Journal of Machine Learning Research 23 (2022) 1-29 Submitted 6/22; Revised 12/22; Published 2/23

FLIP: A Utility Preserving Privacy Mechanism for Time Series

Tucker McElroy TUCKER.S.MCELROY@CENSUS.GOV
Research and Methodology Directorate, U.S. Census Bureau
4600 Silver Hill Road,Washington, D.C. 20233-9100, USA

Anindya Roy ANINDYA@UMBC.EDU
U.S. Census Bureau
Department of Mathematics and Statistics
University of Maryland, Baltimore County
1000 Hilltop Cir, Baltimore, MD 21250

Gaurab Hore GAURABH1@UMBC.EDU
Department of Mathematics and Statistics
University of Maryland, Baltimore County
1000 Hilltop Cir, Baltimore, MD 21250

Editor: Christian Shelton

Abstract

Guaranteeing privacy in released data is an important goal for data-producing agencies. There has been extensive research on developing suitable privacy mechanisms in recent years. Particularly notable is the idea of noise addition with the guarantee of differential privacy. There are, however, concerns about compromising data utility when very stringent privacy mechanisms are applied. Such compromises can be quite stark in correlated data, such as time series data. Adding white noise to a stochastic process may significantly change the correlation structure, a facet of the process that is essential to optimal prediction. We propose the use of all-pass filtering as a privacy mechanism for regularly sampled time series data, showing that this procedure preserves certain types of utility while also providing sufficient privacy guarantees to entity-level time series. Numerical studies explore the practical performance of the new method, and an empirical application to labor force data show the method's favorable utility properties in comparison to other competing privacy mechanisms.

Keywords: all-pass filter; privacy measure; privacy-utility trade-off; spectral density; stationary time series

1. Introduction

Privacy protection in data disclosure has a long history, first being formalized by the U.S. Privacy Act of 1974. Since that time, there have been numerous developments in the control of disclosure risk while processing data. While there are many different formulations of randomized privacy mechanisms, the majority of the algorithms have been developed for independent data. However, given the surging demand for granular published time series in different economic sectors, statistical agencies face the need for proper disclosure avoidance algorithms that are tailored specifically to time series published over a fixed frequency, such as monthly or quarterly.

In Abowd et al. (2012), the authors propose a disclosure avoidance mechanism that relies on the use of noise infusion through a permanent multiplicative noise distortion factor, used for magnitudes, counts, differences, and ratios. That paper is one of the few works on privacy that investigate the perturbation of time series properties that result from application of the privacy mechanism. One of the concerns with the conventional implementation of privacy mechanisms for dependent data, such as a time series, is that the dependence structure may be significantly altered. In most applications with dependent data, the ability to estimate the dependence structure is a key facet of data utilization; for example, optimal forecasting of a time series depends upon knowing the autocorrelation structure. Thus, current privacy mechanisms result in potentially degraded data utility for time series data.

Differential privacy (DP) is the most popular privacy mechanism, developed in a series of papers; see Dwork (2006), Dwork et al. (2006), Dwork and Roth (2014), for example. The concept of differential privacy ensures that the removal or addition of a single database item does not (substantially) affect the outcome of any analysis, thereby letting individual sensitive records avoid identification in a given database. Differential privacy is algorithmically simple, and allows anything learnable via statistical queries to be learned “differentially privately.” The DP framework greatly expands the applicability of output perturbation, a technique for protecting individuals’ privacy by adding a small amount of random noise to the released data. DP may not be appropriate if multiple examples correspond to the same individual; a modified framework that resolves some of the hurdles in traditional privacy mechanisms is the Pufferfish framework (Kifer and Machanvajjhala, 2011).

There is a body of work that uses DP algorithms for dependent and other structured data by using data transformations that make the data amenable to DP algorithms. These methods apply DP in the transformed domain and convert the privatized data back to the time domain. Rastogi and Nath (2010) propose a privacy mechanism called PASTE. The mechanism uses the Fourier perturbation algorithm (FPA), which combines the discrete Fourier transform (DFT) with DP to support time series of count queries while ensuring good utility – in the sense that aggregate queries are accurately reflected by the privatized time series – and non-disclosure of individual data. Ács et al. (2012) propose an optimization of the FPA that allows the release of histograms, where the global sensitivity is 1. They show through experimental evaluation that their scheme improves the ‘query’ utility of the initial FPA by a factor 10. Lyu et al. (2017) applies FPA to a time series on consumption and proposes the wavelet perturbation algorithm (WPA), replacing the DFT with the discrete wavelet transform (DWT). The authors show through experimental results that WPA ensures better ‘query’ utility than FPA. A recent work on differential privacy in time series, Lako et al. (2021), proposes a “clamping Fourier perturbation algorithm” (CFPA). The CFPA uses a so-called clamping mechanism for reducing the sensitivity, thereby reducing the noise introduced in FPA. Lako et al. (2021) also proposes the “clamping wavelet perturbation algorithm” (CWPA), a similar adaptation of WPA.

Some privacy mechanisms for dependent data propose using modifications of the DP mechanism adapted for correlated data. Use of the Pufferfish framework for correlated data is suggested in Song et al. (2017). The authors introduce a generalization of DP, using the Wasserstein mechanism, which applies to any general Pufferfish framework. In Song and

Chaudhuri (2017), the authors propose a modified Pufferfish mechanism, called the Markov Quilt mechanism, and illustrated its advantages compared to conventional DP when applied to time series data. Some papers use cryptographic techniques to infuse privacy in given databases; see Shi et al. (2011a).

There is also a large literature on privacy-preserving data mining in time series; see the survey in Hong et al. (2013). These mechanisms generally rely on perturbation methods such as noise addition, compression-based perturbation, and geometric transformation perturbation. Often the mechanisms in this area discuss privacy under data aggregation. Mechanisms that reduce the risk of privacy leakage incurred by recurring components of an aggregate query, such as sums and counts of the distributed time series are preferred; Shi et al. (2011b) suggests that on receiving the query Q, a user evaluates Q on their own time series, perturbs the result, and sends the perturbed results back to the aggregator. The aggregator combines the perturbed results from all users to produce the final result. They allow both users and the aggregator to be malicious, with the flexibility that at least a fraction of users is honest.

Erdogdu et al. (2015) consider real-time privacy mechanisms where a user would like to continuously release time series data that is correlated with their own private data. Erdogdu et al. (2015) propose a sequential scheme to achieve privacy, and validate the method on synthetic and real-time series data. Their framework minimizes the average distance between the actual and distorted series, while bounding the leakage of the data.

Some privacy mechanisms are based on ideas from the cryptography literature. In Lu et al. (2012) authors recommend an efficient privacy-preserving aggregation method based on the homomorphic Paillier cryptosystem technique, which uses a super-increasing sequence to structure multi-dimensional data. The Paillier cryptosystem can achieve homomorphic properties, widely desirable in many privacy-preserving applications (Sang et al., 2009). There have been other attempts at defining privacy mechanisms for time series data such as OPTSTREAM (Fioretto and Hentenryck (2019)) and VAULT (Stach (2019)). These privacy schemes protect either a single timestamp (event-level), or all the data per user (user-level), or per window (w-event-level) in the time series, considering all timestamps as equally significant. Katsomallos et al. (2022) define a configurable privacy notion, landmark privacy, which differentiates events into significant (landmarks) and regular, achieving better data utility while adequately preserving the privacy of each event. In Stinner (2017) the author describes a disclosure control model based largely on Bayesian decision theory, and defines utility and risk in the Bayesian framework. Fixing an upper bound of the risk, Stinner (2017) tries to maximize the utility.

This body of work does not, in our opinion, properly address the particular privacy-utility concerns for time series data. One of the earlier statistical papers that examine the privacy-utility optimization framework is Wasserman and Zhou (2009); the authors forcefully argue for maintaining data utility while implementing disclosure avoidance algorithms. We concur with that sentiment. However, there is a dearth of privacy mechanisms for regularly observed time series data that maintain data utility – since both privacy and utility must take properties of the temporal dynamics (such as serial correlation) into account. Therefore we propose a new privacy framework (called Linear Incremental Privacy) – as well as new utility conditions, called 'second-order utility' – for regularly sampled time series. (Our methods presume that the data is regularly sampled.) Our second contribution is a novel mechanism for balancing the horns of the privacy-utility dilemma via the device of all-pass filtering. This new framework is implemented and assessed through simulations; the utility properties of competing privacy mechanisms are compared to our proposed method on labor force data.

2. A Privacy-Utility Framework for Time Series

We propose a framework for making perturbations of a sensitive time series, before its release, such that the perturbations strike a balance between the conflicting objectives of retaining data utility and providing privacy protection. The main methodology involves random phase-changed versions of the observed time series, which leave the second-order properties of the series unchanged while providing measurable privacy protection.

We will assume that the attacker has some prior information about the series we are trying to protect. Let $\{\tilde{X}_t\}$ be the series that is sensitive, and which requires protection, and let $\{\tilde{Z}_t\}$ be various auxiliary time series comprising the knowledge that an advanced attack can use to predict the observed series. We will assume the following observation model:

$$\begin{pmatrix} \tilde{X}_t \\ \tilde{Z}_t \end{pmatrix} = \begin{pmatrix} \mu_t^X \\ \mu_t^Z \end{pmatrix} + \begin{pmatrix} X_t \\ Z_t \end{pmatrix}, \tag{1}$$

where $\{X_t, Z_t\}$ are assumed to be jointly stationary with spectral density matrix

$$f_{X,Z}(\lambda) = \begin{pmatrix} f_X(\lambda) & f_{XZ}(\lambda) \\ f_{ZX}(\lambda) & f_Z(\lambda) \end{pmatrix}, \tag{2}$$

and $\{\mu_t^X, \mu_t^Z\}$ are deterministic trend components that can be represented by lower order polynomials in time. We assume that the spectral matrix $f_{X,Z}$ is known to the data-publishing agency, and is also available to an *augury* attack. (We use the term *augury* to denote a scenario that is ideal for the attacker, involving a degree of outside information.) We will first develop our proposal – which is formulated in the frequency domain – for a privacy-utility framework using only the stationary part $\{X_t, Z_t\}$, and then show how to modify the proposal to incorporate deterministic trend factors $\{\mu_t^X, \mu_t^Z\}$.

Before proceeding, we define some of the notations and conventions that will be used throughout. For time series, the braces notation $\{X_t\}$ will denote the entire series, while X_t will denote the value of the series at time t. The associated Roman letter, say X, without the time subscript will denote the data vector. Thus, $\{Z_t\}$ denotes the underlying stationary time series representing the auxiliary information, whereas $Z = (Z_1, \ldots, Z_T)'$ will denote the vector of the attacker's knowledge over the observation period $1, 2, \ldots, T$. Also $z = e^{-i\lambda}$ will denote a unit norm complex variate, where $-\pi < \lambda \leq \pi$ will denote a frequency. For integrals with respect to spectral density f of a function u, we set $\langle u, f \rangle = (2\pi)^{-1} \int_{-\pi}^{\pi} u(\lambda) f(\lambda) d\lambda$ and $\langle u, f \rangle_\pi = \pi^{-1} \int_0^\pi u(\lambda) f(\lambda) d\lambda$. Also, when $u(\lambda) = 1$, we simply denote the integrals as $\langle f \rangle$ and $\langle f \rangle_\pi$, respectively. For a spectral density f, $\tilde{f}$ will denote the normalized density $\tilde{f}(\lambda) = f(\lambda)/\langle f \rangle_\pi$, and thus $\langle \tilde{f} \rangle_\pi = 1$.

