Inferring serial correlation with dynamic backgrounds

Song Wei [1] **Yao Xie** [1] **Dobromir Rahnev** [2]

Abstract

Sequential data with serial correlation and an unknown, unstructured, and dynamic background is ubiquitous in neuroscience, psychology, and econometrics. Inferring serial correlation for such data is a fundamental challenge in statistics. We propose a Total Variation (TV) constrained least square estimator coupled with hypothesis tests to infer the serial correlation in the presence of unknown and unstructured dynamic background. The TV constraint on the dynamic background encourages a piecewise constant structure, which can approximate a wide range of dynamic backgrounds. The tuning parameter is selected via the Ljung-Box test to control the bias-variance trade-off. We establish a non-asymptotic upper bound for the estimation error through variational inequalities. We also derive a lower error bound via Fano's method and show the proposed method is near-optimal. Numerical simulation and a real study in psychology demonstrate the excellent performance of our proposed method compared with the state-of-the-art.

1. Introduction

Serial correlation and serial dependence have been central to time series analysis (Hong, 2010). Modern time-series data from neuroscience, psychology, and economics usually contain both a substantial serial dependence and a non-stationary drift (Akrami et al., 2018; Wexler et al., 2015; Moskowitz et al., 2012; Fischer & Whitney, 2014; Cicchini et al., 2018; McIlhagga, 2008; Rahnev et al., 2015). A well-known example comes from human reaction times, which are thought to be autocorrelated but also drift throughout an experiment (Laming, 1968). The drift can be due to many factors such as becoming better on the task, increased tiredness, and attention or arousal fluctuations. None of these influences take a specific parametric form. While some (e.g., learning or fatigue) are likely to be monotonic, others (e.g., fluctuations in attention) can be expected to waver unpredictably. This non-stationary background drift is thus typically considered a nuisance variable.

It is typically of strong scientific interest to infer the presence and/or assess serial correlation's strength with an unknown and unstructured dynamic background. The magnitude of autocorrelation has direct implications for many scientific theories. For example, Fischer & Whitney (2014) proposed that the human brain creates a "perceptual continuity field" where the subjective percept at one point of time directly influences the percept within a subsequent 15-second window. Such effects are known as "serial dependence" and are an active area of research within psychology and neuroscience. Progress in this and related endeavors depends on one's ability to estimate the magnitude of autocorrelation in certain time series, even in the presence of substantial unstructured drift.

The most popular tool to handle the serial correlation is the autoregressive (AR) time series. However, the presence of even a small drift can induce strong biases in the estimated AR coefficients. For example, unmodeled background drift can masquerade as autocorrelation, as illustrated in the first panel in Figure 1. This issue has been pointed out before by Dutilh et al. (2012), but no solution exists to date. Techniques have been developed for tracking the unknown dynamic background with minimum structural assumptions (Hodrick & Prescott, 1997; Kim et al., 2009; Harchaoui & Lévy-Leduc, 2010) but these approaches do not estimate the serial correlation. *Thus, we currently lack an efficient method to capture the autocorrelation strength in a time series in the presence of highly unstructured dynamic drifts.*

Motivated by this, we consider the following problem. Assume a sequence of observations $x_1, \ldots, x_T$ over time horizon T, which are generated from the underlying non-stationary $\text{AR}(p)$ time series model:

$$x_i = f_i + \sum_{j=1}^{p} \alpha_j x_{i-j} + \varepsilon_i, \quad i = 1, \ldots, T, \qquad (1)$$

where $\varepsilon_1, \ldots, \varepsilon_T$ are i.i.d. sub-Gaussian random noise with zero mean and variance σ_0^2, $\alpha_1, \ldots, \alpha_p$ are AR coefficients, $f_1, \ldots, f_T$ are deterministic dynamic backgrounds

[1]School of Industrial and Systems Engineering, Georgia Tech, [2]School of Psychology, Georgia Tech, Atlanta, GA, USA. Correspondence to: Yao Xie <yao.xie@isye.gatech.edu>, Song Wei <song.wei@gatech.edu>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

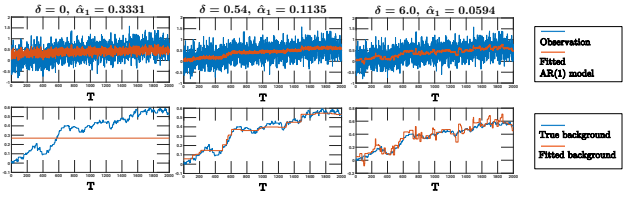


Figure 1. An example showing that proper modeling of dynamic background is important in capturing serial correlation. The estimate $\widehat{\alpha}_1$ is specified on the top of each column, with the ground truth $\alpha_1 = 0.1$; δ is a hyperparameter that controls the data fit and model complexity. In the first panel, we directly fit an AR(1) model while ignoring the dynamic background, leading to overestimating $\widehat{\alpha}_1$. In the third panel, the result overfits the dynamic background, leading to underestimating the AR coefficient. The second panel is the desired result obtained via our method.

and $x_{-p+1}, \ldots, x_0$ are the known history. The goal is to infer the presence and/or estimate the unknown AR coefficients and dynamic backgrounds simultaneously from data. *To ensure our model is general, we do not impose parametric or distributional assumptions on the dynamic backgrounds.*

In this paper, we present a new convex optimization based method to estimate the AR coefficients for sequential data in the presence of unknown dynamic background, coupled with the Ljung-Box test for model diagnosis. We cast the problem as minimizing the least square error with a Total Variation (TV) constraint on the dynamic background, which encourages a piecewise constant structure and can approximate a wide range of unstructured drifts with good precision. We establish performance guarantees for the ℓ_2 recovery error of the coefficients. To efficiently tune hyperparameters to control the bias-variance trade-off, we adopt the Ljung-Box test (Ljung & Box, 1978). Extensive numerical experiments are performed to validate the effectiveness of our proposed method. We also test our method on a real psychology dataset to demonstrate it can infer whether or not there is a statistically significant correlation.

1.1. Related work

Standard time series models (Brockwell et al., 1991) such as autoregressive and moving average models do not include dynamic backgrounds. On the other hand, the classical approach to capture dynamic background usually makes strong structural assumptions such as the linear trend, periodical trend Clark (1987) or hidden Markov model (Hamilton, 1989). Our problem involves a highly unstructured background, which requires new solution approach; moreover, existing theory does not apply because the unstructured dynamic background leads to a non-stationary time series, which does not satisfy the strong-mixing condition. This disables us from using the classic asymptotic analysis.

Recent works for similar problems also use convex optimization to fit the dynamic background while making few structural assumptions. This line of work typically considers solving a least square problem with various penalties or constraints to encourage desired structures on the fitted background, which can approximate the unknown ground truth. For instance, H-P filter (Hodrick & Prescott, 1997) imposed ℓ_2 penalty on the second-order difference to encourage a smooth background; Kim et al. (2009) considered a variant of H-P filter with an ℓ_1 regularization function to capture a piecewise linear background. Another related work (Harchaoui & Lévy-Leduc, 2010) considered change-point detection in the means using least square estimation with TV penalty; since the number of change points is unknown, the work essentially estimates a piecewise constant background. While many advances have been achieved, these existing works have not considered serial correlation together with the dynamic background.

Our proposed method is related to variable fusion (Land & Friedman, 1997) and fused lasso (Tibshirani et al., 2005). Here the unstructured, dynamic background leads to a high-dimensional problem: we have T equations and $T + p$ variables; the optimal solution is not unique. Thus, we borrow the analytical technique in analyzing high-dimensional lasso, particularly the restricted eigenvalue conditions for the design matrix (Bickel et al., 2009; Meinshausen & Yu, 2009; Van De Geer & Bühlmann, 2009) to derive the theoretical results, while further exploiting the special structure of our design matrix.

