Bar-Ilan University

Faculty of Engineering

Differentially Private Statistical Hypothesis Testing

Ziv Chaba, Hila Cohen

Supervisor: Dr. Or Sheffet

October 2024

Acknowledgments

We would like to express our sincere gratitude to our academic supervisor, Dr. Or Sheffet, for his expert guidance, insightful feedback, and continuous assistance throughout this research.
We also wish to extend our thanks to our families for their ongoing support.

Abstract

Existing research has primarily explored differential privacy, in contexts where individuals contribute single data points. Our work extends this scope by examining challenging scenarios in which each individual provides an entire time series, such as financial or economic data. This study addresses the challenge of conducting differentially private statistical hypothesis testing on time series data, with a particular focus on the Dickey-Fuller unit root test. We develop a differentially private mechanism tailored for the Dickey-Fuller test, ensuring that privacy is maintained without compromising the accuracy of statistical inference.

Through a series of experiments, we compare the performance of the differentially private Dickey-Fuller test with its non-private counterpart across various autoregressive models. Our findings demonstrate that the differentially private algorithm can produce results closely aligned with the non-private version, particularly when the privacy budget is optimized and the sample size is sufficiently large. We also find that certain autoregressive models are more robust to the effects of differential privacy. Additionally, we apply our method to a real-world dataset, demonstrating its practical utility in preserving privacy.

Our research contributes to the growing field of privacy-preserving data analysis, offering a novel approach to hypothesis testing in time series data. The results indicate that differential privacy can be effectively integrated into time series analysis, allowing for the protection of individual privacy in sensitive datasets without significantly sacrificing statistical validity.

Table of Contents

Chapter 1

Introduction

The increasing use of large-scale data collection in fields such as economics, finance, healthcare, social sciences and more has led to heightened concerns about privacy. As organizations accumulate vast amounts of personal data for analysis, the risk of individual data being exposed or misused has grown substantially. These privacy concerns are especially pronounced in time series data, where individual participants provide a sequence of data points over time, often revealing sensitive information such as health records, financial transactions, or location data. In response to this challenge, differential privacy has emerged as a powerful framework that enables data analysis while protecting individual privacy.

Differential privacy ensures that the output of a statistical analysis remains largely unchanged whether or not a particular individual's data is included in the dataset. This protection is achieved by introducing a controlled amount of noise into the data, which limits the ability of attackers to infer private details about individuals. The trade-off between privacy and data utility is central to differential privacy. By carefully calibrating the noise, analysts can maintain high data accuracy while adhering to strong privacy guarantees.

While significant progress has been made in developing differentially private algorithms for a variety of statistical methods, most of the existing work has focused on cases where individuals contribute a single data point. However, many real-world applications involve time series data, where each participant provides multiple observations over time. Time series data present unique challenges for privacy, as the temporal dimension can reveal more about an individual's behavior or attributes. For example, in financial data, patterns in spending or investment behavior can be extracted over time, and in health data, fluctuations in vital signs can signal specific medical conditions. The risk of reidentification in time series data is thus higher, making it crucial to develop privacy-preserving methods tailored to this data type.

In this research, we focus on a specific time series problem: the unit root test, particularly the Dickey-Fuller test, which is widely used in econometrics to assess the stationarity of time series data. Stationarity is a key concept in time series analysis, referring to the property that the statistical characteristics of a seriesremain constant over time. Non-stationary data can lead to misleading results in statistical models, making stationarity testing a fundamental step in time series analysis. The Dickey-

Fuller test helps determine whether a time series contains a unit root, indicating non-stationarity. Given the widespread use of the Dickey-Fuller test in fields like economics (e.g., GDP growth, inflation rates), finance (e.g., stock prices, interest rates), and public policy, ensuring the privacy of individuals contributing time series data is of paramount importance.

Several studies such as [4, 6, 8, 10–12], and more, have addressed the challenge of applying differential privacy to hypothesis testing, but as mentioned before, these have primarily focused on scenarios where each individual contributes a single point of data. Such approaches avoid the complexity introduced by datasets that contain a full record or time series for each participant. In this project, we extend this body of work by exploring how differential privacy can be integrated for time series data, with specific focus into the Dickey-Fuller test. Our goal is to ensure that individuals' privacy is protected while still preserving the integrity of the statistical test.

The application of differential privacy to time series data presents several unique challenges. One of the key difficulties is that time series data are often autocorrelated, meaning that consecutive observations are not independent of one another. This violates the assumptions of many traditional algorithms, which typically assume independent and identically distributed (i.i.d.) data points. Moreover, the length of the time series can vary significantly between different cases, introducing additional variability that complicates the task of ensuring privacy.

Our approach leverages the Laplace mechanism, a widely used technique in differential privacy, to introduce noise into the statistical estimators used in the Dickey-Fuller test. Specifically, we apply noise to the mean of the autoregressive coefficients and the t-statistic used in the test. By carefully calibrating the amount of noise based on the sensitivity of the test statistics, we ensure that privacy is preserved without significantly degrading the accuracy of the test results.

This research contributes to the field of privacy-preserving statistical analysis by demonstrating that it is possible to perform unit root testing on time series data while maintaining differential privacy. Our method has broad applicability, as the Dickey-Fuller test is used in a wide range of fields, and the challenges we address are common in many types of time series analyses.

To validate our approach, we conduct extensive experiments on both synthetic and real-world datasets. First, we test our differentially private Dickey-Fuller algorithm on synthetic autoregressive time series, varying key parameters such as the autoregressive coefficients, and the number of individuals contributing data. We compare the results of the differentially private algorithm with the standard, non-private version of the Dickey-Fuller test to assess how closely the private test approximates the original under different privacy budgets. Our findings show that with an appropriate choice of privacy budget, the differentially private test can closely match the performance of the non-private test, particularly when the number of individuals in the dataset is large enough.

In addition to the synthetic experiments, we apply our method to real-world data from the World Bank's GDP dataset. This dataset, which includes time series of GDP growth for numerous countries, provides a rich source of real-world time series data. By applying our differentially private Dickey-Fuller test to this data,

we demonstrate that our method is effective in preserving privacy for real-world applications. The results of our experiments highlight the practical applicability of our approach in real-world settings, where privacy concerns are paramount.

In conclusion, as the collection and use of time series data continue to grow across various domains, the need for robust privacy protections becomes increasingly important. This research advances the understanding of how differential privacy can be effectively applied to time series data while preserving the utility of statistical tests like the Dickey-Fuller test. By ensuring that individuals' data remains private, we contribute to the development of ethical and trustworthy data analysis methods that can be applied in sensitive domains such as economics, healthcare, and finance.

Chapter 2

Preliminaries

This section provides an overview of the fundamental concepts and mathematical tools relevant to our project. These concepts serve as the foundation for our investigation into developing a time series differentially private mechanism for the Dickey-Fuller test.

2.1 Time Series Analysis

A **time series** is a sequence of data points that are indexed in chronological order. Typically, a time series consists of observations recorded at successive intervals that are uniformly spaced in time. Time series analysis is a crucial methodological approach in various fields, including economics, finance, environmental studies, and social sciences, where temporal dynamics play a pivotal role in data interpretation and decision-making. Examples of time series include the daily number of COVID-19 cases, quarterly gross domestic product (GDP) figures, monthly unemployment rates, daily active user counts for social media platforms, and annual averages global temperatures (Fig. 2.1).

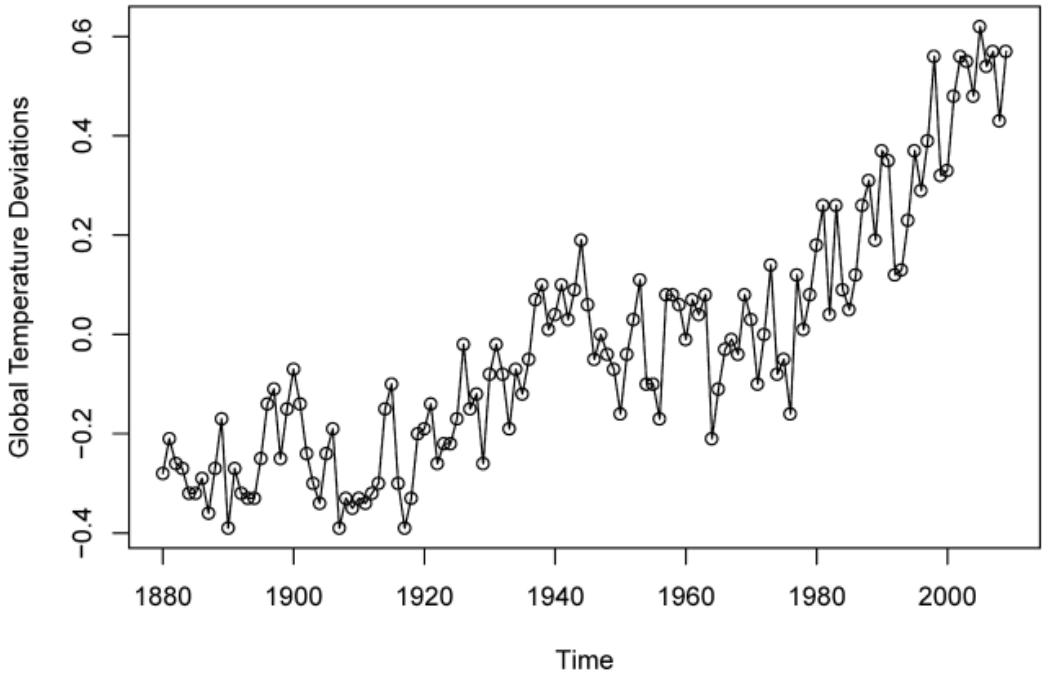

Figure 2.1: Yearly average global temperature deviations (1880–2009) in degrees centigrade [9].

Definition 2.1 (Time Series)**.** A **time series** can be defined as a set of observations $\{y_t\}$, where $t = 1, 2, \ldots, T$ represents the time index and y_t is the observed value at time t.

Following definition 2.1, time series data can exhibit patterns such as trends (long-term movements) or seasonality (periodic fluctuations). Statistical methods are applied to understand, model, and forecast these dynamics.

2.2 T-Test

A **t-test** is a statistical test that evaluates whether there is a significant difference between the means of two groups. It helps determine whether the observed differences between groups (such as treatment versus control or before and after a change) are likely due to actual effects rather than random chance. The test produces a "t-statistic," which helps quantify how large the difference is relative to the variability in the data. It's especially useful when sample sizes are small, and the population's standard deviation is unknown, making it a critical tool in inferential statistics.

The core idea of the t-test is to compare the difference between group means to the variability within the groups. Based on this comparison, the test yields a p-value, which tells us whether the difference is statistically significant. The significance level, denoted by α, is the threshold used to decide whether to reject the null hypothesis (the assumption that there is no difference between the groups). The most commonly used significance level is 0.05 (or 5%), meaning that there is a 5% chance the results

could be due to random variation, rather than a real effect. If the p-value is less than 0.05, the result is considered statistically significant, and the null hypothesis is rejected.

T-tests are widely used in fields such as medicine, psychology, education, economics, and business. For example, in medical research, t-tests might be used to assess whether a new drug significantly improves patient outcomes compared to a placebo. In psychology, they might compare two groups to see if an intervention improves test scores. In education, researchers use t-tests to examine whether different teaching methods result in different levels of student performance. In economics and business, t-tests help compare things like revenue from different marketing strategies or customer satisfaction scores before and after implementing a change.

The t-test offers a robust method for assessing whether observed differences between groups are statistically meaningful or simply the result of random variation, allowing researchers to draw more confident conclusions about the presence of real effects.

2.3 Autoregressive Model

Models for time series data can have many forms and represent different stochastic processes. In time series analysis, **Autoregressive (AR) Model** is one of the fundamental approaches used for modeling time-dependent data. The core idea behind the AR model is that the current value of a time series can be explained as a linear function of its previous values. This assumes that the patterns and dependencies in the past observations of the time series are informative for predicting future values.

Definition 2.2 (Autoregressive (AR) Model of order p)**.** An **Autoregressive (AR) Model of order p, abbreviated AR(p)**, is of the form:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + e_t \tag{2.1}$$

where x_t is a series of observations starts with x_0, which is a fixed constant. The parameters $\phi_1, \phi_2, \ldots, \phi_p$ are constants ($\phi_p \neq 0$). Although it is not necessary yet, we assume that $\{e_t, t = 0, 1, 2, \ldots\}$ is a Gaussian noise series with mean zero and variance σ^2.

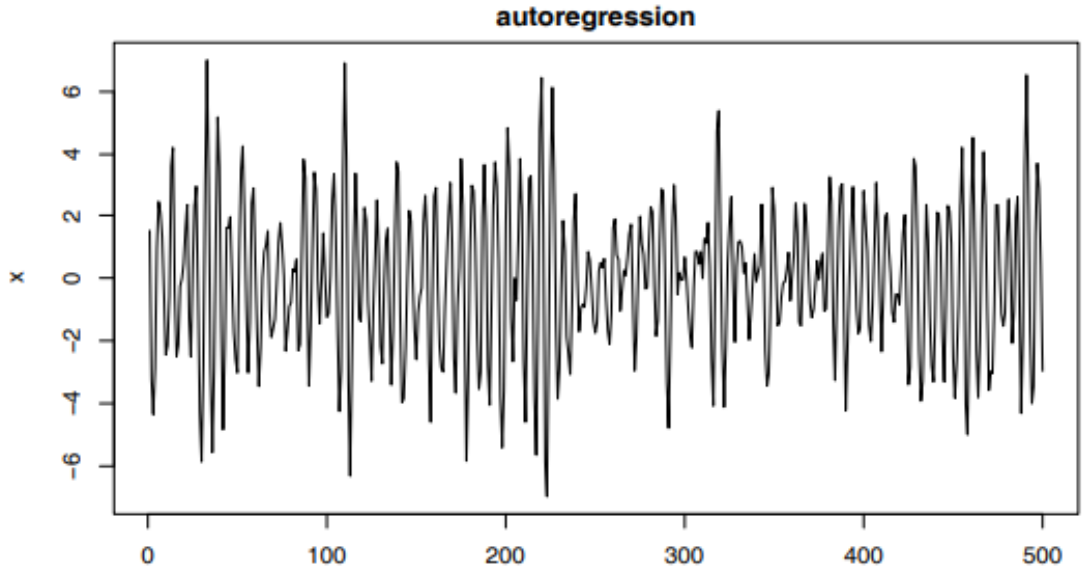

Figure 2.2: Autoregressive time series generated from the model $x_t = x_{t-1} - 0.9x_{t-2} - 2 + e_t$ for $t = 1, 2, \ldots, 500$. This model's equation describes a regression (or prediction) of the current value x_t of a time series based on the two preceding values of the series. Consequently, this model is referred to as autoregression [9].