2.1 Post-privatization Utility Via All-pass Filtering

With every implementation of a disclosure avoidance mechanism, one also has to consider the utility of privatized data – in order to avoid nonsensical results. The utility of privatized data is achieved by maintaining analytical validity for answers to standard queries about the data. For a time series $\{X_t\}$, most common queries will be related to its spectral density f_X, or the autocovariance function (ACVF) of the series, $\gamma_X(h) = \mathrm{Cov}(X_t, X_{t+h}) = \langle z^h, f_X \rangle$. Queries about time series may include features of the marginal distribution such as skewness and kurtosis or the number of modes, or more detailed spectral information measured via higher-order autocumulants. However, the primary utility for regularly spaced time series usually comes from means and autocovariances, which are the basis of forecasting formulas; in this article we define the time series utility objective in terms of the mean and the autocovariances. Preservation of utility after privatization would therefore require that distortion of the ACVF (or spectral density) is minimized. The overall shift in utility can be assessed by a discrepancy between the ACVFs—or the autocorrelation functions (ACF), denoted $\rho_X(h) = \gamma_X(h)\gamma_X(0)^{-1}$—of the original time series $\{X_t\}$ and the published time series $\{\hat{X}_t\}$. Formally, one could measure so-called 'second-order' utility by using a normalized degradation measure

$$U(\{X_t\}, \{\hat{X}_t\}) = 1 - \mathcal{D}(f_X, f_{\hat{X}}),$$

where $\mathcal{D}$ is a discrepancy function taking values in $[0, 1]$, with $\mathcal{D}(f_X, f_{\hat{X}}) = 0$ indicating no loss of second-order utility from privatization. Note that such a degradation measure depends only on the second order properties (i.e., the spectral density) of the original and published series, and does not depend on the sample path.

Remark 1 (Polyspectra) *It is possible to consider higher-order utility via degradation measures associated with polyspectra, but these facets of a stochastic process are less vital than the spectral density, and will not be further pursued here.*

Remark 2 (Noise addition and utility) *As we have seen in Joye and Libert. (2013), Abowd et al. (2012), Shi et al. (2011a), most disclosure avoidance methods for time series have recommended noise infusion (or some related framework). While in the usual independent scenario noise addition only affects the variance of the entity, in time series it attenuates the entire ACF, thereby significantly compromising second-order utility. Consider univariate time series for simplicity, and let $\hat{X}_t = X_t + N_t$ represent the published series, where $\{N_t\}$ is a time series of noise infusion. Since $\{N_t\}$ is generated independently of the stochastic process $\{X_t\}$, we have $\gamma_{\hat{X}}(h) = \gamma_X(h) + \gamma_N(h)$. Moreover, since the infused noise is i.i.d., we have $\gamma_N(h) = 0$ for $|h| > 0$. Therefore, for any lag $|h| > 0$*

$$|\rho_{\hat{X}}(h)| = \left| \frac{\gamma_{\hat{X}}(h)}{\gamma_{\hat{X}}(0)} \right| = \left| \frac{\gamma_X(h)}{\gamma_X(0) + \gamma_N(0)} \right| < C\, |\rho_X(h)|,$$

where $0 < C = SNR/(1 + SNR) < 1$, and $SNR = \gamma_X(0)/\gamma_N(0) > 0$ is the signal-to-noise ratio. Hence, for all $|h| > 0$ we have attenuation $|\rho_{\hat{X}}(h)| < |\rho_X(h)|$. The amount of attenuation is directly related to SNR, and could be substantial even for a moderately large signal-to-noise ratio. Hence, for time series where essential queries depend on the entire ACF, inference based on the released series $\{\hat{X}_t\}$ could be very different from that based on the sensitive series $\{X_t\}$, thereby severely degrading second-order utility.

It follows from Remark 2 that second-order utility is imperilled by noise infusion, no matter what the marginal structure is. Therefore, we seek a disclosure avoidance framework that maintains second-order utility by altering the ACVF as little as possible. We propose using convolution instead of addition; specifically, we propose a special type of linear time-invariant filtering as the primary operation for perturbing a time series.

Definition 1 (All-pass filter) *Let $\{X_t\}$ be a stationary multivariate time series with spectral density matrix f_X. Let B denote the backshift operator (McElroy and Politis, 2020), and let $\hat{X}_t = \Psi(B)X_t = \sum_k \psi_k X_{j-k}$ be a filtered version of $\{X_t\}$, where $\Psi(B) = \sum_k \psi_k B^k$ is a linear time invariant filter. Then the filter Ψ is all-pass if its frequency response function is unitary, i.e., $\Psi(z)\Psi(z^{-1})' = I$, where I is the identity matrix, and $z = e^{-i\lambda}$ for any $\lambda \in [-\pi, \pi]$.*

Because the spectral density matrix of $\{\hat{X}_t\}$ is given by $f_{\hat{X}}(\lambda) = \Psi(z)\, f_X(\lambda)\, \Psi(z^{-1})'$, it follows that if Ψ is all-pass then $\mathrm{tr} f_{\hat{X}}(z) = \mathrm{tr} f_X(\lambda)$ (where tr denotes the trace). At present we are focused on the univariate case where $|\Psi(z)| = 1$, and hence $f_{\hat{X}}(\lambda) = f_X(\lambda)$, i.e., $\mathcal{D}(f_X, f_{\hat{X}}) = 0$. Thus, perturbation of a univariate time series using an all-pass filter gives us complete second-order utility: $U(X, \hat{X}) = 1$. For the univariate case, we can write the all-pass filter using the *cepstral* representation (McElroy and Politis, 2020) as

$$\Psi(z) = \exp(\phi(z)), \quad \phi(z) = \sum_{k \in \mathrm{Z}} \phi_k z^k. \tag{3}$$

The condition that Ψ has to be unitary then implies that $\phi(z)$ is an anti-symmetric Laurent series, i.e., $\phi_k = -\phi_{-k}$ and $\Psi(e^{-i\lambda}) = \exp(-ig(\lambda))$, where $g(\lambda) = 2\sum_{k\geq 1} \phi_k \sin(\lambda k)$. It follows that an all-pass filter can be thought of as a pure phase filter. In a simple all-pass filter, such as $\Psi(z) = z^a$, the filter forces a constant lag shift (by a time units) for the time series. In a more general all-pass filter the phase change is frequency-dependent, and hence the effect of the filter on the time series is much more nuanced; this can be assessed through the phase delay function of the filter – see McElroy and Politis (2020).

In summary, filtering a univariate time series with a general all-pass filter provides a framework for perturbing the data that leaves the spectral density (which encodes the second order characteristics) unaltered; whereas use of the all-pass filter preserves second-order utility, the extent of privacy protection afforded depends upon the phase $g(\lambda)$ of the filter, as well as the stochastic properties of $\{X_t\}$. In what follows we will refer to the all-pass filter Ψ used to perturb as a *privacy mechanism* for $\{X_t\}$.

2.2 Linear Incremental Privacy (LIP)

Noise infusion via addition is common in privacy mechanisms, because the privacy measures are typically defined in terms of the probability of change in individual data records. For time series, measures of privacy that target the marginal distribution of the random variables are not particularly useful; recall that optimal linear forecasts depend upon a process' serial correlation structure, and the marginal distribution plays no direct role (aside from its mean and variance). Put another way, successful attacks can employ not only cross-sectional information – which is the focus in formulations such as DP – but also temporal information, and this latter aspect has received little attention in terms of privacy criteria.

Moreover, algorithms that optimize marginal criteria may well distort the joint dependence structure of a time series, yielding perturbed data that no longer have the correct temporal dynamics (cf. Remark 2) – or worse, have fallacious and spurious dynamics introduced.

This reality motivates our proposal for a new, more suitable privacy measure for a time series; the proposal is based on an incremental notion of privacy, in terms of how prediction accuracy – based on an adversarial information set – is altered by a given privacy mechanism. The prediction formula is moment-based, and hence our definition is moment-based as well. Concisely, suppose the attacker already has some limited capacity to divine (i.e., predict) sensitive information, and we wish to measurably degrade that capability. Since optimal prediction in statistics is classically formulated in terms of the conditional expectation (this follows from using a mean squared error loss; see McElroy and Politis (2020)), we also adopt this paradigm in our treatment below. We begin by considering arbitrary protection frameworks that generate a publishable $\{\hat{X}_t\}$ from sensitive $\{X_t\}$, and then secondly consider filtering mechanisms; finally, we specialize to the case of all-pass filters.

For the next few paragraphs consider a simplified scenario where X, $\hat{X}$, and Z are random vectors. The best estimate — in the sense of mean squared error (MSE) loss — of X given the attacker's information is the conditional expectation, written $E[X|Z]$. If $\hat{X}$ is published, then an updated attack using the additional information in $\hat{X}$ (over and above what the attacker already knew) is $E[X|\hat{X},Z]$. This will be called an *augury* attack if the joint distribution of X, $\hat{X}$, and Z is known. (In practice, such a joint distribution might not be known, making the attacker's task more difficult; hence the augury scenario is worst-case from the standpoint of privacy.) In the case of linear estimators (which are appropriate if the random vectors are Gaussian, and might more generally be employed due to their simplicity if the data is non-Gaussian),

$$E[X|\hat{X},Z] = E[X|Z] + \mathrm{Cov}[X,\hat{X}|Z]\,\mathrm{Var}[\hat{X}|Z]^{-1}\,(\hat{X} - E[\hat{X}|Z]). \tag{4}$$

The second term on the right represents the update to the attack, made available by the publication of $\hat{X}$. We say that $\hat{X}$ is *private* if this update is zero for all variables Z; publication of $\hat{X}$ has not assisted the attacker to predict X. We distinguish between *augury private* and *feasibly private* – the latter is a notion contingent on various possible models of the joint distributions. Computing the MSE, we obtain

$$\mathrm{Var}[X|Z] - \mathrm{Var}[X|\hat{X},Z] = \mathrm{Cov}[X,\hat{X}|Z]\,\mathrm{Var}[\hat{X}|Z]^{-1}\,\mathrm{Cov}[\hat{X},X|Z].$$

The left hand side expresses the conditional variances before and after publication of $\hat{X}$, and the difference is a measure of additional, or incremental, vulnerability for the sensitive data. On the right hand side of the equation we have a non-negative definite matrix – or a non-negative scalar when X is univariate. The quantity equals zero when $\hat{X}$ offers no assistance to the attack. This suggests defining the following quantity as the *privacy measure*:

$$\mathcal{P}(X,\hat{X},Z) = 1 - \frac{\det \mathrm{Cov}[X,\hat{X}|Z]\,\mathrm{Var}[\hat{X}|Z]^{-1}\,\mathrm{Cov}[\hat{X},X|Z]}{\det \mathrm{Var}[X|Z]}.$$

Clearly this privacy measure is not well-defined if $\det \mathrm{Var}[X|Z] = 0$, which corresponds to the rather trivial case in which the attacker already knows the sensitive information –

in such a case privacy is impossible. Otherwise, the privacy measure can be viewed as one minus a multivariate squared conditional correlation, corresponding to the R^2 quantity familiar from linear models.

We now specialize these notions to the case of a univariate time series; we can remove the determinant, but now the conditioning set is the whole time series $\{Z_t\}$:

$$\mathcal{P}(\{X_t\}, \{\hat{X}_t\}, \{Z_t\}) = 1 - \frac{\mathrm{Cov}(X_t, \hat{X}_t | \{Z_t\})^2}{\mathrm{Var}(\hat{X}_t | \{Z_t\}) \mathrm{Var}(X_t | \{Z_t\})}. \tag{5}$$

Although the conditional covariance and conditional variance terms in (5) involve the stochastic process at time t, the privacy measure is not time-dependent because of our standing assumption that $\{X_t\}$ and $\{Z_t\}$ are jointly stationary – this assumption entails that these conditional quantities are the same for all t.