There are two closely related recent works: Xu (2008) used polynomials to approximate the dynamic background, and Zhang et al. (2020) developed an online forecasting algorithm based on least square estimation with ℓ_2 variable fusion constraint. These works do not explicitly consider highly unstructured backgrounds. We compare with both methods via numerical simulations in Section 4 and show the advantage of our approach when there are dynamic, unstructured backgrounds; moreover, we also present a method for hyperparameter selection based on the Ljung-Box test.

2. Proposed Method

2.1. Total Variation constrained least square estimation

Consider a *Total Variation constrained least square estimator* (`TV-LSE`) to estimate the AR coefficients and the dynamic backgrounds simultaneously, which is obtained by solving the following convex optimization problem:

$$\begin{array}{ll} \underset{\alpha_1 \ldots, \alpha_p, f_1, \ldots, f_T}{\text{minimize}} & \frac{1}{2T} \sum_{i=1}^{T} \left(x_i - \sum_{j=1}^{p} \alpha_j x_{i-j} - f_i \right)^2 \\ \text{subject to} & \sum_{i=1}^{T-1} |f_{i+1} - f_i| < \delta, \end{array} \tag{2}$$

where δ is a user-specified hyperparameter (the selection of δ is discussed in Section 2.2).

As discussed for the problem formulation (1), different from the conventional AR model, here we consider an *unknown* and *time-varying background*. Since the number of observa-

tions and the number of parameters both grow at the same rate as the time horizon T increases, we cannot uniquely recover the parameters using the available observations. Thus, we impose a TV constraint on the dynamic background, essentially choosing one solution with the smallest variations. Such an approach can serve as a good approximation to a broad class of unstructured, dynamic backgrounds.

2.2. Hyperparameter tuning procedure

We will show that the choice of the hyper-parameter δ in (2) will critically impact its solution (the recovered dynamic background and the AR coefficient). As illustrated in the first panel in Figure 1, setting $\delta = 0$ will result in a very simple $\text{AR}(1)$ model, but a very biased estimate $\widehat{\alpha}_1$. Clearly, this model under-fits data. On the other hand, the third panel in Figure 1 shows that when δ is too large, the fitted model will have a small empirical loss but overfitted background, which still results in very biased $\widehat{\alpha}_1$. From the second panel in Figure 1, we can see the fitted piecewise constant background faithfully captures the dynamics, and this model yields an accurate $\widehat{\alpha}_1$.

Figure 1 illustrates that δ controls the bias-variance trade-off: a larger δ leads to a smaller fitting error, but an overfitted background, and thus the estimated AR coefficients are biased. Therefore, we cannot use the fitting error to tune δ. Instead, we choose δ by the Ljung-Box test, which can test the model's goodness-of-fit by checking the remaining serial correlation in the residual sequence. This test provides a p-value to quantify the goodness-of-fit. Since larger p-value indicates less remaining serial correlation in the residuals, we select δ with the maximum p-value. Details of parameter tuning procedure can be found in Appendix A.

We want to comment that we cannot use (vanilla) cross-validation (CV) technique to choose δ. The CV splits the data into training data and testing data. Typically, the model for the training data and the test data are identical; thus, CV error can be used to estimate the actual test error. However, here since our background is dynamic and different on the test and the training data, we cannot apply model fitted on training data to the test data to tune hyperparameter.

2.3. Bootstrap confidence interval

Finally, we present two bootstrap methods to construct confidence intervals (CI) for the AR coefficients. For serially correlated data, we cannot use the conventional bootstrap for i.i.d. data (Efron, 1992); instead, we use the following techniques: (i) Wild bootstrap (Wu, 1986), which resamples from the fitted residuals and (ii) a variant (Künsch, 1989) of local moving block bootstrap, which is designed for non-stationary time series. Details are deferred to Appendix A.

3. Non-asymptotic Error Bounds

Now we present the main theoretical results, including the upper and lower bounds for the parameter ℓ_2 recovery errors using (2). Here, we deal with $\text{AR}(1)$ sequences. An extension to $\text{AR}(p)$ sequences can be found in Appendix D.

3.1. Main insights

We start with some necessary definitions. Define a sequence as being *ε-recoverable*, if the ℓ_2 recovery error using (2) is smaller than ε. The collection of recoverable sequences forms the *ε-recoverable region*. We make the following assumptions to ensure the dynamic background does not change too drastically. (i) The background contains at most s changes over the time horizon T, i.e., it consists of at most $s+1$ pieces. In other words, the rate-of-change for the dynamic background is on the order of s/T, and the change does not happen very often. (ii) The magnitude of the each change is upper-bounded by δ_0:

$$|\Delta_i| \le \delta_0, \quad i = 2, \dots, T, \tag{3}$$

where $\Delta_i = f_i - f_{i-1}$, $i = 2, \dots, T$ are one-step changes of the dynamic background.

It is known that the least square estimator is consistent for any stationary AR time-series. Intuitively, the "smaller" the dynamic background is, the "closer" the sequence is to its stationary counterpart. A fundamental question is *What ranges of dynamic backgrounds and AR coefficients can be estimated accurately?* We answer this question via Theorems 1 and 2, which establish the upper and lower bounds of the recovery error that depend on the number of changes s and the size of the change δ_0. However, in our setting, $s(T)$ is non-decreasing with respect to (w.r.t.) the time horizon T. Thus, we cannot expect the recovery error to shrink to zero with increasing T as the usual asymptotic analysis.

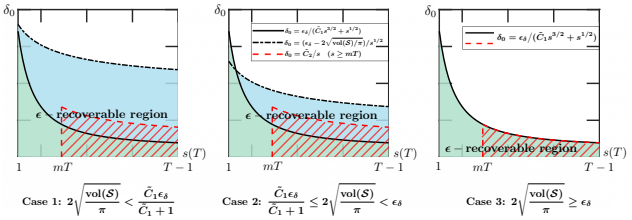


Figure 2. Illustration of ε-recoverable region for fixed T. This region shrinks when hypothesis class $\mathcal{S}$ grows larger. The expressions for the curves in case 1 are the same with case 2. The upper bound (8) will be nearly tight in the region shaded in red lines.

Theorem 1 establishes the sufficient condition to ensure the ℓ_2 recovery error does not exceed ε:

$$\min\left\{\widetilde{C}_1 s^{3/2}\delta_0, 2\sqrt{\text{vol}(\mathcal{S})/\pi}\right\} + s^{1/2}\delta_0 \le \varepsilon - \delta := \varepsilon_\delta, \tag{4}$$

where $\widetilde{C}_1$ is a positive constant, $\mathcal{S}$ is a user-specified set that the true parameter resides in and $\text{vol}(\cdot)$ denotes the volume of a set. Therefore, (4) is a sufficient condition that s and δ_0 of a ε-recoverable sequence needs to satisfy, and thus it defines the boundary for ε-recoverable region as illustrated in

Figure 2. The recoverable region is the union of the blue and the green regions in Figure 2: the green region does not vary but the blue region shrinks with increasing $\mathrm{vol}(\mathcal{S})$ and eventually vanishes once $2\sqrt{\mathrm{vol}(\mathcal{S})/\pi}$ exceeds ε_δ. This can be explained by that when the unknown coefficients reside in a larger $\mathcal{S}$, it is more difficult to recover the true parameters for the same accuracy ε, which leads to a smaller recoverable region. Moreover, as illustrated in Figure 2, in the ε-recoverable region, there exists a collection of instances (the region shaded in red lines) where the best achievable performance (lower bound) meets the upper bound of our proposed estimator over the finite time horizon.

