Министерство образования республики беларусь

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра математического моделирования и анализа данных

ДОРОШКО Ольга Валерьевна

Анализ дискретных временных рядов

Магистерская диссертация

специальность: 1-31 80 09 – «Прикладная математика и информатика»

Научный руководитель

Сталевская Светлана Николаевна

доцент кафедры ММАД ФПМИ,

кандидат. физ-мат. наук

Допущена к защите

«\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_ 2018 г.

Зав. кафедрой ММАД

\_\_\_\_\_\_\_\_\_\_\_\_И.А. Бодягин

доцент, кандидат физико-математических наук

Минск, 2018

Оглавление

[Общая характеристика работы 4](#_Toc516590602)

[Агульная характарыстыка працы 5](#_Toc516590603)

[Abstract 6](#_Toc516590604)

[Введение 7](#_Toc516590605)

[Глава 1 Inar(1) модель 9](#_Toc516590606)

[1.1. Модель авторегрессии 9](#_Toc516590607)

[1.2. Оператор прореживания 12](#_Toc516590608)

[1.3. Модель INAR(1) 13](#_Toc516590609)

[1.4. Свойства модели INAR(1) 15](#_Toc516590610)

[Глава 2 Модель PSINAR(1) 19](#_Toc516590611)

[2.1. Варианты распределений 19](#_Toc516590614)

[2.2. Модель PSINAR(1) 20](#_Toc516590615)

[2.3. Свойства модели PSINAR(1) 26](#_Toc516590616)

[2.4. Частные случаи модели PSINAR(1) 27](#_Toc516590617)

[2.4.1. Процесс Пуассона PoINAR(1) 27](#_Toc516590618)

[2.4.2. Геометрическая GINAR(1) модель 29](#_Toc516590619)

[Глава 3 Оценка параметров модели 32](#_Toc516590620)

[3.1. Оценка Юле-Уокера 32](#_Toc516590621)

[3.2. Условная оценка функции методом наименьших квадратов 33](#_Toc516590622)

[3.3. Оценка максимального правдоподобия 34](#_Toc516590623)

[3.4. Смещение и вариация оценок 34](#_Toc516590624)

[3.5. Прогнозирование модели 35](#_Toc516590625)

[3.5.1. Идентификация модели 35](#_Toc516590626)

[3.5.2. Диагностика модели 36](#_Toc516590627)

[3.5.3. Прогнозирование 36](#_Toc516590628)

[3.6. Экспериментальные исследования 37](#_Toc516590629)

[Глава 4 Модели искажений 42](#_Toc516590630)

[4.1. Модели искажений авторегрессионных временных рядов 42](#_Toc516590631)

[4.2. Модели искажений для INAR(1). 44](#_Toc516590632)

[4.3. Оценка параметров модели при наличии искажений 47](#_Toc516590633)

[4.4. Выбор константы усечения с 51](#_Toc516590634)

[4.5. Экспериментальные исследования 53](#_Toc516590635)

[4.5.1. Сдвиг уровня и временной сдвиг 53](#_Toc516590636)

[4.5.2. Выбросы 55](#_Toc516590637)

[Заключение 62](#_Toc516590638)

[Список литературы 63](#_Toc516590639)

[Приложения 65](#_Toc516590640)

[Приложение A 65](#_Toc516590641)

[Приложение В 71](#_Toc516590642)

# Общая характеристика работы

Магистерская диссертация, 89 с., 32 рис., 11 табл., 2 прил., 27 источников.

**Ключевые слова:** АВТОРЕГРЕССИОННЫЙ ВРЕМЕННОЙ РЯД, ЦЕЛОЧИСЛЕННАЯ АВТОРЕГРЕССИОННАЯ МОДЕЛЬ, ЦЕЛОЧИСЛЕННЫЕ ДАННЫЕ, ИСКАЖЕНИЯ, ВРЕМЕННОЙ СДВИГ, СДВИГ УРОВНЯ, ВЫБРОС, МЕТОД НАИМЕНЬШИХ КВАДРАТОВ, МЕТОД ЮЛЕ-УОКЕРА, МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ, М-ОЦЕНКИ, СМЕЩЕНИЕ, ВАРИАЦИЯ.