So far our privacy measure is agnostic about the mechanism used to produce $\hat{X}$. For example, we could use (5) to assess noise addition; in this case, it follows from the discussion in Remark 2 that $E[\hat{X}_t | \{Z_t\}] = E[X_t | \{Z_t\}]$ (assuming that $\{N_t\}$ and $\{Z_t\}$ are independent, which is natural), and hence $\{\hat{X}_t\}$ is private. However, due to the second-order utility issues with noise addition, we are instead more interested in studying linear filtering mechanisms Ψ; the following result provides a simpler expression for the privacy measure in such a case.

Proposition 1 *Let $\{X_t, Z_t\}$ be jointly stationary with spectral matrix* (2)*. Then the error process $X_t - E[X_t | \{Z_t\}]$ (where we have used an optimal linear predictor for the conditional expectation symbol) is stationary with mean zero and spectral density*

$$f_{X|Z}(\lambda) = f_X(\lambda) - \frac{f_{XZ}(\lambda) f_{ZX}(\lambda)}{f_Z(\lambda)}. \tag{6}$$

Suppose a linear filtering privacy mechanism is employed, i.e., $\hat{X}_t = \Psi(B) X_t$ for a linear filter Ψ. If $\langle f_{X|Z} \rangle > 0$, then

$$\mathcal{P}(\{X_t\}, \{\hat{X}_t\}, \{Z_t\}) = 1 - \frac{\langle \Psi, f_{X|Z} \rangle^2}{\langle \Psi \overline{\Psi}, f_{X|Z} \rangle \, \langle f_{X|Z} \rangle},$$

where $\overline{\Psi(e^{-i\lambda})} = \Psi(e^{i\lambda})$.

The condition in Proposition 1 that $\langle f_{X|Z} \rangle > 0$ is equivalent to saying that X_t cannot be perfectly predicted from $\{Z_t\}$; since $\det f_{X,Z}(\lambda) = f_Z(\lambda) \, f_{X|Z}(\lambda)$, we see that the condition fails if and only if $f_{X,Z}(\lambda)$ is singular for λ in a set of non-zero Lebesgue measure. It is therefore harmless to debar this case, since it corresponds to the attacker already possessing a perfect capacity to predict the sensitive data.

Since the privacy measure $\mathcal{P}(\{X_t\}, \{\hat{X}_t\}, \{Z_t\})$ is based on protecting the residual information in $\{X_t\}$ after its linear prediction using the attacker's information $\{Z_t\}$, we will call the measure **Linear Incremental Privacy**, and denote it by $\mathrm{LIP}(\Psi, f_{X|Z})$. More formally, we have the following definition.

Definition 2 (LIP) *Let $\{X_t, Z_t\}$ be jointly stationary with spectral density $f_{X,Z}(\lambda)$ (2), and let the residual spectral density be $f_{X|Z}(\lambda)$ as defined in* (6). *Then the Linear Incremental Privacy (LIP) of $\{X_t\}$ given $\{Z_t\}$ with respect to the linear filtering mechanism Ψ is defined as*

$$LIP(\Psi, f_{X|Z}) = 1 - \frac{\langle \Psi, f_{X|Z} \rangle^2}{\langle \Psi\overline{\Psi}, f_{X|Z} \rangle \, \langle f_{X|Z} \rangle}.$$

Remark 3 (LIP with an all-pass filter) *In the special case that Ψ is an all-pass filter, LIP has a simplified form. Using the property of an all-pass filter Ψ that $\Psi\overline{\Psi} = 1$, we obtain*

$$LIP(\Psi, f_{X|Z}) = 1 - \frac{\langle \Psi, f_{X|Z} \rangle^2}{\langle f_{X|Z} \rangle^2}. \tag{7}$$

2.3 $\delta-$LIP: a Privacy-Utility Framework

We describe a framework for building a privacy mechanism Ψ with desirable privacy and utility properties. In the augury solution, any all-pass filter Ψ would guarantee perfect second-order utility. Hence Ψ should be chosen to meet the minimum privacy requirement. If there are further considerations, then the class of all-pass filters meeting the privacy requirements can be optimized to ensure the desired properties.

Given that all augury mechanisms Ψ are associated with perfect second-order utility, an immediate approach for selecting an optimum Ψ would be to maximize the privacy measure $\mathrm{LIP}(\Psi, f)$:

$$\Psi_{opt} = \operatorname*{arg\,max}_{\Psi : |\Psi(z)|=1} \mathrm{LIP}(\Psi, f).$$

The optimum solution, if it exists, is usually unique and hence poses a concern in the context of data protection. If the attacker happens to know f, then one would be able to compute the filter coefficients for Ψ_{opt} (since it uniquely depends on f) and invert the computation to get the original values of the sensitive series $\{X_t\}$. It is surprising that for any given residual spectral density $f = f_{X|Z}$ the class of all-pass filters Ψ giving a perfect value of one for the privacy measure $\mathrm{LIP}(\Psi, f)$ is non-empty and can be parametrized by a function class. Hence, the class of optimum privacy solutions Ψ could be randomly sampled from the function class, thereby providing reasonable protection. The following result shows the existence of the solution class.

Theorem 1 *Let $\{X_t, Z_t\}$ be jointly stationary with residual spectral density $f = f_{X|Z}$, defined in* (6). *Let $\hat{X}_t = \Psi(B)X_t$ be the released series using a privacy mechanism Ψ. Define the class of solutions for a perfect value of the privacy measure for a spectral density f as*

$$C_\Psi(f) = \{\Psi(z) = \sum_k \psi_k z^k : |\Psi(z)| = 1, LIP(\Psi, f) = 1\}. \tag{8}$$

For any residual spectral density f, the class $C_\Psi(f)$ is non-empty and contains all mechanisms of the form $\Psi(e^{-i\lambda}) = \exp\{i\pi\tilde{R}(F(\lambda))\}$ where the function R belongs to the class

$$\mathcal{R} = \{R : [0,1] \to [0,1], R(0) = 0, R(x) + R(1-x) = 1, \forall\, x \in [0,1]\}, \tag{9}$$

$\tilde{R}(\lambda) = sgn(\lambda)R(sgn(\lambda)\lambda)$, and $F(\lambda) = \int_0^\lambda \tilde{f}(\omega)d\omega$ with $\tilde{f}(\lambda) = f(\lambda)/\langle f\rangle_\pi$ denoting the normalized spectral density.

The optimum solutions form a large class due to the possible choices of the R function. The functions in the class (9) need not be monotone; a simple class of examples is provided by the cumulative distribution function (CDF) of any symmetric Beta random variable, with parameters a and a (Beta(a, a)) for some positive real number a. More generally one could choose R equal to a mixture of Beta CDFs, e.g.,

$$R(x) = \sum_{j=1}^{J} \frac{\alpha_j}{2} [\text{Beta}(x|a_j, b_j) + \text{Beta}(x|b_j, a_j)],$$

where $0 < \alpha_j < 1, \sum_{j=1}^J \alpha_j = 1$ and $a_1, \ldots, a_J, b_1, \ldots, b_J$ are positive real numbers. To meet additional privacy-utility objectives, more constraints can be put on R, thereby restricting the class (9). Later we will assume that R is differentiable, and will put constraints on the Lipschitz constants and the derivatives.

One could make a randomized choice for the function R over the function class $\mathcal{R}_f$ for each conditional spectral density $f = f_{X|Z}$. However, it is not clear how much variation such a randomized choice would induce in the optimum filter. If the filters are relatively less sensitive to the choice of R, then an attacker with direct knowledge of $f_{X|Z}$ would be able to approximate the filter using any optimum choice corresponding to f.

Following the privacy literature, we propose a privacy budget δ to enrich the class of possible mechanisms. Specifically, for a given spectral density f we consider mechanisms Ψ such that LIP$(\Psi, f) \geq 1 - \delta$ for some $0 \leq \delta < 1$. Thus, we have the following definition for desirable privacy mechanisms when the privacy budget is δ.

Definition 3 ($\delta-$LIP) *A privacy mechanism Ψ is $\delta-$LIP for a given spectral density f and some predetermined privacy budget $0 \leq \delta < 1$ if LIP$(\Psi, f) \geq 1 - \delta$.*

The following result provides a sufficient condition for a privacy mechanism Ψ to be $\delta-$LIP.

Theorem 2 *Let $\{X_t, Z_t\}$ be jointly stationary with residual spectral density $f = f_{X|Z}$ defined in (6). Let h be a spectral density, and suppose the privacy mechanism Ψ_h belongs to the class $\mathcal{C}_\Psi(h)$ defined in (8). Specifically, let $\Psi_h(e^{-i\lambda}) = \exp\{i\pi\tilde{R}(H(\lambda))\}$ where $\tilde{R}(\lambda) = sgn(\lambda)R(|\lambda|)$, $R \in \mathcal{R}$ defined in (9), and $H(\lambda) = \int_0^\lambda \tilde{h}(\omega)d\omega$ is the CDF associated with the normalized spectral density $\tilde{h}(\lambda) = h(\lambda)/\langle h\rangle_\pi$. In addition, assume $R \in \mathcal{R}_L$ is Lipschitz with Lipschitz constant L_R, i.e., R belongs to the subclass*

$$\mathcal{R}_L = \{R \in \mathcal{R} : |R(x) - R(y)| \leq L_R|x - y|, \forall x, y \in [0, 1], L_R > 0\}. \tag{10}$$

For $0 \leq \delta < 1$, let

$$\mathcal{F}_R(f, \delta) = \left\{ h : [0, \pi] \to [0, \infty), \langle h\rangle_\pi = \langle f\rangle_\pi, \sup_{0\leq\lambda\leq\pi} |h(\lambda) - f(\lambda)| \leq \frac{\sqrt{\delta}\langle f\rangle_\pi}{L_R\pi^2} \right\}. \tag{11}$$

Then the privacy mechanism Ψ_h is $\delta-$LIP if $h \in \mathcal{F}_R(f, \delta)$.

A randomized mechanism provides greater protection. The class $\mathcal{F}_R(f_{X|Z}, \delta)$ given in Theorem 2 is a function class, and along with the choice of the R function in the definition of the optimum privacy mechanism provides a sufficiently rich class for randomization of the privacy mechanism. Consider a probability measure P_R supported on the class $\mathcal{R}_L$; given $R \sim P_R$, let $P_{h|R}$ be a conditional probability on $\mathcal{F}_R(h, \delta)$. Then a randomized choice of the privacy mechanism would be a randomly sampled value of $\Psi(e^{-i\lambda}) = \exp\{i\pi\tilde{R}(H(\lambda))\}$, where $\tilde{R}(\lambda) = sgn(\lambda)R(sgn(\lambda)\lambda)$ and $(R, h) \sim P_R \times P_{h|R}$.

Remark 4 *While the class over which the spectral density h used in the construction of the privacy mechanism can be sampled is broader than $\mathcal{F}_R(f, \delta)$, it is important to note that regardless of which function is chosen, the privacy mechanism Ψ_h is still an all-pass filter and will provide full analytical validity.*

As mentioned above, the privacy mechanism will not be useful unless one can guarantee the filter coefficients for the optimum all-pass filter are different for different choices of R and h, particularly when h is chosen in the neighborhood $\mathcal{F}_R(f, \delta)$ of the true spectral density f. The following result shows that even for a subclass of the possible choices of (R, h), the variation in the all-pass filter coefficients can be substantial.

