upper bound for the total variation distance between $\text{Law}(X_n)$ and $\text{Law}(X_n^*)$, $n \in \mathbb{N}$. This bound is significantly sharper than those presented in our earlier paper [35].

Corollary 2.8. *Under the conditions of Theorem 2.7, the following rate estimate holds:*

$$\|\text{Law}(X_n) - \text{Law}(X_n^*)\|_{TV} \leq 2c_1 (1 + \mathbb{E}(V(X_0))) e^{-c_2 n^{1/2}},$$

with the same constants c_1 and c_2 as in Theorem 2.7.

Proof. According to the optimal transportation cost characterization of the total variation distance, we have

$$\frac{1}{2} \|\text{Law}(X_n) - \text{Law}(X_n^*)\|_{TV} \leq \inf_{\kappa \in \mathcal{C}(X_n, X_n^*)} \int_{\mathcal{X} \times \mathcal{X}} \mathbf{1}_{x \neq y} \kappa(dx, dy),$$

where $\mathcal{C}(X_n, X_n^*)$ denotes the set of probability measures on $\mathcal{B}(\mathcal{X} \times \mathcal{X})$ with marginals $\text{Law}(X_n)$ and $\text{Law}(X_n^*)$, respectively. By a slight abuse of notation, we can assume that $(X_n)_{n \in \mathbb{N}}$ and $(X_n^*)_{n \in \mathbb{N}}$ are forward coupled. Using Theorem 2.7, we can estimate the right-hand side further by writing

$$\inf_{\kappa \in \mathcal{C}(X_n, X_n^*)} \int_{\mathcal{X} \times \mathcal{X}} \mathbf{1}_{x \neq y} \kappa(dx, dy) \leq \mathbb{P}(X_n \neq X_n^*) = \mathbb{P}(\tau > n) \leq c_1 (1 + \mathbb{E}(V(X_0))) e^{-c_2 n^{1/2}}$$

which gives the desired inequality. $\square$

In the remaining part of this section, we aim to verify the Central Limit Theorem (CLT) for certain functionals of a Markov chain in a non-stationary random environment. There are multiple approaches to achieve this result. For instance, we can use Corollary A.6, or the results described in Merlevède and Peligrad's recent paper [39]. The advantage of the latter approach is that it not only yields the usual CLT but also the functional CLT, and the conditions on the decay of the mixing properties are weaker. Regardless of the chosen method, we must verify the equality (35), which states that the variance of the sum grows at the same rate as the sum of the variances. Using either Remark A.7 or Lemma 2.3, it can be easily shown that $\text{Var}(S_n)$ grows at most at the same rate as the sum of the variances $\text{Var}(\Phi(X_k))$, $k = 1, 2, \dots, n$. However, it is far from trivial to establish a lower bound for $\text{Var}(S_n)$ to ensure the equality in (35). For this, we follow a new approach employing the law of total variance and the Cramér-Rao bound known from estimation theory and information geometry. For any fixed $n \geq 1$, we treat the conditional distribution of $\underline{X} := (X_1, \dots, X_n)$ given the environment $\underline{Y} := (Y_0, \dots, Y_{n-1})$ as a parametric statistical model, where the environment $\underline{Y}$ is considered as the parameter. For technical reasons, we will restrict our discussion in the remaining part of this section to the case when $\mathcal{Y} = \mathbb{R}^m$ with $m \in \mathbb{N}_+$.

Assumption 2.9. *We assume the existence of a reference Borel measure ν on $\mathcal{B}(\mathcal{X})$ such that, for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$, the measure $Q(y, x, \cdot)$ is absolutely continuous with respect to ν . Furthermore, the parametric family of conditional densities*

$$\mathcal{Y} \ni y \mapsto p_y(z | x) = \frac{dQ(y, x, \cdot)}{d\nu}(z), \quad (x, z) \in \mathcal{X} \times \mathcal{X}$$

and the measurable function $\Phi : \mathcal{X} \rightarrow \mathbb{R}$, playing the role of an estimator, satisfy all the regularity conditions required for the Cramér-Rao inequality (See Theorem 1A on page 147 in [3] or for a more general version, Corollary 5 in [2]).

Lemma 2.10. *Let $(X_n)_{n \in \mathbb{N}}$ be a Markov chain in a random environment $(Y_n)_{n \in \mathbb{N}}$ with parametric kernel $Q : \mathcal{Y} \times \mathcal{X} \times \mathcal{B}(\mathcal{X}) \rightarrow [0, 1]$ and $\Phi : \mathcal{X} \rightarrow \mathbb{R}$ satisfying Assumption 2.9. Suppose that X_0 is a random initial state independent of $\sigma(Y_n, \varepsilon_{n+1} | n \in \mathbb{N})$ with distribution $\text{Law}(X_0)$ being absolutely continuous w.r.t. ν , where ν is as in Assumption 2.9.*

Then, for the variance of the partial sum $S_n = \Phi(X_1) + \dots + \Phi(X_n)$, we have the lower bound

$$\text{Var}(S_n) \geq \sum_{k=0}^{n-1} \mathbb{E} \left[\frac{1}{r(I_k)} \|\partial_{y_k} \mathbb{E}[S_n | \underline{Y}]\|^2 \right], \quad n \geq 1,$$

where $r(I_k)$ denotes the spectral radius of the Fisher information matrix

$$I_k := \mathbb{E} \left[(\partial_y \log p_y(X_{k+1} | X_k))^\top (\partial_y \log p_y(X_{k+1} | X_k)) \Big| \underline{Y} \right].$$

Proof. By the law of total variance, we can write

$$\text{Var}(S_n) = \mathbb{E} [\text{Var}(S_n | \mathcal{F}_{0,n-1}^Y)] + \text{Var}(\mathbb{E}[S_n | \mathcal{F}_{0,n-1}^Y]) \geq \mathbb{E} [\text{Var}(S_n | \mathcal{F}_{0,n-1}^Y)]. \quad (13)$$

For typographical reasons, we use the notations $\text{Var}_{\underline{Y}}(S_n) := \text{Var}(S_n | \underline{Y})$ and $\text{Cov}_{\underline{Y}}(\cdot, \cdot) := \text{Cov}(\cdot, \cdot | \underline{Y})$. The conditional variance of the partial sum can be expressed as follows:

$$\text{Var}_{\underline{Y}}(S_n) = \sum_{k,l=1}^n \text{Cov}_{\underline{Y}}(\Phi(X_k), \Phi(X_l)) = \mathbf{1}^\top \text{Cov}_{\underline{Y}}(\Phi(\underline{X})) \mathbf{1}, \quad (14)$$

where $\text{Cov}_{\underline{Y}}(\Phi(\underline{X}))$ is the conditional covariance matrix of the random vector

$$\Phi(\underline{X}) = (\Phi(X_1), \dots, \Phi(X_n))^\top,$$

$\mathbf{1} = [1, \dots, 1]^\top \in \mathbb{R}^n$ denotes the size n vector of 1-s.

Using the Cramér-Rao inequality, we arrive at the following estimate

$$\text{Cov}_{\underline{Y}}(\Phi(\underline{X})) \geq (\partial_{\underline{y}}\phi(\underline{y}))I(\underline{y})^{-1}(\partial_{\underline{y}}\phi(\underline{y}))^\top, \quad (15)$$

where $\phi : \mathcal{Y}^n \rightarrow \mathbb{R}^n$, $\phi(\underline{y}) = \mathbb{E}[\Phi(\underline{X}) \mid \underline{Y} = \underline{y}]$ and $I(\underline{y})$ is the Fisher information matrix.

Note that $\partial_{\underline{y}}\phi(\underline{y}) \in \text{Lin}(\mathcal{Y}^n, \mathbb{R}^n)$ and $I(\underline{y}) \in \text{Lin}(\mathcal{Y}^n)$ are linear operators with block representation

$$\begin{aligned} \partial_{\underline{y}}\phi(\underline{y}) &= [\partial_{y_0}\phi(\underline{y}), \dots, \partial_{y_{n-1}}\phi(\underline{y})] \\ [I(\underline{y})]_{kl} &= \mathbb{E} \left[\left(\partial_{y_k} \log p(\underline{X} \mid \underline{y}) \right)^\top \left(\partial_{y_l} \log p(\underline{X} \mid \underline{y}) \right) \middle| \underline{Y} = \underline{y} \right] = -\mathbb{E} \left[\partial_{y_k y_l}^2 \log p(\underline{X} \mid \underline{y}) \middle| \underline{Y} = \underline{y} \right], \end{aligned}$$

where $\partial_{y_i}\phi(\underline{y}) \in \text{Lin}(\mathcal{Y}, \mathbb{R}^n)$, $i = 0, \dots, n-1$, and $[I(\underline{y})]_{kl} \in \text{Lin}(\mathcal{Y})$, $k, l = 0, \dots, n-1$.

By the Markov property, for the conditional density of $\underline{X}$ given $\underline{y}$ can be expressed as the product of $\pi(X_0) = \frac{\text{dLaw}(X_0)}{\text{d}\nu}$ and the parametric transition densities $p_{y_{k-1}}(x_k \mid x_{k-1})$, $k = 1, \dots, n$. Consequently, we obtain

$$\log p(\underline{X} \mid \underline{y}) = \log \pi(X_0) + \sum_{k=1}^n \log p_{y_{k-1}}(X_k \mid X_{k-1}),$$

which implies that the Fisher information matrix has a block-diagonal form:

$$[I(\underline{y})]_{kl} = \begin{cases} I_k := \mathbb{E} \left[\left(\partial_{y_k} \log p_{y_k}(X_{k+1} \mid X_k) \right)^\top \left(\partial_{y_k} \log p_{y_k}(X_{k+1} \mid X_k) \right) \middle| \underline{Y} = \underline{y} \right] & \text{if } k = l \\ 0 & \text{if } k \neq l. \end{cases}$$

Observe that $\mathbf{1}^\top \partial_{\underline{y}}\phi(\underline{y}) = \partial_{\underline{y}}\mathbb{E}[S_n \mid \underline{Y} = \underline{y}] \in \text{Lin}(\mathcal{Y}^n, \mathbb{R})$ hence substituting back the above form of the Fisher information matrix into (14) and applying (15) yields

$$\text{Var}_{\underline{Y}}(S_n) \geq \sum_{k=0}^{n-1} \partial_{y_k}\mathbb{E}[S_n \mid \underline{Y}] I_k^{-1} \partial_{y_k}\mathbb{E}[S_n \mid \underline{Y}]^\top \geq \sum_{k=0}^{n-1} \frac{1}{r(I_k)} \|\partial_{y_k}\mathbb{E}[S_n \mid \underline{Y}]\|^2,$$

where $r(I_k)$ refers to the spectral radius of the Fisher operator I_k . At least, by (13), we obtain

$$\text{Var}(S_n) \geq \sum_{k=0}^{n-1} \mathbb{E} \left[\frac{1}{r(I_k)} \|\partial_{y_k}\mathbb{E}[S_n \mid \underline{Y}]\|^2 \right]. \quad (16)$$

□

Combining the above lemma with Corollary A.6 leads to the following significant conclusion.

Corollary 2.11. *In addition to the conditions of Theorem 2.5 and Lemma 2.10, assume that the function $\Phi : \mathcal{X} \rightarrow \mathbb{R}$ satisfies inequality (11) for some exponent $p > 2$. Further, suppose that*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{E} \left[\frac{1}{r(I_k)} \|\partial_{y_k}\mathbb{E}[S_n \mid \underline{Y}]\|^2 \right] > 0.$$

Then it is evident that the sequence $(n/\text{Var}(S_n))_{n \geq 1}$ is bounded. Consequently, by Corollary A.6, it follows that in addition to the results of Theorem 2.5, we also have

$$\frac{1}{\text{Var}(S_n)^{1/2}} S_n \Longrightarrow \mathcal{N}(0, 1), \text{ as } n \rightarrow \infty$$

in distribution. Therefore, the Central Limit Theorem holds.

To derive a lower bound for $\text{Var}(S_n)$, we strongly relied on the Cramér-Rao bound, which requires certain differentiability and regularity conditions. However, these differentiability assumptions can be relaxed, either by applying a version of the Cramér-Rao inequality developed for non-differentiable models [18], or by using some other information inequality (see, e.g., paragraph 3 in Chapter 30 of [3]).

It seems that it might be as difficult to evaluate the lower bound in (16) as to directly verify that

$$\text{Var}(S_n) = O\left(\sum_{k=1}^n \text{Var}(\Phi(X_k))\right).$$

To show that the lower bound on the right-hand side of (16) is tractable, we establish a more explicit sufficient condition under the assumptions that $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}$ and the function f in the iteration (1) satisfies some monotonicity condition.

Proposition 2.12. *Assume that $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}$, moreover the function $f : \mathcal{X} \times \mathcal{Y} \times [0, 1] \rightarrow \mathbb{R}$ defining the iteration (1) is continuously differentiable and monotonically increasing in its first two variables. Besides the assumptions of Lemma 2.10, assume also that*

$$r^* := \sup_{n \in \mathbb{N}} \mathbb{E}[r(I_n)] = \sup_{n \in \mathbb{N}} \mathbb{E}\left[(\partial_y \log p_y(X_{n+1} | X_n))^2\right] < \infty, \quad (17)$$

and there exists a function $g : \mathcal{Y} \times [0, 1] \rightarrow (0, \infty)$ such that

$$\partial_y f(x, y, u) \geq g(y, u), \quad x, y \in \mathbb{R}, u \in [0, 1].$$

Then, for the variance of the partial sums $S_n = X_1 + \dots + X_n$, we have

$$\text{Var}(S_n) \geq \frac{1}{r^*} \sum_{k=0}^{n-1} \mathbb{E}[g(Y_k, \varepsilon_1)]^2.$$

Proof. By the Cauchy-Schwartz inequality, for $k = 0, \dots, n-1$, we have

$$\mathbb{E}[\partial_{y_k} \mathbb{E}[S_n | \underline{Y}]]^2 \leq \mathbb{E}\left[\frac{1}{r(I_k)} (\partial_{y_k} \mathbb{E}[S_n | \underline{Y}])^2\right] \mathbb{E}[r(I_k)]$$

hence using the inequality (17), we arrive at

$$\text{Var}(S_n) \geq \sum_{k=0}^{n-1} \mathbb{E}\left[\frac{1}{r(I_k)} \|\partial_{y_k} \mathbb{E}[S_n | \underline{Y}]\|^2\right] \geq \frac{1}{r^*} \sum_{k=0}^{n-1} \mathbb{E}[\partial_{y_k} \mathbb{E}[S_n | \underline{Y}]]^2. \quad (18)$$

Fix $0 \leq k < n$. Then, by the monotonicity assumption, we have $\partial_{y_k} \mathbb{E}[X_l | \underline{Y}] \geq 0$ for any $l = 1, \dots, n$. Hence, using that the process $(\varepsilon_n)_{n \geq 1}$ is i.i.d. and independent of the environment $(Y_n)_{n \in \mathbb{N}}$, we can estimate as

$$\partial_{y_k} \mathbb{E}[S_n | \underline{Y}] = \sum_{l=1}^n \partial_{y_k} \mathbb{E}[X_l | \underline{Y}] \geq \partial_{y_k} \mathbb{E}[X_{k+1} | \underline{Y}] = \mathbb{E}[\partial_y f(X_k, Y_k, \varepsilon_{k+1}) | \underline{Y}] \geq \mathbb{E}[g(Y_k, \varepsilon_1) | \underline{Y}].$$

Substituting this estimate back into the inequality (18), we obtain the desired result. $\square$

Remark 2.13. *If the environment $(Y_n)_{n \in \mathbb{Z}}$ is stationary, then the lower bound for the variance of $S_n = X_1 + \dots + X_n$ given in Proposition 2.12 simplifies to*

$$\text{Var}(S_n) \geq \frac{n}{r^*} \mathbb{E}[g(Y_0, \varepsilon_1)]^2,$$

so the inequality required in the assumptions of Corollary 2.11 is clearly satisfied.

Using the recent results by Merlevède and Peligrad in [39], we can derive the following functional central limit theorem for certain functionals of a MCRE. This result is significantly stronger than the statement formulated in Corollary 2.11.

Theorem 2.14. *Assume that the conditions of Corollary 2.11 hold, with the exception that instead of the inequality (12) appearing among the conditions of Theorem 2.5, we only require that*

$$\sum_{n \geq 1} n^{2/(p-2)} \alpha^Y(n) < \infty.$$

Furthermore, for $n \geq 1$, let $v_n(t) = \min\{1 \leq k \leq n \mid \text{Var}(S_k) \geq t \text{Var}(S_n)\}$, and for $1 \leq k \leq n$, define

$$\xi_{k,n} = \frac{\Phi(X_k) - \mathbb{E}[\Phi(X_k)]}{\text{Var}(S_n)^{1/2}}.$$

Then the sequence of functions $B_n(t) = \sum_{k=1}^{v_n(t)} \xi_{k,n}$, $t \in (0, 1]$, $n \geq 1$, converges in distribution in $D([0, 1])$ (equipped with the uniform topology) to $(B_t)_{t \in [0, 1]}$, where $(B_t)_{t \in [0, 1]}$ is a standard Brownian motion.