The term "autoregressive" refers to the fact that the model regresses the variable on its own prior values. The AR model is particularly useful when there is evidence of autocorrelation in the data, meaning that the observations are correlated with their own past values. This kind of model is applied when the assumption holds that the relationship between an observation and its previous values is strong enough to help explain and forecast the time series.

A Model for analyzing trend such as seen in the global temperature data in Figure (2.1), is the **random walk with drift model**, which is given by:

$$x_t = \delta + x_{t-1} + e_t \tag{2.2}$$

for $t = 1, 2, \ldots$, with initial condition x_0, where e_t is as before. The constant δ is called drift, and when the drift is zero (i.e., when $\delta = 0$), (2.2) is called simply a **random walk**. The term random walk comes from the fact that, when $\delta = 0$, the value of the time series at time t is the value of the series at $t - 1$ plus a completely random movement determined by e_t. One may rewrite (2.2) as a cumulative sum of white noise variates, that is,

$$x_t = \delta t + \sum_{i=1}^{t} e_i \tag{2.3}$$

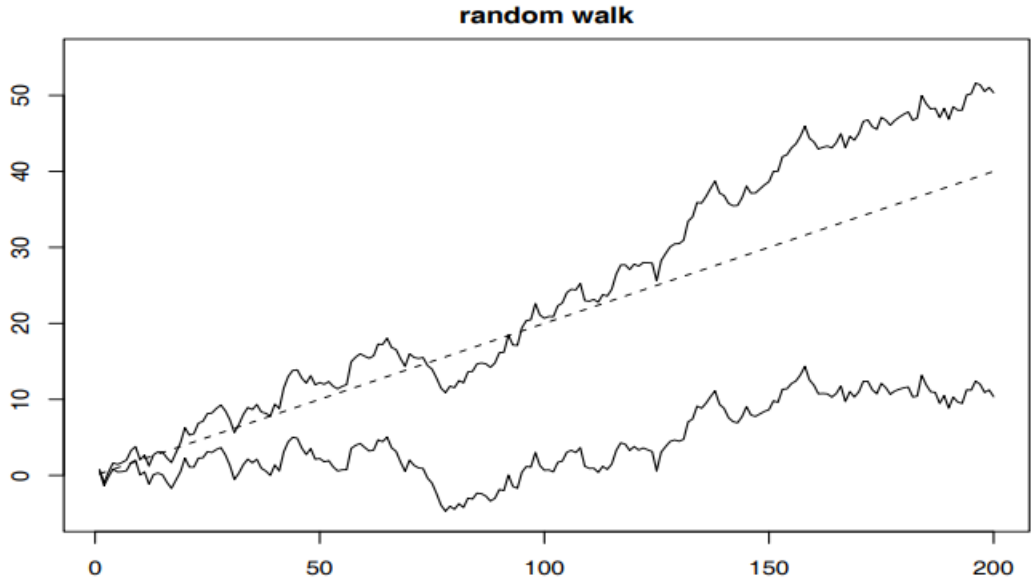

Figure 2.3: Random walk, with drift $\delta = 0.2$ (upper jagged line), without drift, $\delta = 0$ (lower jagged line).

2.4 Stationary Time Series

In the study of time series analysis, the concept of stationarity is fundamental. A **stationary time** series is defined as one that exhibits a consistent statistical structure over time. This stability allows for more reliable forecasting and model fitting, as the underlying processes governing the series do not change. Understanding stationary time series is essential for various applications, where the goal is often to analyze historical data to predict future outcomes.

Definition 2.3 (Strict Stationary)**.** A **Strict Stationary** time series is one for which the probabilistic behavior of any set of values

$$\{x_{t_1}, x_{t_2}, \ldots, x_{t_k}\}$$

is identical to that of the time shifted set

$$\{x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_k+h}\}$$

which means,

$$P\{x_{t_1} \leq c_1, \ldots, x_{t_k} \leq c_k\} = P\{x_{t_1+h} \leq c_1, \ldots, x_{t_k+h} \leq c_k\} \tag{2.4}$$

for all $k = 1, 2, \ldots$, all time points $t_1, t_2, \ldots, t_k$, all numbers $c_1, c_2, \ldots, c_k$, and all time shifts $h = 0, \pm 1, \pm 2, \ldots$.

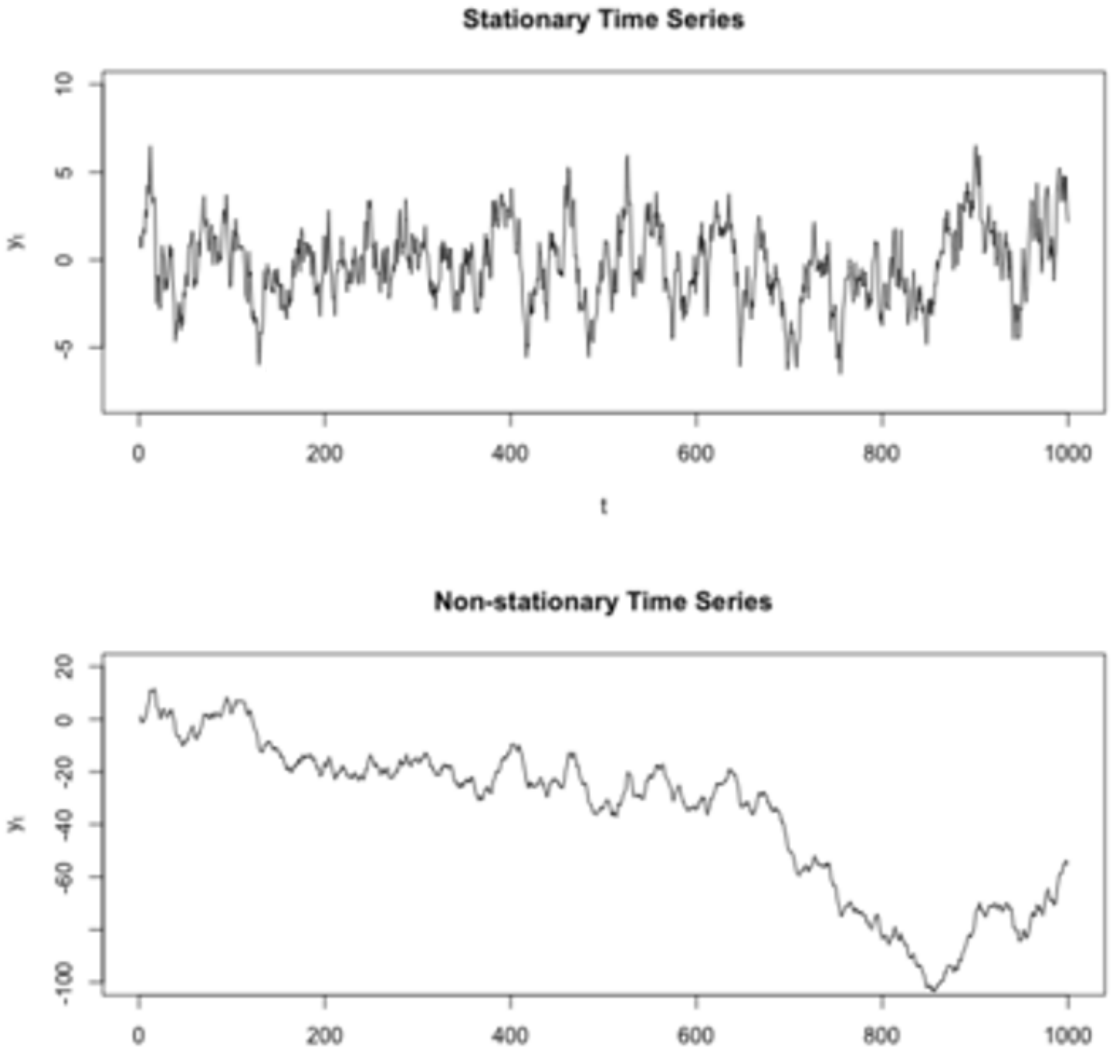

Figure 2.4: Two simulated time series processes are presented; one is stationary (upper) and the other is non-stationary (lower).

Definition 2.4 (Trend-Stationary)**.** A time series x_t is considered **Trend-Stationary** if it can be written as:

$$x_t = \mu_t + y_t$$

where x_t are the observations, μ_t denotes the trend, and y_t is a stationary process. μ_t is also called Deterministic Trend. This component represents the systematic, non-random part of the series, often modeled as a deterministic function of time (e.g., linear, polynomial). It captures the long-term direction or pattern of the data.

The presence of a deterministic trend, μ_t, suggests that while the series may grow or change direction over time, the deviations, y_t, from this trend are stable.

In an autoregressive trend stationary series, the model captures both the trend and the autoregressive nature of the data. Autoregressive trend stationary models are particularly useful in fields like economics and finance, where many time series

exhibit both long-term trends (e.g., economic growth) and short-term dynamics (e.g., cyclical behavior). They help in understanding the underlying patterns and improving forecasting accuracy.

Recognizing the properties of stationarity enables researchers and analysts to apply appropriate techniques to ensure that their models are robust and reliable.

2.5 Unit Root

A **unit root** is a characteristic of a time series that suggests it is non-stationary, meaning its statistical properties (like mean and variance) change over time. Essentially, it implies that past values have a long-lasting influence on future values, unlike stationary series where deviations are temporary.

Identifying unit roots is important because it helps determine whether a time series can be modeled and forecasted reliably. Time series with unit roots often exhibit random walks, where trends can drift indefinitely. This is useful for determining whether data should be differenced (a method of transformation) to make it stationary, which allows for better modeling, as most econometric models assume stationarity.

In economics and finance, many variables such as GDP, stock prices, and interest rates exhibit unit roots. For example, a shock to stock prices can have a permanent effect rather than a temporary one. Identifying and addressing unit roots in such data can prevent flawed conclusions in policy analysis or investment decisions, as it affects how we perceive long-term trends and predict future behavior.

Definition 2.5 (Unit Root)**.** Consider a discrete-time stochastic process $\{x_t, t = 1, 2, 3, \ldots\}$, and suppose that it can be written as an autoregressive process of order p (i.e., AR(p)):

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + e_t$$

Where x_t is a series of observations where x_0 is a fixed constant (for convenience, assume $x_0 = 0$), and $\phi_1, \phi_2, \ldots, \phi_p$ are constants ($\phi_p \neq 0$), and $\{e_t, t = 0, 1, 2, \ldots\}$ is a serially uncorrelated, zero-mean stochastic process with constant variance σ^2. If $m = 1$ is a root of the model's characteristic equation:

$$m^p - m^{p-1}a_1 - m^{p-2}a_2 - \cdots - a_p = 0$$

of multiplicity 1, then the stochastic process has a **Unit Root**.

Most statistical tests and models, like regression analysis, assume stationarity. If a series has a unit root, using these tests without first transforming the data can lead to spurious regression, where relationships between variables appear stronger than they are. This makes it critical to test for unit roots using unit root tests before applying standard techniques.

A **Unit Root Test** provides a way to test whether given observations reflect a random walk (the null case) as opposed to a causal process (the alternative).

By addressing unit roots, analysts can improve the robustness of their models, leading to better forecasts and more informed decision-making. There are several unit root tests, and the Dickey-Fuller test is one of the primary tests employed.

2.6 Dickey-Fuller test

The **Dickey-Fuller test** [5] tests the null hypothesis that a unit root is present in an autoregressive time series model. The basic model that is tested is the autoregressive model

$$Y_t = \rho Y_{t-1} + e_t, \quad t = 1, 2, 3, \ldots \tag{2.5}$$

When $Y_0 = 0$, ρ is a real number, and $\{e_t\}$ is a sequence of independent normal random variables with mean zero and variance σ^2 (which means, $e_t \sim \text{NID}(0, \sigma^2)$).

If $|\rho| < 1$, the time series Y_t converges (as $t \to \infty$) to a stationary time series. If $|\rho| = 1$, the time series is not stationary and the variance of Y_t is $t\sigma^2$. The time series with $|\rho| = 1$ is sometimes called a random walk. If $|\rho| > 1$, the time series is not stationary and the variance of the time series grows exponentially as t increases.

Given n observations $Y_1, \ldots, Y_n$, the maximum likelihood estimator of ρ is the least squares estimator.

$$\hat{\rho} = \left(\sum_{t=1}^{n} Y_{t-1}^2 \right)^{-1} \sum_{t=1}^{n} Y_t Y_{t-1} \tag{2.6}$$

The class of model to which Dickey-Fuller test relates is greater. The two following models are also treated:

$$(\mu) \quad Y_t = \mu + \rho Y_{t-1} + e_t, \quad t = 1, 2, 3, \ldots \tag{2.7}$$

$$(\tau) \quad Y_t = \mu + \beta t + \rho Y_{t-1} + e_t, \quad t = 1, 2, 3, \ldots \tag{2.8}$$

When, as before, $Y_0 = 0$.

We assume that n observations are available for analysis, and define the $n - 1$ dimensional vectors

$$1^T = (1, 1, 1, \ldots, 1)$$

$$t^T = \left(1 - \tfrac{n}{2}, 2 - \tfrac{n}{2}, 3 - \tfrac{n}{2}, \ldots, n - 1 - \tfrac{n}{2}\right)$$

$$Y_t^T = (Y_2, Y_3, Y_4, \ldots, Y_n)$$

$$Y_{t-1}^T = (Y_1, Y_2, Y_3, \ldots, Y_{n-1})$$

Let $U_1 = Y_{t-1}$, $U_2 = (1, Y_{t-1})$, and $U_3 = (1, t, Y_{t-1})$. Then, $\hat{\rho}_\mu$ and $\hat{\rho}_\tau$ are defined as the last entry of the vectors $(U_2^T U_2)^{-1} U_2^T Y_t$ and $(U_3^T U_3)^{-1} U_3^T Y_t$, respectively, and they are the least squares estimators of ρ under eq. (2.7) and (2.8), respectively.

The statistics analogous to the regression t statistics for the test of the null hypothesis, $H_0 : \rho = 1$, for all the models are:

$$\hat{\tau} = (\hat{\rho} - 1) \left(S_{e1}^2 c_1 \right)^{-1/2} \tag{2.9}$$

$$\hat{\tau}_\mu = (\hat{\rho}_\mu - 1) \left(S_{e2}^2 c_2 \right)^{-1/2} \tag{2.10}$$

$$\hat{\tau}_\tau = (\hat{\rho}_\tau - 1)\left(S_{e3}^2 c_3\right)^{-1/2} \tag{2.11}$$

When S_{ek}^2 is defined as the appropriate regression residual mean square:

$$S_{ek}^2 = (n-k-1)^{-1}\left[Y_t^T\left(I - U_k(U_k^T U_k)^{-1}U_k^T\right)Y_t\right], \quad k = 1, 2, 3$$

And c_k is the element of $(U_K^T U_K)^{-1}$ which is located lower-right coordinate.