Theorem 3 *Suppose $\{X_t, Z_t\}$ are jointly stationary with residual spectral density $f = f_{X|Z}$ defined in* (6)*. Assume that the released data is based on a privacy mechanism $\Psi_h(e^{-i\lambda}) = \exp\{i\pi R(H(\lambda))\}$ where the function R is as in* (10)*, satisfying*

$$\sup_{0\le\lambda\le\pi} |\pi\tilde{f}(\lambda) - 1| > \frac{\sqrt{\delta}}{L_R\pi} \tag{12}$$

for some $\delta > 0$, and h is of the form $h(\lambda) = A\,(\tilde{f}(\lambda) + \Delta)$, where $A = \langle f\rangle_\pi/(1 + \pi\Delta)$ and $H(\lambda) = \int_0^\lambda \tilde{h}(t)dt$. Then Ψ_h is $\delta-$LIP if $\Delta \in (0, B]$ for

$$B = \sqrt{\delta}\left(L_R\pi^2 \sup_{0\le\lambda\le\pi} |\pi\tilde{f}(\lambda) - 1| - \pi\sqrt{\delta}\right)^{-1}. \tag{13}$$

Moreover if R has a derivative r such that $r(x) > 0$ for all $x \in (0, 1)$ and $f(\lambda) > 0$ for $0 \le \lambda \le \pi$, then

$$|\Psi_h(e^{-i\lambda}) - \Psi_f(e^{-i\lambda})| \ge \frac{\alpha\pi}{2} Q(\lambda)|F(\lambda) - \lambda/\pi|, \tag{14}$$

where $Q(\lambda) = \min_{x\in L(\lambda)} r(x)$ with $L(\lambda)$ denoting the closed line segment joining $F(\lambda)$ and $(1-\alpha)F(\lambda) + \alpha(\lambda/\pi)$, for any $0 < \lambda < \pi$ and $\alpha = \frac{\Delta\pi}{1+\Delta\pi}$.

If the condition (12) is violated, that would mean that the spectral density f is nearly constant, and so is any shifted version h. In that case, the privacy mechanism should be based on spectral densities h which are in the neighborhood of h but not a constant shift.

If $\Delta \approx 0$, the lower bound on the right-hand side of (14) is close to zero. Therefore, when the sampled density is close to the true density, the pointwise distance in the filter is potentially minimal. If $F(\lambda) \approx \lambda/\pi$, i.e., the true spectral density is essentially constant, then the normalized density $\tilde{h}$ under the constant shift model is approximately equal to $\tilde{f}$.

In this case also there is not much variation in the filter away from the optimum choice. However, in general, the lower bound in (14) shows that the pointwise difference between the frequency response functions of the constructed filter and that of the optimum filter under the true spectral density is bounded away from zero over a large frequency band, thereby providing a sufficient modification to the optimum filter.

Since $f > 0$, for each $\lambda > 0$ the line segment $L(\lambda)$ is a compact sub-interval of the unit interval, and hence $Q(\lambda) > 0$. However, this bound may be extremely small, making the perturbation potentially small. The worst case scenario will be that the derivative r is nearly a point mass at $x = 0.5$, in which case R is nearly a constant almost everywhere. Since the user is free to choose the R function, situations where R is nearly a perfect sigmoidal function with a steep rise at $x = 0.5$ can be avoided. The assumption of a positive derivative of R is not necessary, but is sufficient. In the choices that are recommended in this article, such as symmetric mixtures of the beta CDF, R is a CDF on the unit interval with fully supported density, and hence the assumption that r is positive is easily met. When R is a CDF corresponding to a beta mixture of beta densities symmetric about 0.5, then the density will be monotone on $L(\lambda)$, and we can replace $Q(\lambda)$ by $r(x_\lambda)$ where $x_\lambda = \min\{F(\lambda), (1-\alpha)F(\lambda) + \alpha(\lambda/\pi)\}$.

For computational convenience, the constant shift class $h = \tilde{f} + \Delta$ provides suitable spectral densities h that could be used to construct the $\delta-$LIP privacy filter Ψ_h. To sample such a function conditional on f, one could simply sample the constant Δ from a Uniform density. Specifically, if

$$\Delta \sim \text{Uniform}\,[0, B]\,,$$

then Ψ_h corresponding to $h = \tilde{f} + \Delta$ is a $\delta-$LIP mechanism.

2.4 $\delta-$LIP Mechanism Under Nonstationary Trend Factors

Thus far we have developed the methodology where the sensitive series $\{X_t\}$ and the auxiliary information series $\{Z_t\}$ are jointly stationary. However, in practice, the data is likely to have nonstationary features. While general stochastic trend models are popular, for a given data span the trends can often be approximated by lower order polynomials in time. In the following, we extend the proposed methodology to the deterministic trend model (1), where μ_t^X and μ_t^Z are lower-order polynomials. Note that this framework covers stationary time series with non-zero means. Thus, when the time series is stationary with a constant mean, the proposed filter-based mechanism will leave the mean unchanged.

If a linear filter Ψ leaves a dth order polynomial unchanged, then we will say Ψ is a *dth order trend-invariant* filter. If the time series is trend-stationary (i.e., when the trend is removed, we will be left with a stationary series), we can put constraints on $\phi(z)$ – and hence $\Psi(z)$ – to ensure a *trend-invariant* privacy-utility framework (i.e., preserving lower order polynomial trends along with the autocovariance/autocorrelation function) using a $\delta-$LIP mechanism.

Let $\{X_t\}$ satisfy $(1-B)X_t = U_t$, where $\{U_t\}$ is a stationary series with spectral representation

$$U_t = \int_{-\pi}^{\pi} e^{-it\lambda} d\mathcal{Z}(\lambda),$$

and $\mathcal{Z}(\lambda)$ is an orthogonal increment process (see McElroy and Politis (2020)). Setting $z = e^{-i\lambda}$, the spectral representation of $\{X_t\}$ for $t \geq 0$ is

$$\begin{aligned} X_t &= X_0 + \sum_{j=1}^{t} U_j = X_0 + \sum_{j=1}^{t} \int_{-\pi}^{\pi} z^j d\mathcal{Z}(\lambda) \\ &= X_0 + \int_{-\pi}^{\pi} \sum_{j=1}^{t} z^j d\mathcal{Z}(\lambda) = X_0 + \int_{-\pi}^{\pi} z \frac{z^t - 1}{z - 1} d\mathcal{Z}(\lambda). \end{aligned}$$

Suppose $\hat{X}_t = \Psi(B) X_t = \sum_k \psi_k X_{t-k}$. Then

$$\begin{aligned} \hat{X}_t &= \sum_k \psi_k \left(X_0 + \int_{-\pi}^{\pi} z \frac{z^{t-k} - 1}{z - 1} d\mathcal{Z}(\lambda) \right) \\ &= \sum_k \psi_k X_0 + \int_{-\pi}^{\pi} z \frac{\sum_k \psi_k z^{t-k} - \sum_k \psi_k}{1 - z} d\mathcal{Z}(\lambda) \\ &= \Psi(1) X_0 + \int_{-\pi}^{\pi} z \frac{\Psi(z) z^t - \Psi(1)}{z - 1} d\mathcal{Z}(\lambda). \end{aligned} \tag{15}$$

This last expression is derived following the discussion in Wildi and McElroy (2016). Recall that $\Psi(z) = \exp\{\phi(z)\}$. If $\phi(1) = 0$ (i.e.,$\Psi(1) = 1$), from (15) it follows that only the increments $X_t - X_{t-1}$ get their phase altered by the all-pass filter, while the initial value (and the level) remain unchanged. There is no phase delay at frequency zero. Hence, any linear trend is passed unaltered.

Generalizing to higher order polynomial effects of order d, there is no phase delay at frequency zero provided that the filter $\Psi = \exp\{\phi\}$ has d vanishing derivatives at zero:

$$\frac{\partial^k \phi(z)}{\partial z^k}\Big|_{z=0} = 0, \quad 0 \leq k \leq d. \tag{16}$$

Thus, to construct a dth order trend-invariant utility-preserving $\delta-$LIP mechanism with respect to an original spectral density $f = f_{X|Z}$, we need to put additional constraints on the R function and the sampled spectral density h so that the filter constructed as $\Psi_h(z) = \exp\{\phi(z)\} = \exp\{i\pi R(H(\lambda))\}$ satisfies the derivative condition (16). This would mean that R has vanishing derivatives at $\lambda = 0$, and that h is bounded with the desired number of bounded derivatives at $\lambda = 0$.

Theorem 4 *Let a privacy budget δ be given. Assume all the conditions of Theorem 3 hold. Then any privacy mechanism of the form $\Psi_h(z) = \exp\{i\pi R(H(\lambda))\}$, where the pair (R, h) belong to the class*

$$\mathcal{C}_\delta(R, h, f) = \{(R, h) : R \in \mathcal{R}_L, h \in \mathcal{F}_R(f, \delta), \max_{0 \leq k \leq d} |R^{(k)}(0)| = 0, \max_{0 \leq k \leq d-1} |h^{(k)}(0)| < \infty\}, \tag{17}$$

will be a dth order trend-invariant utility-preserving privacy mechanism. Here the classes $\mathcal{R}_L$ and $\mathcal{F}_R(f, \delta)$ are defined in (10) and (11), respectively, and the derivatives $R^{(k)}(x)$ and $h^{(k)}(\lambda)$ are assumed to be well defined in an open neighborhood of zero.

3. FLIP: A Feasible Version of Linear Incremental Privacy

There are two practical issues to resolve before the LIP framework can be implemented: the calculation of filter coefficients from a given choice of spectral density f (and transform R), and the estimation of f from the data.

3.1 Computation of the Filter Coefficients

We need to compute the filter coefficients $\{\psi_k\}$ of the filter $\Psi_h(z) = \exp\{i\pi R(H(\lambda))\}$. Recall that the odd function $g(\lambda) = -\pi R(H(\lambda)) = i\phi(e^{-i\lambda})$ (defined over the interval $[-\pi, \pi]$) has a Fourier expansion in terms of the cepstral coefficient $\{\phi_k\}$, viz., $g(\lambda) = 2\sum_{k=1}^{\infty} \phi_k \sin(\lambda k)$. Inverting this relation, we have

$$\phi_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(\lambda)\sin(\lambda k)d\lambda = \frac{1}{\pi}\int_{0}^{\pi} g(\lambda)\sin(\lambda k)d\lambda, \quad k = 1, 2, \ldots. \tag{18}$$

These coefficients ϕ_k can be computed for any desired k; in practice, we can terminate the computations when k is sufficiently large such that ϕ_k is negligible (say, machine precision), or up to an order that is computationally feasible.

Next, the filter coefficients ψ_k are obtained by matching coefficients in the expansion $\Psi(z) = \sum_k \psi_k z^k = \exp\sum_k \phi_k z^k$. As $\phi(z)$ is a Laurent series corresponding to an odd sequence, we can write $\phi(z) = \phi^+(z) - \phi^+(z^{-1})$, where $\phi^+(z) = \sum_{k\geq 1}\phi_k z^k$ is a power series. Setting

$$\psi^+(z) = \sum_{k\geq 0}\psi_k^+ z^k = \exp\{\phi^+(z)\}$$
$$\psi^-(z) = \sum_{k\geq 0}\psi_k^- z^k = \exp\{-\phi^+(z)\},$$

we see that the filter can be expressed as

$$\Psi(z) = \exp\{\phi^+(z)\}\cdot\exp\{-\phi^+(z^{-1})\} = \psi^+(z)\cdot\psi^-(z^{-1}), \tag{19}$$

which is a Wiener-Hopf factorization of the Laurent series $\psi(z) = \sum_k \psi_k z^k$. It is clear that we only need a method of computing $\{\psi_1^+, \psi_2^+, \ldots\}$ and $\{\psi_1^-, \psi_2^-, \ldots\}$ from $\{\phi_1, \phi_2, \ldots\}$, and this is provided by the cepstral recursions of McElroy and Politis (2020). In particular, setting $\psi_0^+ = 1$ and $\psi_0^- = 1$, the recursions are

$$\begin{aligned}(j+1)\psi_{j+1}^+ &= \sum_{k=0}^{j}(k+1)\phi_{k+1}\psi_{j-k}^+, \\ (j+1)\psi_{j+1}^- &= -\sum_{k=0}^{j}(k+1)\phi_{k+1}\psi_{j-k}^- \end{aligned}\tag{20}$$

for $j = 0, 1, \ldots$. Finally, collecting the coefficients of z^j in the product (19), the filter coefficients ψ_j can be determined.