Proof. In the proof, we check the conditions of Corollary 2.2 found on page 3 of [39]. These conditions can be divided into two groups: conditions on the moments of the random variables $\xi_{k,n}$, $1 \leq k \leq n$, and conditions on the mixing properties of these variables. We first verify the fulfillment of the conditions in the first group.

By Lemma 2.3, we have $c_1 = \sup_{n \in \mathbb{N}} \mathbb{E}(|\Phi(X_n)|^p) < \infty$, and Lemma 2.10 ensures that $n/\text{Var}(S_n) < c_2$ for some constant $c_2 > 0$ and for all $n \geq 1$. Then, we have

$$\|\xi_{k,n}\|_p = \frac{1}{\text{Var}(S_n)^{1/2}} \|\Phi(X_k) - \mathbb{E}[\Phi(X_k)]\|_p \leq 2n^{-1/2} \left(\frac{n}{\text{Var}(S_n)} \right)^{1/2} \|\Phi(X_k)\|_p \leq 2c_1^{1/r} c_2^{1/2} n^{-1/2},$$

from which it immediately follows that $\max_{1 \leq k \leq n} \|\xi_{k,n}\|_p \rightarrow 0$ as $n \rightarrow \infty$, and furthermore, $\sup_{n \geq 1} \sum_{k=1}^n \|\xi_{k,n}\|_p^2 < \infty$. Comparing this with the remark following Corollary 2.2 in [39], we conclude that condition (3) and the first part of condition (7) related to moments in [39] are satisfied. Additionally, condition (1) is inherently satisfied by the array $\{\xi_{k,n} \mid 1 \leq k \leq n, n \geq 1\}$ due to its definition.

We now proceed to verify the conditions related to the mixing properties. The precise definitions of the weak strong mixing coefficients $\alpha_1(k)$, $\alpha_{1,n}(k)$, and $\alpha_{2,n}(k)$ can also be found in [39]. For our purposes, it suffices to note that the following sequence of inequalities holds:

$$\alpha_{1,n}(k) \leq \alpha_{2,n}(k) \leq 2\alpha^{\Phi(X)}(k), \quad 1 \leq k \leq n, \quad n \geq 1,$$

therefore, to verify the second part of condition (7) in [39] and to satisfy condition (8), it is enough to establish that

$$\sum_{n \geq 1} n^{2/(p-2)} \alpha^{\Phi(X)}(n) < \infty. \quad (19)$$

From here, the proof continues by estimating the mixing coefficient $\alpha^{\Phi(X)}$ and follows exactly the same path as described in the proof of Theorem 1.5 and 2.5. By Lemma 2.4, the chain $(X_n)_{n \in \mathbb{N}}$ admits a random iteration representation satisfying the coupling condition with $b(n) = O(e^{-cn^{1/2}})$, $c > 0$. Therefore, we can estimate

$$\alpha^{\Phi(X)}(n) \leq \alpha^X(n) \leq \alpha^Y(\lfloor n/2 \rfloor + 1) + b(n - \lfloor n/2 \rfloor), \quad n > 2N,$$

where $N > 0$ is as in Lemma 2.4. Considering that $\sum_{n=1}^{\infty} n^{2/(p-2)} b(n - \lfloor n/2 \rfloor) < \infty$, and that the condition $\sum_{n \geq 1} n^{2/(p-2)} \alpha^Y(n) < \infty$ assumed, we obtain that the infinite series appearing in condition (19) is convergent.

With this, we have verified all of the conditions of Corollary 2.2 in [39], from which it follows that the statement formulated in the theorem holds. This completes the proof. $\square$

Remark 2.15. *In the proof of the above Theorem 2.14, we do not use Theorem 2.1 of [39] in its strongest form, but rather we verify only the conditions of Corollary 2.2. It is likely that Theorem 2.14 can be further strengthened and generalized for Markov chains in a random environment satisfying a weak-strong mixing-type condition. However, this is beyond the scope of the present study.*

3 Single server queuing systems

In the early 20th century, Danish engineer Agner Krarup Erlang pioneered what would later be known as queuing theory [17]. His work at a telephone company, where he developed a mathematical model to determine the minimum number of telephones needed to handle calls efficiently, laid the foundation for this field. Today, queuing theory extends far beyond telecommunications, significantly influencing areas like inventory management, logistics, transportation, industrial engineering, and service design. Notably, it plays a key role in reducing costs within product-service design [43].

For simplicity, we focus on single-server queuing systems with infinite buffer and first-in, first-out (FIFO) service discipline (see Figure 1). It's worth noting that more complex queuing systems, such as those with multiple servers, can be analyzed using analogous methods. Let the time between

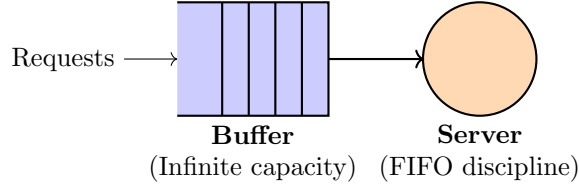


Figure 1: Schematic overview of the single-server queuing system under investigation: Z_n is the time between the arrivals of the n -th and $(n+1)$ -th customers, S_n is the time to serve the n -th customer, and W_n is their waiting time before service.

the arrival of customers $n+1$ and n is denoted by Z_{n+1} , and the service time for customer n is given by S_n , for $n \in \mathbb{N}$. Then, the evolution of the waiting time W_n of customer n can be described by the *Lindley recursion*

$$W_{n+1} = (W_n + S_n - Z_{n+1})_+, \quad n \in \mathbb{N}, \quad (20)$$

with $W_0 := 0$, meaning that we begin with an empty queue.

The ergodic theory of general state space Markov chains allows to treat the case where $(S_n)_{n \in \mathbb{N}}$, $(Z_n)_{n \in \mathbb{N}}$ are i.i.d. sequences, independent of each other. However, dependencies frequently arise in queuing networks when the arrival processes intertwine with the departure processes of other queues. Additionally, factors like complex processing operations, including batching or the presence of multiple distinct customer classes, can introduce intricate interdependencies within the system. Consequently, the renewal process assumption for the arrival process, which makes queueing models amenable to simple analysis, no longer holds.

To the best of our knowledge, Loynes was the first who studied the stability of waiting times under the assumption that the pair (S, Z) is merely stationary and ergodic [37]. Stability of W_n , $n \in \mathbb{N}$ means here that there exist a unique limit distribution of W_n as $n \rightarrow \infty$, whatever the initialization W_0 is. Loynes introduced the terminology categorizing queues as 'subcritical' when $\mathbb{E}(S_0) < \mathbb{E}(Z_0)$, 'critical' when $\mathbb{E}(S_0) = \mathbb{E}(Z_0)$, and 'supercritical' when $\mathbb{E}(S_0) > \mathbb{E}(Z_0)$. Loynes proved for single-server queuing systems that subcritical queues are stable, supercritical queues are unstable and critical queues can be stable, properly substable, or unstable [37].

Building upon Loynes's foundational work, Györfi and Morvai expanded and refined the understanding of these queuing systems in [25]. They extended Loynes' result by proving that for subcritical queues, an even stronger version of stability called *forward coupling* holds also true (c.f. Definition 2.6). Györfi and Morvai's theorem concerning queues in this general setting reads as follows:

Theorem 3.1. *Let $\xi_n = S_n - Z_{n+1}$ and assume that the process $(\xi_n)_{n \in \mathbb{Z}}$ is stationary and ergodic with $\mathbb{E}(S_0) < \mathbb{E}(Z_0)$. Then $(W_n)_{n \in \mathbb{N}}$ is forward coupled with a stationary and ergodic $(W'_n)_{n \in \mathbb{Z}}$ such that $W'_0 = \sup_{n \in \mathbb{N}} Y_n$, where $Y_0 = 0$ and $Y_n = \sum_{k=1}^n \xi_{-k}$, $n \geq 1$.*

While the aforementioned theorems ensure that the distribution of waiting times converges to a well-defined limiting law, they regrettably do not furnish any insights into the properties of this

stationary limit, nor do they shed light on the speed of convergence. When addressing the latter question, the only available rate estimate that comes to our aid is encapsulated in an inequality found in Theorem 4 on page 25 of [4], which is expressed as:

$$\|\text{Law}(W_n) - \text{Law}(W'_0)\|_{TV} \leq \mathbb{P} \left(\min_{0 \leq k \leq n} X_k > \max(W_1, W'_0 + \xi_0) \right), \quad (21)$$

where ξ_n is as in Theorem 3.1, and $(X_n)_{n \in \mathbb{N}}$ is defined as $X_0 = 0$, $X_n = \sum_{k=1}^n \xi_k$, $n \geq 1$. However, it's worth noting that the primary limitation of this formula lies in its practical applicability. Evaluating the probability on the right-hand side of this equation can be a formidable task, rendering it impractical as a concrete and readily usable rate estimate.

If the inter-arrival times are i.i.d. and the sequences S and Z are independent of each other, then the process W can be viewed as a Markov chain in the random environment S , with driving noise Z . Conversely, if the service times are i.i.d., and again S and Z are independent, W can be regarded as a Markov chain in the random environment Z , with driving noise S . Therefore, both of these special cases of queuing systems fall within the theoretical framework outlined in Section 2. In [34, 35], we analyzed such queuing models with an additional Gärtner-Ellis-type condition (see Assumption 3.4 below or Assumption 4.2 in [35]), which is a well-established practice in queuing theory. For instance, in Section 3 of [25], similar conditions are employed to investigate the exponential tail behavior of the limit distribution of queue length when arrivals exhibit weak dependence. Further justification for the applicability of Gärtner-Ellis-type conditions can be found in Remark 4.3 of [35].

In what follows, we revisit the queuing model studied in [35]. Specifically, we consider the case where S and Z are independent, and the latter is an i.i.d. sequence. The reverse case discussed in [34] can be treated analogously. We begin by examining the scenario where the sequence of service times is strictly stationary. Regarding this case, we formulate our standing assumptions.

Assumption 3.2. *There exists an $M > 0$ such that the sequence of service times $(S_n)_{n \in \mathbb{Z}}$ is a strictly stationary process taking values in $[0, M]$. Furthermore, $(S_n)_{n \in \mathbb{Z}}$ is independent of the sequence $(Z_n)_{n \in \mathbb{Z}}$.*

Assumption 3.3. *The inter-arrival times $(Z_n)_{n \in \mathbb{Z}}$ form an i.i.d. sequence of $\mathbb{R}_+$ -valued random variables, and $\mathbb{E}[S_0] < \mathbb{E}[Z_1]$ holds.*

Assumption 3.4. *There exists $\eta > 0$ such that for all $t \in (-\eta, \eta)$, the limit*

$$\Gamma(t) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} e^{t(S_1 + \dots + S_n)}$$

exists and Γ is differentiable on $(-\eta, \eta)$.

Under Assumptions 3.2, 3.3, and 3.4, by Lemma 4.4 of [35], the sequence of waiting times $(W_n)_{n \in \mathbb{N}}$, defined by the Lindley recursion (20), forms a Markov chain in the random environment $(S_n)_{n \in \mathbb{Z}}$ on the state space $\mathcal{X} = \mathbb{R}_+$ with parametric kernel

$$Q(s, w, A) := \mathbb{P} \left[(w + s - Z_1)_+ \in A \right], \quad s \in [0, M], w \in \mathbb{R}_+, A \in \mathcal{B}(\mathbb{R}_+).$$

Furthermore, there exists $\bar{t} > 0$ such that, with the following coefficients:

$$\begin{aligned} V(w) &:= e^{\bar{t}w} - 1, \quad w \geq 0, \\ \gamma(s) &:= \mathbb{E} \left[e^{\bar{t}(s - Z_1)} \right], \quad s \geq 0, \\ K &:= e^{\bar{t}M}, \end{aligned}$$

the drift condition (7) and the long-term contractivity condition (9) are satisfied, meaning

$$[Q(s)V](w) \leq \gamma(s)V(w) + K,$$

and

$$\bar{\gamma} := \limsup_{n \rightarrow \infty} \left(\mathbb{E}^{1/n} \left[K \prod_{k=1}^n \gamma(S_k) \right] \right) < 1.$$

We now introduce an additional assumption on the inter-arrival times, which will be required to establish the minorization condition.

Assumption 3.5. One has $\mathbb{P}(Z_0 \geq \tau) > 0$ for

$$\tau := M + \frac{4}{\frac{1}{\bar{\gamma}^{1/2}} - 1}.$$

This does not impose a significant restriction, since if Z_0 is an unbounded random variable, Assumption 3.5 is automatically satisfied.

By Lemma 4.6 of [35], under Assumptions 3.2, 3.3, 3.4, and 3.5, it can be shown that there exists $\bar{\beta} \in (0, 1)$ such that, for all $s \in [0, M]$, $A \in \mathcal{B}(\mathbb{R}_+)$, and $w \in V^{-1}([0, R(s)])$,

$$Q(s, w, A) \geq (1 - \bar{\beta})\delta_0(A), \text{ where } R(s) = \frac{2K(s)}{\epsilon\gamma(s)}, \quad (22)$$

$\epsilon := \left(\frac{1}{\bar{\gamma}^{1/2}} - 1\right)/2$, and δ_0 is the one-point mass concentrated at 0.

The main result of Chapter 4, which discusses queuing theory applications in [35], is Theorem 4.7. It states that under Assumptions 3.2, 3.3, 3.4, and 3.5, there exists a probability measure μ_* on $\mathcal{B}(\mathbb{R}_+)$, independent of the initial length of the queue, such that

$$\|\text{Law}(W_n) - \mu_*\|_{TV} \leq c_1 e^{-c_2 n^{1/3}}, \quad (23)$$

for some $c_1, c_2 > 0$. Furthermore, if $(S_n)_{n \in \mathbb{Z}}$ is ergodic, then for an arbitrary measurable and bounded function $\Phi : \mathbb{R}_+ \rightarrow \mathbb{R}$,

$$\frac{\Phi(W_0) + \dots + \Phi(W_{n-1})}{n} \rightarrow \int_{\mathbb{R}_+} \Phi(z) \mu_*(dz), \quad (24)$$

in L^p , for all $1 \leq p < \infty$.

Using Theorem 2.7 and Corollary 2.8, we can prove even more. In the queuing model we study, for instance, the sequence of waiting times $(W_n)_{n \in \mathbb{N}}$ is forward coupled with a strictly stationary sequence $(W_n^*)_{n \in \mathbb{N}}$. This does not require assuming the ergodicity of the service time sequence, and we obtain a tractable upper bound on the tail probability of the random coupling time. Finally, we demonstrate a faster convergence rate than that presented in the estimate (23). The following theorem addresses this result.

Theorem 3.6. *Let Assumptions 3.2, 3.3, 3.4, and 3.5 be in force. Then there exists a stationary process $(S_n, W_n^*)_{n \in \mathbb{Z}}$ that satisfies the Lindley recursion (20). Moreover, appropriate versions of the processes $(W_n)_{n \in \mathbb{N}}$ and $(W_n^*)_{n \in \mathbb{N}}$ are forward coupled. For the tail probability of the random coupling time τ , we have the bound*

$$\mathbb{P}(\tau > n) \leq c_1 e^{-c_2 n^{1/2}},$$

for suitable constants $c_1, c_2 > 0$, depending only on the quantities $\bar{\beta}$, $\bar{\gamma}$, and ϵ . Furthermore, for the total variation distance of $\text{Law}(W_n)$ and $\text{Law}(W_n^*)$ we have the following estimate:

$$\|\text{Law}(W_n) - \text{Law}(W_n^*)\|_{TV} \leq 2c_1 e^{-c_2 n^{1/2}},$$

with the same constants c_1 and c_2 .

Proof. The sequence of service times $(S_n)_{n \in \mathbb{N}}$ is strictly stationary, so Assumption 2.2 A) simplifies to the (9) long-term contractivity condition used in our earlier paper [35]. We have already shown above that this holds under the conditions of the theorem. From the minorization condition (22), part B) of Assumption 2.2 follows. Thus, we can apply Theorem 2.7, which states that there exists a stationary process $(S_n, W_n^*)_{n \in \mathbb{Z}}$ that satisfies the Lindley recursion (20). Moreover, appropriate versions of the processes $(W_n)_{n \in \mathbb{N}}$ and $(W_n^*)_{n \in \mathbb{N}}$ are forward coupled. Considering the specific form of the Lyapunov function in the drift condition, $V(w) = e^{tw} - 1$, and the initial condition $W_0 = 0$, we obtain

$$\mathbb{E}(V(W_0)) = 0.$$

Therefore, for the tail probability of the random coupling time τ , we have

$$\mathbb{P}(\tau > n) \leq c_1(1 + \mathbb{E}(V(W_0)))e^{-c_2 n^{1/2}} = c_1 e^{-c_2 n^{1/2}},$$

as stated. Finally, the estimate for the total variation distance between $\text{Law}(W_n)$ and $\text{Law}(W_n^*)$ immediately follows from Corollary 2.8. $\square$

In the remaining part of this section, we relax the assumption that the sequence of service times $(S_n)_{n \in \mathbb{Z}}$ is stationary. We only assume that it is a sequence of weakly dependent variables with sufficiently favorable mixing properties. Accordingly, our assumptions will be as follows.