We can conclude this section:

- **The Dickey-Fuller Test Purpose:** This test is used to determine whether an autoregressive time series has a unit root, indicating non-stationarity.
- **Null Hypothesis:** The series includes a unit root and thus is non-stationary.
- **Alternative Hypothesis:** The series is free from a unit root, and therefore more likely to be stationary.

2.7 Differential Privacy

Differential Privacy (DP) is a robust mathematical framework designed to release statistical information from datasets while safeguarding the privacy of individual data subjects. It allows data holders to share aggregate insights about a group while minimizing the risk of revealing specific information about individuals. This is achieved by adding carefully calibrated noise to statistical outputs, which preserves the utility of the data while limiting inferences that can be made about any single person in the dataset.

In simpler terms, differential privacy acts as a constraint on algorithms used to publish aggregate information from statistical databases, ensuring that private information about individual records remains confidential. For instance, government agencies employ differentially private algorithms to publish demographic data or statistical aggregates, ensuring the confidentiality of survey responses. Similarly, companies can analyze user behavior while controlling the visibility of sensitive information, even to internal analysts. The concept of differential privacy also involves the idea of a *Privacy Budget* or *Privacy Loss*. This budget is quantified using a parameter known as ϵ, which plays a critical role in balancing privacy and accuracy. A smaller ϵ value indicates a stronger privacy guarantee but typically results in reduced accuracy of the data. Conversely, a larger ϵ allows for more accurate results but comes at the cost of weaker privacy protections. By adjusting ϵ, organizations can tailor their privacy measures according to the sensitivity of the data and the context of its use. This trade-off is essential for ensuring that the utility of the information is maintained while still protecting individual identities from potential disclosure.

Before providing an explicit definition of differential privacy, it is essential to define its major components [1].

Definition 2.6 (Probability Simplex)**.** Given a discrete set B, the **Probability Simplex** over B, denoted $\Delta(B)$, is defined to be:

$$\Delta(B) = \{x \in \mathbb{R}^{|B|} : x_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^{|B|} x_i = 1\}$$

Definition 2.7 (Randomized Algorithm)**.** A **randomized algorithm** M with domain A and discrete range B is associated with a mapping $M : A \to \Delta(B)$. On input $a \in A$, the algorithm M outputs $M(a) = b$ with probability $(M(a))_b$ for each $b \in B$.

We are now ready to formally define differential privacy, which guarantees that a randomized algorithm behaves similarly on similar input databases.

Definition 2.8 (Differential Privacy)**.** A randomized algorithm (alternatively, mechanism) A is called (ϵ, δ)-**Differentially Private** if for all two datasets D, D' which differ in one record, and for all group of outputs S, it holds that [7]:

$$\Pr[A(D) \in S] \leq e^{\epsilon} \Pr[A(D') \in S] + \delta$$

where the probability is taken over the randomness used by the algorithm A. If $\delta = 0$, the guarantee is called pure differential privacy, or ϵ-Differential Privacy.

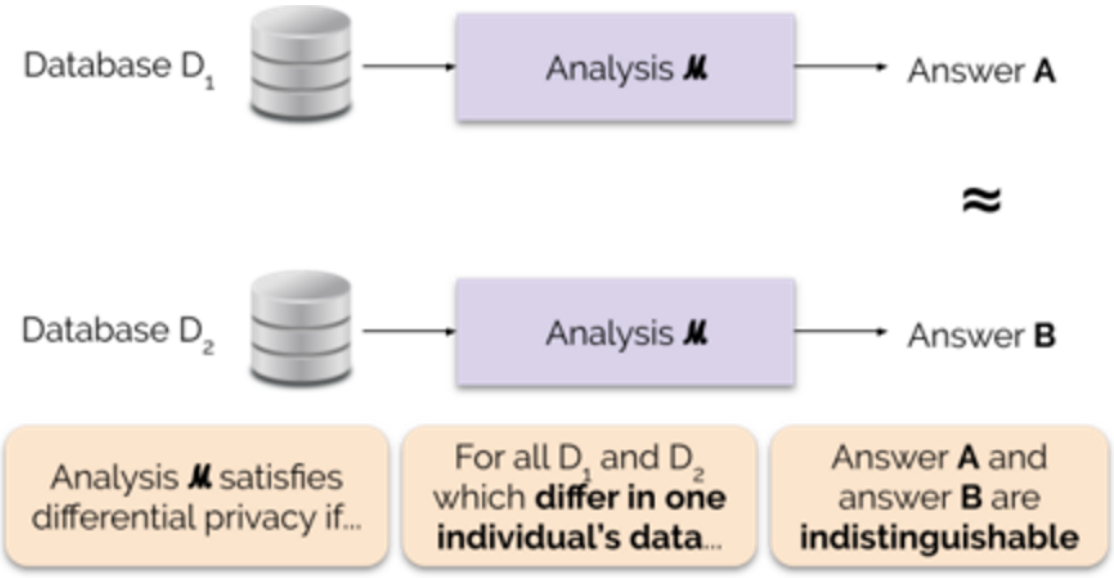

Figure 2.5: An abstract overview of a differential privacy algorithm $\mathcal{M}$

Fact 2.1 (Composition)**.** Let A, B two randomized algorithms and let S be a group of outputs. Let A be a (ϵ_1, δ_1)-differentially private, and B be a (ϵ_2, δ_2)-differentially private. When A, B are running independently, it holds that [3]:

$$\Pr\left[\langle A(D), B(D)\rangle \in S\right] \leq e^{\epsilon_1+\epsilon_2} \Pr\left[\langle A(D'), B(D')\rangle \in S\right] + (\delta_1 + \delta_2)$$

Means, that if A does one type of data analysis, and B does other one, and we publish both of the results together, we still preserve Differential Privacy, with higher "cost".

Fact 2.2 (Post-Processing)**.** Let A be a randomized algorithm. If A is (ϵ, δ)-differentially private, we get that $f(A)$ is also (ϵ, δ)-differentially private for every computation f which is independent of the dataset D.

These facts, which can be proven directly from the definition of differential privacy, reveal the robustness and flexibility of the framework. The first fact, *Composition* (2.1), illustrates how privacy guarantees can be preserved when multiple differentially private algorithms are applied to the same dataset. Specifically, if two randomized algorithms, A and B, each satisfy differential privacy with parameters (ϵ_1, δ_1) and (ϵ_2, δ_2), respectively, then when both algorithms are run independently and their results are published together, the combined process remains differentially private with the parameters $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$. This property demonstrates that differential privacy can still be achieved in complex data processing scenarios, although the privacy guarantee degrades slightly as more analyses are performed. This "privacy cost" is a critical consideration when designing systems that aim to perform multiple data analyses while maintaining privacy protection.

The second fact, *Post-Processing* (2.2), emphasizes that the privacy guarantees of a differentially private algorithm remain intact even after arbitrary post-processing steps. If an algorithm A is (ϵ, δ)-differentially private, then any function f applied to the output of A will also satisfy (ϵ, δ)-differential privacy, given that f does not depend on the original dataset D. This is an important feature because it ensures that downstream computations or decisions made based on the output of a differentially private algorithm cannot introduce additional privacy risks, further reinforcing the stability and reliability of differential privacy in practical applications.

A key concept underlying differential privacy is *Global Sensitivity*. This notion quantifies how much a function's output can change when a single record in the dataset is modified. Formally, the global sensitivity of a function measures the maximum difference in the function's output between any two datasets that differ by only one element. By understanding the global sensitivity of a function, we can determine the appropriate amount of noise to add in order to achieve differential privacy, ensuring that the function's privacy protection is robust against the worst-case changes in the input data. This concept is crucial in controlling the trade-off between privacy and utility, as lower sensitivity allows for less noise and, consequently, more accurate outputs while still preserving privacy.

Definition 2.9 (Global Sensitivity)**.** Let $q : u^n \to \mathbb{R}^d$ be a query function, where u^n is our universe, and $d \geq 0$. The **Global Sensitivity** is defined as:

$$\mathrm{GS}_q = \max_{(D,D')} \|q(D) - q(D')\|_1$$

Where the maximum is taken over any two datasets that differ in one record.

The global sensitivity of a function gives an upper bound on how much we must perturb its output to preserve privacy. Moreover, the global sensitivity of a function

depends on its mathematical structure. For example, when considering the mean function, the global sensitivity depends on the size of the dataset. Specifically, for a dataset of size n, the global sensitivity of the mean query is inversely proportional to n, since the change in a single record has a smaller impact on the average as the number of records increases. Formally, for a function that computes the mean of a dataset, the global sensitivity is $\frac{\Delta}{n}$, where Δ represents the maximum possible difference between individual data points, calculated as the upper bound minus the lower bound. This means that larger datasets, with more records, allow for smaller sensitivity and, consequently, less noise to be added, resulting in more accurate outputs while still preserving differential privacy.

Next, we introduce a fundamental mechanism used to achieve differential privacy, the Laplace mechanism. This mechanism ensures privacy by adding noise drawn from the Laplace distribution to the output of a function computed on the dataset. The magnitude of the noise is calibrated to the sensitivity of the function, ensuring that small changes in the input dataset result in small changes in the output distribution, which satisfies the differential privacy condition.

Definition 2.10 (The Laplace Distribution)**. The Laplace Distribution** (centered at 0) with mean μ and scale b is the distribution with probability density function (PDF):

$$\text{Lap}(x \mid b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

The variance of this distribution is $\sigma^2 = 2b^2$. We will sometimes write $\text{Lap}(b)$ to denote the Laplace distribution with mean 0 and scale b.

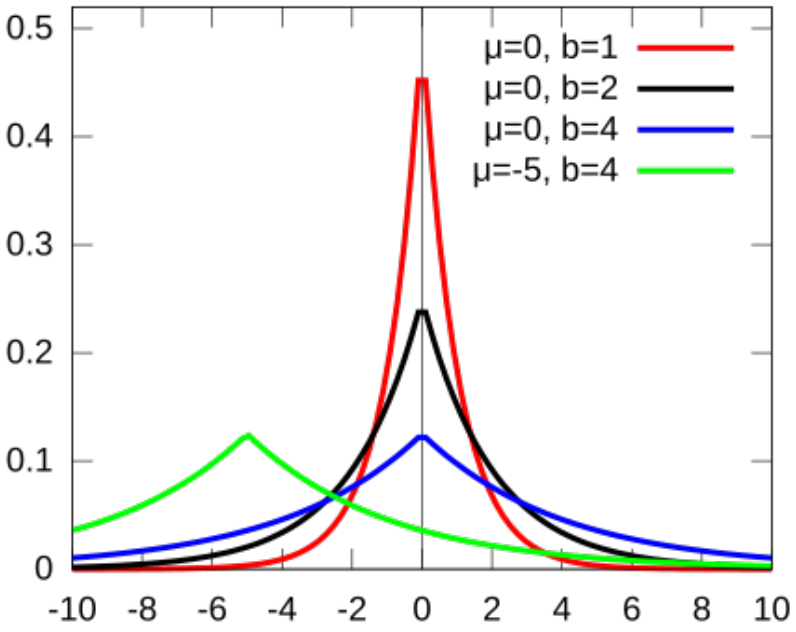

Figure 2.6: Examples of the Laplace probability density functions (PDF).

Definition 2.11 (The Laplace Mechanism [2])**.** Assume a universe $\mathcal{X}$ from which databases are collections of records. Given any function $f : \mathcal{X}^n \to \mathbb{R}^d$, the **Laplace Mechanism** is defined as:

$$M_L(x, f(\cdot), \epsilon) = f(x) + (Y_1, Y_2, \ldots, Y_d)$$

where $Y_1, Y_2, \ldots, Y_d$ are i.i.d. random variables drawn from $\text{Lap}\left(\frac{\text{GS}_f}{\epsilon}\right)$.

As its name suggests, the Laplace Mechanism computes f, and perturbs each coordinate with noise drawn from the Laplace distribution. The scale of the noise is calibrated to the sensitivity of f divided by ϵ of each coordinate.

Chapter 3

The Algorithm

3.1 Overview

As described above, several recent studies have concentrated on the challenge of hypothesis testing under differential privacy, where each individual contributes a single data point. Therefore, we have decided to investigate and address the challenge of conducting differentially private statistical hypothesis testing in scenarios when each individual provides an entire time series.

In this project, we are applying a differentially private method to the Dickey-Fuller test to ensure data privacy while performing this important statistical analysis. This approach allows us to test for a unit root in time series data while protecting the confidentiality of individual data points, balancing privacy concerns with accurate results.

As mentioned above, the Dickey-Fuller test's input is an autoregressive time series, in models corresponding to (2.5), (2.7) and (2.8). As such, our algorithm assumes N participants, where each one provides an autoregressive time series of length T. That means each individual provides T data points.

These T data points may be from any type of autoregressive series data, from all areas of interest, which follows one of the cases that Dickey-Fuller are design to analyze. These autoregressive series forms can be applied across various domains such as economics, finance, climate science, and public health.

For each individual's time series, according to the appropriate model's formula, we calculate the estimators:

- $\hat{\rho}$ - estimator for ρ, the coefficient of Y_{t-1} in each autoregressive time series, which, as mentioned, plays a crucial role in the analysis process.
- t-Value of the hypothesis test.
- Rejection Rate of the Null Hypothesis of the Dickey-Fuller test.

By these steps of analyzing the data, we in practice performing N times the analysis of the Dickey-Fuller test, one per each individual's time series.

After this step, we remain with three vectors of length N, as the number of individuals in the analysis: a vector of N $\hat{\rho}$ estimators, a vector of N t-values, and another vector of N rejection rates of the Dickey-Fuller test.

In the next step, we apply Differential Privacy in the algorithm. After reaching these three vectors of length N, we apply Differentially Private Mean calculation of these three estimators based on the Laplace Mechanism, leading us to three privacy preserving estimators, which represent the result of the analysis. From these we can learn and infer conclusions from the given dataset, detecting whether or not there is a unit root, which can inform us whether the dataset behaves as expected, allowing us to take appropriate actions. These actions might involve adjusting a patient's medical treatment or recommending changes to a country's economic policies to achieve improved GDP performance over time.

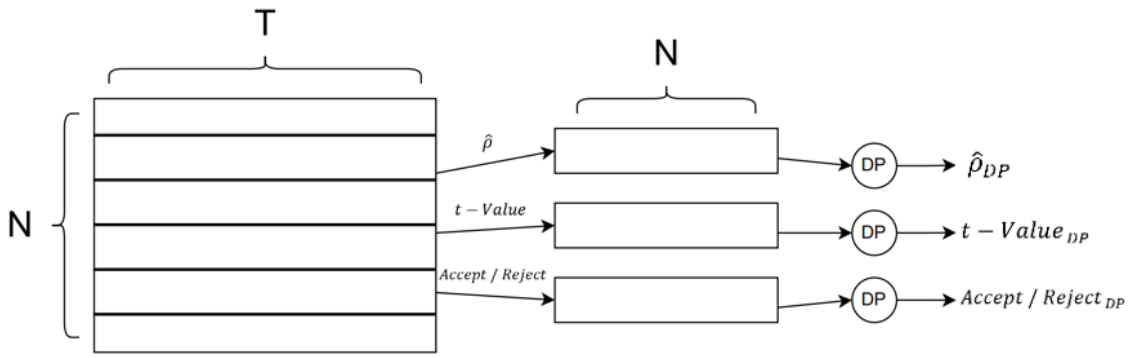

Figure 3.1: Our Algorithm overview – contains the described three major part, which are shown from left to right in this scheme.