The cepstral coefficients ϕ_k are obtained by computing the integrals in (18) numerically. Thus, computing a large number of ϕ_k may be computationally expensive. In practice, a reasonable approximation to the Fourier series of $g(\lambda)$ can be attained by truncating the series at a finite order, say K, and using only finitely many ϕ_k, for $k \leq K$. The error in the approximation of $\phi^+(z) = \sum_{k \geq 1} \phi_k z^k$ by $\phi^+_K(z) = \sum_{1 \leq k \leq K} \phi_k z^k$ can be bounded as a function of K based on the smoothness of ϕ. For example,

$$\|\phi^+ - \phi^+_K\|_\infty \leq (2\pi(2r-1)K^{2r-1})^{-1} \langle g^{(r)} \rangle$$

for a positive integer r, where $g^{(r)}$ is the rth derivative of $g(\lambda)$ with respect to λ. Thus, one could minimize the value of the upper bound over a reasonable range of values for K and choose the smallest value that minimizes the bound over that range. In our simulation, we found that $K = 20$ provided an adequate approximation to the desired cepstral form of the filter. An important point to note is that regardless of the choice of K, the all-pass filter $\Psi_K(z) = \exp\{\phi^+_K(z) - \phi^+_K(z^{-1})\}$ would still provide full analytical validity for the released data.

In practice some truncation of $\Psi(z)$ is needed when applying the filter to a time series sample of length T, because it is not possible to compute infinitely many filter coefficients. If for some $M \geq 1$ a two-sided filter $\Psi(z)$ has length $2M + 1$, with M future and M past data points being filtered, then the filter output $\{\hat{X}_t\}$ will not have values for the first and last M data points in the sample; see McElroy and Politis (2020). This will result in a series of length $(T - 2M)$. In order to have a final released series that is of the same length as that of the original sensitive series, we use forecasts and backcasts to extend the series by M consecutive observations (on both the beginning and end of the sample), and then apply the filter on the extended series to obtain a final series of length T. Specifically, if $(X_1, \ldots, X_T)$ is the observed series, then we create an extended series $(X^E_{-M+1}, \ldots, X^E_0, X_1, \ldots, X_T, X^E_{T+1}, \ldots, X^E_{T+M})$, where X^E_t an estimate of the minimum mean squared error linear projection (i.e., the conditional expectation $E(X_t | X_1, \ldots, X_T)$ if the process is Gaussian). The forecasts can be obtained using the spectral density f_X already computed for the privacy mechanism.

3.2 Spectral Density Estimation

The $\delta-$LIP framework is based on knowing the true spectral density f; however, it is typically necessary to estimate the spectral density (although it's possible that the data publishing agency may have historical values from prior modeling). We refer to the application of the $\delta-$LIP framework, when working with a spectral density estimate $\hat{f}$, as a Feasible $\delta-$LIP, or $\delta-$FLIP. That is, a privacy mechanism Ψ is $\delta-$FLIP for a spectral density estimate $\hat{f}$, and some predetermined privacy budget $0 \leq \delta < 1$, if $\text{LIP}(\Psi, \hat{f}) \geq 1 - \delta$. This is just Definition 3, where $\hat{f}$ is used for the spectral density. Since the resulting mechanism Ψ depends on $\hat{f}$, and not the true f, privacy becomes an empirical measure; this definition is sensible, because the prediction paradigm that lies behind our privacy framework must also depend on spectral density estimates, and not on the true unknown spectrum.

As regards utility, recall that because we are using an all-pass filter for our privacy mechanism, second-order utility is preserved automatically *for any* choice of $\phi(z)$ with anti-symmetric coefficients; therefore, any statistical error made in the estimation of the spectral density matrix $f_{X,Z}$ has no bearing on second-order utility whatsoever.

To obtain a data-based estimate of the spectral density $f_{X,Z}$ of the stationary part, one could detrend the data by fitting a dth order polynomial and use the residual as approximation for $\{X_t, Z_t\}$. Typically for a trend estimation problem, estimation of a deterministic trend would change the time series properties of the error process $\{X_t, Z_t\}$. However, here the trend is removed only at this step for the estimation of the spectral density of the stationary portion – see (1) and (2). Because we are implementing a dth order trend-invariant filter, the obtained privacy filter Ψ can be directly applied to the observations $\tilde{X}_t$ (or their forecast-extended version, for which we need to extend the trend at either end of the series).

From the residual time series one can estimate $f_{X,Z}$ by using a parametric model class or by nonparametric methods (see discussion in McElroy and Politis (2020)), as summarized in the two options[1] below:

- Option 1: Estimate $f_{X,Z}$ with a model-based estimator, such as the spectral matrix for a Vector Autoregression of order p, or VAR(p). Assuming a parametric model is correctly specified, estimate the model parameters using maximum likelihood (or a similar procedure). Once the model has been fitted to the data, plug into the expression for the model spectral matrix and obtain $\hat{f}_{X,Z}$.

- Option 2: Use a nonparametric estimator of $f_{X,Z}$ based upon the sample autocovariances, viz. $\hat{\gamma}_k = T^{-1} \sum_{t=1}^{T-k} W_t W_{t+k}'$ for $k \geq 0$, where $W_t = (X_t, Z_t)'$. For a dth order trend-invariant filter, the nonparametric estimator has to be chosen to have bounded derivative at $\lambda = 0$ up to order d. We use a flat-top taper (McElroy and Politis, 2020) with threshold $C = 1/T$ to obtain an estimate of $f_{X,Z}$ based on the sample autocovariances.

3.3 Implementation of FLIP

We propose using the following steps for an implementation of the $\delta-$FLIP mechanism.

1. Estimate the spectral density matrix $f_{X,Z}$ (2) based on the observed data.

2. Compute the conditional spectral density $f = f_{X|Z}(\lambda)$ from $f_{X,Z}$ using (6).

3. For privacy budget $\delta \geq 0$ and for desired order d of the trend, choose $R \in \mathcal{R}_L$ with the desired number of derivatives at zero. Compute B in (13). Sample $\Delta \sim$ Uniform$[0, B]$, and construct the sample spectral density

$$h(\lambda) = A(\tilde{f}(\lambda) + \Delta),$$

where $A = \langle f \rangle_\pi / [1 + \pi\Delta]$. Compute the function $g(\lambda) = -\pi R(H(\lambda))$.

4. Compute the coefficients ϕ_k (for $k \leq K$) in the cepstral representation of the all-pass filter (3) using (18).

5. Fix truncation value M, and compute the all-pass filter coefficients ψ_j for $-M \leq j \leq M$ from (20).

[1] We also studied a third option, where a model is first selected using a model selection procedure (such as AIC) within a class of parametric models (such as a Vector Autoregression), and then the required quantities are computed based on the fitted model. The performance was similar to that of the parametric and nonparametric options considered below, and hence we do not report the results.

6. Use the estimated $f_{X,Z}$ to extend the series with M forecasts and M backcasts, denoted via $\{\tilde{X}_t^E\}$. If a trend model has been used, employ the estimated trend values to add to the extension on either end of the sample.

7. Apply the filter $\{\psi_k\}_{-M}^{M}$ to obtain the values of the privatized series

$$\hat{X}_t = \sum_{j=-M}^{M} \psi_j \tilde{X}_{t-j}^E, \; t = 1, \ldots, T.$$

4. Numerical Illustrations

In this section we present the results of a modest simulation study, as well as some real data analysis. The simulated data are a bivariate time series, with the first component $\{X_t\}$ being the sensitive time series of interest; the second component $\{Z_t\}$, an auxiliary correlated time series, is assumed to be what the attacker has access to. We examine different scenarios associated with varying degrees of correlation between the sensitive series and the auxiliary series. The real data are from the Quarterly Workforce Indicator (QWI) database published by U.S. Census Bureau; a detailed description of the data is given in Example 3 below. We also provide comparisons with other time series privacy mechanisms.

4.1 Performance of FLIP

Example 1: FLIP mechanism with $\delta = 0$.

In the first example we investigate a FLIP solution with $\delta = 0$ (i.e., maximum privacy), and report the operating characteristics of the FLIP mechanism. Note that in this case the randomization only occurs through the selection of an R function; the spectral density h used in the construction of the privacy mechanism has to be the true density $f = f_{X|Z}$. The sensitive series and the attacker's series are assumed to be jointly described by a bivariate VAR(1). Specifically, if $\{X_t\}$ is the sensitive series and $\{Z_t\}$ represents the attacker's knowledge, then the data generating model is

$$\begin{pmatrix} X_t \\ Z_t \end{pmatrix} = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix} \begin{pmatrix} X_{t-1} \\ Z_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ \zeta_t \end{pmatrix},$$

where $(\epsilon_t, \zeta_t)'$ are independently and identically distributed as a bivariate normal with mean zero and covariance matrix Σ equal to a scalar σ^2 times the two dimensional identity matrix, i.e., $(\epsilon_t, \zeta_t)' \sim N_2(\mathbf{0}, \sigma^2 \mathbf{I})$. We parameterize the process in terms of the cross-correlation $\rho = \text{Corr}(X_t, Z_t)$, and also assume the $\text{Var}(X_t) = \text{Var}(Z_t) = v$. This allows us to examine the impact of the attacker's knowledge, summarized in terms of the cross-correlation ρ, on the proposed privacy-utility framework. Thus, the stationary covariance matrix of (X_t, Z_t) is equal to

$$\gamma_0 = v \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

where for stationarity γ_0 must satisfy the Riccati equations $\gamma_0 = \Phi \gamma_0 \Phi^T + \Sigma$, and Φ is the coefficient matrix. Thus, $\gamma_0 - \Sigma$ must be positive definite; in order for this to happen, the matrix $\gamma_0 - \Sigma$ must have non-negative determinant – the condition reduces to $v \geq \sigma^2/(1-\rho)$.

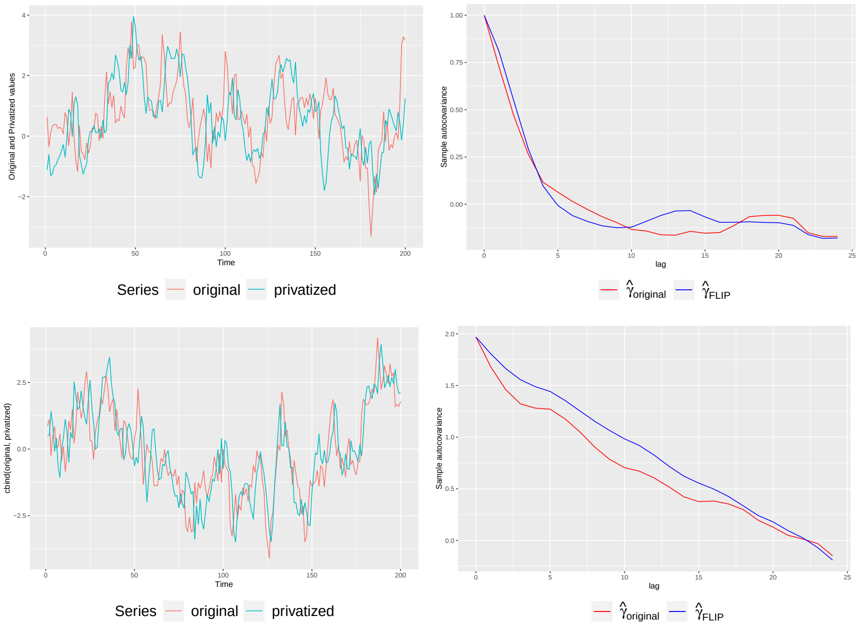

Figure 1: Original series and privatized version. Top left: Sample paths of the detrended series for the case of low cross-correlation. Top right: Sample autocovariance plots for the case of low cross-correlation. Bottom left: Sample paths of the detrended series for the case of high cross-correlation. Bottom right: Sample autocovariance plots for the case of high cross-correlation ($\hat{\gamma}_{original}$ for the actual series, $\hat{\gamma}_{FLIP}$ for the privatized series).