Assumption 3.7. *We assume that the inter-arrival time sequence $(Z_n)_{n \geq 1}$ is i.i.d. and independent of the service time sequence $(S_n)_{n \in \mathbb{N}}$. Moreover, we also assume that $\mathbb{E}(Z_0^2) < \infty$.*

Assumption 3.8. *We assume that the service time sequence $(S_n)_{n \in \mathbb{N}}$ takes values in $\mathcal{Y} = [0, M]$. For $t \geq 0$, we define the function*

$$\Lambda(t) = \limsup_{n \rightarrow \infty} \sup_{j \in \mathbb{N}} \frac{1}{n} \log \mathbb{E} \left[\exp \left(t \sum_{k=0}^n (S_{k+j} - Z_{k+j+1}) \right) \right].$$

Assume that there exists a parameter $\bar{t} > 0$ such that $\Lambda(\bar{t}) < 0$.

Remark 3.9. *In Lemma 4.4 of [35], under Assumption 3.4, we showed that there exists a parameter $\bar{t} > 0$ such that in Assumption 3.8, $\Lambda(\bar{t}) < 0$. Although such a condition is quite standard in both queuing theory and large deviation theory, verifying the differentiability of the function $t \mapsto \Gamma(t)$ in Assumption 3.4 can be challenging in practice. Moreover, the proof of Lemma 4.4 in [35] does not generalize to the case of a non-stationary sequence $(S_n)_{n \in \mathbb{N}}$, since the definition of the function Λ includes a supremum, which prevents the application of convexity-based arguments.*

For fixed $n, j \in \mathbb{N}$, define

$$\lambda_{n,j}(t) = \frac{1}{n} \log \mathbb{E} \left[\exp \left(t \sum_{k=0}^n (S_{k+j} - Z_{k+j+1}) \right) \right].$$

If $\limsup_{n \rightarrow \infty} \mathbb{E}(S_n) < \mathbb{E}(Z_1)$, it can be shown that for sufficiently large n , $\lambda'_{n,j}(0) < 0$. Additionally, if the mixing coefficients $\alpha^S(n)$ decay sufficiently fast to zero, it can be established that $\lambda''_{n,j}(0) < \infty$. However, this alone does not ensure the existence of a parameter $\bar{t} > 0$ such that $\Lambda(\bar{t}) < 0$.

The goal of the next proposition is to provide a sufficient condition under which the otherwise difficult-to-verify requirement on $\Lambda(t)$ in Assumption 3.8 holds. This condition relies on a mixing property stronger than α -mixing, known as ψ -mixing. In analogy with the dependence measure $\alpha(\mathcal{G}, \mathcal{H})$ introduced in Section 1, we define

$$\psi(\mathcal{G}, \mathcal{H}) = \sup \left\{ \left| \frac{\mathbb{P}(G \cap H)}{\mathbb{P}(G)\mathbb{P}(H)} - 1 \right| : \mathbb{P}(G) > 0, \mathbb{P}(H) > 0, G \in \mathcal{G}, H \in \mathcal{H} \right\}. \quad (25)$$

Furthermore, for an arbitrary sequence of random variables $(W_t)_{t \in \mathbb{Z}}$, the sequence of ψ -mixing coefficients is defined as

$$\psi^W(n) = \sup_{j \in \mathbb{Z}} \psi(\mathcal{F}_{-\infty, j}^W, \mathcal{F}_{j+n, \infty}^W), \quad n \in \mathbb{N}.$$

We say that the process $(W_n)_{n \in \mathbb{Z}}$ is ψ -mixing if $\psi^W(n) \rightarrow 0$ as $n \rightarrow \infty$.

It is straightforward to show that every ψ -mixing process is also α -mixing, but the converse does not hold. Moreover, ψ -mixing is one of the strongest forms of strong mixing conditions, implying several others such as α -, β -, ϕ -, and ρ -mixing. Nevertheless, the class of ψ -mixing processes remains of considerable interest. For instance, in the case of strictly stationary, finite-state Markov chains, the notions of α - and ψ -mixing coincide (See Theorem 3.1 in [8]).

To formulate the next proposition, we need the notion of stochastic ordering. A real-valued random variable W_1 is said to be stochastically smaller than another random variable W_2 , denoted by $W_1 \preceq W_2$, if

$$\mathbb{P}(W_1 > x) \leq \mathbb{P}(W_2 > x), \quad x \in \mathbb{R},$$

or, equivalently,

$$\mathbb{E}[u(W_1)] \leq \mathbb{E}[u(W_2)],$$

for every monotonically increasing function $u : \mathbb{R} \rightarrow \mathbb{R}$.

Proposition 3.10. *Let the sequence of inter-arrival times $(Z_n)_{n \geq 1}$ be i.i.d. and independent of the service time sequence $(S_n)_{n \in \mathbb{N}}$, which is assumed to be ψ -mixing and takes values in $\mathcal{Y} = [0, M]$. Suppose furthermore that there exists a random variable S^* such that $\mathbb{E}[S^*] < \mathbb{E}[Z_1]$, moreover $S_n \preceq S^*$ for all $n \in \mathbb{N}$.*

Then there exists $\bar{t} > 0$ such that

$$\Lambda(\bar{t}) < 0,$$

where $\Lambda(t)$ is the rate function defined in Assumption 3.8.

Proof. Let $j \in \mathbb{N}$ and $n \geq 1$ be arbitrary but fixed, and define $W_k = e^{t(S_{k-1} - Z_k)}$, $k = 1, \dots, n$. Our goal is to estimate $\mathbb{E}[W_{j+1} \cdots W_{j+n}]$. For integers p, q satisfying $pq \leq n < p(q+1)$, which allows us to write

$$W_{j+1} \cdots W_{j+n} = \eta_1 \cdots \eta_p \Delta_n,$$

where $\eta_k = \prod_{l=0}^{q-1} W_{j+k+lp}$, $k = 1, \dots, p$, and $\Delta_n = W_{j+pq+1} \cdots W_{j+n}$.

Using $\Delta_n \leq e^{tM(p-1)}$ and Hölder-inequality, we have

$$\mathbb{E}[W_{j+1} \cdots W_{j+n}] \leq e^{tM(p-1)} \prod_{k=1}^p \mathbb{E}^{1/p}[\eta_k^p]. \quad (26)$$

For any fixed k , by independence, we have

$$\mathbb{E}[\eta_k^p] = \mathbb{E} \left[\prod_{l=0}^{q-1} e^{tp(S_{j+k-1+lp} - Z_{j+k+lp})} \right] = \mathbb{E} [e^{-tpZ_1}]^q \mathbb{E} \left[\prod_{l=0}^{q-1} e^{tpS_{j+k-1+lp}} \right].$$

Using layer cake representation, and that $S_n \preceq S^*$, for $n \in \mathbb{N}$, we can write

$$\begin{aligned} \mathbb{E} \left[\prod_{l=0}^{q-1} e^{tpS_{j+k-1+lp}} \right] &= \int_{[0, \infty)^q} \mathbb{P}(e^{tpS_{j+k-1+lp}} \geq \tau_l, l = 0, \dots, q-1) d\tau_0 \dots d\tau_{q-1} \\ &\leq (1 + \psi^S(p))^{q-1} \prod_{l=0}^{q-1} \int_{[0, \infty)} \mathbb{P}(e^{tpS_{j+k-1+lp}} \geq \tau_l) d\tau_l \\ &= (1 + \psi^S(p))^{q-1} \prod_{l=0}^{q-1} \mathbb{E} [e^{tpS_{j+k-1+lp}}] \leq (1 + \psi^S(p))^{q-1} \mathbb{E} [e^{tpS^*}]^q. \end{aligned}$$

To sum up, and taking the supremum in j , we obtain the following estimate:

$$\sup_{j \in \mathbb{N}} \mathbb{E}[W_{j+1} \cdots W_{j+n}] \leq e^{tM(p-1)} (1 + \psi^S(p))^{q-1} \mathbb{E}[e^{tp(S^* - Z_1)}]^q. \quad (27)$$

Since $\frac{d}{dh} \mathbb{E}[e^{h(S^* - Z_1)}] \big|_{t=0} = \mathbb{E}[S^* - Z_1] < 0$, for arbitrary m satisfying $\mathbb{E}[S^* - Z_1] < m < 0$ there exists $h_0 > 0$ such that

$$\mathbb{E} [e^{h(S^* - Z_1)}] < 1 + mh, \quad 0 < h < h_0.$$

Let $h'_0 \in (0, h_0)$ be fixed. By the ψ -mixing property of the sequence $(S_n)_{n \in \mathbb{N}}$, there exists $p_0 \geq 1$ such that

$$\zeta := (1 + \psi^S(p_0))(1 + mh'_0) < 1.$$

Using (27) with $t = \bar{t} := h'_0/p_0$ and $q_n = \lfloor \frac{n}{p_0} \rfloor$, $n \geq 1$, we obtain

$$\sup_{j \in \mathbb{N}} \mathbb{E}[W_{j+1} \cdots W_{j+n}] \leq e^{h'_0 M} \zeta^{q_n} \leq e^{h'_0 M} \zeta^{-1} \zeta^{n/p_0},$$

and thus

$$\sup_{j \in \mathbb{N}} \frac{1}{n} \log \mathbb{E} \left[\exp \left(\bar{t} \sum_{k=0}^n (S_{k+j} - Z_{k+j+1}) \right) \right] \leq \frac{n+1}{np_0} \log \zeta + \frac{1}{n} \log(e^{h'_0 M} \zeta^{-1})$$

which immediately implies that

$$\Lambda(\bar{t}) = \limsup_{n \rightarrow \infty} \sup_{j \in \mathbb{N}} \frac{1}{n} \log \mathbb{E} \left[\exp \left(\bar{t} \sum_{k=0}^n (S_{k+j} - Z_{k+j+1}) \right) \right] \leq \frac{1}{p_0} \log \zeta < 0.$$

□

It is conceivable that the requirement on $\Lambda(t)$ in Assumption 3.8 could be verified under mixing conditions on the sequence $(S_n)_{n \in \mathbb{N}}$ that are weaker than ψ -mixing. In particular, [10] addresses large deviation theorems under strong mixing. The authors showed that for stationary processes with a hyper-exponential mixing rate—specifically, $\alpha(n) \ll \exp(-n(\log n)^{1+\delta})$ for some $\delta > 0$ —the existence of a limit analogous to the one in the definition of the rate function Λ is ensured (see Theorem 1 in [10]). However, they also provided counterexamples based on Doeblin recurrent, irreducible Markov chains with countable state spaces to demonstrate that this decay condition on the mixing coefficient cannot be significantly weakened.

Lemma 3.11. *Under Assumption 3.7, the sequence of waiting times $(W_n)_{n \in \mathbb{N}}$, defined by the Lindley recursion (20), forms a Markov chain in the random environment $(S_n)_{n \in \mathbb{Z}}$ on the state space $\mathcal{X} = \mathbb{R}_+$ with the parametric kernel*

$$Q(s, w, A) := \mathbb{P}[(w + s - Z_1)_+ \in A], \quad s, w \in \mathbb{R}_+, \quad A \in \mathcal{B}(\mathbb{R}_+).$$

For the Lyapunov function $V(w) = e^{tw} - 1$, $w \in \mathbb{R}_+$, the drift condition holds for any choice of $t > 0$:

$$[Q(s)V](w) \leq \gamma(s)V(w) + K(s)$$

with $\gamma(s) = K(s) = \mathbb{E}[e^{t(s-Z_1)}]$, $s \geq 0$.

Proof. The sequence of waiting times $(W_n)_{n \in \mathbb{N}}$, governed by the Lindley recursion (20), constitutes a non-linear autoregressive process of the form (1). In this framework, $(S_n)_{n \in \mathbb{N}}$ represents the sequence of exogenous covariates, while $(Z_n)_{n \in \mathbb{N}}$ plays the role of the noise process $(\varepsilon_n)_{n \in \mathbb{N}}$. Furthermore, by Assumption 3.7, $(Z_n)_{n \in \mathbb{N}}$ is i.i.d. and independent of $(S_n)_{n \in \mathbb{N}}$, and thus the identity in the form (2) holds. Consequently, $(W_n)_{n \in \mathbb{N}}$ can be interpreted as a Markov chain in a random environment defined by $(S_n)_{n \in \mathbb{N}}$, with the corresponding parametric kernel given by

$$Q(s, w, A) = \mathbb{P}((s + w - Z_1) \in A)_+, \quad s, w \in \mathbb{R}_+, \quad A \in \mathcal{B}(\mathbb{R}_+).$$

Let $t > 0$ be arbitrary and define $V(w) = e^{tw} - 1$. Using the inequality $e^{(x)_+} - 1 \leq e^x$ for all $x \in \mathbb{R}$, we obtain the drift condition:

$$\begin{aligned} [Q(s)V](w) &= \mathbb{E}[V((w + s - Z_1)_+)] = \mathbb{E}[e^{t(w+s-Z_1)_+} - 1] \\ &\leq \mathbb{E}[e^{t(w+s-Z_1)}] = \mathbb{E}[e^{t(s-Z_1)}] (e^{tw} - 1) + \mathbb{E}[e^{t(s-Z_1)}] \\ &= \gamma(s)V(w) + K(s), \end{aligned}$$

where $\gamma(s) = K(s) = \mathbb{E}[e^{t(s-Z_1)}]$, $s \geq 0$ which completes the proof. □

The following lemma, which ensures the fulfillment of part B) of Assumption 2.2, essentially coincides with Lemma 4.6 in [35]. We include the proof here solely for the sake of completeness.

Lemma 3.12. Assume that $\mathbb{P}(Z_1 \geq M + \tau) > 0$ holds for some $\tau > 0$. Then, there exists $\bar{\beta} \in (0, 1)$ such that, for all $s \in [0, M]$ and $w \in [0, \tau]$,

$$Q(s, w, A) \geq (1 - \bar{\beta})\delta_0(A), \quad A \in \mathcal{B}(\mathbb{R}_+),$$

where δ_0 denotes the Dirac measure concentrated at 0.

Proof. Let $s \in [0, M]$, $w \in [0, \tau]$, and $A \in \mathcal{B}(\mathbb{R}_+)$. We can write

$$\begin{aligned} Q(s, w, A) &= \mathbb{P}([w + s - Z_1]_+ \in A) \geq \mathbb{P}([w + s - Z_1]_+ = 0) \delta_0(A) \\ &= (1 - \mathbb{P}(s + w - Z_1 > 0)) \delta_0(A) \\ &\geq (1 - \mathbb{P}(M + \tau - Z_1 > 0)) \delta_0(A), \end{aligned}$$

which shows that the desired inequality holds for any choice of $\bar{\beta}$ satisfying

$$\mathbb{P}(Z_1 < M + \tau) < \bar{\beta} < 1.$$

□

Let Assumptions 3.7 and 3.8 be in force, and let $\bar{t} > 0$ be as specified in Assumption 3.8. Then, by Lemma 3.11, for the Lyapunov function $V(w) = e^{\bar{t}w} - 1$, $w \geq 0$, the drift condition (7) holds with $\gamma(s) = K(s) = \mathbb{E}[e^{\bar{t}(s-Z_1)}]$, for $s \in \mathcal{Y} = [0, M]$. Furthermore, since $\gamma(s) = K(s) \leq e^{\bar{t}M}$, we have

$$\mathbb{E} \left[K(S_j) \prod_{k=1}^n \gamma(S_{k+j}) \right] \leq e^{\bar{t}M(n+1)},$$

for all $j \in \mathbb{N}$ and $n \geq 1$, hence the integrability condition in Assumption 2.2 is satisfied. Moreover, by Assumptions 3.7 and 3.8,

$$\bar{\gamma} = \limsup_{n \rightarrow \infty} \sup_{j \geq -1} \mathbb{E}^{1/n} \left[\exp \left(\bar{t} \sum_{k=0}^n (S_{k+j} - Z_{k+j+1}) \right) \right] = e^{\Lambda(\bar{t})} < 1,$$

which implies that part A) of Assumption 2.2 also holds.