3.2 Pseudo-Code

The Pseudo-Code of the algorithm can be formed as follows (See also Appendix (A) for the code):

Pseudo-Code: Differentially Private Autoregressive Time Series Dickey-Fuller Test Analysis

Input:

- Series_Matrix[N][T]: A matrix where each row corresponds to the autoregressive time series data of length T for each of the N individuals, from one of the forms Dickey-Fuller treats.

Output:

- $\hat{\rho}_{DP}$ Estimator: Differentially Private Mean of the $\hat{\rho}$ estimators of the N individuals.
- t-Value Estimator: Differentially Private Mean of the t-Values estimators of the hypothesis testing of Dickey-Fuller of the N individuals.

- Rejection Rate: Differentially Private Mean of the rejection rates of the Null Hypothesis of the Dickey-Fuller test of the N individuals.

Procedure:

1. **Initialize Empty Vectors:**
   - Rho_Estimators[N]: Vector to store the $\hat{\rho}$ estimators for each individual.
   - t_Values[N]: Vector to store the t-Values of the hypothesis test for each individual.
   - Rejection_Rates[N]: Vector to store the rejection rates of the Dickey-Fuller test for each individual.
2. **For Each Individual** $i = 1$ **to** N:
   - Extract Corresponding Time Series:
     - $Y_i \leftarrow$ Series_Matrix$[i]$
   - Perform Dickey-Fuller Test:
     - Calculate Rho_i: Estimation of ρ, the coefficient of $Y_{i(t-1)}$, from the autoregressive time series Y_i, based on the formula of the corresponding autoregressive rule.
     - Calculate t_Value_i: Perform calculation of the t-Value of this hypothesis testing, based on the formula of the corresponding autoregressive rule.
     - Calculate $Rejection\_Rate_i$: Assess the rejection rate of the null hypothesis from the Dickey-Fuller test, regarding the Y_i time series.
   - **Store Results:**
     - Rho_Estimators$[i] \leftarrow Rho_i$: Bounded between 0 to 1.3, $\Delta = 1.3$
     - t_Values$[i] \leftarrow$ t_Value$_i$: Bounded between -20 to 20, $\Delta = 40$
     - Rejection_Rates$[i] \leftarrow$ Rejection_Rate$_i$: Bounded between 0 to 1, $\Delta = 1$
3. **Apply Differential Privacy Mechanism:**
   - Rho_Mean_DP $\leftarrow \frac{1}{N}\sum_{i=1}^{N}$ Rho_Estimators$[i]$ + Lap $\left(\frac{3.9}{\epsilon N}\right)$
   - t_Value_Mean_DP $\leftarrow \frac{1}{N}\sum_{i=1}^{N}$ t_Values$[i]$ + Lap $\left(\frac{120}{\epsilon N}\right)$
   - Rejection_Rate_Mean_DP $\leftarrow \frac{1}{N}\sum_{i=1}^{N}$ Rejection_Rates$[i]$ + Lap $\left(\frac{3}{\epsilon N}\right)$
4. **Return:**
   - Rho_Mean_DP: Differentially private mean of the $\hat{\rho}$ estimators.
   - $t\_Value\_Mean\_DP$: Differentially private mean of the t-Values.
   - $Rejection\_Rate\_Mean\_DP$: Differentially private mean of the rejection rates.

End of Procedure

Claim. The described algorithm is $(\epsilon, 0)$-*Differentially Private.*
Proof. Let D and D' be two neighbors datasets. Hence, the three vectors we created are different on one record, and therefore they are also neighbors. Consequently, we can use for each of these metrics the Laplace mechanism, with global sensitivity (GS) of $\frac{\Delta}{n}$, i.e. Lap $\left(\frac{\Delta}{\epsilon N}\right)$. Pay attention that $\Delta_{\hat{\rho}} = 1.3$ (its values are bounded between 0 to 1.3), $\Delta_{t-Value} = 40$ (its values are bounded between -20 to 20) and $\Delta_{RejectionRate} = 1$ (its values are bounded between 0 to 1). Each of the three metrics has a privacy budget of $\epsilon' = \frac{\epsilon}{3}$, which yields $GS_{\hat{\rho}} = \frac{3.9}{\epsilon N}$, $GS_{t-Value} = \frac{120}{\epsilon N}$ and $GS_{RejectionRate} = \frac{3}{\epsilon N}$. From fact (2.1) we can conclude that the total privacy loss of the algorithm is ϵ, indeed. □

In the next section we explain in details the experiments that were done.

Chapter 4

The Experiments

4.1 Overview

Our major goal is to understand how Differentially Private Statistical Hypothesis Testing can applied when each individual gives a whole time series as his input data, instead only single data point. In order to examine this, we have chosen to investigate the Dickey-Fuller test performance under the constraint of preserving the privacy of each individual, and in which each individual in this test innately gives a whole autoregressive time series. The Dickey-Fuller test can treat three types of autoregressive time series, and in our experiments, we called models (2.5), (2.7) and (2.8) as model N (the native model), model C (similar to the native model, but with adding of a constant) and model CT (similar to model C, but with adding of a time-dependent trend). The Dickey-Fuller test, being a statistical hypothesis test, was conducted with a significance level of 0.05 to determine whether to reject the null hypothesis.

In order to achieve this ultimate goal, we perform four experiments.

In the first experiment we check the performance of the Dickey-Fuller test when we have multiple individuals, instead of one as the regular version, and compare the base test's performance to its differentially private version. In particular, we examine the Dickey-Fuller test's performance in all of its three autoregressive versions, with synthetic dataset, when each time the dataset fits the model that we check. The comparison is done between our results to the results of *Dickey and Fuller* in their original paper [5].

The second experiment builds on the first one. In the second experiment, we check how well the Dickey-Fuller test performs when we use the estimators that belong to the other models than the model of the dataset, and we do this for all three kinds of datasets.

In the third experiment, we seek to assess the robustness of the three autoregressive models with respect to differential privacy by comparing their performance to that of the standard Dickey-Fuller test. Additionally, we aim to explore the relationship between the sample size, denoted as N, and the privacy budget, represented by ϵ. Understanding this relationship is crucial for estimating the required privacy loss associated with a specified number of participants in the test. This will enable us to

achieve accurate test performance while ensuring the privacy of each participant is maintained.

The first three experiments are conducted on synthetic data.

In the fourth experiment, we analyze the Dickey-Fuller Test and compare its performance to that of its differentially private counterpart using a collection of data obtained from real-world observations rather than synthetic data.

As described, we make a comparison between the performance of the differentially private version to the regular non-differentially private algorithm, and one of our main focuses will be on which version of autoregressive time series dataset the differentially private algorithm performs the most similar to the non-differentially private algorithm. Moreover, we perform each test in the first and second experiments with a changing number of individuals (N). In addition, we checked the results with a grid of values of ρ, the coefficient of Y_{t-1}, that as mentioned in the Preliminaries (2.6), affects the test's result. The exact grid of values of ρ that we have chosen was examined by *Dickey and Fuller* in their article, $\{0.8, 0.82, 0.85, 0.88, 0.9, 0.92, 0.95, 0.99, 1, 1.02, 1.05\}$. Moreover, each experiment was checked with various values of privacy budget of the algorithm, ϵ, in order to see its effect on the results. We hypothesize that a larger privacy budget will yield results more closely resembling those obtained without the implementation of differential privacy. In addition, we assume each individual gives $T = 100$ data points of an autoregressive dataset.

4.2 Experiment 1

Goal: Evaluating the performance of the Dickey-Fuller test with multiple individuals and Differential Privacy.

Methodology: We have checked the Dickey-Fuller test with combination of groups of

$$N = \{20, 25, 30, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300\}$$

individuals, when each individual give three types of autoregressive time series who fits each of the three models that Dickey-Fuller test treats. We are checking these three types of autoregressive time series with different coefficient of Y_{t-1}, ρ, in the autoregressive model's formula

$$\rho = \{0.8, 0.82, 0.85, 0.88, 0.9, 0.92, 0.95, 0.99, 1, 1.02, 1.05\}$$

which is the grid of parameters that Dickey-Fuller originally checked in their research. We check also the results when performing the Differential Privacy algorithm with multiple privacy budget (privacy loss) values

$$\epsilon = \{0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40\}$$

In the autoregressive models (2.7) and (2.8), we defined the constant value as $\mu = 0.5$, and in model (2.8), we defined the trend's coefficient as $\beta = 0.2$. Moreover, as mentioned, we assume in this experience that each individual gives a time series of

length $T = 100$, and a statistical significance level of $\alpha = 0.05$. Because of the fact that we handle a dimension of probability here, we repeat all of these examinations 50 times and analyze the standard deviation (STD) of them.

Results: First, we can see that the STD of the values represented in the graphs is small so the results over the 50 repetitions of this experiment were close to each other, meaning they behave in the same way.

Second, we can see in the graphs of the Rejection Rates, that in all three models, and also for the Differentially Private Dickey-Fuller algorithm, the minimum of rejection rates is at $\rho = 1$, as described in the research of Dickey and Fuller. This means, that at $\rho = 1$ it is the most likely to accept the null hypothesis and declare the time series as a series that includes a unit root and thus non-stationary, as described in the Preliminaries (2.6).

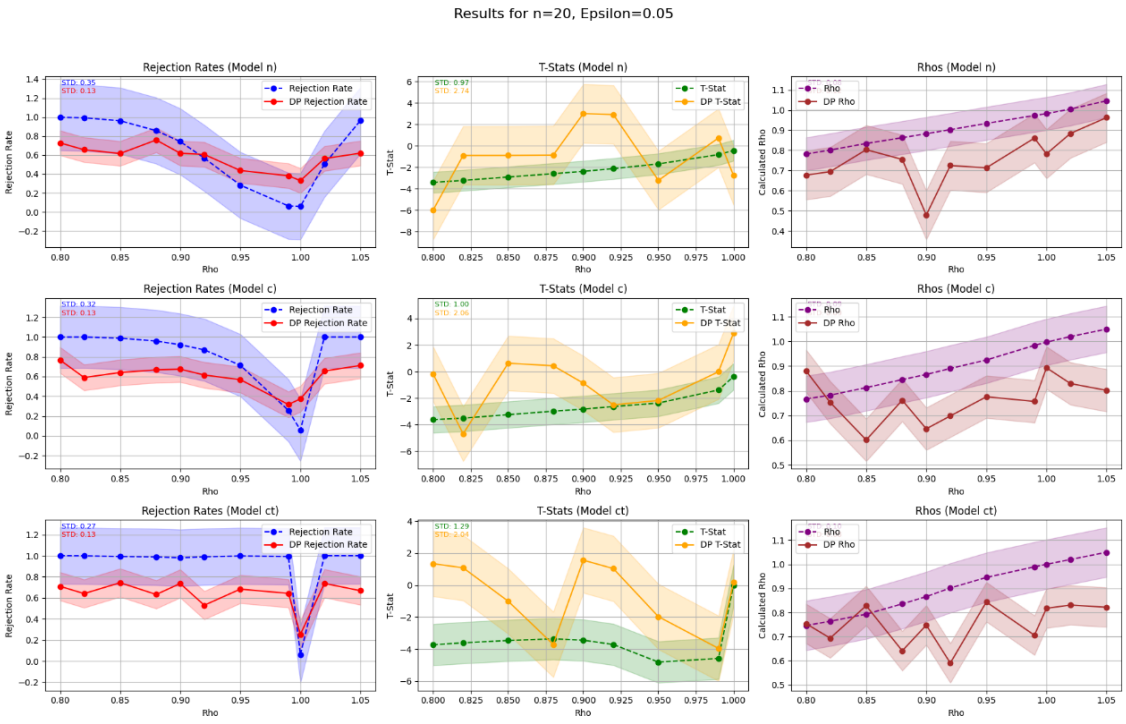

Figure 4.1: Results of the First Experiment for all three types of autoregressive models with and without Differential Privacy (mentioned as DP), with small amount $N = 20$ of individuals, and a small privacy budget $\epsilon = 0.05$. We can see that for a small number of individuals and a small privacy budget, the results of the Differentially Private algorithm are very noisy and not so accurate compared to the non-Differentially Private algorithm.

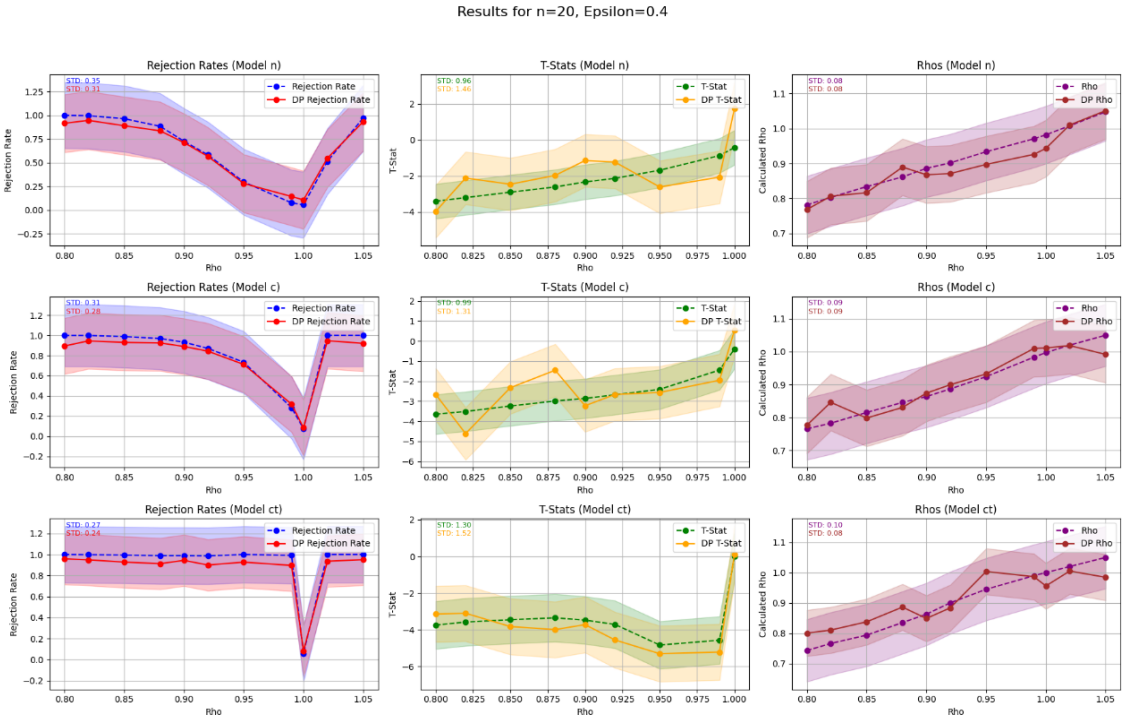

Figure 4.2: Results of the First Experiment for all three types of autoregressive models with and without Differential Privacy (mentioned as DP), with small amount $N = 20$ of individuals, and a high privacy budget $\epsilon = 0.40$. We can see that the results of the Differentially Private algorithm are closer to the non-Differentially Private algorithm, and are less noisy and more accurate compared to the case of Figure (4.1).