Here we set $v = \sigma^2/(1-\rho) + 1$ to make $\gamma_0 - \Sigma$ positive definite. The coefficient matrix is then solved from the Riccati equations as $\Phi = (\gamma_0 - \Sigma)\gamma_0^{-1/2}$, where $\gamma_0^{-1/2}$ is a matrix square-root of γ_0^{-1}. This parameterization provides a VAR(1) with cross-correlation equal to ρ and error variances equal to σ^2. For the present example we chose $\sigma^2 = 0.5$. The cross correlation was chosen to be either 0.1 or 0.7, to allow for cases representing, respectively, either a low or high degree of information possessed by the attacker. We generated 500 Monte Carlo replicates of the bivariate series in each case, with sample size $T = 200$.

We implemented the steps for performing FLIP on each of the generated series, and the results are presented in Figure 1. For implementing FLIP we chose the truncation order for the all-pass filter to be $K = 25$ for the cepstral representation and to be $M = 45$ for the filter coefficients. The forecast extension was done using both options as described earlier in the steps of the FLIP algorithm. We only show the results for option one, since both options yielded similar results. Figure 1 shows a typical original sensitive series with the privatized version for the two different cross-correlation values (top left and bottom left panels). The figure also shows the sample autovariance values for the original and the privatized series for the two cases (top right and bottom right panels).

While each privatized series retains the chief dynamic features of the sensitive series, the trajectories are substantially different over the observation window, showing that predictions based on the privatized version will be of little use in discerning the sensitive series.

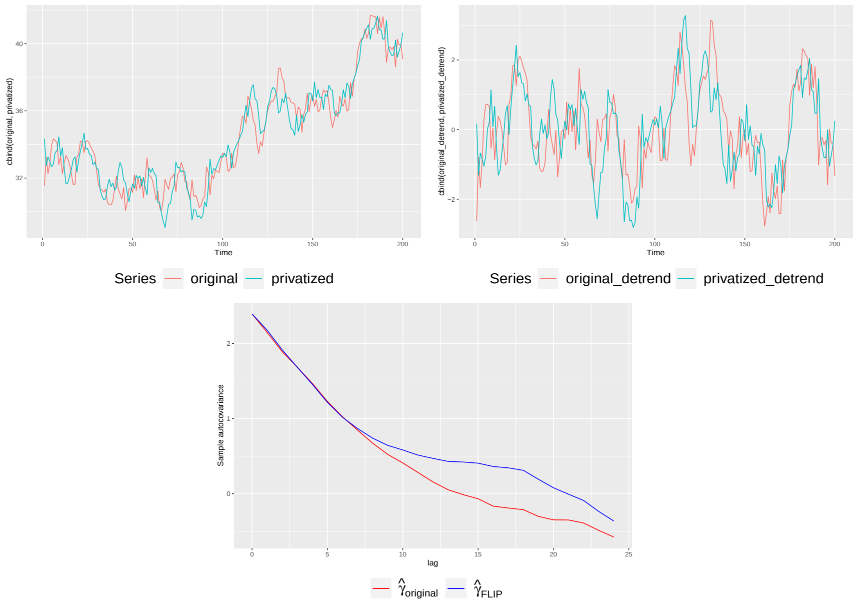

Figure 2: Original series along with privatized version for a linear trend model. Left: Sample paths with trend. Right: Sample paths without trend. Bottom: Sample autocovariance plots for both series ($\hat{\gamma}_{original}$ for the actual series, $\hat{\gamma}_{FLIP}$ for the privatized series).

The sample ACVF plots show that most of the sample autocovariances for the sensitive and privatized series are close, and thus second-order utility is retained to a large degree. The minor discrepancies in the autocovariance functions arise because the filter — due to the truncation of filter coefficients at a finite lag M — is only approximately all-pass. The agreement between the ACVF of the original series and the privatized series therefore improves with a larger sample size, when longer filter lengths can be used. We examined the squared distance between the sample autocorrelations of the original series and the FLIP-ped series, measured via $D_{ACF} = H^{-1}\sum_{h=0}^{H}(\rho_h - \hat{\rho}_h)^2$, where ρ_h and $\hat{\rho}_h$ are the lag h sample autocorrelations for $\{X_t\}$ and $\{\hat{X}_t\}$, respectively. In our numerical results we set $H = 24$. The variation in D_{ACF} around zero — based on 500 replications for the two different cross-correlation settings — was quite low, indicating the high second-order utility of the FLIP-ped series. The average value (over 500 replications) of the sample version of the privacy measure (5) was greater than 0.99 for both values of the cross-correlation between X_t and Z_t.

Example 2: **$\delta-$FLIP mechanism for $\delta > 0$.**

In the next example we consider a model with a linear trend, and demonstrate the usefulness of the $\delta-$FLIP algorithm in leaving the trend unchanged while perturbing the stationary disturbance around the trend. We consider a bivariate process with time-varying mean, as in (1), where the stationary part $\{X_t, Z_t\}$ is the VAR(1) process of the previous example, and the linear trends for the sensitive series and auxiliary series are

$$\mu_t^X = 30 + 0.05t, \qquad \mu_t^Z = 10 + 0.06t.$$

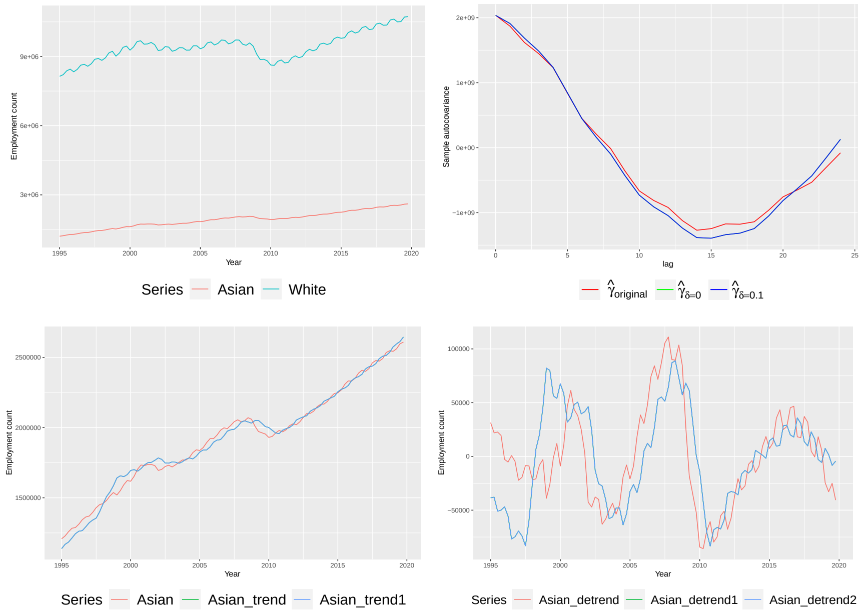

Figure 3: Top left: Employment counts for Asians and Whites in California. Top right: Sample autocovariance for original Asian employment count ($\hat{\gamma}_{original}$) and privatized versions ($\delta = 0$ corresponds to $\hat{\gamma}_{\delta=0}$ and $\delta = 0.1$ corresponds to $\hat{\gamma}_{\delta=0.1}$). Bottom left: Asian employment count (*Asian*) and privatized versions ($\delta = 0$ corresponds to *Asian trend* and $\delta = 0.1$ corresponds to *Asian trend1*). Bottom right: Detrended Asian employment count (*Asian detrend*) and privatized versions ($\delta = 0$ corresponds to *Asian detrend1* and $\delta = 0.1$ corresponds to *Asian detrend2*). Note: The graph with the two series are shown in their original scale.

The privacy budget δ is chosen to be 0.1. The data is detrended using ordinary least squares, by regressing each series $\{X_t\}$ and $\{Z_t\}$ separately on $\{1, t\}$. Then the FLIP algorithm is applied to the least squares residual series to obtain an all-pass filter that will leave any trend of order $d = 1$ unchanged. The estimated trend then is added back to the filtered least squares residuals to obtain the final privatized series.

The left panel of Figure 2 shows the original sensitive series along with the filtered version for a typical series generated using the assumed model. The middle panel shows the least squares residual series along with the perturbed series after application of a $\delta-$FLIP filter. The right panel shows the sample autocorrelation of the least squares residual series along with that of the perturbed series. The results of the simulation on the least squares residuals were very similar to those obtained in Example 1 using the stationary series, and hence are not reported here.

Example 3: FLIP-ping Quarterly Workforce Indicators.

To illustrate how the proposed FLIP methodology can be used on a regularly sampled time series that is published by a statistical agency, we examine Quarterly Workforce Indicator (QWI) data published by the U.S. Census Bureau. The indicators are based on different jobs and work location administrative data from 49 states, and are available quarterly; see

Abowd and Vilhuber (2011) for more details about the construction and publication of the data. All data were extracted from the QWI Explorer website QWI on 17th March 2022 at 10:45 pm. We focus upon the quarterly indicator "Beginning of Quarter Employment: Count" (or *employment count* for short) for the states of California and Maryland, with an observation period of Q1 1995 through Q4 2019.

The QWI are available at different levels of aggregation, and the granularity of finer cross-tabulation of these data — based on factors such as geographic region, age, race, sex and education category — may lead to disclosure of sensitive information on demographic and economic details of local labor markets. While higher level aggregation are less susceptible to disclosure risk, finer partitions of the data need to be protected. Below, we illustrate how the FLIP mechanism can be used to perturb lower level aggregation of the QWI series while retaining important time series features, so as to make the public use data dynamically consistent. We use published data, and hence there is no risk of disclosure. Thus, this is a data-driven example constructed for illustrative purposes.

1. California Employment Count Across Races: Here we examine the QWI employment count for California, where the data is cross-tabulated by race. No race or ethnic group constitutes a pure majority of California's population: 39% of state residents are Latino, 35% are White, 15% are Asian American or Pacific Islander, 5% are Black, 4% are multiracial, and fewer than 1% are Native American or Alaska Natives, according to the 2020 Census. We choose quarterly employment count of Whites as the series known to the attacker, i.e., the $\{Z_t\}$ series in our proposal. The count of the Asian group is chosen as the series $\{X_t\}$ that needs to be protected from attackers. Below, we report the results of applying the $\delta-$FLIP mechanism to the employment count data for $\delta = 0$ and $\delta = 0.1$.

Case: $\delta = 0$: The values of the two series are in millions, but for implementation of the FLIP methodology we work with the standardized version of both the series, obtained by subtracting the sample mean and dividing by the sample standard deviation. Because a linear trend seems to be inadequate to capture the stable growth pattern for both $\{X_t\}$ and $\{Z_t\}$, we assume $d = 3$, i.e. the trend is a polynomial of order three. As in Example 2, we use the residuals from the trend fit as the $\{X_t\}$ and $\{Z_t\}$ series; also we set $K = 25$ and $M = 25$. The beta mixture in the definition of the R function uses two components. The measure D_{ACF} is calculated as described in Example 1, yielding the value 0.0016. The privacy measure evaluated for the series is 0.9988; results are plotted in Figure 3. As intended, there are substantial differences between the stationary parts of the observed series and the perturbed series, whereas the sample autocovariance estimates are nearly identical.