Suppose that for some $0 < r < 1/\bar{\gamma}$ and $\tau := \frac{1}{\bar{t}} \log \left(1 + \frac{2}{r} \right)$,

$$\mathbb{P}(Z_1 \geq M + \tau) > 0.$$

We now verify part B) of Assumption 2.2. We have $R(s) = \frac{2K(s)}{r\gamma(s)} = 2/r$ since $\gamma(s) = K(s)$ for all $s \in [0, M]$. By Lemma 3.12, there exists $\bar{\beta} \in (0, 1)$ such that for all $s \in [0, M]$ and $w \in [0, \tau]$,

$$Q(s, w, A) \geq (1 - \bar{\beta})\delta_0(A), \quad A \in \mathcal{B}(\mathbb{R}_+),$$

where δ_0 denotes the Dirac measure concentrated at 0. Note that $w \in [0, \tau]$ if and only if $V(w) \in [0, 2/r] = [0, R(s)]$, for $s \in [0, M]$. Therefore, the minorization condition (8) holds with $\kappa_R(s, \cdot) = \delta_0(\cdot)$ for $s \in [0, M]$. Furthermore, the minorization coefficient $\beta(R(s), s) = \bar{\beta} \in (0, 1)$ is independent of s , hence part B) of Assumption 2.2 is also satisfied.

Theorem 3.13. Assume that the sequences $(S_n)_{n \in \mathbb{N}}$ and $(Z_n)_{n \geq 1}$ satisfy Assumptions 3.7 and 3.8, moreover for some $0 < r < 1/\bar{\gamma}$,

$$\mathbb{P} \left(Z_1 \geq M + \frac{1}{\bar{t}} \log \left(1 + \frac{2}{r} \right) \right) > 0,$$

where $\bar{t}$ is as in Assumption 3.8, and $\bar{\gamma} = \exp(\Lambda(\bar{t}))$.

Then for the sequence of waiting times $(W_n)_{n \in \mathbb{N}}$, defined by the Lindley recursion (20), the following statements hold.

A) If the sequence of service times $(S_n)_{n \in \mathbb{N}}$ is strongly mixing, then the L^1 weak law of large numbers holds:

$$\frac{1}{n} \sum_{k=1}^n [W_k - \mathbb{E}(W_k)] \xrightarrow{L^1} 0, \quad \text{as } n \rightarrow \infty.$$

B) If there exists $c > 0$ and $\kappa > 1$ such that $\alpha^S(n) \leq cn^{-\kappa}$, $n \geq 1$, then

$$\frac{1}{n} \sum_{k=1}^n [W_k - \mathbb{E}(W_k)] \xrightarrow{a.s.} 0, \quad \text{as } n \rightarrow \infty.$$

C) If the sequence of mixing coefficients $(\alpha^S(n))_{n \in \mathbb{N}}$ satisfies the condition

$$\sum_{n \in \mathbb{N}} (n+1)^2 \alpha^S(n)^\delta < \infty$$

for some exponent $\delta \in (0, 1)$, then there exists a constant $\sigma > 0$ such that

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{\sqrt{n}} \left| \sum_{k=1}^n (W_k - \mathbb{E}(W_k)) \right| \geq a \right) \leq \int_{\mathbb{R}} \mathbb{1}_{[-a, a]^c}(\sigma t) \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \quad a > 0.$$

Proof. From Lemma 3.11, it follows that the sequence of waiting times $(W_n)_{n \in \mathbb{N}}$ is a Markov chain in a random environment $(S_n)_{n \in \mathbb{N}}$, which satisfies the drift condition (7) with $V(w) = e^{tw} - 1$ and $\gamma(s) = K(s) = \mathbb{E}[e^{t(s-Z_0)}]$.

Given that $S_n \in [0, M]$ for all $n \in \mathbb{N}$, it holds that $\mathbb{E}[K(S_j) \prod_{k=1}^n \gamma(S_{k+j})] \leq e^{tM(n+1)}$, which implies that the integrability condition in Assumption 2.2 is trivially satisfied, i.e.,

$$\mathbb{E} \left[K(S_j) \prod_{k=1}^n \gamma(S_{k+j}) \right] < \infty \quad \text{for all } j \in \mathbb{N} \text{ and } n \geq 1.$$

Moreover, by Assumption 3.8, the parameter $t > 0$ can be chosen such that

$$\bar{\gamma} = \limsup_{n \rightarrow \infty} \sup_{j \in \mathbb{N}} \mathbb{E}^{1/n} \left[\prod_{k=0}^n \gamma(S_{k+j}) \right] < 1,$$

which ensures that part A) of Assumption 2.2 is satisfied.

Notice that, since $\gamma = K$, hence by Lemma 3.12, the parametric kernel Q satisfies the minorization condition (8) with

$$R(s) = \frac{2K(s)}{r\gamma(s)} = \frac{2}{r}, \quad \beta(R(s), s) = \bar{\beta} < 1, \quad \text{and } \kappa_{R(s)}(s, A) = \delta_0(A).$$

We can conclude that part B) of Assumption 2.2 also holds.

In Theorem 2.5, the inequality (11) is satisfied with $\Phi = \text{id}_{[0, \infty)}$ for any $p > 1$, while the deterministic initial condition $W_0 = 0$ obviously ensures $\mathbb{E}V(W_0) < \infty$.

We have thus shown that all conditions of Theorem 2.5 are met, from which the assertions of the present theorem follow. $\square$

In what follows, we will demonstrate that the seemingly complex and technical conditions of Lemma 2.10 and Theorem 2.14 boil down to easily verifiable and natural conditions concerning the sequence $(S_n)_{n \in \mathbb{N}}$ and the inter-arrival time distribution Z_1 . Utilizing this framework, we establish the functional central limit theorem for the sequence of waiting times $(W_n)_{n \in \mathbb{N}}$.

Assume that the law of Z_1 is absolutely continuous with respect to the Lebesgue measure, i.e., $\text{Law}(Z_1)(dz) = f_{Z_1}(z) dz$, where $f_{Z_1}(z) = 0$ for $z \leq 0$. For any Borel set $A \in \mathcal{B}(\mathbb{R}_+)$, we can express $Q(s, w, A)$ as follows:

$$\begin{aligned} Q(s, w, A) &= \int_{[0, \infty)} \mathbb{1}_A((w + s - z)_+) f_{Z_1}(z) dz \\ &= \mathbb{1}_A(0) \mathbb{P}(Z_1 > w + s) + \int_0^{w+s} \mathbb{1}_A(w + s - z) f_{Z_1}(z) dz \\ &= \mathbb{1}_A(0) \mathbb{P}(Z_1 > w + s) + \int_{[0, \infty)} \mathbb{1}_A(z) f_{Z_1}(w + s - z) dz \\ &= \int_{\mathcal{X}=\mathbb{R}_+} \mathbb{1}_A(z) (\mathbb{P}(Z_1 > w + s) \mathbb{1}_{\{0\}}(z) + f_{Z_1}(w + s - z) \mathbb{1}_{(0, \infty)}(z)) d\nu(z), \end{aligned}$$

where $\nu(dz) = \delta_0(dz) + dz$.

To sum up, we obtained that there exists a Borel measure ν on $\mathcal{B}(\mathcal{X})$, such that $Q(s, w, \cdot) \ll \nu$ for all $(s, w) \in \mathcal{Y} \times \mathcal{X}$, where $\mathcal{X} = \mathbb{R}_+$ and $\mathcal{Y} = [0, M]$. Moreover, the transition densities are given by

$$p_s(z | w) = \frac{dQ(s, w, \cdot)}{d\nu}(z) = \mathbb{P}(Z_1 > w + s) \mathbb{1}_{\{0\}}(z) + f_{Z_1}(w + s - z) \mathbb{1}_{(0, \infty)}(z). \quad (28)$$

For a fixed $n \geq 1$, let $\underline{w} = [w_1, \dots, w_n]$ and $\underline{s} = [s_0, \dots, s_{n-1}]$, and define the likelihood function as

$$p(\underline{w} | \underline{s}) = \prod_{k=1}^n p_{s_{k-1}}(w_k | w_{k-1}). \quad (29)$$

If the density function $f_{Z_1} : \mathbb{R} \rightarrow [0, \infty)$ is continuously differentiable everywhere, then at every point $\underline{w} \in [0, \infty)^n$ where $p(\underline{w} | \underline{s}) > 0$, the derivative $\partial_{s_i} \log p(\underline{w} | \underline{s}) = \partial_{s_i} p_{s_i}(w_{i+1} | w_i)$ for $i = 0, \dots, n-1$ exists and is finite. Furthermore, we have

$$\frac{\partial}{\partial s_i} \int_{\mathcal{X}^n} (w_1 + \dots + w_n) p(\underline{w} | \underline{s}) d\underline{w} = \int_{\mathcal{X}^n} (w_1 + \dots + w_n) \partial_{s_i} p(\underline{w} | \underline{s}) d\underline{w}, \quad (30)$$

hence the regularity conditions required for the Cramér-Rao inequality are satisfied.

Lemma 3.14. *Assume that $f_{Z_1} : [0, \infty) \rightarrow [0, \infty)$ is continuously differentiable everywhere, and*

$$\int_{[0, \infty)} \frac{f'_{Z_1}(z)^2}{f_{Z_1}(z)} dz < \infty.$$

Then there exists a constant $r^ > 0$ such that for any $n \geq 1$, and $\Sigma_n = \sum_{k=1}^n W_k$, the following inequality holds:*

$$\text{Var}(\Sigma_n) \geq \frac{1}{r^*} \sum_{k=0}^{n-1} \mathbb{P}(S_k > Z_1)^2.$$

Proof. Fix $n \geq 1$ and let $\underline{s} = [s_0, \dots, s_{n-1}] \in [0, M]^n$. Since $(S_n)_{n \in \mathbb{N}}$ is a scalar-valued process, the Fisher information matrix is diagonal. Therefore, using (28), for $k = 0, \dots, n-1$, we have

$$\begin{aligned} r(I(s_k)) &= [I(\underline{s})]_{kk} = \mathbb{E} \left[(\partial_{s_k} \log p_{s_k}(W_{k+1} | W_k))^2 \middle| \underline{S} = \underline{s} \right] \\ &= \mathbb{E} \left[\int_{\mathbb{R}} \frac{f_{Z_1}(W_k + s_k)^2}{\mathbb{P}(Z_1 > W_k + s_k)} \mathbb{1}_{\{0\}}(z) + \frac{f'_{Z_1}(W_k + s_k - z)^2}{f_{Z_1}(W_k + s_k - z)} \mathbb{1}_{(0, \infty)}(z) (\delta_0(dz) + dz) \middle| \underline{S} = \underline{s} \right] \\ &= \mathbb{E} \left[\frac{f_{Z_1}(W_k + s_k)^2}{\mathbb{P}(Z_1 > W_k + s_k)} + \int_{[0, \infty)} \frac{f'_{Z_1}(W_k + s_k - z)^2}{f_{Z_1}(W_k + s_k - z)} dz \middle| \underline{S} = \underline{s} \right] \\ &\leq \sup_{z \in [0, \infty)} \frac{f_{Z_1}(z)^2}{\mathbb{P}(Z_1 > z)} + \int_{[0, \infty)} \frac{f'_{Z_1}(z)^2}{f_{Z_1}(z)} dz. \end{aligned}$$

By l'Hôpital's rule, we have $\lim_{z \rightarrow \infty} \frac{f_{Z_1}(z)^2}{\mathbb{P}(Z_1 > z)} = \lim_{z \rightarrow \infty} -f'_{Z_1}(z) = 0$, which implies that

$$\sup_{z \in [0, \infty)} \frac{f_{Z_1}(z)^2}{\mathbb{P}(Z_1 > z)} < \infty.$$

Therefore, there exists an upper bound for $r(I(s_k))$ that is independent of both k and n ; let this bound be denoted by r^* .

Since the mapping $(s, w, z) \mapsto (s + w - z)_+$, defining the Lindley recursion (20), is monotonic in both s and w , we observe that for each $1 \leq j \leq n$ and $0 \leq k \leq n-1$, the derivative $\partial_{s_k} \mathbb{E}[W_j | \underline{S}] \geq 0$. Therefore,

$$\begin{aligned} \|\partial_{s_k} \mathbb{E}[\Sigma_n | \underline{S}]\|^2 &= \left(\sum_{j=1}^n \partial_{s_k} \mathbb{E}[W_j | \underline{S}] \right)^2 \geq (\partial_{s_k} \mathbb{E}[W_{k+1} | \underline{S}])^2 = \mathbb{P}(W_k + S_k - Z_{k+1} > 0 | \underline{S})^2 \\ &\geq \mathbb{P}(S_k - Z_1 > 0 | \underline{S})^2. \end{aligned}$$

By Lemma 2.10 and the Cauchy-Schwartz inequality, using the bounds obtained above, we get

$$\begin{aligned}\text{Var}(\Sigma_n) &\geq \sum_{k=0}^{n-1} \mathbb{E} \left[\frac{1}{r(I_k)} \|\partial_{s_k} \mathbb{E}[\Sigma_n | \underline{S}]\|^2 \right] \\ &\geq \sum_{k=0}^{n-1} \frac{1}{r^*} \mathbb{E} [\mathbb{P}(S_k - Z_1 > 0 | \underline{S})^2] \geq \frac{1}{r^*} \sum_{k=0}^{n-1} \mathbb{P}(S_k > Z_1)^2,\end{aligned}$$

which completes the proof. $\square$

Theorem 3.15. *Beyond the conditions established in Theorem 3.13, suppose that the function $f_{Z_1} : [0, \infty) \rightarrow [0, \infty)$ is continuously differentiable everywhere, and that*

$$\int_{[0, \infty)} \frac{f'_{Z_1}(z)^2}{f_{Z_1}(z)} dz < \infty.$$

Additionally, assume that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{P}(S_k > Z_1)^2 > 0, \quad (31)$$

and that for some exponent $\delta > 0$, the series of mixing coefficients $(\alpha^S(n))_{n \in \mathbb{N}}$ satisfies

$$\sum_{n \geq 1} n^\delta \alpha^S(n) < \infty.$$

Let $\Sigma_n = W_1 + \dots + W_n$, $v_n(t) = \min\{1 \leq k \leq n \mid \text{Var}(\Sigma_k) \geq t \text{Var}(\Sigma_n)\}$, and for $1 \leq k \leq n$, define

$$\xi_{k,n} = \frac{W_k - \mathbb{E}[W_k]}{\text{Var}(\Sigma_n)^{1/2}}.$$

Then the sequence of functions $B_n(t) = \sum_{k=1}^{v_n(t)} \xi_{k,n}$, for $t \in (0, 1]$ and $n \geq 1$, converges in distribution in $D([0, 1])$ (equipped with the uniform topology) to a standard Brownian motion B .

Proof. In the proof of Theorem 3.13, we showed that under these conditions, the sequence $(W_n)_{n \in \mathbb{N}}$ forms a Markov chain in a random environment, which satisfies the conditions of Theorem 2.5. Additionally, Lemma 3.14, in conjunction with the condition (31), guarantees that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{E} \left[\frac{1}{r(I_k)} \|\partial_{s_k} \mathbb{E}[S_n | \underline{S}]\|^2 \right] > 0.$$

With the Lyapunov function $V(w) = e^{\bar{t}w-1}$, the inequality (11) that appears in the conditions of Theorem 2.5 holds for $\Phi = \text{id}_{\mathbb{R}_+}$ and any $p > 1$. Consequently, the sequence $(\alpha^S(n))_{n \in \mathbb{N}}$ satisfies the condition regarding the mixing coefficients in Theorem 2.14.

Overall, we conclude that the conditions of Theorem 2.14 are met, which leads to the desired result. $\square$

The conditions in Assumptions 3.7 and 3.8 can most likely be significantly weakened. We conjecture that the boundedness condition on the service times $(S_n)_{n \in \mathbb{N}}$ and the assumption on the tail probabilities of Z_1 are of a purely technical nature and may be removed in future work. Theorem 1.8 in [34], for instance, addresses the case when both S and Z are unbounded, $(S_n)_{n \in \mathbb{N}}$ is i.i.d., and so is $(Z_n)_{n \geq 1}$. In that setting, the main challenge is that the condition on the minorization coefficient in part B) of Assumption 2.2 is no longer satisfied. However, the results of our paper [35] can still be applied under appropriate assumptions on the environment. One such condition, as formulated in Theorem 1.8 of [34], is that Z_1 has a Gumbel-like tail:

$$\mathbb{P}(Z_1 \geq z) \leq C_1 \exp(-C_2 e^{C_3 z}), \quad \text{for some } C_1, C_2, C_3 > 0.$$

Answering such questions, as well as investigating the case when the service times $(S_n)_{n \in \mathbb{N}}$ are i.i.d. and the inter-arrival times $(Z_n)_{n \geq 1}$ are weakly dependent, will be the subject of our future research.

Acknowledgment

I extend my sincere gratitude to Miklós Rásonyi for his invaluable assistance and insightful discussions on the subject matter, which greatly contributed to the development of this paper. I would also like to thank the two anonymous reviewers for their careful reading and constructive comments, which significantly improved the quality of the manuscript.