Объект исследования – дискретные временные ряды, в том числе модели INAR(1) и частный случай модели, когда инновационная последовательность представляет степенной ряд, PSINAR(1), в том числе при наличии искажений.

Цель работы – построение оценок параметров модели PSINAR(1) в случае отсутствия искажений и построение м-оценок при наличии искажений, а также прогнозирование.

Методы исследования – математические методы анализа данных.

Результатами работы являются выведенные свойства и оценки данных моделей, а также функционал, реализованный на языке R, для идентификации модели, построения ее оценок, исследования модели на наличие искажений, а также дальнейшее предсказание последовательности.

Областью применения являются разнообразного вида очереди, в которых применима целочисленная модель.

**Структура магистерской диссертации:**

Введение дает краткое описание и экскурс в историю проблемы. В главе 1 описывается временной ряд AR(p) и его свойства. Вводится параметр биномиального прореживания и на его основании модель INAR(1) и ее свойства, по аналогии с AR(1). В главе 2 рассмотрена модель INAR(1) с инновациями на основе степенных рядов, представлены свойства модели, а так же частные случаи. В главе 3 приведены методы оценки параметров модели INAR(1), их модификации для приведенных в главе 2 частных случаев модели и себя краткие сведения по идентификации модели и ее прогнозированию. Для оценки самих методов введены понятия смещения и вариации оценки, проведены ряд экспериментов. В главе 4 представлены различные модели искажений временных рядов. Проанализированы свойства оценок, приведенных в главе 3, в случае, когда модель имеет искажения, а также введены новые методы оценки параметров, позволяющие их игнорировать.

# Агульная характарыстыка працы

Магістарская дысертацыя, 89 с., 32 мал., 11 табл., 2 прым., 27 крыніц.

**Ключавыя словы**: АЎТАРЭГРЭСІЙНЫ ЧАСОВЫ ШЭРАГ, ЦЭЛАЛІКАВЫЯ АЎТАРЭГРЭСІЙНАЯ МАДЭЛЬ, ЦЭЛАЛІКАВЫЯ ДАДЗЕНЫЯ, СКАЖЭННI, ЧАСОВАЙ ЗРУХ, ЗРУХ ШКОДНАГА, ВЫКІД, МЕТАД НАЙМЕНШЫХ КВАДРАТАЎ, МЕТАД ЮЛІ-УОКЕРА, МЕТАД МАКСІМАЛЬНАГА ПРАЎДАПАДАБЕНСТВА, М-АЦЭНКІ, ЗРУШЭННЕ, ВАРЫЯЦЫІ.

Аб'ект даследавання - дыскрэтныя часовыя шэрагі, у тым ліку мадэлі INAR (1) і прыватны выпадак мадэлі, калі інавацыйная паслядоўнасць ўяўляе спаважнаю шэраг, PSINAR (1), у тым лiку пры наяўнасцi скажэнняў.

Мэта работы - пабудова адзнак параметраў мадэлі PSINAR (1) у выпадку адсутнасці скажэнняў і пабудова м-адзнак пры наяўнасці скажэнняў, а таксама прагназаванне.

Метады даследавання - матэматычныя метады аналізу дадзеных.

Вынікамі працы з'яўляюцца выведзеныя ўласцівасці і ацэнкі дадзеных мадэляў, а таксама функцыянал, рэалізаваны на мове R, для ідэнтыфікацыі мадэлі, пабудовы яе ацэнак, даследаванні мадэлі на наяўнасць скажэнняў, а таксама далейшае прадказанне паслядоўнасці.