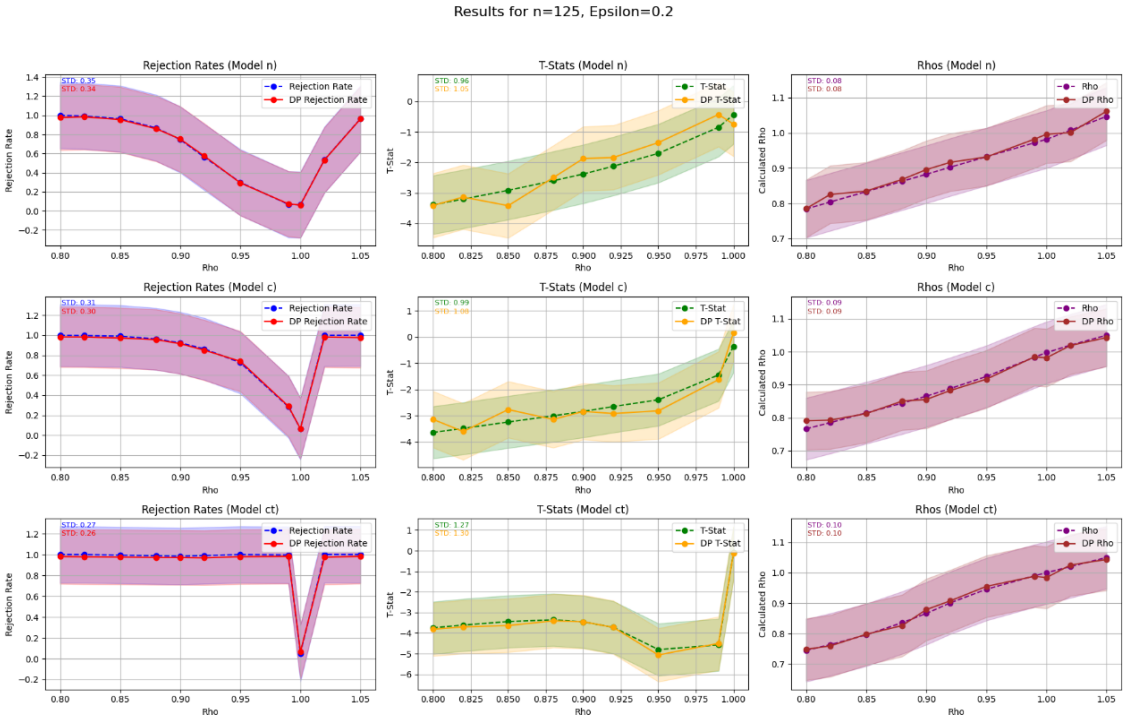

Figure 4.3: Results of the First Experiment for all three types of autoregressive models with and without Differential Privacy (mentioned as DP), with a medium amount $N = 125$ of individuals, and a medium privacy budget $\epsilon = 0.20$. We can see that the results of the Differentially Private algorithm are even less noisy and more accurate compared to the non-Differentially Private algorithm, and compared to the cases of Figure (4.1) and Figure (4.2).

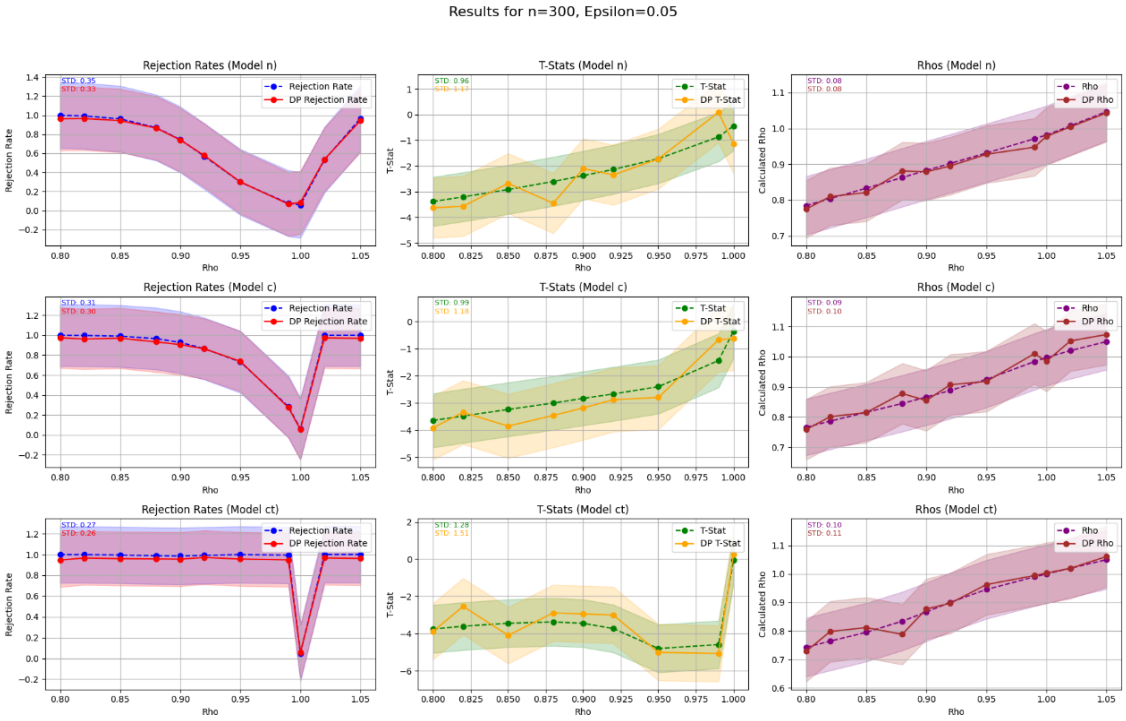

Figure 4.4: Results of the First Experiment for all three types of autoregressive models with and without Differential Privacy (mentioned as DP), with a large amount $N = 300$ of individuals, and a small privacy budget $\epsilon = 0.05$. We can see that the results of the Differentially Private algorithm are even less noisy and more accurate compared to the non-Differentially Private algorithm, and compared to the cases of Figure (4.1), in which we also had a small privacy budget of $\epsilon = 0.05$.

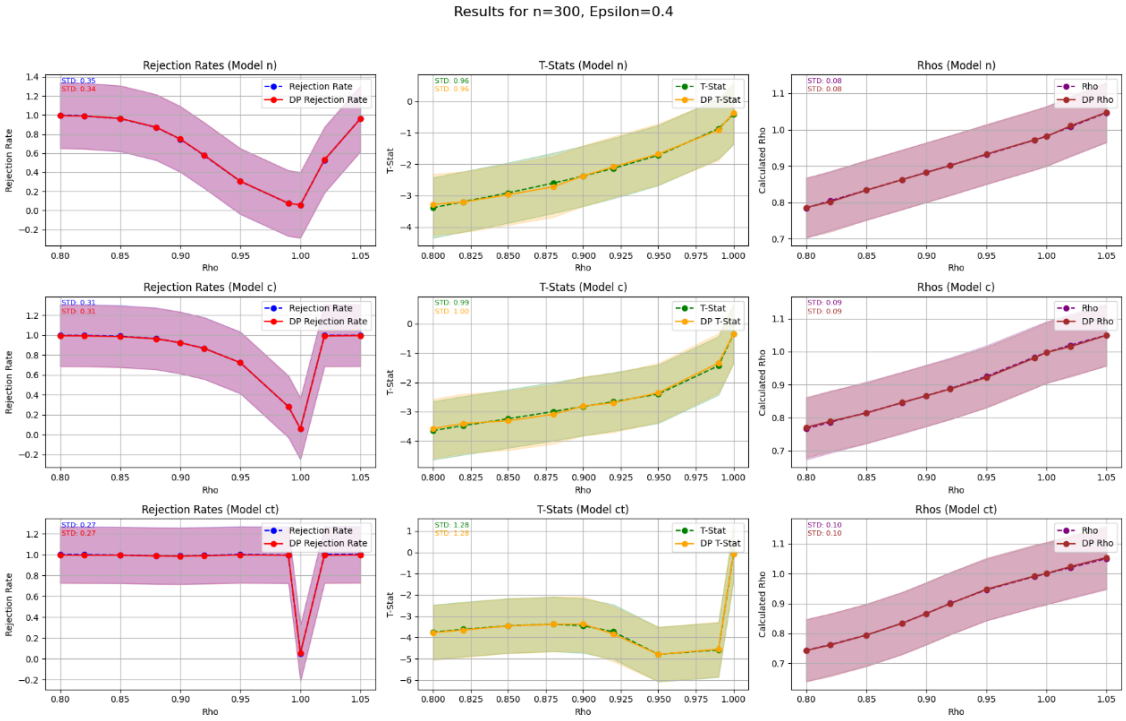

Figure 4.5: Results of the First Experiment for all three types of autoregressive models with and without Differential Privacy (mentioned as DP), with a large amount $N = 300$ of individuals, and a large privacy budget $\epsilon = 0.40$. We can see that the results of the Differentially Private algorithm are the least noisy and have the most accurate – almost a perfect fit, compared to all the above cases, and are the closest to the non-Differentially Private algorithm . We can see that for this number of individuals and this value of privacy budget, we can achieve similar results as the non-Differentially Private algorithm, so we can trust this method.

We can see that, as expected, the graphs of the Differentially Private Dickey-Fuller test algorithm (in red, yellow, and burgundy respectively) is getting closer to the graphs of regular test algorithm (in blue, green, and purple respectively), as the privacy budget (privacy loss) ϵ is growing. On the other hand, we can see that when the number of individuals, N, is growing, the graphs of the Differentially Private Dickey-Fuller test algorithm and the non-Differentially Private algorithm are also getting closer, meaning we can trust the results of the Differentially Private algorithm, which preserves individuals' privacy, as the regular algorithm, and therefore one would prefer to use it instead of the regular version. These phenomena will also be discussed in later section. Furthermore, the rejection rate curve exhibits the steepest slope at $\rho = 1$ for model CT, followed by model C, while model N displays the least pronounced slope. Notably, although the $t - Statistic$ curve (i.e. t-Value, designated as the $T - Stas$, in the middle) for models N and C consistently demonstrate an upward trajectory, the CT model exhibits a decline for the range of $0.90 < \rho < 0.95$. However, this is followed by a substantial increase for the interval of $0.95 < \rho < 1.00$.

4.3 Experiment 2

Goal: Assessing the robustness of the Dickey-Fuller test when applied to estimators from alternative autoregressive models across all three dataset types, and comparing the results with those obtained from the corresponding model-specific estimators to examine the test's adaptability and effectiveness under different conditions.

Methodology: We defined the same grid of parameters as previous experiment. First, we check what happens to the Dickey-Fuller test, both in regular version and Differential Private version, when we use the dataset in the form of the first type (2.5) for all three models. Then, we did the same for the two other models (2.7) and (2.8).

As in the previous experiment, we repeat all the examinations in this experiment 50 times.

Results: First, we applied dataset that fits model n (2.5) for all three models.

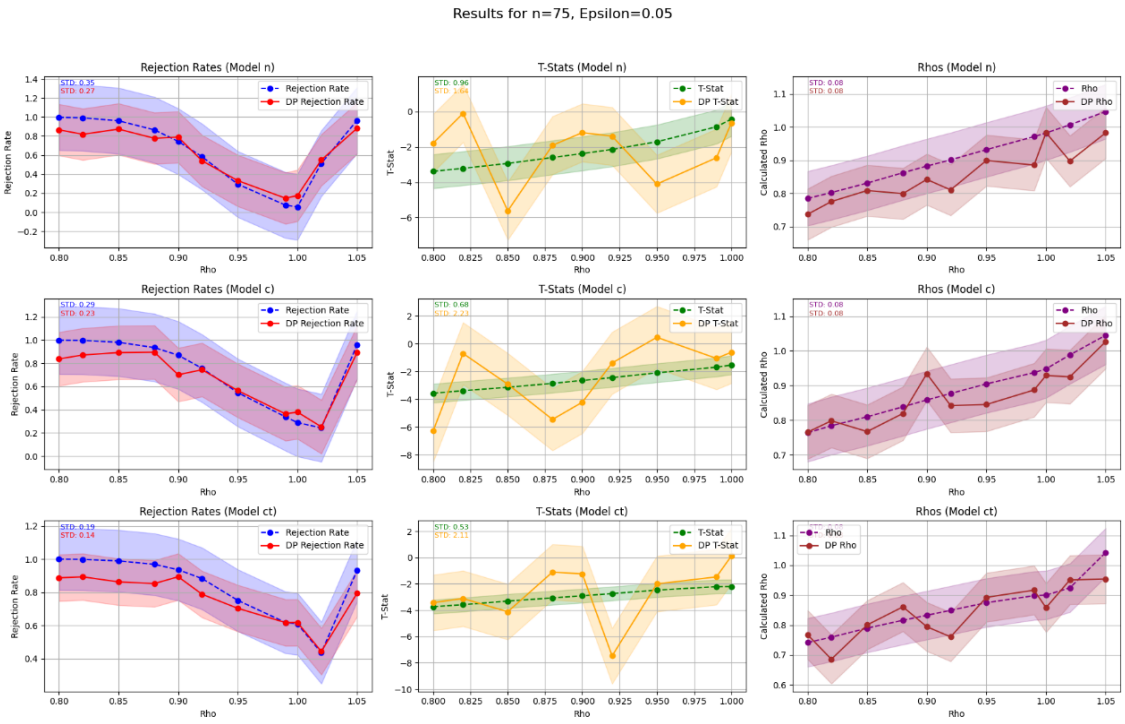

Figure 4.6: Results of the Second Experiment for dataset from the first type of autoregressive applied on for models, with and without Differential Privacy (mentioned as DP), with small amount $N = 75$ of individuals, and a small privacy budget $\epsilon = 0.05$. We can see that similarly to the first experiment, for a small number of individuals and a small privacy budget, the results of the Differentially Private algorithm are noisy and not so accurate compared to the non-Differentially Private algorithm.