3.2. Preliminaries

Denote the observation $x_{i:j} = (x_i, \ldots, x_j)^{\mathrm{T}}$, where the superscript $^{\mathrm{T}}$ denotes vector/matrix transpose. Given $x_{1:T}$ and known history x_0, we aim to estimate coefficient vector $\beta = (\alpha_1, \mu, \Delta_2, \ldots, \Delta_T)^{\mathrm{T}} \in \mathbb{R}^{T+1}$, where $\mu = f_1, \Delta_i = f_i - f_{i-1}, i = 2, \ldots, T$. Further denote the random noise vector by $\varepsilon_{1:T} = (\varepsilon_1, \ldots, \varepsilon_T)^{\mathrm{T}}$, and the random *design matrix* by $\mathbb{X} = (x_{0:T-1}, L) \in \mathbb{R}^{T\times(T+1)}$, where $L \in \mathbb{R}^{T\times T}$ is the lower triangular matrix with non-zeros entries all being ones. For notational simplicity, we rewrite (1) as

$$x_{1:T} = \mathbb{X}\beta + \varepsilon_{1:T}. \tag{5}$$

Denote $\Delta = (0, 0, \Delta_2, \ldots, \Delta_T)^{\mathrm{T}}$ and $\|\cdot\|_q$ to be ℓ_q vector norm. Here, we slightly abuse the notation and assume β (instead of Δ) has at most s non-zero entries. Since the TV constraint only encourages a sparse structure on Δ, Δ will have at most $s-2$ non-zero entries:

$$\mathcal{B} = \{\Delta : \|\Delta\|_0 = s-2 \in \{1, \ldots, T-1\},\ \|\Delta\|_\infty \le \delta_0\}.$$

The space for the unknown true coefficients is defined as

$$\Theta_T = \{\beta : (\alpha_1, \mu) \in \mathcal{S}', \Delta \in \mathcal{B}\}. \tag{6}$$

The hypothesis class $\mathcal{X}$, i.e. the set from which we want to estimate the coefficients, is defined as

$$\mathcal{X} = \{\beta : (\alpha_1, \mu) \in \mathcal{S},\ \|\Delta\|_1 < \delta\},$$

where

$$\mathcal{S} = \{(\alpha_1, \mu) : \alpha_1^2 + \mu^2 \le \delta_s^2\}.$$

Here, δ_s is a user-specified parameter, which specifies the size of hypothesis class $\mathcal{X}$. We assume $\mathcal{X}$ contain the ground truth, i.e., $\mathcal{S}' \subset \mathcal{S}$.

Our goal is to estimate the unknown $\beta \in \Theta_T$ by $\widehat{\beta}_T \in \mathcal{X}$ through solving the following convex optimization problem:

$$\widehat{\beta}_T = \arg\min_{\beta\in\mathcal{X}} \frac{1}{2T}\|x_{1:T} - \mathbb{X}\beta\|_2^2. \tag{7}$$

Note that this is slightly different from (2) (which only has constraint $\|\Delta\|_1 < \delta$). However, since δ_s is typically set to be sufficiently large such that the solution will not occur on the boundary of $\mathcal{S}$, (7) leads to the same estimate as (2).

3.3. Upper bound

Here, we start with a non-asymptotic upper bound for the ℓ_2 recovery error for the model coefficients and derive the condition for recoverable sequences in (4).

Theorem 1 (Upper Bound on ℓ_2 estimation error)**.** For $\widehat{\beta}_T$ defined by (7), for any $A_1 > 1$, $A_2 > \sqrt{A_1}$ and $A_3 > 0$, and for any selected hyperparameter δ, with probability at least $1 - (2T)^{1-A_1} - (2T)^{1-A_2^2/A_1} - 2(2T)^{-A_3^2/A_1^2}$, we have

$$\|\widehat{\beta}_T - \beta\|_2 \le \min\Big\{\widetilde{C}_1\sqrt{s}\max\{s\delta_0, \delta\}, 2\sqrt{\mathrm{vol}(\mathcal{S})/\pi}\Big\} + \delta + \sqrt{s}\delta_0, \tag{8}$$

where $\widetilde{C}_1 = \widetilde{C}_1(A_1, A_2, A_3)$ is a positive constant.

Note that (8) implies smaller s and δ_0 will lead to smaller error. To ensure the upper bound is less than a pre-specified $\varepsilon > 0$, we choose δ at most $\varepsilon/(1 + \widetilde{C}_1\sqrt{T})$, which leads to the condition for ε-recoverable sequences in (4).

3.4. Lower bound

Now we lower bound the ℓ_2 recovery error using triangle inequality and then improve it via Fano's method.

A naive lower bound can be derived by triangle inequality:

$$\|\widehat{\beta}_T - \beta\|_2 \ge \|\widehat{\Delta} - \Delta\|_2 \ge \|\Delta\|_2 - \|\widehat{\Delta}\|_2 \ge \|\Delta\|_2 - \delta,$$

where $\|\widehat{\Delta}\|_2 \le \delta$ due to the TV constraint. However, since we usually do not know $\|\Delta\|_2 = O(\sqrt{s}\delta_0)$ *a priori*, we cannot set δ close to $\|\Delta\|_2$ to make sure the best achievable performance. Besides, to control the worst performance (i.e. upper bound), δ is $O(\varepsilon/\sqrt{T})$. Therefore, for series $x_{1:T}$ with a large $\|\Delta\|_2$, we should expect that the recovery error is $\Theta(\sqrt{s}\delta_0)$. However, this type of series is typically outside the recoverable region. We are more interested in the lower bound for those instances with much smaller $\|\Delta\|_2$. For a certain type of series (which satisfies assumptions (9) and (10) below), we can obtain a tighter lower bound by Fano's method as follows:

Theorem 2 (Lower bound on ℓ_2 estimation error)**.** If there exist $\widetilde{C}_2 > 0$ and $0 < m < M \le 1$ such that

$$s(t) \in [mt, Mt], \quad t = 1, \ldots, T_0, \tag{9}$$

$$s(t)\delta_0(t) \le \widetilde{C}_2, \quad t = 1, \ldots, T_0, \tag{10}$$

then for any $C_6 \in (0, 1)$ and for any estimator $\widetilde{\beta}_T$, we have

$$\sup_{\beta\in\Theta_T} \mathrm{pr}\left(\|\widetilde{\beta}_T - \beta\|_2 \ge C_2\right) \ge 1 - C_6, \tag{11}$$

and

$$C_2 \ge \frac{1}{2}\exp\left\{-\frac{C_3 + C_4\widetilde{C}_2 M + C_5\widetilde{C}_2^2 M^2 + (\log 2)/T}{C_6 m}\right\}, \tag{12}$$

where C_3, C_4, C_5 are positive constants dependent on δ_s.

Under assumptions (9) and (10), the naive lower bound will be of order $O(1/\sqrt{T})$ and therefore the lower bound C_2 (constant order) will be tighter. We denote it by C_{lower}. Besides, (8) ensures the upper bound will be at most $2\sqrt{\text{vol}(\mathcal{S})/\pi}$ plus a $O(1/\sqrt{T})$ term. Thus, we can also obtain a constant order upper bound C_{upper}. In this special case, the ℓ_2 estimation error will stay within a constant order interval $\|\widehat{\beta}_T - \beta\|_2 \in [C_{\text{lower}}, C_{\text{upper}}]$ with constant probability for $t = 1, \ldots, T_0$. We illustrate this constant order interval in the region shaded in green in Figure 3.