Вобласцю ўжывання з'яўляюцца разнастайнага выгляду чарзе, у якіх дастасоўная цэлалікавых мадэль.

**Структура магістарскай дысертацыі:**

Ўвядзенне дае кароткае апісанне і экскурс у гісторыю праблемы. У главе 1 апісваецца часовай шэраг AR (p) і яго ўласцівасці. Ўводзіцца параметр биномиального прарэджвання і на яго падставе мадэль INAR (1) і яе ўласцівасці, па аналогіі з AR (1). У чале 2 разгледжана мадэль INAR (1) з інавацыямі на аснове ступенных шэрагаў, прадстаўленыя ўласцівасці мадэлі, а так жа прыватныя выпадкі. У раздзеле 3 прыведзены метады ацэнкі параметраў мадэлі INAR (1), іх мадыфікацыі для прыведзеных у чале 2 прыватных выпадкаў мадэлі i кароткія звесткі па ідэнтыфікацыі мадэлі і яе прагназаванню. . Для ацэнкі саміх метадаў ўведзеныя паняцці зрушэння і варыяцыі ацэнкі, праведзены шэраг эксперыментаў. У разьдзеле 4 прадстаўлены розныя мадэлі скажэнняў часовых шэрагаў. Прааналізаваны ўласцівасці ацэнак, прыведзеных у раздзеле 3, у выпадку, калі мадэль мае скажэнні, а таксама ўведзеныя новыя метады ацэнкі параметраў, якія дазваляюць іх ігнараваць.

# Abstract

The master's thesis, 89 pages, 32 figures., 11 tabl., 2 app., 27 literature references.

**Keywords:** AUTOREGRESSIVE TIME SERIES, INTEGER-VALUED AUTOREGRESSIVE MODEL, INTEGER-VALUED DATA, DISTORTION, TRANSIENT SHIFT, LEVEL SHIFT, SPIKY OUTLIER, LEAST SQUARES METHOD, YULE-WALKER METHOD, MAXIMUM LIKELIHOOD METHOD, M-ESTIMATOR, BIAS, VARIATION.

Object of study – discrete time series, including INAR(1) and a particular case of the model when the innovation sequence represents a power series, PSINAR (1), including in the presence of distortions.

The purpose of the work is constructing PSINAR (1) model parameter estimation in the absence of distortion and constructing m-estimation in the presence of distortions, as well as forecasting.

Methods of research - mathematical methods of data analysis.

The results of the work are the derived properties and estimation of these models, as well as the functional implemented in R, for identifying the model, constructing its estimation, investigating the model for distortions, and further predicting the sequence.

The scope of application is a diverse type of queue, in which an integer-valued model is applicable.

**The structure of the master's thesis:**

The introduction gives a brief description and an insight into the history of the problem. Chapter 1 describes the AR (p) time series and its properties. The parameter of binomial thinning is introduced and on its basis the model INAR (1) and its properties, by analogy with AR (1). In Chapter 2, the model INAR (1) with innovations based on power series is presented, the properties of the model are presented, as well as special cases. Chapter 3 provides methods for estimating the parameters of the model INAR (1), their modifications for the particular cases of the model presented in Chapter 2 and brief information on model identification and forecasting. To estimate the methods themselves, the concepts of displacement and variation of the estimation were introduced, and a number of experiments were carried out. Chapter 4 presents various models of time series distortions. The properties of the estimates given in Chapter 3 are analyzed, in the case when the model has distortions, and new methods for estimating parameters are introduced, which allow them to be ignored.

# Введение

Одной из основных задач, которые ставит перед собой современная экономическая наука, является анализ и прогнозирование разного рода динамики развития всех видов рынков. Анализ временных рядов – совокупность математико-статистических методов анализа, предназначенных для выявления структуры временных рядов и для их прогнозирования. Выявление структуры временного ряда необходимо для построения математической модели явления, которое является источником анализируемого временного ряда. Прогноз будущих значений временного ряда используется для эффективного принятия решений.