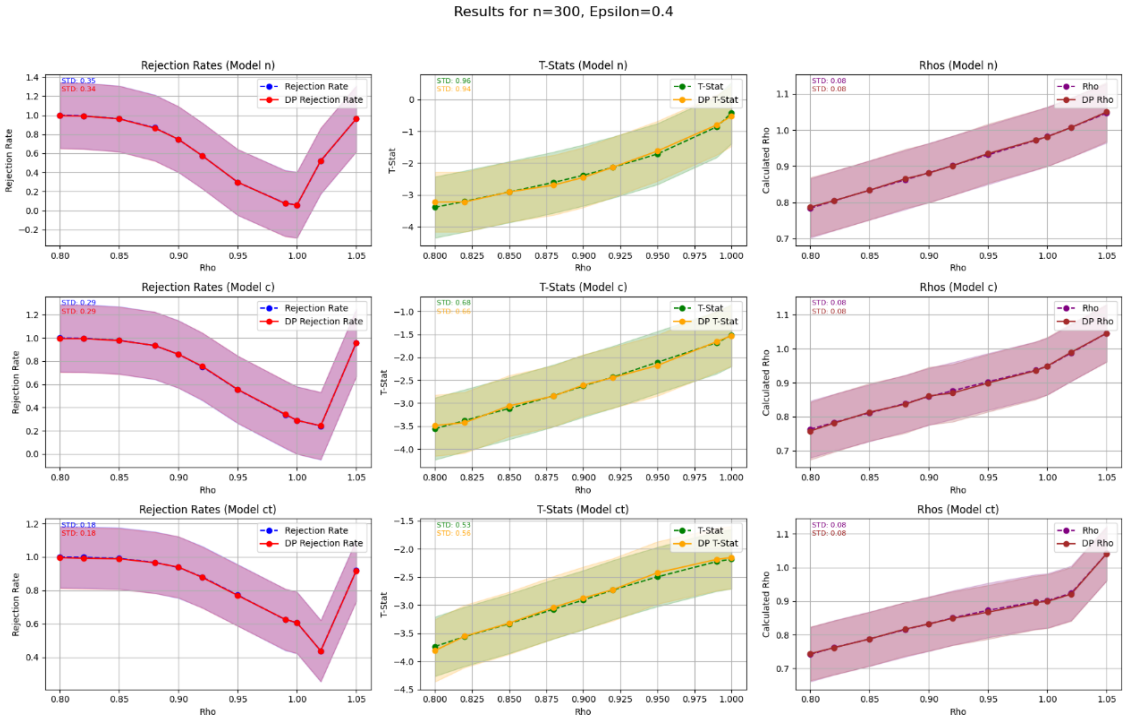

Figure 4.7: Results of the Second Experiment for dataset from the first type of autoregressive applied on for models, with and without Differential Privacy (mentioned as DP), with large amount $N = 300$ of individuals, and a large privacy budget $\epsilon = 0.40$. We can see that similarly to the first experiment, for a large number of individuals and a large privacy budget, similarly to Figure (4.5), that we achieve almost a perfect accuracy compared to the regular, non-differentially private algorithm.

The results presented indicate that, for model N (2.5), from which the data was generated, we observe behavior consistent with the first experiment, as anticipated. Notably, we find that for models C (2.7) and CT (2.8), the minimum value for the graph of the rejection rates occurs at $\rho = 1.02$, rather than at $\rho = 1.00$ as observed in the initial experiment. This is a noteworthy finding, particularly given that models C and CT incorporate a constant (μ) and a constant with trend (μ,β), respectively, in as an addition to model N. It suggests that for models c and ct, the constraints $\mu \neq 0$ and $\beta \neq 0$ must be satisfied for their estimators to predict accurately.

In addition, we applied dataset that fits model C (2.7) for all three models:

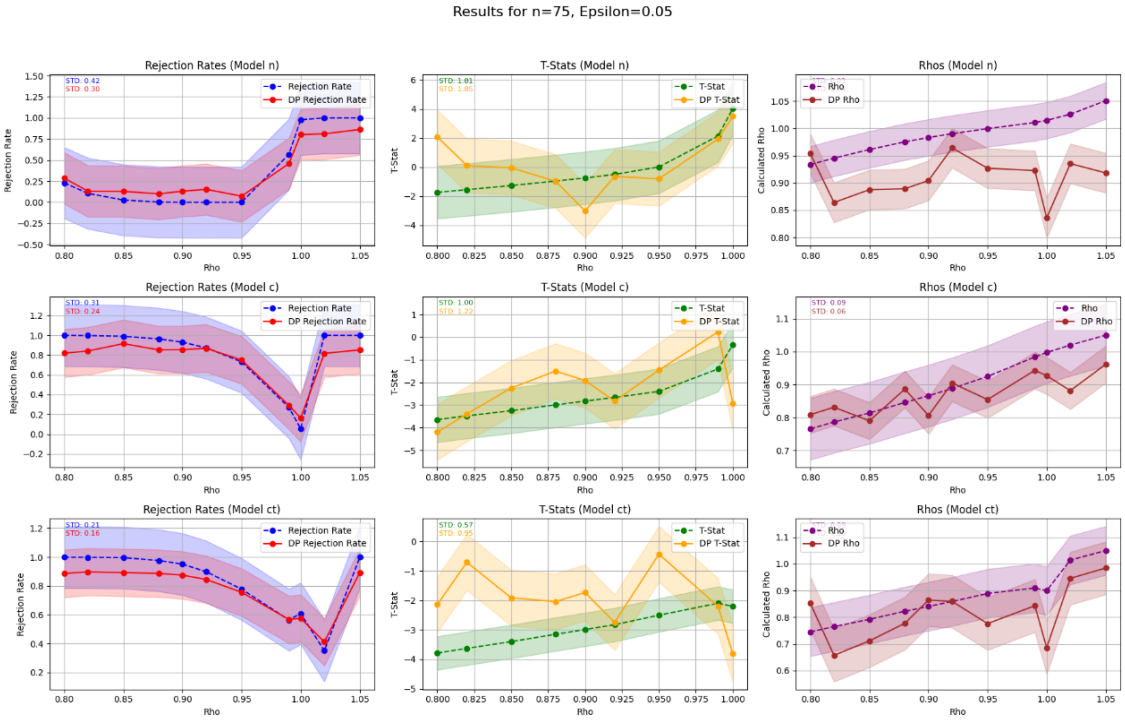

Figure 4.8: Results of the Second Experiment for dataset from the second type of autoregressive applied on for models, with and without Differential Privacy (mentioned as DP), with small amount $N = 75$ of individuals, and a small privacy budget $\epsilon = 0.05$. As in the first experiment, a small number of individuals and privacy budget yield noisy, less accurate results.

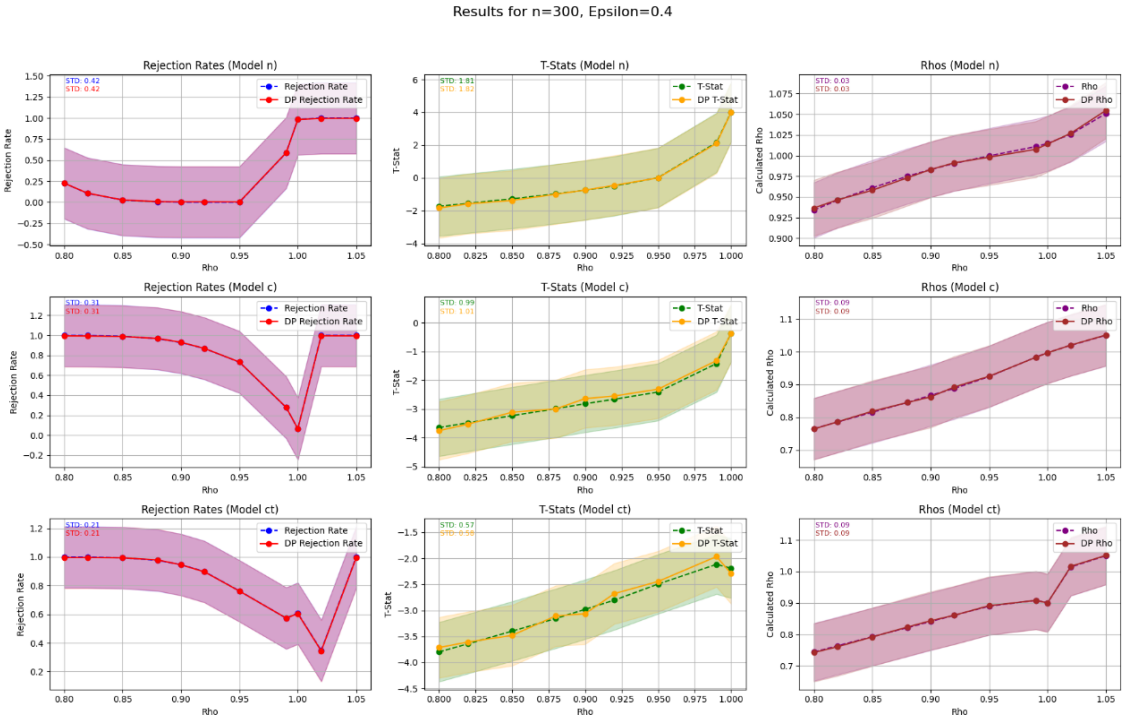

Figure 4.9: Results of the Second Experiment for dataset from the second type of autoregressive applied on for models, with and without Differential Privacy (mentioned as DP), with large amount $N = 300$ of individuals, and a large privacy budget $\epsilon = 0.40$. As in the first experiment and shown in Figure (4.5), a large number of individuals and a large privacy budget yield nearly perfect accuracy compared to the non-Differentially Private algorithm.

These results indicate that model C (2.7), from which the data was generated, exhibits behavior consistent with the first experiment, as expected. In contrast, for model CT (2.8), the minimum value for the graph of rejection rates occurs at $\rho = 1.02$, rather than at $\rho = 1.00$ as seen in the initial experiment. This is a significant finding, especially since model ct includes a trend (β), respectively, in an addition to model C. It suggests that the constraint $\beta \neq 0$ must be satisfied for model ct's estimators to predict accurately. A similar phenomenon was observed when applying datasets which fit model N to the estimators of all three models.

We also observe that under these settings, model N produces results that are essentially meaningless, as its graph of rejection rates resembles a sigmoid function, achieving a rejection rate of approximately 100% at $\rho = 1.00$, rendering it ineffective. Specifically, its estimators are unable to make any predictions when provided with data from the second type (2.7). This highlights the limitations of model N in this context, as it is unable to adapt to the characteristics of the model C datasets (2.7), leading to a complete lack of predictive capability in this case.

Moreover, we applied dataset that fits model CT (2.8) for all three models:

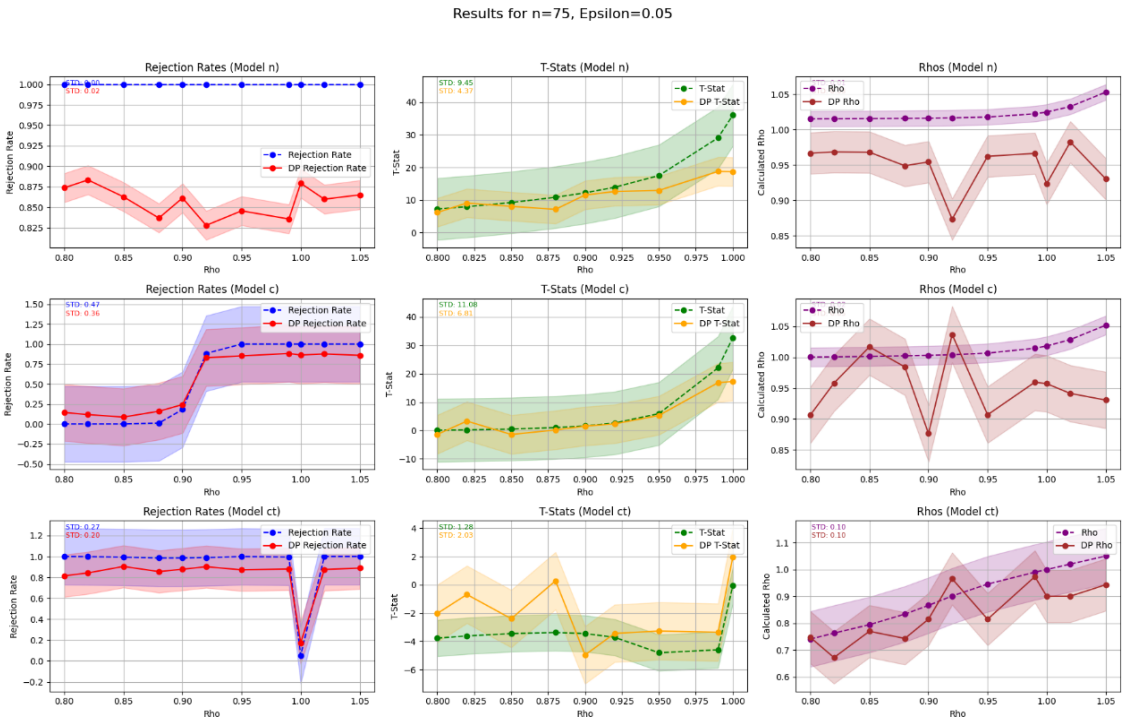

Figure 4.10: Results of the Second Experiment for dataset from the third type of autoregressive applied on for models, with and without Differential Privacy (mentioned as DP), with small amount $N = 75$ of individuals, and a small privacy budget $\epsilon = 0.05$. As in the first experiment, a small number of individuals and privacy budget lead to noisy, less accurate outcomes.

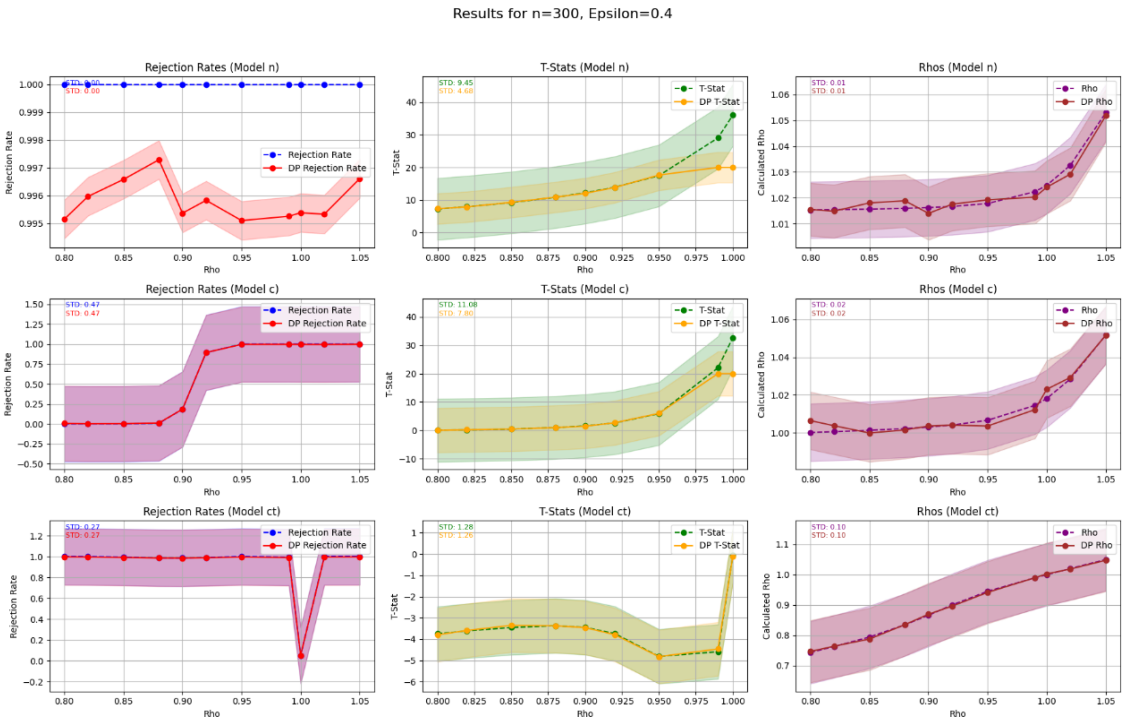

Figure 4.11: Results of the Second Experiment for dataset from the third type of autoregressive applied on for models, with and without Differential Privacy (mentioned as DP), with large amount $N = 300$ of individuals, and a large privacy budget $\epsilon = 0.40$. As in the first experiment and illustrated in Figure (4.5), a large number of individuals and a substantial privacy budget result in nearly perfect accuracy compared to the non-Differentially Private algorithm.