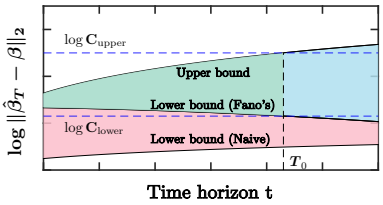


Figure 3. Illustration of the naive and Fano's lower bounds for ℓ_2 error under assumptions (9) and (10). The trajectories of ℓ_2 estimation error for those instances covered by the red lines in Figure 2 stay within the region shaded in green.

Figure 3 illustrates that the upper bound stays close to Fano's lower bound on a finite time horizon, which shows the near-optimality of `TV-LSE`. Besides, the green region in Figure 3 illustrates the constant probability estimation error trajectories for those instances within the red region in Figure 2.

3.5. Proof outline

We now sketch the proof for the main theorems. Detailed proofs for Theorems 1 and 2 as well as Propositions 1 and 2 are deferred to Appendix C.

The proof of Theorem 1 is largely based on *Restricted Eigenvalue* (RE) condition and *Variational Inequality* (VI). Consider the penalized form of (7)

$$\widehat{\beta}_T = \arg\min_{\beta \in \mathbb{R}^{T+1}} \frac{1}{2T}\|x_{1:T} - \mathbb{X}\beta\|_2^2 + \lambda\|\Delta\|_1, \qquad (13)$$

where λ is the tuning parameter. By Lagrangian duality, we can show that (13) is equivalent to (7). It is known that (Wainwright, 2019) there is a one-to-one correspondence between δ and λ: if $\widehat{\beta}_T = \widehat{\beta}_T(\lambda)$ minimizes (13), then it also minimizes (7) with $\delta = \|\widehat{\Delta}\|_1$.

Formulation (13) links our problem to high-dimensional lasso (Wainwright, 2019). This connection motivates us to invoke RE condition due to Bickel et al. (2009); Van De Geer & Bühlmann (2009) for the design matrix to bound ℓ_2 estimation error, since the RE condition is the weakest known sufficient condition according to Raskutti et al. (2010). Although there have been works verifying RE conditions for $x_{0:T-1}$ (Loh & Wainwright, 2011; Basu & Michailidis, 2015; Wu & Wu, 2016) or L (Harchaoui & Lévy-Leduc, 2010), we cannot directly apply them to our setting. Specifically, expanding $x_{0:T-1}$ with a square matrix L in $\mathbb{X}$ leads to the rank-deficiency of $\mathbb{X}^{\mathrm{T}}\mathbb{X}$, thus simply exploring the structure of $x_{0:T-1}$ cannot address the problem.

Partition the index set $\{1, \ldots, T+1\}$ into three disjoint parts $I_i (i = 1, 2, 3)$, where $I_1 = \{1, 2\}$, I_2 is the indices for non-zero Δ_i's and I_3 is the indices for zeros in β. By using an index set I as the subscript of a vector, we keep all entries with indices from I intact and zero out entries with indices from its complement I^{c}. Since the ℓ_1 constraint does not encourage sparsity on α_1 and μ, we modify the definition of the REs as follows

$$\phi_{\min}(u) = \min_{e \in R_1} \frac{\|\mathbb{X}e\|_2}{\sqrt{T}\|e\|_2}, \quad \phi_{\max}(u) = \max_{e \in R_2} \frac{\|\mathbb{X}e\|_2}{\sqrt{T}\|e\|_2}, \qquad (14)$$

where $R_1 = \{e : 2 = \|e_{I_1}\|_0 \le \|e\|_0 \le u\}$ and $R_2 = \{e : 1 \le \|e\|_0 \le u,\ \|e_{I_1}\|_0 = 0\}$.

Remark 1. The smallest RE can be understood as follows: take columns of $\mathbb{X}$ indexed by 1, 2 and another $u-2$ indices from ${I_1}^{\mathsf{c}}$ to form a new matrix $\widetilde{\mathbb{X}}$, then $\phi_{\min}(u)$ is the smallest among eigenvalues of all possible $\widetilde{\mathbb{X}}^{\mathrm{T}}\widetilde{\mathbb{X}}/T$'s. Similarly, $\phi_{\max}(u)$ is the largest eigenvalue of $\widetilde{\mathbb{X}}^{\mathrm{T}}\widetilde{\mathbb{X}}/T$, where $\widetilde{\mathbb{X}}$ is composed of u columns of $\mathbb{X}$ with indices chosen from ${I_1}^{\mathsf{c}}$.

In the following analysis, we apply a recently developed technique based on variational inequality (Juditsky & Nemirovski, 2019; Juditsky et al., 2020) to establish the upper bound. Consider the gradient field of the objective in (7):

$$F_{x_{1:T}}(z) = \big(A[x_{1:T}]\, z - \ a[x_{1:T}]\big)/T,$$

where $a[x_{1:T}] = \mathbb{X}^{\mathrm{T}}(\sum_{i=1}^{T} x_i x_{i-1}, x_{1:T}^{\mathrm{T}})^{\mathrm{T}}$ and $A[x_{1:T}] = \mathbb{X}^{\mathrm{T}}\mathbb{X}$. This vector filed is affine and monotone, since we can verify the symmetric matrix $A[x_{1:T}]$ is positive semi-definite. The minimizer of (7), i.e. $\widehat{\beta}_T$, is in fact the solution to the following VI:

$$\text{find } z \in \mathcal{X} : \langle F_{x_{1:T}}(w), w - z\rangle \ge 0,\ \forall w \in \mathcal{X}.$$

Moreover, the ground truth β is zero of $\widetilde{F}_{x_{1:T}}(z)$ and solution to the following VI:

$$\text{find } z \in \mathcal{X} : \langle \widetilde{F}_{x_{1:T}}(w), w - z\rangle \ge 0,\ \forall w \in \mathcal{X},$$

where $\widetilde{F}_{x_{1:T}}(z) = \big(A[x_{1:T}]\, z - \ A[x_{1:T}]\, \beta\big)/T$. We can see that $F_{x_{1:T}}(z)$ and $\widetilde{F}_{x_{1:T}}(z)$ only differ in the following constant term:

$$\eta = F_{x_{1:T}}(\beta) - \widetilde{F}_{x_{1:T}}(\beta) = (A[x_{1:T}]\, \beta - \ a[x_{1:T}])/T.$$

Intuitively, the difference between the above two VIs should reflect the difference between the solutions to those two variational inequalities, i.e. our estimator $\widehat{\beta}_T$ and the ground truth β. We will show how to bound the ℓ_2 estimation error using $\|\eta\|_\infty$ in the following theorem.

The following proposition establishes the error bound for the AR coefficient and the initial background, combined.