Существуют две основные цели анализа временных рядов: (1) определение природы ряда и (2) прогнозирование – предсказание будущих значений временного ряда по настоящим и прошлым значениям. Обе эти цели требуют, чтобы модель ряда была идентифицирована и формально описана. Как только модель определена, становится возможным интерпретировать рассматриваемые данные с ее помощью, а далее и экстраполировать ряд на основе найденной модели, т.е. предсказать его будущие значения.

Мы часто сталкиваемся с временными рядами, характеристики которых несовместимы с подходом непрерывного моделирования. Целочисленные вариационные временные ряды встречаются во многих контекстах, часто в виде количества событий или отдельных лиц в последовательных интервалах или в последовательных точках времени. Примерами этого являются число клиентов, ожидающих обслуживания, ежедневное число отсутствующих работников в фирме, число занятых линий в телефонной сети, отмечаемое каждый час, число несчастных случаев на производстве каждый месяц, количество сделок акций, произведенных за один день и т.д.

За последние три десятилетия вырос интерес к дискретным временным рядам, и было предложено несколько моделей для стационарных процессов с дискретным маргинальным распределением. В [1] предложена первая неотрицательная целочисленная авторегрессионная (INAR (1)) модель. В [23] предложено определение моделей авторегрессии на основании биномиального оператора прореживания. В работе [10] вводится новая стационарная целочисленная авторегрессия первого порядка с нулевыми завышенными инновациями Пуассона.

В целом, детальные исследования были проведены не только по формулировке модели, но и по свойствам [20], оценкам [12], тестам [13] и асимптотические распределения модельных оценок [7] для разных дискретных предельных распределений.

В главе 1 вводится понятие авторегрессионного временного ряда, устанавливаются некоторые свойства модели AR(p). На основании авторегрессионного временного ряда и, введенного параметра биномиального прореживания, вводится модель INAR(1) и ее свойства, по аналогии с AR(1). Так же приводится сравнение двух введенных моделей с одинаковыми математическим ожиданием и дисперсией.

В главе 2 рассмотрена модель INAR(1) с инновациями на основе степенных рядов, представлены свойства модели, а так же частные случаи.

В главе 3 приведены методы оценки параметров модели INAR(1), их модификации для приведенных в главе 2 частных случаев модели. Для оценки самих методов введены понятия смещения и вариации оценки. Представленные в данной главе экспериментальные исследования позволяют получить наглядное представление о свойствах полученных оценок.

В главе 4 представлены различные модели искажений временных рядов. Более подробно рассмотрены некоторые виды искажений для модели PSINAR(1). Проанализированы свойства оценок, приведенных в главе 3, в случае, когда модель имеет искажения, а также введены новые методы оценки параметров, позволяющие их игнорировать.

# Inar(1) модель

## Модель авторегрессии

Авторегрессионная модель – модель стационарного процесса, выражающего значение показателя в виде линейной комбинации конечного числа предшествующих значений этого показателя и аддитивной случайной составляющей.

В процессе анализа реальных экономических явлений понятие стационарности может быть лишь удобной абстракцией для применения статистических моделей. Количество уровней, включенных в правую часть уравнения авторегрессии, определяет порядок уравнения.

Ее построению предшествует оценка наличия автокорреляции в изучаемом ряду.

Автокорреляция в рядах динамики - это зависимость между уровнями ряда или зависимость между исходным динамическим рядом и тем же рядом, но смещенным на определенный временной интервал, называемый лагом. Автокорреляция первого порядка (first-order autocorrelation) оценивает степень зависимости между соседними значениями временного ряда. Автокорреляция второго порядка (second-order autocorrelation) оценивает тесноту связи между значениями, разделенными двумя временными интервалами, и т.д.