The results indicate that model CT (2.8), from which the data was generated, exhibits behavior consistent with the first experiment, as expected.

We also find that model C (2.7) yields results that are essentially meaningless, as its graph of rejection rates takes on a sigmoid shape, reaching nearly 100% rejection at $\rho = 1.00$, rendering it ineffective. Its estimators cannot make any predictions when provided with data from the third type (2.8). This illustrates the limitations of model C in this context, as it unsuccessfully accommodates the characteristics of the model CT datasets (2.8), resulting in a complete lack of predictive capability.

Additionally, in the non-differentially private algorithm, the autoregressive model N (2.5) is consistently rejected 100% of the time for every value of ρ. The differentially private algorithm exhibits similar behavior, producing random values—due to the application of the Laplace mechanism for mean calculation—that approach a rejection rate of 100%. This indicates that model N cannot be adapted to the dataset of model CT, leading to meaningless estimator values in this scenario.

4.4 Experiment 3

Goal: This experiment has two key objectives. First, we aim to determine which of the three autoregressive models exhibits the greatest robustness to the effects of the differential privacy mechanism, compared to their performance in the standard Dickey-Fuller test. Second, we seek to explore the relationship between sample size, denoted as N, and the privacy budget, represented by ϵ. Understanding this relationship is crucial for estimating the required privacy loss associated with a specified number of participants in the test, enabling us to achieve accurate test performance while preserving the privacy of each participant.

Methodology: We defined the same grid of parameters as the first experiment. As in the previous experiments, we repeat all the examinations in this experiment 50 times.

Results: In order to examine which model exhibits the best robustness to the effects of the differential privacy mechanism, we measured the Mean Squared Error (MSE) between the differentially and the non-differentially private algorithm for each of the performance metrics displayed in the graphs – Rejection Rates, T-Stats (T-Statistic, or T-Value) and $\hat{\rho}$, the estimator of ρ.

We conducted this analysis for each of the three models applicable to the Dickey-Fuller test and for every combination of N, the number of individuals in the test, and ϵ, the privacy budget of the differentially private algorithm.

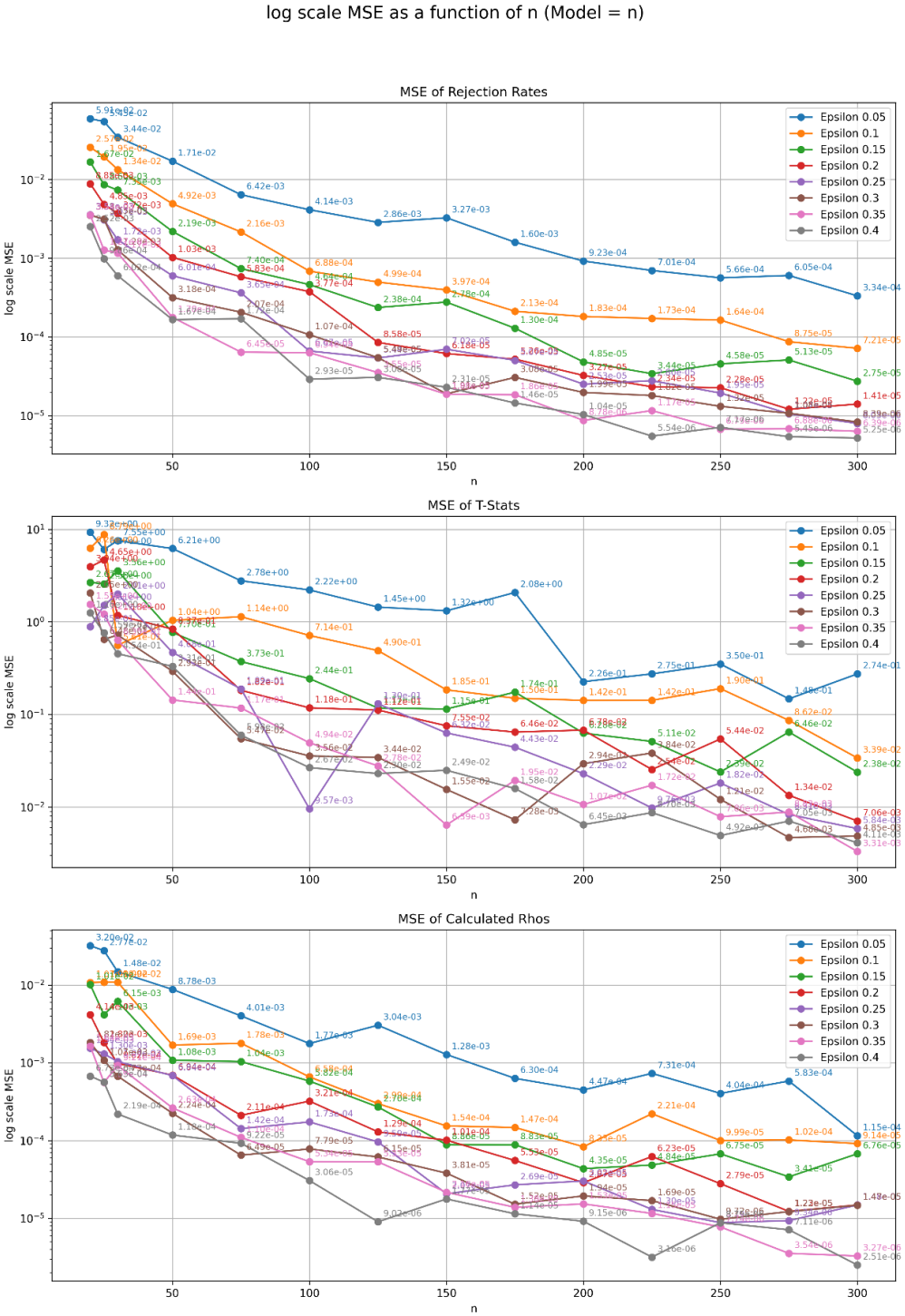

Figure 4.12: Results of the Third Experiment, for model N, the first type of autoregressive models handled by Dickey-Fuller test (2.5). The graphs are shown in log scale of the Mean Squared Error (vertical axis), and these values are shown for every number of individuals, N (horizontal axis), and for every privacy budget ϵ in the grid of values that we check.

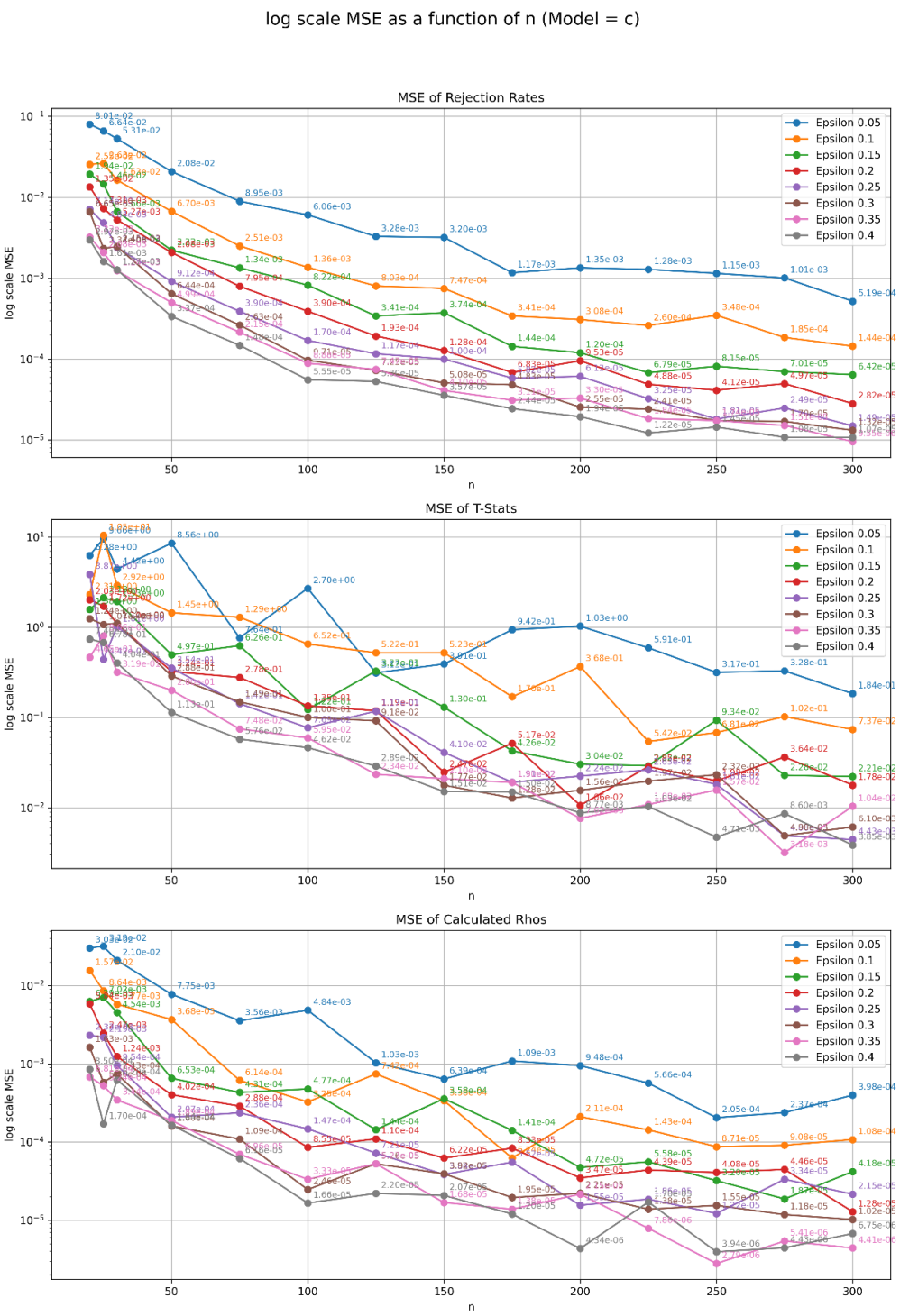

Figure 4.13: Results of the Third Experiment, for model C, the second type of autoregressive models handled by Dickey-Fuller test (2.7). The graphs display the Mean Squared Error (MSE) on a logarithmic scale on the vertical axis. These values are shown for each sample size N on the horizontal axis, as well as for every privacy budget ϵ within the range of values examined.

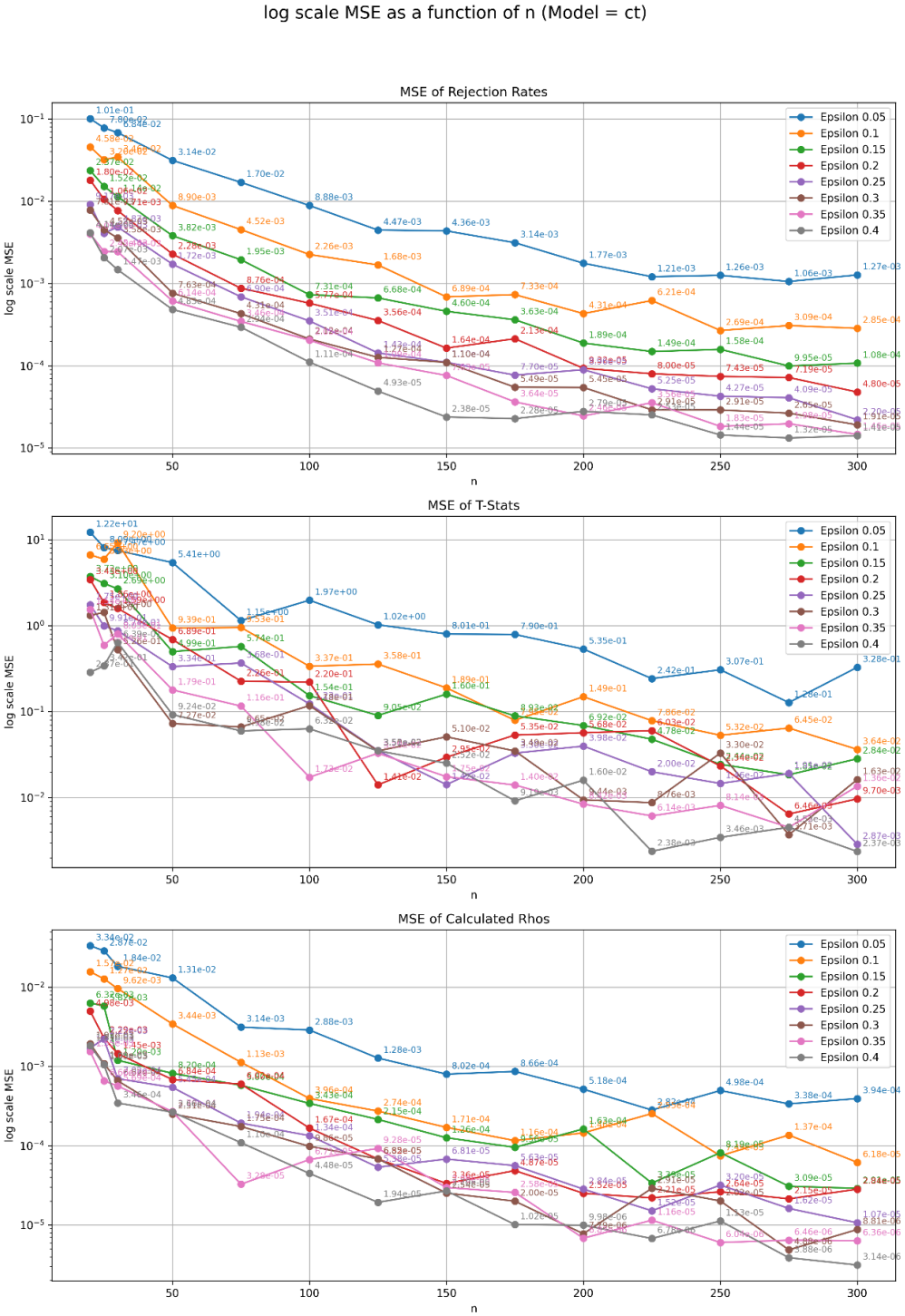

Figure 4.14: Results of the Third Experiment, for model CT, the third type of autoregressive models handled by Dickey-Fuller test (2.8). The graphs illustrate the Mean Squared Error (MSE) on a logarithmic scale along the vertical axis. The values are represented for each sample size N on the horizontal axis, alongside every privacy budget ϵ within the examined range of values.