Declaration of generative AI and AI-assisted technologies in the writing process

The author(s) would like to declare that during the preparation of this work, ChatGPT 3.5 was utilized to enhance the language and readability of the paper. As a non-native English speaker, the author(s) sought to improve the quality of the English language presentation. It is important to emphasize that the ideas and content of this article represent the author's original intellectual contributions. After utilizing this tool, the author(s) thoroughly reviewed and edited the content as necessary and take(s) full responsibility for the final publication.

References

- [1] A. AGOSTO, G. CAVALIERE, D. KRISTENSEN, AND A. RAHBK, *Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX)*, Journal of Empirical Finance, 38 (2016), pp. 640–663. Recent developments in financial econometrics and empirical finance.
- [2] J. F. BERCHER, *On generalized Cramér-Rao inequalities, and an extension of the Shannon-Fisher-Gauss setting*, New Perspectives on Stochastic Modeling and Data Analysis, (2014), pp. 19–35.
- [3] A. A. BOROVKOV, *Mathematical Statistics*, CRC Press, 1999.
- [4] A. A. BOROVKOV, *Stochastic Processes in Queuing Theory*, Nauka, 1972. In Russian.
- [5] A. A. BOROVKOV, *Ergodicity and Stability of Stochastic Processes*, John Wiley & Sons, 1998.
- [6] A. A. BOROVKOV AND S. G. FOSS, *Stochastically recursive sequences and their generalizations*, Limit theorems for random processes and their applications (Russian), 20 (1993), pp. 32–103.
- [7] L. BOTTOU, F. E. CURTIS, AND J. NOCEDAL, *Optimization methods for large-scale machine learning*, SIAM Review, 60 (2018), pp. 223–311.
- [8] R. C. BRADLEY, *Basic properties of strong mixing conditions. a survey and some open questions*, Probability Surveys, 2 (2005), pp. 107–144.
- [9] R. C. BRADLEY, *Central limit theorems under weak dependence*, Journal of Multivariate Analysis, 11 (1981), pp. 1–16.
- [10] W. BRYC AND A. DEMBO, *Large deviations and strong mixing*, Annales de l’IHP Probabilités et statistiques, 32(4) 1996, pp. 549–569.
- [11] R. COGBURN, *The ergodic theory of Markov chains in random environments*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 66 (1984), pp. 109–128.
- [12] ———, *On direct convergence and periodicity for transition probabilities of Markov chains in random environments*, The Annals of Probability, (1990), pp. 642–654.
- [13] Z. M. DEBALY AND L. TRUQUET, *Iterations of dependent random maps and exogeneity in nonlinear dynamics*, Econometric Theory, 37 (2021), pp. 1135–1172.
- [14] P. DIACONIS AND D. FREEDMAN, *Iterated random functions*, SIAM Review, (1999), pp. 45–76.

- [15] P. DOUKHAN, *Mixing. In: Mixing. Lecture Notes in Statistics*, Springer, New York, NY, vol 85, 1994.
- [16] M. EKSTRÖM, *A general central limit theorem for strong mixing sequences*, Statistics & Probability Letters, 94 (2014), pp. 236–238.
- [17] A. K. ERLANG, *The theory of probabilities and telephone conversations*, Nyt Tidsskrift for Matematik B, 20 (1909), pp. 33–39.
- [18] T. ESPINASSE AND P. ROCHET, *Cramér-Rao inequality for non-differentiable models*, Comptes Rendus Mathématique, 350 (2012), 711–715.
- [19] D. FELSMANN, *Stability of general state space Markov chains*, Master’s thesis, Eötvös Loránd University, 2022. (in Hungarian).
- [20] S. FOSS AND T. KONSTANTOPOULOS, *Extended renovation theory and limit theorems for stochastic ordered graphs*, Markov Process. Related Fields, 9 (2003), pp. 413–468.
- [21] B. GERENCSÉR AND M. RÁSONYI, *Invariant measures for multidimensional fractional stochastic volatility models*, Stochastics and Partial Differential Equations: Analysis and Computations, (2022), pp. 1–33.
- [22] B. GERENCSÉR AND M. RÁSONYI, *On the ergodicity of certain Markov chains in random environments*, Journal of Theoretical Probability, (2023), pp. 1–33.
- [23] P. GORGI AND S. J. KOOPMAN, *Beta observation-driven models with exogenous regressors: a joint analysis of realized correlation and leverage effects*, Tinbergen Institute Discussion Papers 20-004/III, Tinbergen Institute, Jan. 2020.
- [24] L. GYÖRFI, A. LOVAS, AND M. RÁSONYI, *On the strong stability of ergodic iterations*, arXiv preprint, (2023), p. 2304.04657.
- [25] L. GYÖRFI AND G. MORVAI, *Queueing for ergodic arrivals and services*, in Limit theorems in probability and statistics, I. Berkes, et al., ed., vol. 2, 2002, pp. 127–141. Fourth Hungarian colloquium on limit theorems in probability and statistics, Balatonlelle, Hungary, 1999.
- [26] Y. HAFOUTA, *On the functional CLT for slowly mixing triangular arrays*, arXiv preprint arXiv:2111.05807, (2021).
- [27] P. HALL AND C. C. HEYDE, *Martingale limit theory and its application*, Academic press, 2014.
- [28] B. HANSEN, *A weak law of large numbers under weak mixing*, <https://users.ssc.wisc.edu/~bhansen/papers/wlln.pdf> (2019).
- [29] N. HERRNDORF, *A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables*, The Annals of Probability, 12 (1984), pp. 141 – 153.
- [30] M. IOSIFESCU, *Iterated function sytems. a critical survey*, Mathematical Reports, (2009), pp. 181–229.
- [31] Y. KIFER, *Perron-Frobenius theorem, large deviations, and random perturbations in random environments*, Math. Zeitschrift, 222 (1996), pp. 677–698.
- [32] Y. KIFER, *Limit theorems for random transformations and processes in random environments*, Transactions of the American Mathematical Society, 350 (1998), pp. 1481–1518.
- [33] T. LINDVALL, *Lectures on the coupling method*, Courier Corporation, 2002.
- [34] A. LOVAS AND M. RÁSONYI, *Ergodic theorems for queueing systems with dependent inter-arrival times*, Operations Research Letters, 49 (2021), pp. 682–687.

- [35] A. LOVAS AND M. RÁSONYI, *Markov chains in random environment with applications in queuing theory and machine learning*, Stochastic Processes and their Applications, 137 (2021), pp. 294–326.
- [36] A. LOVAS AND M. RÁSONYI, *Functional central limit theorem and strong law of large numbers for stochastic gradient Langevin dynamics*, Appl Math Optim, (2023)), p. 78.
- [37] R. M. LOYNES, *The stability of a queue with non-independent inter-arrival and service times*, Mathematical Proceedings of the Cambridge Philosophical Society, 58 (1962), pp. 497–520.
- [38] D. L. MCLEISH, *A maximal inequality and dependent strong laws*, The Annals of probability, (1975), 3(5), pp. 829–839.
- [39] F. MERLEVEDE AND M. PELIGRAD, *Functional clt for nonstationary strongly mixing processes*, Statistics & Probability Letters, 156 (2020), p. 108581.
- [40] F. MERLEVÈDE AND M. PELIGRAD, AND E. RIO, *A Bernstein type inequality and moderate deviations for weakly dependent sequences*, Probab. Theory Relat. Fields 151 (2011), pp. 435–474.
- [41] F. MERLEVÈDE, M. PELIGRAD, AND S. UTEV, *Functional clt for martingale-like nonstationary dependent structures*, Bernoulli, 25 (2019), pp. 3203–3233.
- [42] S. P. MEYN AND R. L. TWEEDIE, *Markov chains and stochastic stability*, Springer-Verlag, 1993.
- [43] D. S. NGUYEN, *Application of queuing theory in service design*, in 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2017, pp. 837–840.
- [44] S. OREY, *Markov chains with stochastically stationary transition probabilities*, The Annals of Probability, 19 (1991), pp. 907–928.
- [45] E. RIO, *The functional law of the iterated logarithm for stationary strongly mixing sequences*, The Annals of Probability, 23(3) (1995), pp. 1188–1203
- [46] M. ROSENBLATT, *A central limit theorem and a strong mixing condition*, Proceedings of the national Academy of Sciences, 42 (1956), pp. 43–47.
- [47] T. SEPPALAINEN, *Large deviations for Markov chains with random transitions*, The Annals of Probability, 22 (1994), pp. 713–748.
- [48] O. STENFLO, *Markov chains in random environments and random iterated function systems*, Trans. American Math. Soc., 353 (2001), pp. 3547–3562.
- [49] L. TRUQUET, *Ergodic properties of some Markov chains models in random environments*, 2021.
- [50] L. TRUQUET, *Strong mixing properties of discrete-valued time series with exogenous covariates*, Stochastic Processes and their Applications, 160 (2023), pp. 294–317.
- [51] . R. E. WELSCH, *Limit laws for extreme order statistics from strong-mixing processes*, The Annals of Mathematical Statistics, 43(2) (1972), pp. 439–446.
- [52] H. WHITE, *Asymptotic Theory for Econometricians*, Emerald Group Publishing Limited, 2000.

A Brief survey of key results on α -mixing sequences

In this section, we present a selection of theorems from the literature on α -mixing sequences. These results encompass fundamental topics such as the law of large numbers, and the central limit theorem. However, we do not address all relevant topics—for instance, the distribution of extreme values in α -mixing sequences or concentration inequalities. For a detailed treatment of these subjects, we refer the interested reader to [51] and [40].

By integrating the theorems presented here with the transition of mixing results from Sections 1 and 2, we establish a robust theoretical framework. This framework enables the statistical analysis of non-linear autoregressive processes with exogenous covariates and Markov chains in random environments. Given the vast literature on α -mixing processes, a comprehensive review is beyond the scope of this paper. Instead, we focus on key results that are particularly useful for the statistical analysis of weakly dependent sequences. For readers seeking a more in-depth overview, we recommend Doukhan [15], which, with its detailed references and literature survey, provides an excellent starting point.

In theorems concerning α -mixing sequences, the key condition is typically about how rapidly the mixing coefficient sequence $(\alpha^W(n))_{n \in \mathbb{N}}$ decays to zero. In the econometric literature, the term *size* is frequently used to characterize this behavior (cf. Definition 3.45 in [52]). However, since the definition is not uniform, despite its ability to make theorems more concise and elegant, we avoid using *size* to prevent misunderstandings.

In the context of stationary processes, it is well-known that strong mixing implies ergodicity ensuring the applicability of the strong law of large numbers. For non-stationary, heterogeneously distributed α -mixing sequences, McLeish [38] established the following version of the strong law of large numbers.

Theorem A.1 (McLeish, 1975). *Consider a sequence of $\mathbb{R}$ -valued random variables $(W_n)_{n \in \mathbb{N}}$ with $\mathbb{E}W_n = 0$, $n \in \mathbb{N}$ and with α -mixing coefficients satisfying $\alpha^W(n) \leq cn^{-\frac{r}{r-2}}$, $n \in \mathbb{N}$, for $c > 0$ and $r > 2$. Suppose that for some p such that $r/2 < p \leq r < \infty$,*

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}^{2/r}|W_n|^p}{n^{2p/r}} < \infty.$$

Under these conditions,

$$\frac{1}{n} \sum_{k=1}^n W_n \xrightarrow{a.s.} 0, \quad n \rightarrow \infty.$$

Remark A.2. *The condition related to the decay of the moments in Theorem A.1*

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}^{2/r}|W_n|^p}{n^{2p/r}} < \infty$$

is automatically satisfied if $\sup_{n \in \mathbb{N}} \mathbb{E}|W_n|^p < \infty$.

In the context of ergodic theory, we distinguish between ergodic, weakly mixing, and strongly mixing processes. We classify a collection of random variables of the form $\{\xi_{n,i} \mid 1 \leq i \leq n\}$ as weakly mixing if

$$\frac{1}{n} \sum_{k=1}^n \alpha^{\xi_{n,\cdot}}(k) \rightarrow 0, \quad n \rightarrow \infty,$$

where $\alpha^{\xi_{n,\cdot}}(\cdot)$ denotes the α -mixing coefficient corresponding to the n th row of the array.

Hansen established the weak law of large numbers, more precisely L^1 -law of large numbers, for heterogenous weak mixing processes and arrays [28]. A notable and valuable characteristic of this result is the absence of a stationarity assumption, thereby broadening the applicability of the result. The elegance of Hansen's proof lies in its simplicity, leveraging the standard representation of the variance of the truncated mean as the weighted Cesàro sum of covariances, and bounds the latter using the mixing inequality for bounded random variables.

As noted by Hansen [28], strong mixing implies weak mixing, and weak mixing, in turn, implies ergodicity, moreover this nesting is strict. However, for any process $(W_n)_{n \in \mathbb{N}}$, the sequence

of mixing coefficients $(\alpha^W(n))_{n \in \mathbb{N}}$ is monotonically decreasing. Consequently, their Cesàro sum converges to zero if and only if $\alpha^W(n) \rightarrow 0$ as $n \rightarrow \infty$. As a result, the classes of weakly mixing and strongly mixing random sequences coincide. Thus, for weakly dependent sequences, Hansen's theorem can be stated as follows:

Theorem A.3 (Hansen, 2019). *Consider a strongly mixing $\mathbb{R}$ -valued process $(W_n)_{n \in \mathbb{N}}$, and define the sequence of partial sums $S_n := \sum_{k=1}^n W_k$, $n \geq 1$. Additionally, suppose that the condition*

$$\lim_{B \rightarrow \infty} \sup_{n \geq 1} \frac{1}{n} \sum_{k=1}^n \mathbb{E}(|W_k| \mathbf{1}(|W_k| \geq B)) = 0 \quad (32)$$

holds. Then, we have

$$\frac{S_n}{n} - \frac{\mathbb{E}(S_n)}{n} \xrightarrow{L^1} 0, \quad n \rightarrow \infty.$$

Proof. Hansen stated and proved this theorem, in a bit more general setting, for triangular arrays of weakly mixing variables (See Theorem 1 on page 4 in [28]). For making the explanation self-contained, we present the proof of this simpler version here.

Without the loss of generality we can assume that $\mathbb{E}(W_n) = 0$, $n \in \mathbb{N}$. Let $\varepsilon > 0$ be arbitrary and choose $B > 0$ such that

$$\sup_n \frac{1}{n} \sum_{k=1}^n \mathbb{E}(|W_k| \mathbf{1}(|W_k| > B)) < \varepsilon. \quad (33)$$

Let us introduce

$$\begin{aligned} W'_n &= W_n \mathbf{1}(|W_n| \leq B) - \mathbb{E}(W_n \mathbf{1}(|W_n| \leq B)) \\ W''_n &= W_n \mathbf{1}(|W_n| > B) - \mathbb{E}(W_n \mathbf{1}(|W_n| > B)). \end{aligned}$$

Obviously, $W_n = W'_n + W''_n$, $n \in \mathbb{N}$ hence by the triangle inequality and (33), we have

$$\frac{1}{n} \mathbb{E} \left| \sum_{k=1}^n W_k \right| \leq \frac{1}{n} \mathbb{E} \left| \sum_{k=1}^n W'_k \right| + \frac{2}{n} \sum_{k=1}^n \mathbb{E}(|W_k| \mathbf{1}(|W_k| > B)) < \frac{1}{n} \mathbb{E} \left| \sum_{k=1}^n W'_k \right| + 2\varepsilon. \quad (34)$$

Furthermore, W'_n satisfies the bound $|W'_n| \leq 2B$, and for its mixing coefficient $\alpha^{W'}(n) \leq \alpha^W(n)$, $n \in \mathbb{N}$ holds, consequently by the mixing inequality for bounded variables (cf. Theorem A.5 in [27] or the proof of Lemma A.1. in [36]),

$$|\mathbb{E}(W'_k W'_l)| = |\text{Cov}(W'_k, W'_l)| \leq 16B^2 \alpha^W(|k-l|).$$

By Jensen's inequality, we can estimate

$$\mathbb{E}^2 \left| \sum_{k=1}^n W'_k \right| \leq \sum_{k,l=1}^n \mathbb{E}(W'_k W'_l) \leq 16B^2 \sum_{k,l=1}^n \alpha^W(|k-l|) = 16B^2 n \left(\alpha^W(0) + 2 \sum_{k=1}^n \alpha^W(k) \right).$$

Substituting this into (34) yields

$$\frac{1}{n} \mathbb{E} \left| \sum_{k=1}^n W_k \right| < 4B \left(\frac{\alpha^W(0)}{n} + \frac{2}{n} \sum_{k=1}^n \alpha^W(k) \right)^{1/2} + 2\varepsilon,$$

where the upper bound tends to 2ε as $n \rightarrow \infty$ since $(W_n)_{n \in \mathbb{N}}$ is strongly mixing, and thus $\limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left| \sum_{k=1}^n W_k \right| < 2\varepsilon$ holds for arbitrary $\varepsilon > 0$ which completes the proof. $\square$

Remark A.4. *The average uniform integrability condition (32) is automatically satisfied if the process $(W_n)_{n \in \mathbb{N}}$ has a uniformly bounded moment $\sup_{n \in \mathbb{N}} \mathbb{E}(|W_n|^r) < \infty$ for some $r > 1$.*

Our objective is to establish the (functional) central limit theorem for certain functionals of the sequence of iterates $(X_n)_{n \in \mathbb{N}}$ when $(Y_n)_{n \in \mathbb{N}}$ is merely α -mixing and stationarity is not assumed. While we studied this problem in the context of stochastic gradient Langevin dynamics [36], our focus was limited to stationary data streams. Our approach relied significantly on Corollary 2 in [29], which offers broad applicability, extending even to non-stationary processes. However, the condition $\lim_{n \rightarrow \infty} n^{-1} \mathbb{E}(S_n^2) = \sigma^2$ required by this corollary is not generally met in cases when the exogenous regressor $(Y_n)_{n \in \mathbb{N}}$ is non-stationary.