Наличие автокорреляции проверяется на основе коэффициентов автокорреляции. Величина временного лага определяет порядок коэффициента автокорреляции. После подтверждения наличия автокорреляции в динамическом ряду, может идти речь о построении авторегрессионной модели.

По большому счету формулу автокорреляционной функции, которая применяется в авторегрессионных моделях, использовать нельзя: она применима исключительно к стационарным рядам. Кроме этого, ее крайне сложно получить, а математическое ожидание, которое используется при расчете АКФ найти попросту невозможно. Существует огромное количество различных хитростей и тонкостей, когда суперпозицию аналитических функций называют «моделью локального процесса» - на ее основании получают достаточно сомнительную оценку автокорреляционной функции. [24]

Таким образом, приходим к модели, которая называется процессом авторегрессии порядка p – AR(p) (от английского «AutoRegressive» – авторегрессионная

Пусть на вероятностном пространстве определен временной ряд вида

|  |  |
| --- | --- |
|  | (1.1) |

где - порядок авторегрессии, – вектор коэффициентов авторегрессии, , – дискретный «белый шум» (последовательность независимых одинакова распределенных гауссовских случайных величин, ).

Временной ряд вида (1.1) называют временным рядом авторегрессии порядка (краткое обозначение [24],[26].

Если ввести оператор запаздывания (лага)

|  |  |
| --- | --- |
|  | (1.2) |

то уравнение авторегрессии можно представить в виде:

следовательно

где

|  |  |
| --- | --- |
|  | (1.3) |

есть характеристический полином модели.

Свойства авторегрессионных временных рядов [24], [26]:

1. Для того, чтобы процесс был стационарным, достаточно, чтобы нули характеристического полинома (1.3) лежали вне единичного круга, в частности:

* для : процесс стационарен ;
* для : процесс стационарен

1. Так как ряд конечен, то процесс всегда обратим;
2. Если процесс является стационарным (выполнено свойство 1), то его математическое ожидание равно математическому ожиданию , в данной модели
3. Автоковариационная функция процесса удовлетворяет рекуррентному соотношению

|  |  |
| --- | --- |
|  | (1.4) |

а автокорреляционная функция – соотношению

|  |  |
| --- | --- |
|  | (1.5) |

Для нахождения автокорреляционной функции также можно использовать систему уравнений Юла-Уокера [24],[25]:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | |  |  |  | +… |  | |  |  |  | +… |  | | . | . | . | … | . | | . | . | . | . | | . | . | . | . | |  |  |  | +… |  | | (1.6) |

В частности:

* для :
* для : а остальные значения находятся из рекурентного соотношения (1.5);

1. Для дисперсии процесса справедливо равенство [24]:

|  |  |
| --- | --- |
|  | (1.7) |

В частности:

* для :
* для :

Частным случаем модели , определенной (1.1), является модель с , определяемая следующим соотношением:

|  |  |
| --- | --- |
|  | (1.8) |

## Оператор прореживания

Модель INAR(1) впервые была появилась в публикации McKenzie [16], где он представил несколько AR(1)-подобных моделей для целочисленных временных рядов. В тот момент было важным заметить, что обычная модель AR(1) в виде (1.8) не может быть применена к целочисленным данным в связи с оператором умножения в ее представлении, что было названо “multiplication problem”. Поэтому McKenzie предложил использовать различные механизмы для “прореживания”. Одним из решений этой проблемы стал биномиальный оператор прореживания [22], который и был использован для определения модели INAR(1), вместо обычного оператора умножения.

Пусть X - неотрицательная целочисленная случайная величина; тогда для любого оператор прореживания “ ” определяется формулой

,

где - последовательность независимыx одинаково распределенных случайных величин Бернулли, не зависящие от X, такие, что

.

При этом имеет биномиальное распределение с условными параметрами и , и безусловными параметрами , где , и , где .

Очевидно, что принимает только целые значения из интервала [0,X].

Можно также заметить, что

Откуда можно сделать вывод, что биномиальный оператор усечения имеет практически тот же смысл, что и оператор умножения в данном случае.