Case: $\delta \neq 0$: We repeat the above procedure ($\delta = 0$), but now using $\delta = 0.1$. The D_{ACF} is 0.0026; the privacy measure is 0.9982. Figure 3 show the original series and the perturbed series along with their sample autocovariance values. The top left panel shows the two series together. The top right panel shows the sample autocovariances of the original and the two privatized series. The bottom left panel shows the Asian employment count series along with the two privatized series, whereas the bottom right panel shows those for the detrended version. In the case of $\delta = 0.1$ the perturbations are somewhat—though not substantially—more than those in the $\delta = 0$ case, and the sample ACF of the perturbed series is more distinct from that of the original series.

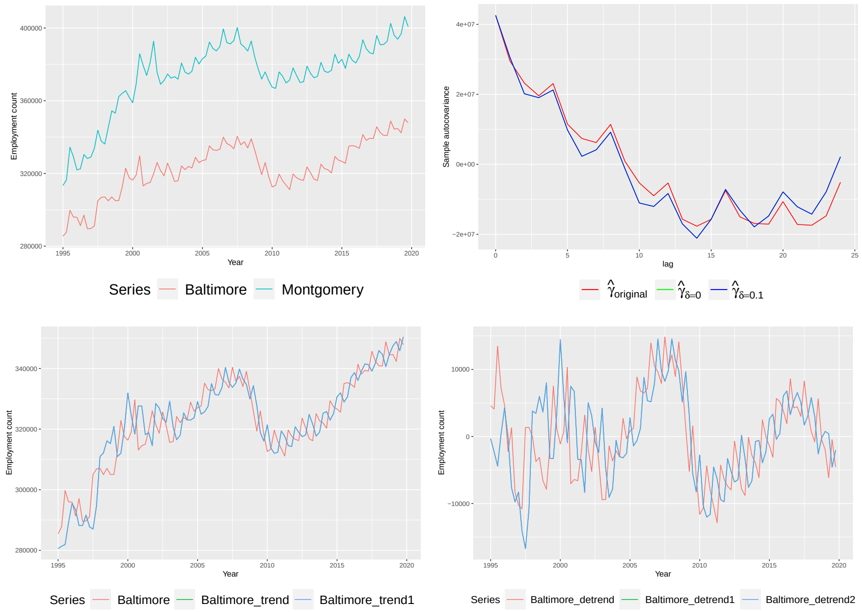

Figure 4: Top left: Employment counts for Baltimore county and Montgomery county in Maryland. Top right: Sample autocovariance for original Baltimore county employment count ($\hat{\gamma}_{original}$) and privatized versions ($\delta = 0$ corresponds to $\hat{\gamma}_{\delta=0}$ and $\delta = 0.1$ corresponds to $\hat{\gamma}_{\delta=0.1}$). Bottom left: Baltimore county employment count (*Baltimore*) and privatized versions ($\delta = 0$ corresponds to *Baltimore trend* and $\delta = 0.1$ corresponds to *Baltimore trend1*). Bottom right: Detrended Baltimore county employment count (*Baltimore detrend*) and privatized versions ($\delta = 0$ corresponds to *Baltimore detrend1* and $\delta = 0.1$ corresponds to *Baltimore detrend2*). Note: The graph with the two series are shown in their original scale.

2. Maryland Employment Count Across Counties: We repeat the exercise with the Maryland employment count QWI, examined across different counties. We choose Montgomery county as the $\{Z_t\}$ series, keeping Baltimore county as the $\{X_t\}$ series; again we use a third order polynomial trend. In Figure 4 we plot the sample autocovariances, the time series sample paths, and the detrended series for both the original data and the privatized ($\delta = 0$ and $\delta = 0.1$) versions. The top left panel shows the two series together. There is substantial contemporaneous correlation between the two series. That explains the fact that the perturbation amounts for $\delta = 0$ and $\delta = 0.1$ are not very different, resulting in two almost identical privatized series. The top right panel shows the sample autocovariance of the original and the two privatized series. The bottom left panel shows the original series along with the two privatized series, whereas the bottom right panel shows the sample autocovariance for the original Baltimore county series and for two privatized versions ($\delta = 0$ and $\delta = 0.1$).

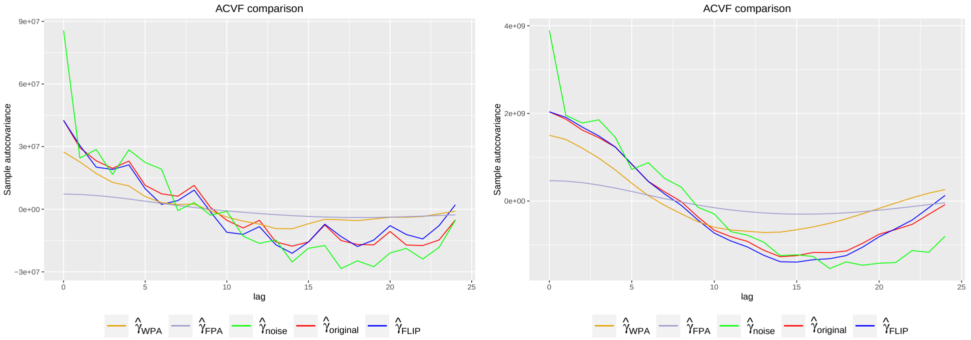

Figure 5: ACVF comparison between the actual series ($\hat{\gamma}_{original}$), FLIP series ($\hat{\gamma}_{FLIP}$), FPA series ($\hat{\gamma}_{FPA}$), WPA series ($\hat{\gamma}_{WPA}$), and noise added series ($\hat{\gamma}_{noise}$) for Baltimore data (left panel) and Asian data (right panel).

4.2 Comparison of FLIP with Existing Privacy Preserving Mechanisms

In this subsection we compare the results obtained using FLIP with those of other privacy mechanisms, such as noise addition, FPA, and CPA. We remark that the modifications of FPA to CFPA—-and of WPA to CWPA—are not pertinent to our framework; moreover, other privacy mechanisms such as those based on data mining or cryptographic methods are not comparable, and hence are omitted. The same applies to methods such as those based on the Pufferfish mechanism.

Since the privacy definitions vary across the different mechanisms (both FPA and WPA rely on different domains of differential privacy, resulting in incomparable privacy measures; whereas FLIP is predicated on a linear filtering privacy mechanism), it is not appropriate to compare the performance of these methods with respect to any single measure. Therefore, the individual methods are applied with parameters tuned to values within their respective recommended ranges; each mechanism is tuned such that it has adequate privacy according to its own privacy metric. We then compare the resulting second-order utility of each mechanism. We compare Gaussian noise addition, FPA, and WPA to FLIP when applied to the QWI data, for both California and Maryland. For FPA and WPA, we perform a discrete Fourier transform and Wavelet transform on the time series. We keep the first K coefficients, and generate the noisy version of that by applying Laplace noise to the coefficient. Then we get back the perturbed series from the ϵ-differentially private Fourier and Wavelet coefficients.

From Figure 5, it is evident that these alternative privacy methods provide more distortion to the autocovariance function compared to FLIP, and hence their second-order utility is impaired to a greater degree than that from the application of FLIP. This is expected, since these alternative time series privacy mechanisms are focused mainly upon a privacy objective, and are not constructed with a privacy-utility trade-off in mind.

5. Discussion and Future Scope

We have proposed a framework for disclosure avoidance of time series data. The methodology is particularly suitable for regularly spaced time series data where the data publisher anticipates the amount of sensitive information that may be available to the attacker in the form of an auxiliary time series that is cross-correlated with the sensitive time series. Specifically, we demonstrate that

- Noise addition alters serial correlation structure, and thereby mars second-order utility;
- All-pass filtering preserves the second-order structure of a time series, and hence maintains second-order utility;
- Linear Incremental Privacy (LIP) can assess privacy protection for stationary time series data;
- Feasible LIP (or FLIP), which uses all-pass filtering with a specified privacy budget $\delta \geq 0$, is a mechanism that can protect time series data.

The novelty of the FLIP proposal consists in how privacy protection is measured (as the inability to improve the attacker's knowledge), as well as providing—unlike conventional randomized privacy mechanisms—full analytical validity, in terms of preserving the mean and autocovariance of the data process.

Although our focus is on stationary time series, we have provided extensions to the all-pass filter design that accommodate non-stationary trend effects. We remark that here we have focused on the first and second order structure of a time series, i.e., the mean function and serial correlation pattern; it is natural to consider higher order structures of the time series, possibly assessed through higher order polyspectra, and investigate an analogue of an all-pass filter that may also preserve such features. In future work we plan to research further extensions to non-stationarity and higher-order dynamics.

We emphasize that our method presumes time series data that is regularly sampled, as is common in official statistics. For irregularly sampled data it may be possible to apply our methodology if the time index is warped; this is a topic of future research. Also, methods such as FPA and WPA do not account for auxiliary information Z. The generalization of these methods to include auxiliary information should also be investigated.

Appendix A. Proofs

Proof [Proof of Proposition 1] Let the lag j autocovariance matrix for the multivariate process $\{X_t, Z_t\}$ be denoted

$$\gamma_j = \begin{pmatrix} \gamma_X(j) & \gamma_{XZ}(j) \\ \gamma_{XZ}(-j) & \gamma_Z(j) \end{pmatrix},$$

where $\gamma_{XZ}(j) = E(X_{t+j} Z_t)$ and $\gamma_{ZX}(j) = E(Z_{t+j} X_t) = \gamma_{XZ}(-j)$. The corresponding cross-spectral densities will be denoted by f_{XZ} and f_{ZX}. The linear projection of X_t on $\{Z_t\}$ is given by

$$E(X_t | \{Z_t\}) = \sum_k \pi_k Z_{t-k}$$

for some coefficients $\{\pi_k\}$ to be determined. This is the best guess (in the sense of minimum MSE among linear predictors) of the original time series at time point t made by the attacker based on his knowledge of $\{Z_t\}$, and from the normal equations it follows that

$$\gamma_{XZ}(h) = E(X_t Z_{t-h}) = E(E[X_t|\{Z_t\}]Z_{t-h}) = \sum_k \pi_k \gamma_Z(h-k)$$

for any integer h. Taking the Fourier transform yields $f_{XZ}(\lambda) = \pi(e^{-i\lambda}) f_Z(\lambda)$, and hence

$$\pi(e^{-i\lambda}) = f_{XZ}(\lambda)/f_Z(\lambda). \tag{21}$$

Next, the autocovariance function of $X_t - E(X_t|\{Z_t\})$ is

$$\begin{aligned}
\mathrm{Cov}(X_t, X_{t-j}|\{Z_t\}) &= \mathrm{Cov}(X_t - E(X_t|\{Z_t\}), X_{t-j} - E(X_{t-j}|\{Z_t\})) \\
&= \gamma_X(j) - \sum_l \pi_l \gamma_{XZ}(l+j) - \sum_k \pi_k \gamma_{XZ}(k-j) \\
&\quad + \sum_k \sum_l \pi_k \pi_l \gamma_Z(j+l-k)
\end{aligned}$$

for any integer j, and taking the Fourier transform yields (6) via (21). Similar calculations yield

$$\mathrm{Cov}(X_t, \hat{X}_t|\{Z_t\}) = \langle \Psi, f_{X|Z}\rangle \quad \text{and} \quad \mathrm{Cov}(\hat{X}_t, \hat{X}_t|\{Z_t\}) = \langle \Psi\overline{\Psi}, f_{X|Z}\rangle,$$

and the formula for the privacy measure follows so long as $f_{X|Z} > 0$.