Very recently there was major progress on the functional CLT for non-stationary mixing sequences. In [41] Merlevède, Peligrad and Utev answered the question raised by Ibragimov concerning the CLT for triangular arrays of non-stationary weakly dependent variables under the Lindeberg condition (cf. page 1 in [26]). Subsequently, Merlevède and Peligrad proved the functional CLT for triangular arrays satisfying a dependence condition weaker than the standard strong mixing condition, termed the weak strong mixing condition [39]. Both of these results require the condition:

$$\sum_{k=1}^n \text{Var}(W_k) = O(\text{Var}(S_n)) \quad (35)$$

which is difficult to verify in general. A key contribution of our paper is the derivation of a Cramér-Rao lower bound for the variance of partial sums in (35), facilitating a functional CLT when $(X_n)_{n \in \mathbb{N}}$ forms a Markov chain in a random environment (See Section 2).

Essentially, aside from certain moment conditions, Theorem 1 in Ekström's paper [16] mandates only the verification that the α -mixing coefficients exhibit a sufficiently rapid decrease:

Theorem A.5 (Ekström, 2014). *Let $\{\xi_{n,i} \mid 1 \leq i \leq d_n, n \in \mathbb{N}\}$ be an array of $\mathbb{R}$ -valued random variables with $\mathbb{E}\xi_{n,i} = 0$, $1 \leq i \leq d_n$, $n \in \mathbb{N}$, and define $S_n = \sum_{k=1}^{d_n} \xi_{n,k}$, $n \in \mathbb{N}$. Assume that for some $r > 0$,*

$$i. \sup_{n \in \mathbb{N}} \max_{1 \leq i \leq d_n} \mathbb{E}|\xi_{n,i}|^{2+r} < \infty, \text{ and}$$

$$ii. \sup_{n \in \mathbb{N}} \sum_{k=0}^{\infty} (k+1)^2 (\alpha^{\xi_{n,\cdot}}(k))^{\frac{r}{4+r}} < \infty,$$

where $\alpha^{\xi_{n,\cdot}}(\cdot)$ denotes the strong mixing coefficient corresponding to the n th row in the array.

Then the distributions $\text{Law}(d_n^{-1/2} S_n)$ and $\mathcal{N}(0, \text{Var}(d_n^{-1/2} S_n))$ are weakly approaching, that is for any bounded continuous function $g : \mathbb{R} \rightarrow \mathbb{R}$,

$$\mathbb{E} \left[g \left(d_n^{-1/2} S_n \right) \right] - \int_{\mathbb{R}} g \left(\text{Var}(d_n^{-1/2} S_n)^{1/2} t \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Proof. For the proof, we refer the reader to [16]. $\square$

Corollary A.6. *Let $(W_n)_{n \in \mathbb{N}}$ be a sequence of $\mathbb{R}$ -valued zero mean random variables. Suppose there exists $r > 0$ such that $\sup_{n \in \mathbb{N}} \mathbb{E}|W_n|^{2+r} < \infty$ and $\sum_{k=0}^{\infty} (k+1)^2 (\alpha^W(k))^{\frac{r}{4+r}} < \infty$ holds. Under these conditions, the distributions of $n^{-1/2} S_n$ and $\mathcal{N}(0, \text{Var}(n^{-1/2} S_n))$ are weakly approaching.*

Moreover, if the sequence $(n/\sigma_n^2)_{n \geq 1}$ is bounded, where $\sigma_n^2 = \mathbb{E}S_n^2$, $n \in \mathbb{N}$, then $\text{Law}(S_n/\sigma_n)$ converges weakly to the standard normal distribution.

Proof. The first part of the statement is immediately follows from Theorem A.5 with $\xi_{n,i} = \frac{\sqrt{n}}{\sigma_n} W_i$, $1 \leq i \leq d_n = n$, $n \geq 1$, since $\xi_{n,\cdot}$ and W have the same α -mixing coefficient for every n .

As for the second part, let $\xi_{n,i} = \frac{\sqrt{n}}{\sigma_n} W_i$, $1 \leq i \leq n$, $n \geq 1$. Easily seen that $\max_{1 \leq i \leq n} \mathbb{E}|\xi_{n,i}|^{2+r} \leq (n/\sigma_n^2)^{1+r/2} \mathbb{E}|W_i|^{2+r}$ hence by Theorem A.5, for any bounded continuous function $g : \mathbb{R} \rightarrow \mathbb{R}$,

$$\mathbb{E} \left[g \left(n^{-1/2} \sum_{k=1}^n \xi_{n,k} \right) \right] - \int_{\mathbb{R}} g(t) \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \rightarrow 0, \text{ as } n \rightarrow \infty,$$

which completes the proof. $\square$

The next important remark in its original form can be found in Ekström's paper (c.f. Remark 1 on page 1 in [16]).

Remark A.7. *Assume that for $r > 2$, $\sup_{n \in \mathbb{N}} \mathbb{E}|W_n|^r < \infty$ and $\sum_{k=0}^{\infty} (\alpha^W(k))^{1-\frac{2}{r}} < \infty$. Then the variance of $n^{-1/2} S_n$ is bounded.*

B Counterexample to long-term contractivity condition

In this point we present an example for a stationary stochastic process $(Y_n)_{n \in \mathbb{N}}$ and a function $\gamma : \mathcal{Y} \rightarrow (0, \infty)$, outlined in Balázs Felsmann's Master's thesis, where $\mathbb{E}(\gamma(Y_0)) < 1$, and despite the favorable mixing properties of $(Y_n)_{n \in \mathbb{N}}$, the long-term contractivity condition (9) fails to hold.

Let $(Z_n)_{n \in \mathbb{Z}}$ be a sequence of i.i.d. Bernoulli variables with $\mathbb{P}(Z_0 = 0) = \mathbb{P}(Z_0 = 1) = 1/2$, and define the process

$$Y_n = Z_n + Z_{n-1}, \quad n \in \mathbb{Z},$$

which takes its values in $\mathcal{Y} = \{0, 1, 2\}$. Clearly, $(Y_n)_{n \in \mathbb{N}}$ is a stationary process such that Y_n and Y_m are independent for $|m - n| > 1$ hence $\alpha^Y(n) = 0$ for $n \geq 2$. We consider a function $\gamma : \mathcal{Y} \rightarrow (0, 1)$, where $\gamma(i) = \gamma_i$, $i = 0, 1, 2$ will be specified later. Let us introduce the following sequences.

$$a_n = \mathbb{E} \left(\prod_{k=1}^n \gamma(Y_k) \right), \quad b_n = \mathbb{E} \left(\mathbb{1}_{\{Z_n=0\}} \prod_{k=1}^n \gamma(Y_k) \right), \quad c_n = \mathbb{E} \left(\mathbb{1}_{\{Z_n=1\}} \prod_{k=1}^n \gamma(Y_k) \right), \quad n \in \mathbb{N}. \quad (36)$$

Clearly, we have $a_n = b_n + c_n$, $n \in \mathbb{N}$, moreover we can write

$$\begin{aligned} b_{n+1} &= \mathbb{E} \left(\mathbb{1}_{\{Z_{n+1}=0\}} \mathbb{1}_{\{Z_n=0\}} \prod_{k=1}^{n+1} \gamma(Y_k) \right) + \mathbb{E} \left(\mathbb{1}_{\{Z_{n+1}=0\}} \mathbb{1}_{\{Z_n=1\}} \prod_{k=1}^{n+1} \gamma(Y_k) \right) \\ &= \frac{1}{2}(\gamma_0 b_n + \gamma_1 c_n), \end{aligned}$$

and similarly

$$\begin{aligned} c_{n+1} &= \mathbb{E} \left(\mathbb{1}_{\{Z_{n+1}=1\}} \mathbb{1}_{\{Z_n=0\}} \prod_{k=1}^{n+1} \gamma(Y_k) \right) + \mathbb{E} \left(\mathbb{1}_{\{Z_{n+1}=1\}} \mathbb{1}_{\{Z_n=1\}} \prod_{k=1}^{n+1} \gamma(Y_k) \right) \\ &= \frac{1}{2}(\gamma_1 b_n + \gamma_2 c_n), \end{aligned}$$

hence the linear recursion

$$\begin{bmatrix} b_{n+1} \\ c_{n+1} \end{bmatrix} = \frac{1}{2} \Gamma \begin{bmatrix} b_n \\ c_n \end{bmatrix} \quad (37)$$

holds, where $\Gamma = \begin{bmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_2 \end{bmatrix}$, and $b_0 = c_0 = 1/2$. Thus for $(a_n)_{n \in \mathbb{N}}$, we have

$$a_n = \frac{1}{2^{n+1}} [1, 1] \Gamma^n \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

By setting $\gamma_0 = 3$, $\gamma_1 = \gamma_2 = 0$, we obtain $a_0 = 1$ and $a_n = \frac{1}{2}(3/2)^n$, for $n \geq 1$, thus $a_n \rightarrow \infty$, as $n \rightarrow \infty$. For $\varepsilon \in (0, 1/4)$, let us define $\gamma_\varepsilon : \mathcal{Y} \rightarrow (0, \infty)$, $\gamma_\varepsilon(y) = \gamma(y) + \varepsilon$, $y = 0, 1, 2$. So, we have $\mathbb{E}(\gamma_\varepsilon(Y_0)) = 3/4 + \varepsilon < 1$, on the other hand

$$\mathbb{E}^{1/n} \left(\prod_{k=1}^n \gamma_\varepsilon(Y_k) \right) \geq \mathbb{E}^{1/n} \left(\prod_{k=1}^n \gamma(Y_k) \right) = \frac{3}{2} 2^{-1/n}, \quad n \geq 1,$$

and thus

$$\liminf_{n \rightarrow \infty} \mathbb{E}^{1/n} \left(\prod_{k=1}^n \gamma_\varepsilon(Y_k) \right) \geq \frac{3}{2}$$

which means that the long-term contractivity condition fails to hold in this situation.

C Coupling condition for MCREs

In this section, our primary objective is to establish an upper bound for the non-coupling probabilities as in Definition 1.1. Our approach refines the proof of Lemma 3.10 in [36]. To achieve this, we introduce the following lemma, which is identical to Lemma 7.4 in [35]. It describes the consequences of the drift condition satisfied by $Q(y_{k-1}) \dots Q(y_l)$, where $\mathbf{y} \in \mathcal{Y}^{\mathbb{N}}$ and $k, l \in \mathbb{N}$, $l < k$ are arbitrary and fixed.

Lemma C.1. For $x \in \mathcal{X}$, $y \in \mathcal{Y}^{\mathbb{N}}$ and $k, l \in \mathbb{N}$, $l < k$, we have

$$[Q(y_{k-1}) \dots Q(y_l)V](x) \leq V(x) \prod_{r=l}^{k-1} \gamma(y_r) + \sum_{r=l}^{k-1} K(y_r) \prod_{j=r+1}^{k-1} \gamma(y_j).$$

Proof. We proceed by induction. Let $x \in \mathcal{X}$ and $l \in \mathbb{N}$ be arbitrary but fixed. For the base case $k = l + 1$, we have

$$[Q(y_l)V](x) \leq \gamma(y_l)V(x) + K(y_l), \quad (38)$$

which follows directly from the drift condition (7).

Induction hypothesis: Assume that the inequality holds for some $k > l$, with fixed $x \in \mathcal{X}$ and $l \in \mathbb{N}$:

$$[Q(y_{k-1}) \dots Q(y_l)V](x) \leq V(x) \prod_{r=l}^{k-1} \gamma(y_r) + \sum_{r=l}^{k-1} K(y_r) \prod_{j=r+1}^{k-1} \gamma(y_j). \quad (39)$$

Induction step: We verify that inequality (39) holds for $k + 1$. By the drift property for $Q(y_k)$ we have

$$[Q(y_k)V](x) \leq \gamma(y_k)V(x) + K(y_k).$$

Operators $V \mapsto [Q(y)V]$, $y \in \mathcal{Y}$ are linear, monotone and for $V \equiv 1$ $[Q(y)V] \equiv 1$, $y \in \mathcal{Y}$. Therefore, taking into account that successive applications of kernels are evaluated in reverse order (see the remark following Definition 2.1), we can write

$$\begin{aligned} [Q(y_k) \dots Q(y_l)V](x) &= [Q(y_{k-1}) \dots Q(y_l)[Q(y_k)V]](x) \\ &\leq \gamma(y_k)[Q(y_{k-1}) \dots Q(y_l)V](x) + K(y_k), \end{aligned}$$

thus by applying the induction hypothesis (39), we obtain

$$\begin{aligned} [Q(y_k) \dots Q(y_l)V](x) &\leq \gamma(y_k) \left[V(x) \prod_{r=l}^{k-1} \gamma(y_r) + \sum_{r=l}^{k-1} K(y_r) \prod_{j=r+1}^{k-1} \gamma(y_j) \right] + K(y_k) \\ &= V(x) \prod_{r=l}^k \gamma(y_r) + \sum_{r=l}^k K(y_r) \prod_{j=r+1}^k \gamma(y_j) \end{aligned}$$

which completes the proof. $\square$

In our earlier papers (See Lemma 7.1 in [35] and Lemma 3.9 in [36]), we relied on special cases of the following lemma. It is also a variant of Lemma 6.1 in [22], albeit in a somewhat broader context. Such representations of parametric kernels satisfying the minorization condition (8) can be deemed standard. For the sake of completeness and to ensure self-containment of our explanation, we present and prove it in its most general form.

Lemma C.2. Let $R > 0$ be arbitrary and suppose the parametric kernel $Q : \mathcal{Y} \times \mathcal{X} \times \mathcal{B}(\mathcal{X}) \rightarrow [0, 1]$ satisfies the minorization condition given by (8). Then there exists a measurable mapping $T^R : \mathcal{X} \times \mathcal{Y} \times [0, 1] \rightarrow \mathcal{X}$ such that

$$Q(y, x, A) = \int_{[0,1]} \mathbb{1}_{T^R(x,y,u) \in A} du,$$

for all $x \in \mathcal{X}$, $A \in \mathcal{B}(\mathcal{X})$ and $y \in \mathcal{Y}$. Furthermore, for any fixed $y \in \mathcal{Y}$, there exists a Borel set $U = U(y) \in \mathcal{B}([0, 1])$ with Lebesgue measure $\text{Leb}_1(U) \geq 1 - \beta(R, y)$ such that for $u \in U$,

$$T^R(x_1, y, u) = T^R(x_2, y, u), \quad x_1, x_2 \in \bar{V}^{-1}([0, R]). \quad (40)$$

Proof. We proceed as in Lemma 7.1 in [35], following the proof of Lemma 6.1 in [22]. The case of countable $\mathcal{X}$ is straightforward and thus omitted. For the uncountable case, we can assume, by the Borel isomorphism theorem, that $\mathcal{X} = \mathbb{R}$ and $\mathcal{B}(\mathcal{X}) = \mathcal{B}(\mathbb{R})$ is the standard Borel σ -algebra of $\mathbb{R}$.