Но эти операторы имеют и отличия в некоторых свойствах. Например, в том, что использование оператора умножения не является случайной операцией, в то время как оператор биномиального прореживания является таковым.

В подтверждение этому можно увидеть, что

В качестве интерпретации биномиального оператора прореживания можно рассмотреть популяцию размера X в момент времени t. Если рассмотреть эту популяцию в момент времени t+1, то ее размер уже может быть меньше, потому что некоторые особи могут погибнуть в промежуток времени от t до t+1. Если особи выживают независимо друг от друга и если вероятность их выживания в момент времени от t до t+1 равна , то число выживших особей будет равно значению .

Из определения оператора “ ” ясно, что:

* .

## Модель INAR(1)

Используя определенный выше оператор биномиального прореживания, McKenzie [16] и Al-Osh и Alzaid [1] представили целочисленный авторегрессионный процесс первого порядка (INAR (1)).

Целочисленный авторегрессионный процесс первого порядка INAR(1) определяется как:

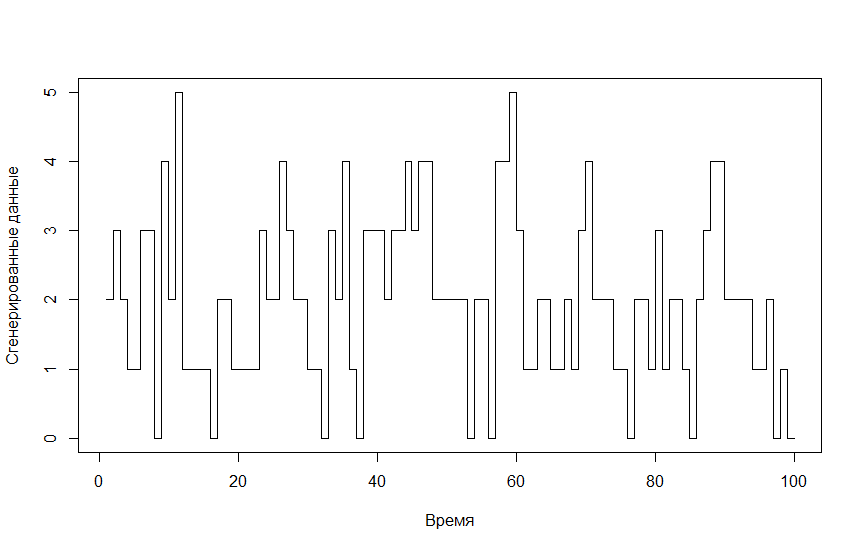
|  |  |
| --- | --- |
| , | (1.9) |

где и последовательность независимых и одинаково распределенных неотрицательных целочисленных случайных величин с математическим ожиданием и дисперсией .

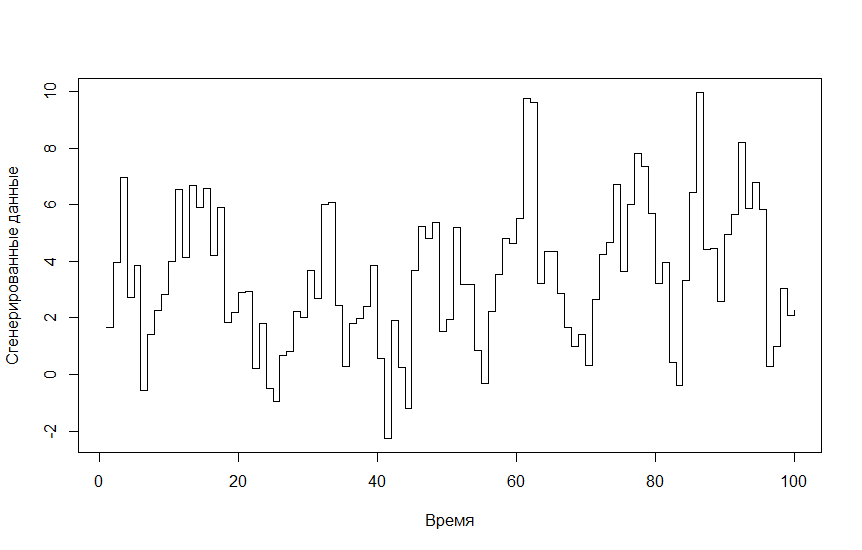
Формулу (1.9) исходя из интерпретации биномиального оператора прореживания можно описать как