■

Proof [Proof of Theorem 1] Let $\Psi(z) = \exp\{\phi(\lambda)\} = \exp\{\sum_k \phi_k z^k\}$ be the cepstral representation (3) of Ψ. Since $|\Psi(z)| = 1$ for all $z = e^{-i\lambda}$, we have $\phi_k = -\phi_{-k}$ and $\Psi(e^{-i\lambda}) = \exp\{ig(\lambda)\}$, where $g(\lambda) = -2\sum_{k\geq 1} \phi_k \sin(\lambda k)$. Thus we want to find odd functions g with expansion $-2\sum_{k\geq 1}\phi_k \sin(\lambda k)$ that will provide a perfect privacy solution. For such an odd function, perfect privacy would necessarily imply $\langle \cos(g), f\rangle_\pi = 0$. Now assume that g has the form $g(\lambda) = \pi R(F(\lambda))$ for $\lambda \in [0,\pi]$, where $F(\lambda) = \int_0^\lambda \tilde{f}(\omega)d\omega$ and $R : [0,1] \to [0,1]$. We need to show that an R function with $\langle \cos(\pi R(F), f\rangle_\pi = 0$ exists. Consider any $R \in \mathcal{R}_f$. By the change of variable $x = F(\lambda)$, $\langle \cos(\pi R(F)), f\rangle_\pi = \int_0^1 \cos(\pi R(x))dx$. Since $R(x) = 1 - R(1-x)$ and $\cos(\theta) = -\cos(\pi - \theta)$ for $\theta \in [0,\pi]$, we have $\langle \cos(\pi R(F)), f\rangle_\pi = 0$. Extending R to [-1, 0] via $R(-x) = -R(x)$, and choosing $g(\lambda) = \pi R(F(\lambda))$ for $\lambda \in [-\pi,\pi]$, we obtain the result. ■

Proof [Proof of Theorem 2] Let h be a spectral density belonging to $\mathcal{F}_R(f,\delta)$ for some $R \in \mathcal{R}$. Then for $z = e^{-i\lambda}$

$$\begin{aligned}
|\Psi_h(z) - \Psi_f(z)| &= |\exp\{i\pi R(H(\lambda))\} - \exp\{i\pi R(F(\lambda)\}| \\
&= |1 - \exp\{i\pi(R(H(\lambda)) - R(F(\lambda)))\}| \\
&\leq \pi|R(H(\lambda)) - R(F(\lambda))| \\
&\leq \pi L_R |H(\lambda) - F(\lambda)| \\
&\leq \pi^2 L_R \sup_{0\leq\lambda\leq\pi} |h(\lambda) - f(\lambda)|/\langle f\rangle_\pi \\
&= \sqrt{\delta},
\end{aligned}$$

where we have used the fact that $|1 - e^x| \leq |x|$ for any complex number x. Because Ψ_h is an all-pass filter, by Remark 3 LIP(Ψ, f) takes the form (7), and

$$\begin{aligned}
\mathrm{LIP}(\Psi_h, f) &= 1 - \frac{\langle \Psi_h, f\rangle^2}{\langle f\rangle^2} \\
&= 1 - \frac{\langle \Psi_f + \Psi_h - \Psi_f, f\rangle^2}{\langle f\rangle^2} \\
&= 1 - \frac{[\langle \Psi_f, f\rangle + \langle \Psi_h - \Psi_f, f\rangle]^2}{\langle f\rangle^2} \\
&= 1 - \frac{\langle \Psi_h - \Psi_f, f\rangle^2}{\langle f\rangle^2} \\
&\geq 1 - \frac{\langle |\Psi_h - \Psi_f|, f\rangle^2}{\langle f\rangle^2} \\
&\geq 1 - (\sqrt{\delta})^2 \\
&= 1 - \delta,
\end{aligned}$$

where the fourth equality is obtained by using the fact $\langle \Psi_f, f\rangle = 0$ (this is shown in the proof of Theorem 1). ■

Proof [Proof of Theorem 3] We will show that Ψ_h is $\delta-$LIP by showing that h belongs to the class (11). From the definition it is clear that $\langle h\rangle_\pi = \langle f\rangle_\pi$ Also,

$$\begin{aligned}
|h - f| &= |\langle f\rangle_\pi(\tilde{f} + \Delta)/(1 + \pi\Delta) - f| \\
&= (1 + \pi\Delta)^{-1}|f + \Delta\langle f\rangle_\pi - f - f\pi\Delta| \\
&= (1 + \pi\Delta)^{-1}\Delta\langle f\rangle_\pi|\pi\tilde{f} - 1|.
\end{aligned}$$

From Theorem 2, for the privacy mechanism Ψ_h associated with h to be $\delta-$LIP we need

$$\sup_{0\leq\lambda\leq\pi} |h(\lambda) - f(\lambda)| \leq \frac{\sqrt{\delta}\langle f\rangle_\pi}{L_R\pi^2}.$$

Thus, Ψ_h is $\delta-$LIP if

$$\frac{\Delta}{1 + \pi\Delta} \sup_{0\leq\lambda\leq\pi} |\pi\tilde{f}(\lambda) - 1| \leq \frac{\sqrt{\delta}}{L_R\pi^2}.$$

or

$$\frac{\Delta}{1 + \pi\Delta} \leq \frac{\sqrt{\delta}}{L_R\pi^2 \sup_{0\leq\lambda\leq\pi} |\pi\tilde{f}(\lambda) - 1|}.$$

The inequality is satisfied by at least one Δ provided the RHS is strictly less than one and in that case any Δ such that $\Delta \leq B$ will satisfy the inequality. Since (12) holds, we have Ψ_h is $\delta-$LIP if $\Delta \leq B$.

Using the bound $|e^x - 1| \geq \frac{1}{2}x$ for any complex number x, we have

$$|\Psi_h(e^{-i\lambda}) - \Psi_f(e^{-i\lambda})| \geq \frac{\pi}{2}|R(H(\lambda)) - R(F(\lambda))|. \quad (22)$$

Here $H(\lambda) = \int_0^\lambda \tilde{h}(\omega)d\omega$, where $\tilde{h} = (1+\Delta\pi)^{-1}(\tilde{f}(\lambda)+\Delta)$. Hence $H(\lambda) = (1+\Delta\pi)^{-1}(F(\lambda)+\Delta\lambda) = (1-\alpha)F(\lambda) + \alpha(\lambda/\pi)$ where $0 < \alpha = \frac{\Delta\pi}{1+\Delta\pi}$. Thus, $|H(\lambda) - F(\lambda)| = \alpha|F(\lambda) - \lambda/\pi|$. Then from (22) and the mean value theorem, we obtain the theorem's final result. ■

Proof [Proof of Theorem 4] Since $\phi(z) = i\pi R(H(\lambda))$ and $H(0) = 0$, we can differentiate successively, using the chain rule, at $\lambda = 0$, and the condition (16) holds. Hence the result follows. ■

References

QWI Explorer, U.S. Census Bureau. `https://qwiexplorer.ces.census.gov/`. Accessed: 03-17-2022 at 10:45 pm.

J. M. Abowd and L. Vilhuber. National estimates of gross employment and job flows from the quarterly workforce indicators with demographic and industry detail. *Journal of econometrics*, 161(1):82–99, 2011.

J. M. Abowd, K. Gittings, K. L. McKinney, B. E. Stephens, L. Vilhuber, and S. Woodcock. Dynamically consistent noise infusion and partially synthetic data as confidentiality protection measures for related time series. *US Census Bureau Center for Economic Studies Paper No. CES-WP-12-13, Available at SSRN: https://ssrn.com/abstract=2159800 or http://dx.doi.org/10.2139/ssrn.2159800*, 2012.

C. Dwork. Differential privacy. *International Colloquium on Automata, Languages and Programming, part II (ICALP)*, 2006.

C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9:211–407, 2014.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. *Theory of Cryptography Conference(TCC)*, pages 265–284, 2006.

M. A. Erdogdu, N. Fawaz, and A. Montanari. Privacy-utility tradeoff for time-series with application to smart-meter data. *Association for the Advancement of Artificial Intelligence*, 2015.

F. Fioretto and P. V. Hentenryck. Optstream: Releasing time series privately. *Journal of Artificial Intelligence Research*, 2019.

S.K. Hong, K. Gurjar, H.S. Kim, and Y.S. Moon. A survey on privacy preserving time-series data mining. *International Conference on Intelligent Computational Systems (ICICS)*, 2013.

M. Joye and B. Libert. A scalable scheme for privacy-preserving aggregation of time-series data. *Financial Cryptography and Data Security*, 78-79:111–125, 2013.

M. Katsomallos, K. Tzompanaki, and D. Kotzinos. Landmark privacy: Configurable differential privacy protection for time series. *Conference on Data and Application Security and Privacy (CODASPY)*, 2022.

D. Kifer and A. Machanvajjhala. Pufferfish: A framework for mathematical privacy definitions. *ACM Transactions on Database Systems*, 2011.

F. L. Lako, P. Lajoie-Mazenc, and M. Laurent. Privacy-preserving publication of time-series data in smart grid. *Security and Communication Networks*, 2021.

R. Lu, X Liang, X. Li, X. Lin, and X. Shen. Eppa: An efficient and privacy-preserving aggregation scheme for secure smart grid communications. *IEEE Transactions on Parallel and Distributed Systems*, pages 1621–1631, 2012.

L. Lyu, Y. W. Law, J. Jin, and M. Palaniswami. Privacy-preserving aggregation of smart metering via transformation and encryption. *IEEE Trustcom/BigDataSE/ICESS, pp. 472–479, IEEE, Sydney, Australia*, 2017.

T. S. McElroy and D. N. Politis. *Time Series: A First Course with Bootstrap Starter*. CRC Press, 2020.

V. Rastogi and S. Nath. Differentially private aggregation of distributed time-series with transformation and encryption. *International Conference on Management of Data, ACM SIGMOD*, pages 735–746, 2010.

Y. Sang, H. Shen, and H. Tian. Privacy-preserving tuple matching in distributed databases. *IEEE Transactions on Knowledge and Data Engineering, 21(12)*, page 1767–1782, 2009.

E. Shi, HTH. Chan, E. Rieffel, R. Chow, and D. Song. Privacy-preserving aggregation of time-series data. *Annual NDDS*, 2011a.

E. Shi, T-H. H. Chan, and E. Rieffel. Privacy-preserving aggregation of time-series data. *In Proc. of the Network and Distributed System Security Symposium, San Diego, California*, 2011b.

S. Song and K. Chaudhuri. Composition properties of inferential privacy for time-series data. *arXiv:1707.02702*, 2017.

S. Song, Y. Wang, and K. Chaudhuri. Pufferfish privacy mechanisms for correlated data. *arXiv:1603.03977*, 2017.

C. Stach. Vault: A privacy approach towards high-utility time series data. *International Conference on Emerging Security Information, Systems and Technologies, pp. 41–46*, 2019.

M. Stinner. Disclosure control and random tabular adjustment. *SSC Annual Meeting, Survey Methods Section*, 2017.

L Wasserman and S. Zhou. A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105:375–389, 2009.

M. Wildi and T. S. McElroy. Optimal real-time filters for linear prediction problems. *J. Time Ser. Econom.*, pages 155–192, 2016.

G. Ács, C. Castelluccia, and R. Chen. Differentially private histogram publishing through lossy compression. *IEEE International Conference on Data Mining*, 2012.