First, it is evident that for lower sample sizes N, the Mean Squared Error (MSE) is larger. This illustrates that a smaller number of individuals, combined with the inherent randomness introduced by the principles of differential privacy, results in noisier outcomes compared to the non-differentially private, standard Dickey-Fuller test. Similarly, a smaller privacy budget ϵ also corresponds to a higher MSE. Conversely, as anticipated, increasing both N and ϵ leads to a decrease in the MSE.

In general, we can see that the MSE values for all of the three metrics, which displayed above are low, so using the differentially private algorithm is trusted and should be preferred. Furthermore, by comparing the MSE graphs of the Rejection Rates of the three models, we can see that we overall get the minimal value of MSE for model N, then for model C, and for model CT the MSE values are the greatest. The main importance of the Dickey-Fuller test is its null hypothesis rejection rates and therefore we can declare model N to be the most robust for such differential privacy application.

Overall, the Mean Squared Error (MSE) values for all three metrics displayed above are relatively low, indicating that the use of the differentially private algorithm is reliable and should be preferred. Additionally, a comparison of the MSE graphs for the Rejection Rates across the three models reveals that model N achieves the lowest MSE, followed by model C, while model CT exhibits the highest MSE values. Given that the primary significance of the Dickey-Fuller test lies in its ability to accept or reject the null hypothesis, we can conclude that model N is the most robust option for such differential privacy application.

Next, we want to assert the connection between the numbers of individuals, N, to the chosen privacy budget for the test, ϵ. It indicates from the above graphs, that we can see a clear inverse relationship between the parameters N,ϵ and their reflected MSE values. For example, for model n (2.5) with $N = 50$, $\epsilon = 0.40$ we receive $MSE_{N,\epsilon} = 1.67 \times 10^{-4}$, and with $N = 200$, $\epsilon = 0.10$, the inverse relationship, we receive $MSE_{N,\epsilon} = 1.83 \times 10^{-4}$. Such relationship holds for every inverse relationship between the parameters N,ϵ, across all three models and metrics. Understanding this connection is crucial, as it can guide one in making wisely and informed decisions regarding the privacy budget based on the number of participants expected to engage in the test.

4.5 Experiment 4

Goal: Comparing the performance of the Dickey-Fuller Test with its differentially private counterpart using a dataset from real-world observations.

Methodology: We have chosen to examine our Differentially Private algorithm's performance with the World Bank GDP Dataset,
(https://data.worldbank.org/indicator/NY.GDP.MKTP.CD).

The World Bank GDP dataset is a comprehensive and highly valuable time series dataset that offers deep insights into the economic performance of nearly every country and region (for example, East Asia and Pacific) worldwide.

GDP, or Gross Domestic Product, is one of the most important indicators used to measure a country's economic activity. It represents the total monetary value of all goods and services produced within a country's borders over a specific time period, typically measured annually. As a fundamental gauge of an economy's health, GDP is widely used by policymakers, economists, and analysts to assess economic growth, compare the economic performance of different countries, and guide decisions on fiscal and monetary policies.

In the context of time series analysis, GDP data often exhibit a trend stationary process. Over time, most countries experience a long-term upward trend in GDP, primarily driven by factors such as population growth, technological advancements, and capital accumulation. This long-term growth trend reflects the natural progression of economies as they expand due to increases in productivity, investment, and innovation. However, while GDP generally trends upward, it is also subject to cyclical fluctuations driven by business cycles. These cycles are characterized by alternating periods of expansion (growth) and contraction (recession), which occur due to a variety of economic forces such as demand shocks, supply shocks, changes in consumer confidence, and policy interventions.

GDP time series data is also influenced by external shocks, such as financial crises, wars, pandemics, or major policy changes. These events can cause significant short-term deviations from the long-term growth path. However, in many cases, economies tend to revert to their growth trend over time, exhibiting a degree of mean reversion. For instance, after recessions or economic downturns, economies often recover, returning to their previous growth trajectory. This characteristic of reverting to a trend after deviations makes GDP data suitable for certain types of time series models, particularly trend stationary models.

When analyzed through the lens of econometrics, GDP is often modeled using autoregressive (AR) time series models. These models are particularly useful because they account for the persistence in the data—that is, the tendency for current GDP values to be influenced by previous values. Autoregressive models are designed to capture the idea that past economic performance has a strong influence on future outcomes. For many countries and regions, GDP exhibits a structure where it follows a deterministic trend over time, with cyclical fluctuations around that trend. This makes GDP a prime candidate for AR models, which can incorporate both the trend and the cyclical components to provide more accurate forecasts and insights into economic dynamics.

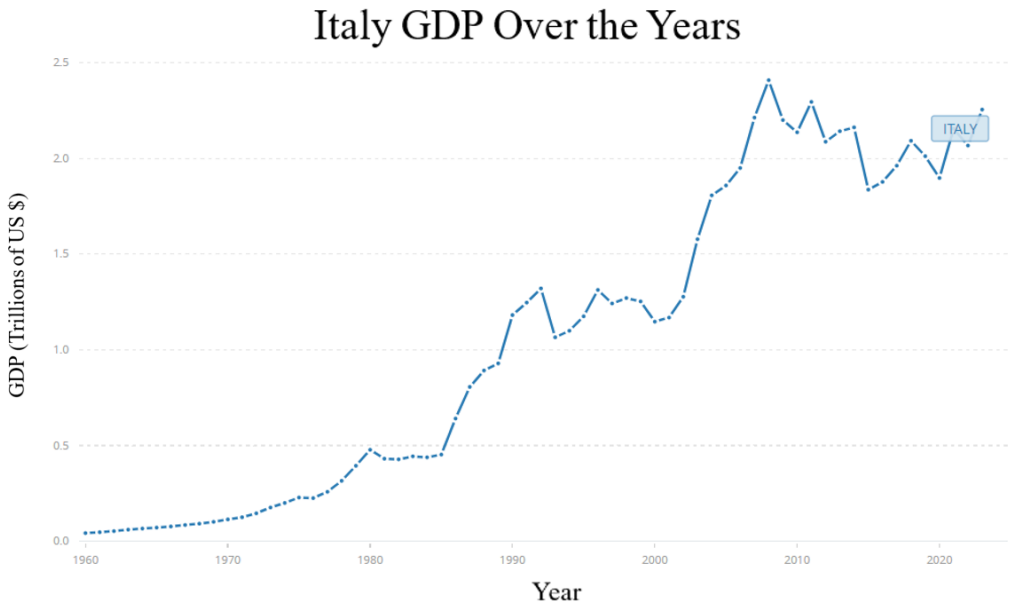

Figure 4.15: The GDP growth of Italy over the years. The trend-stationary behavior is reflected in the graph.

However, not all countries exhibit such smooth autoregressive trends. In some developing or politically unstable nations, GDP may be far more volatile and subject to erratic fluctuations, making trend stationarity harder to observe. For these countries, shocks like political instability, natural disasters, or severe economic crises may lead to structural breaks in the data, disrupting the otherwise steady trend.

The dataset was obtained in *.csv* format and subsequently underwent preprocessing to align with the requirements of our analysis. This preprocessing involved the removal of countries or regions with incomplete data, resulting in a final dataset comprising 134 countries and regions for which time series data is available. For the experiment, we selected a privacy budget of $\epsilon = 0.40$, informed by its demonstrated stability in previous experiments.

Results: From our examination of the graphs of the countries and regions in the dataset, we determined that the appropriate model for this experiment is model CT (2.8). Consequently, we decided to perform both the differentially private and non-differentially private algorithms on this dataset using model CT. We calculated the estimator of ρ ($\hat{\rho}$), and based on the results from the first experiment (4.2), specifically those involving 125 individuals, as shown in Figure (4.3)—which is the closest number to the 134 we have in this experiment—we gathered the $RejectionRate$ of the null hypothesis for this experiment. The results are presented in the table below:

Table 4.1: Summarization of the rho estimator $\hat{\rho}$ of the null hypothesis for both the differentially private and non-differentially private algorithms regards model CT (2.8).

Dickey-Fuller Test Algorithm	Regular	Differentially-Private
$\hat{\rho}$	0.953	0.943

As mentioned, based on Figure (4.3), we can conclude that the $RejectionRate$ of the null hypothesis for this dataset is close to 100%. Therefore, as expected, there is no unit root, and we can infer that the dataset is trend-stationary.

Chapter 5

Conclusions

In this project, we have developed and evaluated a novel approach for performing differentially private statistical hypothesis testing on time series data, specifically focusing on the Dickey-Fuller unit root test. Our work addresses a significant gap in existing research, which has largely been limited to single-point data contributions per individual. By extending the scope to allow individuals to contribute entire time series, we tackled the unique challenges of balancing privacy concerns with the need for accurate statistical inference.

We applied differential privacy mechanisms to the autoregressive models underlying the test, and through a series of rigorous experiments, we demonstrated how the differentially private version of the Dickey-Fuller test can maintain a high level of accuracy, even as privacy protections are enhanced. The results of our experiments reveal important insights into the interplay between differential privacy, sample size, privacy budgets, and the performance of the Dickey-Fuller test.

The first key finding from our research is that the differentially private Dickey-Fuller test remains robust across varying privacy budgets and sample sizes. As we expected, our results indicate that as the privacy budget, ϵ, increases, the performance of the differentially private test approaches that of the non-private test. This trade-off between privacy and accuracy is a crucial aspect of any differentially private algorithm, and our results show that for practical purposes, a well-chosen privacy budget allows for strong privacy guarantees without significantly compromising the utility of the results.

Another significant observation is that the performance of the differentially private test improves with larger sample sizes. As we increased the number of individuals in our experiments, we observed a marked decrease in the noise introduced by the privacy mechanism, leading to more accurate results. This is consistent with theoretical expectations, as larger datasets reduce the sensitivity of the average. This finding suggests that differential privacy mechanisms are particularly well-suited for large-scale time series analyses, where the number of participants is substantial enough to mitigate the privacy-accuracy trade-offs.

Furthermore, our experiments comparing different autoregressive models (Model N (2.5), model C (2.7), and model CT (2.8)) highlighted the robustness of certain models in the presence of differential privacy. Specifically, model N demonstrated the

greatest resilience to privacy noise, followed by model C and model CT. This suggests that in real-world applications where differential privacy is a concern, selecting the appropriate autoregressive model is critical for achieving the best possible balance between privacy and accuracy.

In our final experiment using real-world GDP data from the World Bank, we further validated the applicability of our differentially private Dickey-Fuller test. The results confirmed that the test could detect trend stationarity in a real-world data, even with privacy guarantees. This demonstrates the practical value of our approach for analyzing sensitive economic data, where privacy is often a key concern.

In conclusion, our research provides a comprehensive framework for conducting differentially private hypothesis testing on time series data, with a focus on the Dickey-Fuller unit root test. We have shown that it is possible to preserve individual privacy while maintaining the integrity of statistical results, particularly in scenarios where participants provide entire time series rather than individual data points. Although the Dickey-Fuller test is designed for specific types of autoregressive time series, this work has significant implications for fields such as economics, finance, public health, and any domain where time series data is used and privacy is a concern.

For future work, several avenues of exploration remain. One important area is to extend the application of differential privacy to other types of time series models and hypothesis tests. Additionally, further research could investigate the optimization of privacy budgets based on specific dataset characteristics, enabling more tailored privacy-accuracy trade-offs. Finally, exploring the integration of additional advanced differential privacy techniques could provide even stronger guarantees in time series data environments, where privacy risks are particularly acute.

By advancing the understanding of how differential privacy can be applied to time series data, our research opens the door to more privacy-preserving analytical tools, ensuring that individuals' data can be used in meaningful ways without compromising their confidentiality.

Bibliography

[1] Cynthia Dwork and Aaron Roth, *The Algorithmic Foundations of Differential Privacy*, Theoretical Computer Science, pp. 16-17, 31-32, 2014.

[2] Cynthia Dwork, Frank Mcsherry, Kobbi Nissim, and Adam Smith, *Calibrating noise to sensitivity in private data analysis*, TCC, 2006.

[3] Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N. Rothblum, *Differential privacy under continual observation*, ACM, pp. 715-724, 2010.

[4] Daniel Kifer and Ryan M. Rogers., *A new class of private chi-square tests*, Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, pp. 991–1000, 2017.

[5] David A. Dickey and Wayne A. Fuller, *Distribution of the Estimators for Autoregressive Time Series With a Unit Root*, Journal of the American Statistical Association, pp. 427-431, 1979.

[6] Kazuya Kakizaki, Jun Sakuma, and Kazuto Fukuchi, *Differentially private chi-squared* Proceedings of the 34th International Conference on Machine Learning, pp. 1761–1770, 2017.

[7] Kobbi Nissim, Sofya Raskhodnikova and Adam Smith *Smooth Sensitivity and Sampling in Private Data Analysis*, 2011.

[8] Marco Gaboardi, Hyun-Woo Lim, Ryan M Rogers, and Salil P Vadhan, *Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing*, ICML, pp. 2111–2120, 2016.

[9] Shumway Robert H., Stoffer David S., *Time series analysis and its applications: with R examples (3rd ed.)*, Springer, pp. 5, 13-15, 22-23, 84-85, 2010.

[10] Yue Wang, Daniel Kifer, Jaewoo Lee, and Vishesh Karwa, *Statistical approximating distributions under differential privacy* The Journal of Privacy and Confidentiality, pp. 1-33, 2018.

[11] Yue Wang, Jaewoo Lee, and Daniel Kifer, *Revisiting differentially private hypothesis tests for categorical data*, arXiv preprint arXiv:1511.03376, 2015.

[12] Zachary Campbell, Andrew Bray, Anna Ritz, and Adam Groce, *Differentially private ANOVA testing* Proceedings of the 2018 International Conference on Data Intelligence and Security, pp. 281–285, 2018.

Appendix A

GitHub

The code for all the experiments can be found in this GitHub Repository (https://github.com/Ziv33/DickeyFullerTimeSeriesDP).