Proposition 1 (Upper Bound on ℓ_2 estimation error for α_1 and μ)**.** For $\widehat{\beta}_T$ defined by (7), for A_1, A_2, A_3 and tuning parameter δ in Theorem 1, there exists $k \in (0,1)$ such that $k\lambda = O\left((\log T)^{3/2}/T^{1/2}\right)$, with probability at least $1-(2T)^{1-A_1}-(2T)^{1-A_2^2/A_1}-2(2T)^{-A_3^2/A_1^2}$, we have

$$\sqrt{(\widehat{\alpha}_1-\alpha_1)^2+(\widehat{\mu}-\mu)^2} \leq \max\left\{\frac{4\sqrt{2}}{\kappa^2}\left(\frac{\|\eta\|_\infty}{1-k}+\kappa C_{\delta,\delta_0,s}\right), \delta+s\delta_0\right\}, \quad (15)$$

where δ_0 is the magnitude for one-step changes of dynamic background in (3) and

$$\kappa = \sqrt{\phi_{\min}(2)} - \frac{k\sqrt{2(T-1)}}{(1-k)\sqrt{T}},$$
$$C_{\delta,\delta_0,s} = \frac{2s\delta_0\sqrt{T-1}}{(1-k)\sqrt{T}} + \sqrt{\frac{(s-2)(T-s+1)}{T}}(s\delta_0+\delta).$$

Here, $\phi_{\min}(\cdot)$ is defined in (14). Moreover, we have

$$\|\eta\|_\infty \leq C_0(\log T)^{3/2}/T^{1/2},$$

where $C_0 = C_0(A_1, A_2, A_3)$ is a positive constant.

Next, we establish lower bound using Fano's method.

Proposition 2 (Lower bound on ℓ_2 estimation error)**.** For any estimator $\widetilde{\beta}_T$ and constant $C_2 > 0$, we have

$$\sup_{\beta\in\Theta_T} \text{pr}\left(\|\widetilde{\beta}_T-\beta\|_2 \geq C_2\right) \geq 1- \frac{C_3T + C_4\delta_0(T)\sum_{t=2}^T s(t) + C_5\delta_0^2(T)\sum_{t=2}^T s^2(t) + \log 2}{s(T)\log(1/2C_2)},$$

where C_3, C_4, C_5 are positive constants dependent on δ_s.

We can show that Theorem 2 follows from the above proposition. The key steps in proving this above proposition are to (i) find a large enough ε-packing of Θ_T and (ii) upper bound the KL divergence over this packing.

4. Numerical Experiments

In this section, we perform comprehensive numerical simulations to (i) show that `TV-LSE` works well in practice; (ii) validate our theoretical findings regarding algorithm performance; (iii) compare with existing methods; (iv) demonstrate the good performance of the two proposed bootstrap methods. Recall that our work's primary focus is to estimate the AR coefficients. Thus, we will focus on this in the following four experiments.

Experiment 1. First, we show that our proposed estimation method can accurately recover α_1 from non-stationary AR(1) time series under various settings: $\alpha_1 \in \{0.05, 0.1\}$, $\sigma_0^2 \in \{0.1, 0.2\}$ $\delta_0 \in \{0.05, 0.1\}$ and $T = 5000$. The dynamic background is generated by $f_i = \sum_{k=1}^i \delta_0(U_k - 0.5), i = 1,\ldots,T$, where $U_k \in [0,1], k = 1,\ldots,T$ is a sequence of i.i.d. uniform random numbers. As discussed above, the accuracy depends on both s and δ_0. Here, we consider an extreme case: $s = T+1$ (which is supposed to be the most challenging case). Moreover, we also present the results with δ selected by `DW` test (Durbin & Watson, 1992) as an alternative. (Details on the `DW` test can be found in Section B.1 in the Appendix B.) The convex program (2) is solved by the `cvx` package (Grant & Boyd, 2014) and we tune the hyperparameter δ by Ljung-Box (`LB`) test and Durbin-Watson (`DW`) test, respectively. We repeat the experiment 20 times for each setting, and plot the mean square error (MSE) of $\widehat{\alpha}_1$, p-value of `LB` and `DW` test with varying δ's in Figure 4.

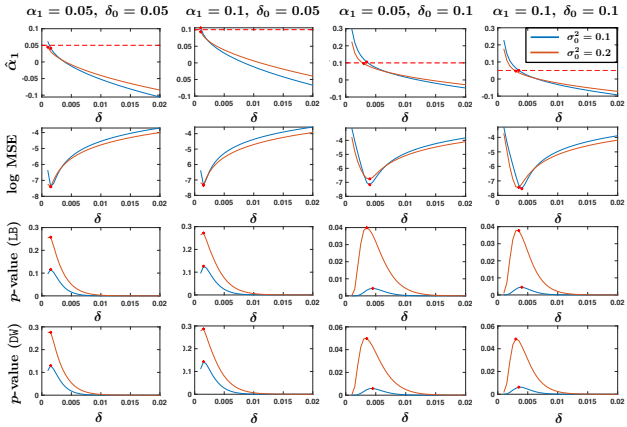
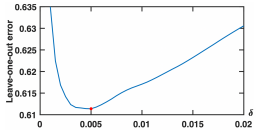


Figure 4. Performance of `TV-LSE` when δ increases. The experimental setting is on the top of each column. The red dashed line denotes the ground truth $\alpha_1 = 0.1$. Note that the δ selected by our proposed tuning procedure (which leads to the maximum p-value of `LB` or `DW` test) gives the best estimate $\widehat{\alpha}_1$.

The results in Figure 4 show that $\widehat{\alpha}_1$ decreases when δ increases, and there is a specific value of δ that leads to the smallest MSE for estimating α_1. In the figure, the red dots in the first two rows correspond to the best achievable δ's in MSE. In the last two rows, those red dots indicate the δ's selected by our proposed tuning procedure. The exact values of those red dots are summarized in Table 1 in Appendix E. The results demonstrate that (i) the best achievable δ's regarding the accuracy and MSE are roughly the same; (ii) our proposed tuning procedure based on both the `LB` test and `DW` test can select the best δ.

Remark 2 (Cross-validation for parameter tuning)**.** We now performed a type of leave-one-out cross-validation (CV). Note that we cannot use vanilla CV because of the presence of the dynamic, unstructured background and serial correlation, as explained above. Instead, in each trial, we leave out one observation and treat it as missing-value (we have to deal with it anyways using "imputation" in real-data).

We then fit the model and obtain one CV error using the predicted value for the left-out observation. We repeat these T times, where T is the length of the sequence, to obtain an estimate of the CV error. This will enable us to perform the CV to estimate a good choice of the hyperparameter δ, which has the smallest CV error.



From our results above, the optimal δ chosen this way (the red dot) is very close to (though not as good as) the optimal δ we had in the third column in Figure 4.

Experiment 2. Next, we validate our theoretical findings for AR(1) case. The dynamic background is generated in the same way as the previous example. Besides, Figure 4 shows that the p-value w.r.t. δ is unimodal, which enables us to use the Golden-section search (tolerance $\varepsilon = 0.04$) to tune δ efficiently. Details on the Golden-section search and this modified tuning procedure can be found in Appendix B.2. We also show how the estimate $\widehat{\alpha}_1$ behaves with changing s, by setting $\alpha_1 = 0.1, \sigma_0^2 = 0.1$, $\delta_0 = 0.1$ and repeating the same estimation procedure 20 times for each $s \in \{5, 35, \ldots, 3005\}$. The mean and standard deviation of $\widehat{\alpha}_1$ over 20 trials w.r.t. s in an errorbar plot are plotted Figure 5.

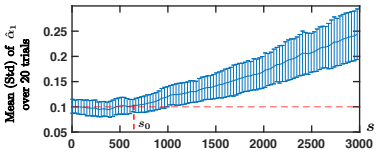


Figure 5. Algorithmic behavior w.r.t. s. The red dashed horizontal line is the ground truth $\alpha_1 = 0.1$. The estimate starts to deteriorate when s exceeds $s_0 \approx 650$ and becomes worse with larger s.