It is easy to see that if $\beta(R, y) = 0$ for some $y \in \mathcal{Y}$ in the minorization condition (8), then for any $A \in \mathcal{B}(\mathcal{X})$ and $x \in \bar{V}^{-1}([0, R])$, both $Q(y, x, A) \geq \kappa_R(y, A)$ and $Q(y, x, \mathcal{X} \setminus A) \geq \kappa_R(y, \mathcal{X} \setminus A)$ hold simultaneously. Consequently, we have

$$Q(y, \cdot, A)|_{\bar{V}^{-1}([0, R])} = \kappa_R(y, A), \quad A \in \mathcal{B}(\mathcal{X}).$$

For $x \in \bar{V}^{-1}([0, R])$, $A \in \mathcal{B}(\mathcal{X}) = \mathcal{B}(\mathbb{R})$, and $y \in \mathcal{Y}$ let

$$q_R(y, x, A) := \begin{cases} \frac{1}{\beta(R, y)} [Q(y, x, A) - (1 - \beta(R, y))\kappa_R(y, A)] & \text{if } \beta(R, y) \neq 0 \\ 0 & \text{if } \beta(R, y) = 0 \end{cases}$$

Additionally, introduce the pseudoinverses of the corresponding cumulative distribution functions as follows:

$$\begin{aligned} Q^{-1}(y, x, z) &:= \inf\{r \in \mathbb{Q} \mid Q(y, x, (-\infty, r]) \geq z\} \\ \kappa_R^{-1}(y, z) &:= \inf\{r \in \mathbb{Q} \mid \kappa_R(y, (-\infty, r]) \geq z\} \\ q_R^{-1}(y, x, z) &:= \inf\{r \in \mathbb{Q} \mid q_R(y, x, (-\infty, r]) \geq z\}. \end{aligned}$$

There exists a measurable mapping $\chi : [0, 1] \rightarrow [0, 1]^2$ such that the pushforward measure $\chi_*(dx)$ equals $dx dy$. In other words, for every Borel set $B \in \mathcal{B}([0, 1]^2)$, $\text{Leb}_2(B) = \text{Leb}_1(\bar{\chi}^{-1}(B))$, where Leb_k refers to the Lebesgue measure on $[0, 1]^k$, $k = 1, 2$. Finally, we define

$$T^R(x, y, u) = \begin{cases} \mathbb{1}_{\chi(u)_1 \geq \beta(R, y)} \kappa_R^{-1}(y, \chi(u)_2) + \mathbb{1}_{\chi(u)_1 < \beta(R, y)} q_R^{-1}(y, x, \chi(u)_2) & \text{if } V(x) \leq R \\ Q^{-1}(y, x, \chi(u)_2) & \text{if } V(x) > R. \end{cases}$$

It is evident that $x \mapsto T^R(x, y, u)$ is constant on $\bar{V}^{-1}([0, R])$ whenever $\chi(u)_1 \geq \beta(R, y)$, implying (40) with $U = \bar{\chi}^{-1}([\beta(R, y), 1] \times [0, 1])$.

Furthermore, for any fixed $r \in \mathbb{R}$, $y \in \mathcal{Y}$ and $x \in \bar{V}^{-1}([0, R])$ by the change of variable formula and the definitions of T^R and q_R , we can write

$$\begin{aligned} \int_{[0, 1]} \mathbb{1}_{\{T^R(x, y, u) \leq r\}} du &= \int_0^1 \int_0^1 \mathbb{1}_{\{\mathbb{1}_{s \geq \beta(R, y)} \kappa_R^{-1}(y, t) + \mathbb{1}_{s < \beta(R, y)} q_R^{-1}(y, x, t) \leq r\}} ds dt \\ &= (1 - \beta(R, y)) \int_0^1 \mathbb{1}_{\kappa_R^{-1}(y, t) \leq r} dt + \beta(R, y) \int_0^1 \mathbb{1}_{q_R^{-1}(y, x, t) \leq r} dt \\ &= (1 - \beta(R, y)) \kappa_R(y, (-\infty, r]) + \beta(R, y) q_R(y, x, (-\infty, r]) \\ &= Q(y, x, (-\infty, r]). \end{aligned}$$

Similarly, for $x \notin \bar{V}^{-1}([0, R])$, we have

$$\int_{[0, 1]} \mathbb{1}_{T^R(x, y, u) \leq r} du = \int_0^1 \int_0^1 \mathbb{1}_{Q^{-1}(y, x, s) \leq r} ds dt = \int_0^1 \mathbb{1}_{Q^{-1}(y, x, s) \leq r} ds = Q(y, x, (-\infty, r]).$$

thus the claimed identity holds.

It remains only to show that T^R is measurable with respect to sigma algebras $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathcal{Y}) \otimes \mathcal{B}([0, 1])$ and $\mathcal{B}(\mathbb{R})$. Indeed, T^R is a composition of measurable functions. $\square$

By Assumption 2.2 B), there exists $0 < r < 1/\bar{\gamma} - 1$ such that $\bar{\beta} := \sup_{y \in \mathcal{Y}} \beta(R(y), y) < 1$ holds with $R(y) := \frac{2K(y)}{r\gamma(y)}$. We define the measurable mapping

$$(x, y, u) \mapsto f(x, y, u) := T^{R(y)}(x, y, u), \quad x \in \mathcal{X}, y \in \mathcal{Y}, u \in [0, 1]. \quad (41)$$

Let $(\varepsilon_t)_{t \in \mathbb{N}}$ be a sequence of i.i.d. variables uniformly distributed on $[0, 1]$ such that sigma algebras $\mathcal{F}_{0, \infty}^\varepsilon$ and $\sigma(Y_t, X_t, t \in \mathbb{N})$ are independent. Furthermore, for $s \in \mathbb{N}$ and $x \in \mathcal{X}$, let us introduce the family of auxiliary processes

$$Z_{s,t}^{x,\mathbf{y}} = \begin{cases} x & \text{if } t \leq s \\ f(Z_{s,t-1}^{x,\mathbf{y}}, y_{t-1}, \varepsilon_t) & \text{if } t > s, \end{cases} \quad (42)$$

where $\mathbf{y} = (y_0, y_1, \dots) \in \mathcal{Y}^\mathbb{N}$ can be any fixed trajectory. Clearly, for $\mathbf{Y} = (Y_n)_{n \in \mathbb{N}}$, the process $Z_{s,t}^{X_s, \mathbf{Y}}$, $t \geq s$ is a version of $(X_t)_{t \geq s}$. In the forthcoming part of the section, we will prove that this process satisfies the coupling condition. First, we will show that for any fixed $x \in \mathbb{N}$ and $\mathbf{y} \in \mathcal{Y}^\mathbb{N}$, the process $Z_{s,t}^{x,\mathbf{y}}$, $t \geq s$ is a *Harris recurrent* time-inhomogeneous Markov chain. The next lemma provides a quenched version of the coupling condition, controlling the coupling time between iterations starting from different initial values.

Lemma C.3. *Let $x_1, x_2 \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{Y}^\mathbb{N}$ be arbitrary but fixed. Then for $0 < m < n$, we have*

$$\mathbb{P}(Z_{0,n}^{x_1, \mathbf{y}} \neq Z_{0,n}^{x_2, \mathbf{y}}) \leq \bar{\beta}^m + (1+r)^{\lfloor \frac{n-1}{m} \rfloor} \left\{ \frac{V(x_1) + V(x_2)}{2} \prod_{j=0}^{\lfloor \frac{n-1}{m} \rfloor} \gamma(y_j) + \sum_{k=0}^{m-1} \sum_{l=k}^{\lfloor \frac{n-1}{m} \rfloor} K(y_l) \prod_{j=1}^{\lfloor \frac{n-1}{m} \rfloor} \gamma(y_{l+j}) \right\}.$$

Proof. For the sake of brevity, we employ a more concise notation: $Z_n^i = Z_{0,n}^{x_i, \mathbf{y}}$, $i = 1, 2$, and $n \in \mathbb{N}$, moreover we introduce

$$\bar{Z}_n := (Z_n^1, Z_n^2), \quad \|\bar{Z}_n\| := \max(V(Z_n^1), V(Z_n^2))$$

and the sequence of successive visiting times

$$\sigma_0 := 0, \quad \sigma_{k+1} = \inf \{n > \sigma_k \mid \|\bar{Z}_n\| \leq R(y_n)\}, \quad k \in \mathbb{N}$$

that are obviously $\mathcal{F}_{1, \infty}^\varepsilon$ -stopping times. Note that on $\{\|\bar{Z}_n\| > R(y_n)\}$ we have

$$\gamma(y_n)(V(Z_n^1) + V(Z_n^2)) + 2K(y_n) \leq (1+r)\gamma(y_n)(V(Z_n^1) + V(Z_n^2)) \quad (43)$$

and thus for $k \geq 1$ and $s \geq 1$, by the Markov inequality, we obtain

$$\begin{aligned} \mathbb{P}(\sigma_{k+1} - \sigma_k > s \mid \mathcal{F}_{1, \sigma_k}^\varepsilon) &\leq \mathbb{E} \left(\mathbb{P}(C_{\sigma_k+s} \mid \bar{Z}_{\sigma_k+s-1}) \prod_{j=1}^{s-1} \mathbb{1}_{C_{\sigma_k+j}} \mid \mathcal{F}_{1, \sigma_k}^\varepsilon \right) \\ &\leq \frac{(1+r)\gamma(y_{\sigma_k+s-1})}{R(y_{\sigma_k+s})} \mathbb{E} \left((V(Z_{\sigma_k+s-1}^1) + V(Z_{\sigma_k+s-1}^2)) \prod_{j=1}^{s-2} \mathbb{1}_{C_{\sigma_k+j}} \mid \mathcal{F}_{1, \sigma_k}^\varepsilon \right), \end{aligned}$$

where C_n is the shorthand notation for the event $\{\|\bar{Z}_n\| > R(y_n)\}$, $n \in \mathbb{N}$.

By the tower rule, we can write

$$\begin{aligned} \mathbb{E} \left((V(Z_{\sigma_k+s-1}^1) + V(Z_{\sigma_k+s-1}^2)) \prod_{j=1}^{s-2} \mathbb{1}_{C_{\sigma_k+j}} \mid \mathcal{F}_{1, \sigma_k}^\varepsilon \right) &= \\ &= \mathbb{E} \left(\mathbb{E} [V(Z_{\sigma_k+s-1}^1) + V(Z_{\sigma_k+s-1}^2) \mid \mathcal{F}_{1, \sigma_k+s-2}^\varepsilon] \prod_{j=1}^{s-2} \mathbb{1}_{C_{\sigma_k+j}} \mid \mathcal{F}_{1, \sigma_k}^\varepsilon \right). \end{aligned}$$

Using the Markov property of $(\bar{Z}_n)_{n \in \mathbb{N}}$ and the drift property of the parametric kernel (7), we have

$$\begin{aligned} \mathbb{E} [V(Z_{\sigma_k+s-1}^1) + V(Z_{\sigma_k+s-1}^2) | \mathcal{F}_{1, \sigma_k+s-2}^\varepsilon] &= \mathbb{E} [V(Z_{\sigma_k+s-1}^1) + V(Z_{\sigma_k+s-1}^2) | \bar{Z}_{\sigma_k+s-2}] \\ &= [Q(y_{\sigma_k+s-2})V](Z_{\sigma_k+s-2}^1) + [Q(y_{\sigma_k+s-2})V](Z_{\sigma_k+s-2}^2) \\ &\leq \gamma(y_{\sigma_k+s-2}) (V(Z_{\sigma_k+s-2}^1) + (Z_{\sigma_k+s-2}^2)) + 2K(y_{\sigma_k+s-2}). \end{aligned}$$

Now, inequality (43) yields

$$\mathbb{E} [V(Z_{\sigma_k+s-1}^1) + V(Z_{\sigma_k+s-1}^2) | \mathcal{F}_{1, \sigma_k+s-2}^\varepsilon] \mathbf{1}_{C_{\sigma_k+s-2}} \leq (1+r)\gamma(y_{\sigma_k+s-2}) (V(Z_{\sigma_k+s-2}^1) + (Z_{\sigma_k+s-2}^2)).$$

Finally, we arrive at

$$\begin{aligned} \mathbb{E} \left((V(Z_{\sigma_k+s-1}^1) + V(Z_{\sigma_k+s-1}^2)) \prod_{j=1}^{s-2} \mathbf{1}_{C_{\sigma_k+j}} \middle| \mathcal{F}_{1, \sigma_k}^\varepsilon \right) &\leq \\ (1+r)\gamma(y_{\sigma_k+s-2}) \mathbb{E} \left((V(Z_{\sigma_k+s-2}^1) + V(Z_{\sigma_k+s-2}^2)) \prod_{j=1}^{s-3} \mathbf{1}_{C_{\sigma_k+j}} \middle| \mathcal{F}_{1, \sigma_k}^\varepsilon \right). \end{aligned}$$

Iteration of this argument in $s-2$ steps leads to the following estimation:

$$\begin{aligned} \mathbb{P}(\sigma_{k+1} - \sigma_k > s \mid \mathcal{F}_{1, \sigma_k}^\varepsilon) &\leq \frac{(1+r)^{s-1} \prod_{j=1}^{s-1} \gamma(y_{\sigma_k+j})}{R(y_{\sigma_k+s})} \mathbb{E} (V(Z_{\sigma_k+1}^1) + V(Z_{\sigma_k+1}^2) | \bar{Z}_{\sigma_k}) \\ &\leq \frac{r(1+r)^{s-1} \prod_{j=1}^s \gamma(y_{\sigma_k+j})}{2K(y_{\sigma_k+s})} [\gamma(y_{\sigma_k}) (V(Z_{\sigma_k}^1) + V(Z_{\sigma_k}^2)) + 2K(y_{\sigma_k})] \\ &\leq (1+r)^s K(y_{\sigma_k}) \prod_{j=1}^s \gamma(y_{\sigma_k+j}), \end{aligned}$$

where we used that $V(Z_{\sigma_k}^1) + V(Z_{\sigma_k}^2) \leq 2R(y_{\sigma_k})$, and $K(\cdot) \geq 1$.

Along similar lines, we can show that

$$\begin{aligned} \mathbb{P}(\sigma_1 > s) &\leq \frac{(1+r)^{s-1} \prod_{j=1}^{s-1} \gamma(y_j)}{R(y_s)} [\gamma(y_0) (V(x_1) + V(x_2)) + 2K(y_0)] \\ &\leq (1+r)^s \prod_{j=1}^s \gamma(y_j) \left[\frac{\gamma(y_0)}{2} (V(x_1) + V(x_2)) + K(y_0) \right]. \end{aligned}$$

Clearly, for any $0 < m \leq n$, on the event $\{\sigma_m > n\}$ we have $\{\sigma_{k+1} - \sigma_k > \lfloor n/m \rfloor\} \cap \{\sigma_k \leq \lfloor n/m \rfloor\}$ for some $k = 0, 1, \dots, m-1$ hence by the union bound and the estimates we obtain for

the time elapsed between consecutive visits, we can write

$$\begin{aligned}
\mathbb{P}(\sigma_m > n) &\leq \mathbb{P}\left(\bigcup_{k=0}^{m-1} \{\sigma_{k+1} - \sigma_k > \lfloor n/m \rfloor\} \cap \{\sigma_k \leq k \lfloor n/m \rfloor\}\right) \\
&\leq \sum_{k=0}^{m-1} \mathbb{P}(\{\sigma_{k+1} - \sigma_k > \lfloor n/m \rfloor\} \cap \{\sigma_k \leq k \lfloor n/m \rfloor\}) \\
&= \sum_{k=0}^{m-1} \sum_{l=k}^{k \lfloor n/m \rfloor} \mathbb{P}(\sigma_{k+1} - \sigma_k > \lfloor n/m \rfloor \mid \sigma_k = l) \mathbb{P}(\sigma_k = l) \\
&\leq \mathbb{P}(\sigma_1 > \lfloor n/m \rfloor) + \sum_{k=1}^{m-1} \sum_{l=k}^{k \lfloor n/m \rfloor} \mathbb{P}(\sigma_{k+1} - \sigma_k > \lfloor n/m \rfloor \mid \sigma_k = l) \\
&\leq \left[\frac{\gamma(y_0)}{2} (V(x_1) + V(x_2)) + K(y_0) \right] (1+r)^{\lfloor n/m \rfloor} \prod_{j=1}^{\lfloor n/m \rfloor} \gamma(y_j) \\
&\quad + (1+r)^{\lfloor n/m \rfloor} \sum_{k=1}^{m-1} \sum_{l=k}^{k \lfloor n/m \rfloor} K(y_l) \prod_{j=1}^{\lfloor n/m \rfloor} \gamma(y_{l+j}).
\end{aligned}$$

Next, we estimate the probability of no-coupling on events when small sets are visited at least m -times. By Lemma C.2 and the choice of $R(y)$ in the definition of f in (41) and (42), for each $j = 1, \dots, m$, $x \mapsto f(x, y_j, \varepsilon_{\sigma_j+1})$ is constant on the level set $\bar{V}^{-1}([0, R(y_{\sigma_j})])$ with probability at least $1 - \bar{\beta}$ independently of $\mathcal{F}_{0, \sigma_j}^\varepsilon$ thus no-coupling happens with probability at most $\bar{\beta}$. Therefore, we can estimate:

$$\mathbb{P}(Z_{0,n}^{x_1, \mathbf{y}} \neq Z_{0,n}^{x_2, \mathbf{y}}, \sigma_m < n) \leq \bar{\beta}^m.$$

Finally, we combine this upper bound with that one what we got for the tail probability of the visiting times, and obtain

$$\begin{aligned}
\mathbb{P}(Z_{0,n}^{x_1, \mathbf{y}} \neq Z_{0,n}^{x_2, \mathbf{y}}) &\leq \mathbb{P}(Z_{0,n}^{x_1, \mathbf{y}} \neq Z_{0,n}^{x_2, \mathbf{y}}, \sigma_m < n) + \mathbb{P}(\sigma_m > n-1) \\
&\leq \bar{\beta}^m + \left[\frac{\gamma(y_0)}{2} (V(x_1) + V(x_2)) + K(y_0) \right] (1+r)^{\lfloor \frac{n-1}{m} \rfloor} \prod_{j=1}^{\lfloor \frac{n-1}{m} \rfloor} \gamma(y_j) \\
&\quad + (1+r)^{\lfloor \frac{n-1}{m} \rfloor} \sum_{k=1}^{m-1} \sum_{l=k}^{k \lfloor \frac{n-1}{m} \rfloor} K(y_l) \prod_{j=1}^{\lfloor \frac{n-1}{m} \rfloor} \gamma(y_{l+j}) \\
&= \bar{\beta}^m + (1+r)^{\lfloor \frac{n-1}{m} \rfloor} \left\{ \frac{V(x_1) + V(x_2)}{2} \prod_{j=0}^{\lfloor \frac{n-1}{m} \rfloor} \gamma(y_j) + \sum_{k=0}^{m-1} \sum_{l=k}^{k \lfloor \frac{n-1}{m} \rfloor} K(y_l) \prod_{j=1}^{\lfloor \frac{n-1}{m} \rfloor} \gamma(y_{l+j}) \right\}
\end{aligned}$$

which completes the proof. $\square$

Proof of Lemma 2.4. The process $Z_{0,n}^{X_0, \mathbf{Y}}$, $n \in \mathbb{N}$ is a version of $(X_n)_{n \in \mathbb{N}}$ thus to simplify the notation in this proof, we may and will redefine $(X_n)_{n \in \mathbb{N}}$ as $X_n := Z_{0,n}^{X_0, \mathbf{Y}}$. Furthermore, we indicate the dependence of Z on the driving noise by writing $Z_{s,t}^{x, \mathbf{y}, \varepsilon}$ instead of $Z_{s,t}^{x, \mathbf{y}}$, where ε refers to $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots)$. We also introduce the usual left shift operation: $(S\mathbf{y})_j = y_{j+1}$ and similarly $(S\varepsilon)_j = \varepsilon_{j+1}$, $j \in \mathbb{N}$.