The results in Figure 5 show that indeed when s and δ_0 are inside the recoverable region, the estimation error is small, and it will grow with an increasing s. Moreover, the error remains small for relatively small s, but once s exceeds s_0 the error starts to increase; that's when the non-stationary series is not in the ε-recoverable region. This observation agrees with our non-asymptotic bounds on estimation error. Moreover, we conduct similar experiments for AR(2) case to validate these findings for a more general case; the results can be found in Appendix E.

Experiment 3. We compare `TV-LSE` with Zhang et al. (2020). In the following, we refer to their method as the "ℓ_2 variant", since it is obtained by solving the convex program with same objective function as (2) except for a different constraint: $\sum_{i=1}^{T-1}(f_{i+1} - f_i)^2 < \delta$. Again, $\delta \geq 0$ is the tuning parameter. We should mention Zhang et al. (2020) did not have a systematic way to tune δ and here we enhance their method by adding our statistical test based hyperparameter tuning as well.

The piecewise linear dynamic background is generated by $f_i = \sum_{k=1}^{i} \delta_0(U_k - 0.5), i = 1, \ldots, T$. Here $U_k = u_i$ for all $k \in \{k_i, \ldots, k_{i+1}\}, i = 0, \ldots, s-1$, where $0 = k_0 < k_1 < \cdots < k_s = T$, $k_1, \ldots, k_{s-1}$ are randomly selected from $\{1, \ldots, T-1\}$ and $u_i \in [0, 1], i = 0, \ldots, s-1$, is a sequence of i.i.d. uniform random numbers. We consider two settings: (1) $s = 1500$, $\delta_0 = 0.05$, $\|\Delta\|_1 = 25.2$; (2) $s = 100$, $\delta_0 = 0.1$, $\|\Delta\|_1 = 46.9$. Here, s denotes the number of changes in the slope; the one step difference vector is not sparse.

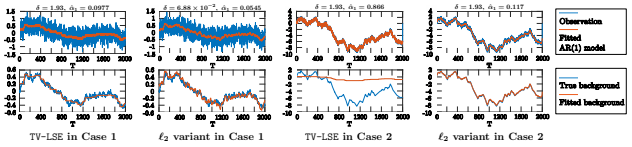


Figure 6. Comparison of proposed method and the ℓ_2 variant in Zhang et al. (2020). We investigated two cases. In Case 1, the dynamic background oscillates more but a with smaller magnitude, whereas in Case 2, the dynamic background is smoother but has larger one-step difference. Our piecewise constant fitted background better captures the dynamics in Case 1, whereas the ℓ_2 variant can better approximate the dynamics in Case 2.

Figure 6 shows that in Case 1, `TV-LSE` yields a very accurate $\widehat{\alpha}_1$, even though the dynamic background drastically oscillates. This is because the one-step changes are small in magnitude, and therefore, a constant can still serve as a good approximation within some short time window, i.e., this type of sequence is still within the recoverable region. The ℓ_2 variant yields a biased estimate for α_1, which is probably the reason that Zhang et al. (2020) focus on relatively smooth and structured dynamics.

In Case 2, even though the dynamic background is smoother than the previous example, the dynamic background changes drastically (large $\|\Delta\|_1$). Thus, in this case, the piecewise constant function is a poorer approximation to the dynamic background. This type of sequence is outside the recoverable region, and `TV-LSE` may not work well for those sequences. Nevertheless, the ℓ_2 variant, together with our proposed hyperparameter tuning procedure, performs well in recovering the serial dependence and serves as an alternative to our proposed estimator. This result agrees with Zhang et al. (2020), where they demonstrated the good performance of this ℓ_2 variant on relatively structured dynamics, since ℓ_2 constraint can lead to a smooth background. In addition, we should mention a polynomial approximation method used in (Xu, 2008) does not perform well in fitting unstructured dynamics and cannot compete with these two aforementioned non-parametric methods. Comparison with this polynomial method can be found at Appendix E.

Experiment 4. Finally, we compare the CIs obtained via two

bootstrap methods. We adopt the following experimental setting: $\alpha_1 = 0.1, \sigma_0^2 = 0.1, T = 1000$. The dynamic drift is piecewise constant with $\delta_0 = 0.1$, $s = 100$. The bootstrap replication is $N = 100$; we use standard normal random numbers as v_t's in residual-based wild bootstrap (`WB`); for local block bootstrap (`LBB`), we choose block size $b = 20$ and local neighborhood length $B = 50$. We illustrate one replication result by plotting the histogram of $\widehat{\alpha}_1$'s from bootstrap samples in Figure 7.

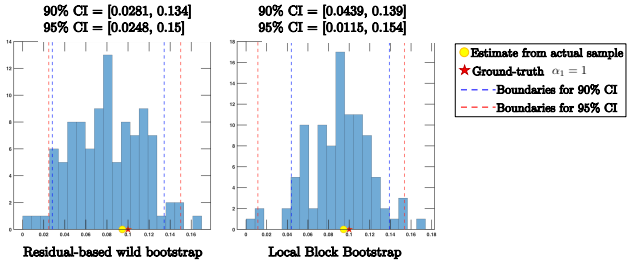


Figure 7. Histogram of $\widehat{\alpha}_1$'s from `WB` samples (left) and `LBB` samples (right). We can see that the `LBB` method yields a shorter CI but lower coverage accuracy.

From 50 repetitions of the above procedure, we find that: (i) the coverage accuracy of 90% and 95% CIs: 0.84 and 0.90 for `WB` and 0.84 and 0.88 for `LBB`; (ii) the average lengths of 90% and 95% CIs: 0.10 and 0.12 for `WB` and 0.095 and 0.114 for `LBB`. The coverage accuracy is slightly lower than the theoretical value since $T = 1000$ is relatively small. The comparison indicates that `LBB` tends to yield shorter CIs but has slightly lower coverage accuracy.

5. Real-data Study

To validate its performance, we apply `TV-LSE` to real data from a psychological experiment. Consider a reaction time (RT) dataset collected from human subjects. The data are taken from a publicly available database introduced by Rahnev et al. (2020) with 149 individual datasets with human data on different tasks. Here we only analyze a single dataset named `Maniscalco_2017_expt1` chosen based on the fact that it has RT data included and features a large number of trials per subject (Maniscalco et al., 2017). Here we use the data from an experiment where 28 human subjects made a series of 1000 perceptual judgments over a period of about one hour. Participants were seated in front of a computer and made their responses using a standard keyboard. The task, which is standard in the field, consisted of deciding whether a briefly presented (33 ms) noisy sinusoidal grating was oriented clockwise or counterclockwise from vertical. Subjects responded as quickly as possible but without sacrificing accuracy. The experimenters recorded each judgment's RT (that is, the time from the onset of the visual stimulus to the button press used to indicate the subject's response), thus creating a time series of 1000 values for each subject.

We first pre-process the raw data by dealing with missing values and obvious outliers. To be precise, we treat RTs that exceed 10 times the interquartile range (i.e., the difference between 75th and 25th percentiles) as outliers and the rest as normal observations. Since naively omitting missing data in time series data will break the serial correlation, we use the median of the normal observations to impute those missing values. The same median is used to replace all outliers. We propose a data-adaptive procedure to tune δ by applying the `LB` test on the logarithm of original residuals since they are strongly right-skewed. We plot the results for all 28 subjects in Figure 17. Due to space limitation, we put Figure 17 as well as details on why we choose logarithm transform in Appendix E. Here, we present three representative subjects in Figure 8. The CIs are constructed via bootstrapping with the same bootstrapping parameters in our simulation.

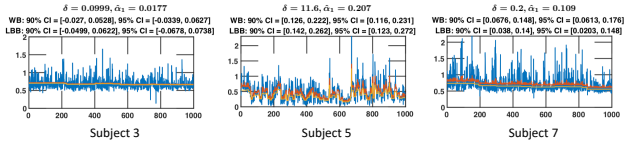


Figure 8. Experimental results on reaction times for subjects 3 (left), 5 (middle) and 7 (right). On the top of each figure, we report hyperparameter δ, the estimate $\widehat{\alpha}_1$ and its CI.

Overall, we can observe that `TV-LSE` can faithfully capture the underlying dynamics. More specifically, we make four distinct observations:

(1) It is clear that there is a substantial background drift. Even in the raw data without any modeling, the drift can often be observed but is even more apparent after applying `TV-LSE` of recovering it. Further, the drift does not have big one-step changes, which is exactly the type of dynamical drift that `TV-LSE` can capture well.

(2) The background drift has a complex shape and varies significantly from person to person. As illustrated in Figure 8, while for some subjects, the RT series appears to be monotonically decreasing (e.g., subjects 4, 6, 7, 8, 11, 12, 15, 16, 26, and 28) or even close to stationary (e.g., subjects 3 and 19), the remaining subjects exhibit complex trends without any obvious pattern. These differences between subjects demonstrate that the trends need to be identified on the individual time series level and cannot make strong structural assumptions about the dynamic drift. Instead, to be able to capture real data, the dynamic background has to be modeled with minimal structural assumptions.

(3) Our method of fitting the background drift recovers reasonable estimate of the AR coefficient α_1. Specifically, α_1 is positive or close to 0 for most subjects, which is expected given the extensive previous literature on RT (Laming, 1968). Nevertheless, `TV-LSE` recovers a negative α_1 for subject 18, which could indicate that the RT series is not universally positively autocorrelated as assumed before and suggests the need for more detailed investigations on this

issue. Further, TV−LSE appears to provide a good fit for the empirically observed RT data across individual subjects, and the size of the hyperparameter δ tends to be larger for time series that visually appear to be less stationary. Thus, TV−LSE recovers both α_1 and δ well, and provides a useful description of the data dynamics.

(4) Our method provides a substantial improvement over the AR(1) model that is typically used to recover the AR coefficient in psychology and neuroscience. As shown in Figure 9, the AR(1) model leads to very high and clearly inflated estimates of α_1 because the model confuses the dynamical drift for an autocorrelation. Overall, TV−LSE performs very well on real data from experiments where it is likely to be applied in the future and is a major advance over the standard AR(1) model.

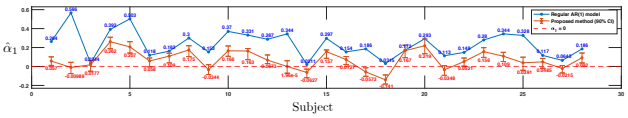


Figure 9. Comparison between TV−LSE and regular AR(1) model for all 28 subjects. The errorbar is 90% CI computed using the LBB method. Regular AR(1) model typically yields larger $\widehat{\alpha}_1$, since it confuses the dynamic drifts as autoregressive effect.

Remark 3. One of the key features of our proposed method is simultaneous estimation of unstructured background and AR coefficients, and it is different from piecewise linear method to estimate the background. Nevertheless, we compare our method with the piecewise linear method (the R package) by first fixing piecewise linear functions to the sequence as a linear trend and then subtracting the estimated trend to estimate the coefficient via least-square estimate (we also tune the parameter to optimize the performance). We call this method "two-step" method. The result is shown in Figure 10, and we note that this approach does not lead to a good result. Clearly, the fitted background does not appropriately capture the slowly time-varying nature of the "background" and instead results in an implausible "overly zig-zag" pattern (e.g., subject 5). As a result, the estimated autoregressive coefficient turns to be much larger than we had before and seems to be too large to be reasonable.

Remark 4. We should note that there are breaks between consecutive 100 observations (Maniscalco et al., 2017). Considering that the observations separated by breaks may not necessarily have serial correlation, we now try a different approach when fitting the model. We break the data into 10 segments (each of length 100). For each subject, we fix α_1 (but not background f_i's) and then solve for a new objective function:

$$\begin{aligned} \min \quad & \tfrac{1}{20T}\textstyle\sum_{k=1}^{10}\sum_{i=1}^{T}(x_i^{(k)} - \alpha_1 x_{i-1}^{(k)} - f_i^{(k)})^2 \\ \text{s.t.} \quad & \textstyle\sum_{k=1}^{10}\sum_{i=1}^{T-1}|f_{i+1}^{(k)} - f_i^{(k)}| < \delta, \end{aligned}$$

where $T = 100$ and k is the segment index. As reported in Figure 11, the result is very similar to what we had before in Figure 8 (in terms of estimated α_1 and background). This is reasonable since essentially we just removed the variation penalty for the nine pairs of observations that straddle different blocks.

Apart from this, we also tried a different imputation approach by linearly interpolating the missing values using two adjacent observations (rather than using a global median). The comparison between previous results and new results is reported in Figure 12 below. We can see they are very similar.

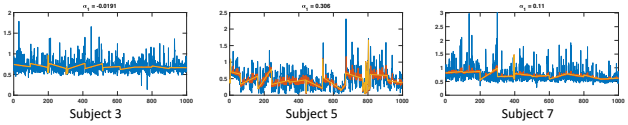


Figure 10. Two-step method: (1) Linear segmentation to estimate trend and (2) LSE to estimate AR coefficient.

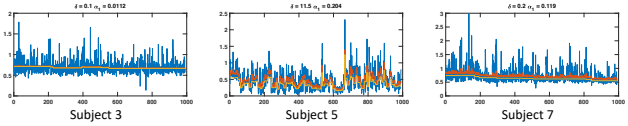


Figure 11. Simultaneous method: TV−LSE coupled with linear interpolation on 10 segments/time series.

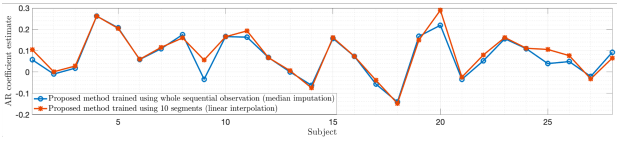


Figure 12. Comparison between previous (blue) and new (red) results using different fitting procedures for all 28 subjects.

6. Summary

In this work, we propose a least square estimation with TV constraint on the dynamic background. This method approximates the dynamic background via a piecewise constant one and therefore can deal with a wide range of highly unstructured dynamics. Two main features of our proposed method are (i) highly unstructured dynamic background and (ii) simultaneous estimation of dynamic background and AR coefficients. We should remark here the dynamic background is very different from Gaussian noise; it captures unpredictable events, such as unexpected distractions or fluctuations in fatigue. Such variations can be both very large and vary substantially across subjects, as can be seen in the subject-level data above. Besides, we should also clarify that the goal is not prediction but accurately inferring the AR coefficients. Given that the background is assumed to be unstructured, predicting future values is indeed impossible — who knows what will be next distraction? Nevertheless, given the estimation of AR coefficients, we can still predict the next observation "minus" the background. Last but not least, we consider low-dimensional case since the data in our targeted application is low dimensional. However, it is important and interesting to subsequently extend this work to a high dimensional setting.

References

Akrami, A., Kopec, C. D., Diamond, M. E., and Brody, C. D. Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature*, 554(7692):368–372, feb 2018. doi: 10.1038/nature25510. URL http://www.nature.com/doifinder/10.1038/nature25510.

Basu, S. and Michailidis, G. Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43(4):1535–1567, 2015.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009.