For arbitrary but fixed $j \in \mathbb{N}$ and $x \in \mathcal{X}$, we can write

$$\mathbb{P}(Z_{j,j+n}^{X_j, \mathbf{Y}, \varepsilon} \neq Z_{j,j+n}^{x, \mathbf{Y}, \varepsilon}) = \mathbb{P}(Z_{0,n}^{X_j, S^j \mathbf{Y}, S^j \varepsilon} \neq Z_{0,n}^{x, S^j \mathbf{Y}, S^j \varepsilon})$$

thus by the tower rule and Lemma C.3, for $0 < m < n$, we have

$$\begin{aligned} \mathbb{P}(Z_{0,n}^{X_j, S^j \mathbf{Y}, S^j \varepsilon} \neq Z_{0,n}^{x, S^j \mathbf{Y}, S^j \varepsilon}) &= \mathbb{E} \left[\mathbb{P} \left(Z_{0,n}^{X_j, S^j \mathbf{Y}, S^j \varepsilon} \neq Z_{0,n}^{x, S^j \mathbf{Y}, S^j \varepsilon} \middle| \mathcal{F}_{0,\infty}^Y \vee \sigma(X_j) \right) \right] \leq \bar{\beta}^m + \\ (1+r)^{\lfloor \frac{n-1}{m} \rfloor} &\mathbb{E} \left[\frac{\mathbb{E}[V(X_j) | \mathcal{F}_{0,\infty}^Y] + V(x)}{2} \prod_{r=0}^{\lfloor \frac{n-1}{m} \rfloor} \gamma(Y_{r+j}) + \sum_{k=0}^{m-1} \sum_{l=k}^{\lfloor \frac{n-1}{m} \rfloor} K(Y_{l+j}) \prod_{r=1}^{\lfloor \frac{n-1}{m} \rfloor} \gamma(Y_{l+r+j}) \right]. \end{aligned} \quad (44)$$

As for $\mathbb{E}[V(X_j) | \mathcal{F}_{0,\infty}^Y]$, by Lemma C.1, we obtain

$$\begin{aligned} \mathbb{E}[V(X_j) | \mathcal{F}_{0,\infty}^Y] &= \mathbb{E}[[Q(Y_{j-1}), \dots, Q(Y_0)V](X_0) | \mathcal{F}_{0,\infty}^Y] \\ &\leq \mathbb{E}(V(X_0)) \prod_{i=0}^{j-1} \gamma(Y_i) + \sum_{r=0}^{j-1} K(Y_r) \prod_{i=r+1}^{j-1} \gamma(Y_i), \end{aligned} \quad (45)$$

where we used that the initial state X_0 is independent of $\sigma(Y_n, \varepsilon_{n+1} | n \in \mathbb{N})$.

For $n > 1$, we choose $m_n = \frac{n-1}{\lfloor n^{1/2} \rfloor - 1}$. It is easy to check that $m_n \geq \lfloor n^{1/2} \rfloor$ for $n > 1$ hence $\bar{\beta}^{m_n} \leq \bar{\beta}^{\lfloor n^{1/2} \rfloor}$. Furthermore, let us fix $\gamma' > \bar{\gamma}$ such that $\tilde{\gamma} := (1+r)\gamma' < 1$. Then by Assumption 2.2 A), there exists $N \in \mathbb{N}$ such that

$$\sup_{j \geq -1} \mathbb{E} \left[K(Y_j) \prod_{k=1}^n \gamma(Y_{k+j}) \right] \leq (\gamma')^n, \quad n \geq N,$$

where for our convenience, we employ the convention $K(Y_{-1}) := 1$. Using (45), for $\lfloor n^{1/2} \rfloor \geq N$, we obtain

$$\begin{aligned} \mathbb{E} \left[\mathbb{E}[V(X_j) | \mathcal{F}_{0,\infty}^Y] \prod_{r=0}^{\lfloor n^{1/2} \rfloor - 1} \gamma(Y_{r+j}) \right] &\leq \mathbb{E}(V(X_0)) \mathbb{E} \left[\prod_{i=0}^{\lfloor n^{1/2} \rfloor + j - 1} \gamma(Y_i) \right] + \sum_{r=0}^{j-1} \mathbb{E} \left[K(Y_r) \prod_{i=r+1}^{\lfloor n^{1/2} \rfloor + j - 1} \gamma(Y_i) \right] \\ &\leq \mathbb{E}(V(X_0)) (\gamma')^{\lfloor n^{1/2} \rfloor + j} + \sum_{r=0}^{j-1} (\gamma')^{\lfloor n^{1/2} \rfloor + j - 1 - r} \\ &\leq (\gamma')^{\lfloor n^{1/2} \rfloor} \left(\mathbb{E}(V(X_0)) + \frac{1}{1 - \gamma'} \right). \end{aligned}$$

Similarly, for $\lfloor n^{1/2} \rfloor - 1 \geq N$, we get

$$\mathbb{E} \left[V(x) \prod_{r=0}^{\lfloor n^{1/2} \rfloor - 1} \gamma(Y_{r+j}) \right] \leq (\gamma')^{\lfloor n^{1/2} \rfloor} V(x),$$

moreover

$$\begin{aligned} \sum_{k=0}^{m_n-1} \sum_{l=k}^{\lfloor n^{1/2} \rfloor - 1} \mathbb{E} \left[K(Y_{l+j}) \prod_{r=1}^{\lfloor n^{1/2} \rfloor - 1} \gamma(Y_{l+r+j}) \right] &\leq \sum_{k=0}^{m_n-1} \sum_{l=k}^{\lfloor n^{1/2} \rfloor - 1} (\gamma')^{\lfloor n^{1/2} \rfloor - 1} \\ &= m_n \left[\frac{(m_n - 1)(\lfloor n^{1/2} \rfloor - 2)}{2} + 1 \right] (\gamma')^{\lfloor n^{1/2} \rfloor - 1}. \end{aligned}$$

To sum up, and taking into account that $m_n = O(n^{1/2})$, we obtain

$$\begin{aligned} \mathbb{P}(Z_{j,j+n}^{X_j, \mathbf{Y}, \varepsilon} \neq Z_{j,j+n}^{x, \mathbf{Y}, \varepsilon}) &\leq \bar{\beta}^{n^{1/2}-1} + \frac{(1+r)^{\lfloor n^{1/2} \rfloor - 1}}{2} (\gamma')^{\lfloor n^{1/2} \rfloor} \left(V(x) + \mathbb{E}(V(X_0)) + \frac{1}{1 - \gamma'} \right) \\ &\quad + cn^{3/2} (1+r)^{\lfloor n^{1/2} \rfloor - 1} (\gamma')^{\lfloor n^{1/2} \rfloor - 1} \\ &\leq \bar{\beta}^{n^{1/2}-1} + \left[\frac{(V(x) + \mathbb{E}(V(X_0)) + \frac{1}{1 - \bar{\gamma}})}{2} + cn^{3/2} \right] \tilde{\gamma}^{\lfloor n^{1/2} \rfloor - 1} \end{aligned}$$

for some $c > 0$, whenever $n \geq (N + 2)^2$ which implies the desired estimate. $\square$

The proof of Theorem 2.7 follows a similar approach to that of Lemma 2.4, with a few key differences. The most significant distinction to keep in mind is that the distribution of X_0^* heavily depends on the whole trajectory $(Y_n)_{n \in \mathbb{Z}}$.

Lemma C.4. *Under the conditions of Theorem 2.7, we have*

$$\limsup_{n \rightarrow \infty} \mathbb{E}^{1/n} \left[V(X_0^*) \prod_{k=0}^{n-1} \gamma(Y_k) \right] \leq \bar{\gamma}.$$

Proof. Without the loss of generality, for our convenience, we can assume that the i.i.d. driving noise is a double-sided infinite process $(\varepsilon_n)_{n \in \mathbb{Z}}$, as well. Let $x \in \mathcal{X}$ be arbitrary and deterministic. Then by Corollary 1 and the subsequent Note in [49], $\text{Law}(Z_{0,m}^{x,\mathbf{Y}}) = \text{Law}(Z_{-m,0}^{x,\mathbf{Y}}) \rightarrow \text{Law}(X_0^*)$, as $m \rightarrow \infty$ in total variation hence

$$\mathbb{E} \left[V(X_0^*) \prod_{k=0}^{n-1} \gamma(Y_k) \right] = \lim_{M \rightarrow \infty} \lim_{m \rightarrow \infty} \mathbb{E} \left[\min \left(M, V(Z_{-m,0}^{x,\mathbf{Y}}) \prod_{k=0}^{n-1} \gamma(Y_k) \right) \right]. \quad (46)$$

By Lemma C.1, we can write

$$\begin{aligned} \mathbb{E} \left[V(Z_{-m,0}^{x,\mathbf{Y}}) \prod_{k=0}^{n-1} \gamma(Y_k) \middle| \mathcal{F}_{-\infty,\infty}^Y \right] &= [Q(Y_{-1}) \dots Q(Y_{-m})V](x) \prod_{k=0}^{n-1} \gamma(Y_k) \\ &\leq V(x) \prod_{r=-m}^{n-1} \gamma(Y_r) + \sum_{r=-m}^{-1} K(Y_r) \prod_{j=r+1}^{n-1} \gamma(Y_j) \end{aligned}$$

thus by the tower rule, we have

$$\mathbb{E} \left[\min \left(M, V(Z_{-m,0}^{x,\mathbf{Y}}) \prod_{k=0}^{n-1} \gamma(Y_k) \right) \right] \leq V(x) \mathbb{E} \left[\prod_{r=-m}^{n-1} \gamma(Y_r) \right] + \sum_{r=-m}^{-1} \mathbb{E} \left[K(Y_r) \prod_{j=r+1}^{n-1} \gamma(Y_j) \right]. \quad (47)$$

By Assumption 2.2, for any fixed $\bar{\gamma} < \gamma' < 1$, exists $N \in \mathbb{N}$ such that

$$\mathbb{E} \left(K(Y_0) \prod_{j=1}^n \gamma(Y_j) \right) \leq (\gamma')^n, \quad n \geq N,$$

and thus we can further estimate the right-hand side in (47) as follows:

$$\begin{aligned} \mathbb{E} \left[\min \left(M, V(Z_{-m,0}^{x,\mathbf{Y}}) \prod_{k=0}^{n-1} \gamma(Y_k) \right) \right] &\leq V(x)(\gamma')^{n+m} + \sum_{r=-m}^{-1} (\gamma')^{n-1-r} \\ &\leq V(x)(\gamma')^{n+m} + \frac{(\gamma')^n}{1 - \gamma'}. \end{aligned}$$

Substituting this estimate back into (46) yields

$$\limsup_{n \rightarrow \infty} \mathbb{E}^{1/n} \left[V(X_0^*) \prod_{k=0}^{n-1} \gamma(Y_k) \right] \leq \gamma'.$$

Now, taking the limit $\gamma' \downarrow \bar{\gamma}$ gives the desired inequality. $\square$

Proof of Theorem 2.7. We consider the processes $Z_{0,n}^{X_0, \mathbf{Y}}$ and $Z_{0,n}^{X_0^*, \mathbf{Y}}$, $n \in \mathbb{N}$, where the latter is a version of $(X_n^*)_{n \in \mathbb{N}}$. For the coupling time, along similar lines as in the proof of Lemma 2.4, by using Lemma C.3, we obtain

$$\begin{aligned} \mathbb{P}(\tau > n) &\leq \mathbb{P}\left(Z_{0,n}^{X_0, \mathbf{Y}} \neq Z_{0,n}^{X_0^*, \mathbf{Y}}\right) \\ &\leq \bar{\beta}^m + \frac{(1+r)^{\lfloor \frac{n-1}{m} \rfloor}}{2} \mathbb{E} \left[\left(\mathbb{E}[V(X_0^*) | \mathcal{F}_{-\infty, \infty}^Y] + \mathbb{E}(V(X_0)) \right) \prod_{r=0}^{\lfloor \frac{n-1}{m} \rfloor} \gamma(Y_r) \right] \\ &\quad + (1+r)^{\lfloor \frac{n-1}{m} \rfloor} \sum_{k=0}^{m-1} \sum_{l=k}^{\lfloor \frac{n-1}{m} \rfloor} \mathbb{E} \left[K(Y_l) \prod_{r=1}^{\lfloor \frac{n-1}{m} \rfloor} \gamma(Y_{l+r}) \right]. \end{aligned}$$

where $0 < m < n$. Again, as in the proof of Lemma 2.4, let us fix $m_n = \frac{n-1}{\lfloor n^{1/2} \rfloor - 1}$, $n > 1$, and $\gamma' > \bar{\gamma}$ such that $\tilde{\gamma} := (1+r)\gamma' < 1$. By Assumption 2.2 and Lemma C.4, we can choose $N \in \mathbb{N}$ be so large that such that

$$\mathbb{E} \left[K(Y_0) \prod_{k=1}^n \gamma(Y_k) \right] \leq (\gamma')^n \text{ and } \mathbb{E} \left[V(X_0^*) \prod_{k=0}^{n-1} \gamma(Y_k) \right] \leq (\gamma')^n, \quad n \geq N.$$

For $\lfloor n^{1/2} \rfloor \geq N$, we have

$$\begin{aligned} \mathbb{P}(\tau \geq n) &\leq \bar{\beta}^{\lfloor n^{1/2} \rfloor} + \frac{(1+r)^{\lfloor n^{1/2} \rfloor - 1}}{2} (1 + \mathbb{E}(V(X_0))) (\gamma')^{\lfloor n^{1/2} \rfloor} \\ &\quad + (1+r)^{\lfloor n^{1/2} \rfloor - 1} \sum_{k=0}^{m_n-1} \sum_{l=k}^{\lfloor n^{1/2} \rfloor - 1} (\gamma')^{\lfloor n^{1/2} \rfloor - 1} \\ &\leq \bar{\beta}^{\lfloor n^{1/2} \rfloor} + \frac{1 + \mathbb{E}(V(X_0))}{2} \tilde{\gamma}^{\lfloor n^{1/2} \rfloor} + \tilde{\gamma}^{\lfloor n^{1/2} \rfloor - 1} \sum_{k=0}^{m_n-1} (k(\lfloor n^{1/2} \rfloor - 2) + 1), \end{aligned}$$

where $\sum_{k=0}^{m_n-1} (k(\lfloor n^{1/2} \rfloor - 2) + 1) = O(n^{3/2})$ which completes the proof. $\square$