Brockwell, P. J., Davis, R. A., and Fienberg, S. E. *Time series: theory and methods: theory and methods*. Springer Science & Business Media, 1991.

Cicchini, G. M., Mikellidou, K., and Burr, D. C. The functional role of serial dependence. *Proceedings of the Royal Society B: Biological Sciences*, 285 (1890):20181722, nov 2018. ISSN 0962-8452. doi: 10.1098/rspb.2018.1722. URL http://www.ncbi.nlm.nih.gov/pubmed/30381379https://royalsocietypublishing.org/doi/10.1098/rspb.2018.1722.

Clark, P. K. The cyclical component of us economic activity. *The Quarterly Journal of Economics*, 102(4):797–814, 1987.

Durbin, J. and Watson, G. S. Testing for serial correlation in least squares regression. I. In *Breakthroughs in Statistics*, pp. 237–259. Springer, 1992.

Dutilh, G., Van Ravenzwaaij, D., Nieuwenhuis, S., Van der Maas, H. L., Forstmann, B. U., and Wagenmakers, E. J. How to measure post-error slowing: A confound and a simple solution. *Journal of Mathematical Psychology*, 56(3):208–216, 2012. ISSN 00222496. doi: 10.1016/j.jmp.2012.04.001. URL http://dx.doi.org/10.1016/j.jmp.2012.04.001.

Efron, B. Bootstrap methods: another look at the jackknife. In *Breakthroughs in Statistics*, pp. 569–593. Springer, 1992.

Fischer, J. and Whitney, D. Serial dependence in visual perception. *Nature Neuroscience*, 17(5):738–43, mar 2014. ISSN 1097-6256. doi: 10.1038/nn.3689. URL http://dx.doi.org/10.1038/nn.3689.

Grant, M. and Boyd, S. CVX: Matlab software for disciplined convex programming, version 2.1, 2014.

Hamilton, J. D. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pp. 357–384, 1989.

Harchaoui, Z. and Lévy-Leduc, C. Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association*, 105(492):1480–1493, 2010.

Hodrick, R. J. and Prescott, E. C. Postwar US business cycles: an empirical investigation. *Journal of Money, Credit, and Banking*, pp. 1–16, 1997.

Hong, Y. Serial correlation and serial dependence. In *Macroeconometrics and Time Series Analysis*, pp. 227–244. Springer, 2010.

Juditsky, A., Nemirovski, A., Xie, L., and Xie, Y. Convex recovery of marked spatio-temporal point processes. *arXiv preprint arXiv:2003.12935*, 2020.

Juditsky, A. B. and Nemirovski, A. Signal recovery by stochastic optimization. *Automation and Remote Control*, 80(10):1878–1893, 2019.

Kim, S.-J., Koh, K., Boyd, S., and Gorinevsky, D. ℓ_1 trend filtering. *SIAM review*, 51(2):339–360, 2009.

Künsch, H. R. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3): 1217–1241, 1989.

Laming, D. R. J. *Information theory of choice-reaction times*. Academic Press, New York, 1968.

Land, S. R. and Friedman, J. H. Variable fusion: A new adaptive signal regression method. Technical report, Department of Statistics, Carnegie Mellon University, 1997.

Ljung, G. M. and Box, G. E. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

Loh, P.-L. and Wainwright, M. J. High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. In *Advances in Neural Information Processing Systems*, pp. 2726–2734, 2011.

Maniscalco, B., McCurdy, L. Y., Odegaard, B., and Lau, H. Limited cognitive resources explain a trade-off between perceptual and metacognitive vigilance. *Journal of neuroscience*, 37(5):1213–1224, 2017.

McIlhagga, W. Serial correlations and 1/f power spectra in visual search reaction times. *Journal of Vision*, 8(9):5–5, jul 2008. ISSN 1534-7362. doi: 10.1167/8.9.5. URL http://jov.arvojournals.org/Article.aspx?doi=10.1167/8.9.5.

Meinshausen, N. and Yu, B. Lasso-type recovery of sparse representations for high-dimensional data. *The Annals of Statistics*, 37(1):246–270, 2009.

Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. Time series momentum. *Journal of Financial Economics*, 104(2):228–250, may 2012. ISSN 0304405X. doi: 10.1016/j.jfineco.2011.11.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X11002613.

Paparoditis, E. and Politis, D. N. Local block bootstrap. *Comptes Rendus Mathematique*, 335(11):959–962, 2002.

Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M., and Lau, H. Confidence Leak in Perceptual Decision Making. *Psychological Science*, 26(11):1664–1680, 2015. ISSN 0956-7976. doi: 10.1177/0956797615595037. URL http://pss.sagepub.com/lookup/doi/10.1177/0956797615595037.

Rahnev, D., Desender, K., Lee, A. L. F., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova, P., Atlas, L. Y., Balcı, F., Bang, J. W., Bègue, I., Birney, D. P., Brady, T. F., Calder-Travis, J., Chetverikov, A., Clark, T. K., Davranche, K., Denison, R. N., Dildine, T. C., Double, K. S., Duyan, Y. A., Faivre, N., Fallow, K., Filevich, E., Gajdos, T., Gallagher, R. M., de Gardelle, V., Gherman, S., Haddara, N., Hainguerlot, M., Hsu, T.-Y., Hu, X., Iturrate, I., Jaquiery, M., Kantner, J., Koculak, M., Konishi, M., Koß, C., Kvam, P. D., Kwok, S. C., Lebreton, M., Lempert, K. M., Ming Lo, C., Luo, L., Maniscalco, B., Martin, A., Massoni, S., Matthews, J., Mazancieux, A., Merfeld, D. M., O'Hora, D., Palser, E. R., Paulewicz, B., Pereira, M., Peters, C., Philiastides, M. G., Pfuhl, G., Prieto, F., Rausch, M., Recht, S., Reyes, G., Rouault, M., Sackur, J., Sadeghi, S., Samaha, J., Seow, T. X. F., Shekhar, M., Sherman, M. T., Siedlecka, M., Skóra, Z., Song, C., Soto, D., Sun, S., van Boxtel, J. J. A., Wang, S., Weidemann, C. T., Weindel, G., Wierzchoń, M., Xu, X., Ye, Q., Yeon, J., Zou, F., and Zylberberg, A. The confidence database. *Nature Human Behaviour*, 4(3):317–325, mar 2020. ISSN 2397-3374. doi: 10.1038/s41562-019-0813-1. URL http://www.nature.com/articles/s41562-019-0813-1.

Raskutti, G., Wainwright, M. J., and Yu, B. Restricted eigenvalue properties for correlated gaussian designs. *The Journal of Machine Learning Research*, 11:2241–2259, 2010.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 67(1):91–108, 2005.

Van De Geer, S. A. and Bühlmann, P. On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics*, 3:1360–1392, 2009.

Wainwright, M. J. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

Wexler, M., Duyck, M., and Mamassian, P. Persistent states in vision break universality and time invariance. *Proceedings of the National Academy of Sciences*, 112(48):14990–5, nov 2015. ISSN 0027-8424. doi: 10.1073/pnas.1508847112. URL http://www.ncbi.nlm.nih.gov/pubmed/26627250.

Wu, C.-F. J. Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, 14(4):1261–1295, 1986.

Wu, W.-B. and Wu, Y. N. Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electronic Journal of Statistics*, 10(1):352–379, 2016.

Xu, K.-L. Bootstrapping autoregression under non-stationary volatility. *The Econometrics Journal*, 11(1):1–26, 2008.

Zhang, K., Ng, C. T., and Na, M. H. Real time prediction of irregular periodic time series data. *Journal of Forecasting*, 39(3):501–511, 2020.