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | = | |  | + | | . |
| популяция в момент времени t | | количество выживших особей в момент времени t-1 | | | число мигрировавших особей | |

В качестве иллюстрации того, как последовательность процесса INAR(1) отличается от аналогичной в стандартном AR(1) процессе, была сгенерирована выборка. Для сравнения двух выборок будем считать, что оба процесса имеют одинаковые математическое ожидание и дисперсию. На Рисунок 1.1 представлена выборка размером 100, элементов генерируемых по модели (1.9)0, в которой принято считать случайными величинами Бернулли с и имеет распределение Пуассона с параметром . На Рисунок 1.2 показано, примерный путь выборки стандартной модели AR (1), в которой случайные величины с нормальным распределением с математическим ожиданием и дисперсией .



**Рисунок 1.1** - **Целочисленные значения временных рядов, с распределением Бернулли (0,5) и Пуассона (1.0)**



**Рисунок 1.2****– Непрерывный временной ряд с нормальным распределением (2, 2)**

Из сравнения Рисунок 1.1 и Рисунок 1.2 видно, что в то время как реализации процесса INAR (1) принимают неотрицательные целочисленные значения, реализации AR (1) - рациональные значения, некоторые из которых отрицательны. Как следствие из этого наблюдения и требования к стационарности для двух процессов, видно из Рисунок 1.1, что реализация процесса INAR (1) имеет много значений вблизи среднего (2). Однако того же нельзя сказать по Рисунок 1.2.

## Свойства модели INAR(1)

Маргиналы распределения модели (1.9) могут быть выражены с точки зрения последовательности {} как:

|  |  |
| --- | --- |
|  | (1.10) |

Для того, чтобы сгенерировать стационарный процесс {} с определенным маргинальным распределением по формуле (1.9), необходимо сначала определить инновационный процесс . Для обеспечения стационарности необходимо, чтобы производящая функция (probability generating function (p.g.f.)) процесса была равна

|  |  |
| --- | --- |
|  | (1.11) |

Откуда следует, что распределение однозначно определяется распределением и наоборот.

Условное среднее и вариация принимают вид

|  |  |
| --- | --- |
|  | (1.12) |
|  | (1.13) |

и обе являются линейными функциями от .

Так как условное среднее является линейной функцией аргумента , то INAR(1) принадлежит классу условных линейных AR(1) моделей.

Заметим также, что условная вариация модели INAR(1) отличается от условной вариации для модели AR(1), так как она изменяется с течением времени (имеет место условная гетероскедастичность).

Среднее значение и дисперсия процесса {}, как определено в (1.9), являются простыми:

|  |  |
| --- | --- |
| + | (1.14) |

и

|  |  |
| --- | --- |
|  | (1.15) |

Из (1.14) и (1.15) видно, что стационарность второго порядка требует, чтобы

|  |  |
| --- | --- |
|  | (1.16) |

и

|  |  |
| --- | --- |
|  | (1.17) |

Теперь, зная математическое ожидание и дисперсию, можно получить индекс дисперсии (index of dispersion):

|  |  |
| --- | --- |
|  | (1.18) |

где

|  |  |
| --- | --- |
|  | (1.19) |

Автоковариационная функция стационарного INAR(1) процесса {} вычисляется по формуле:

|  |  |
| --- | --- |
|  | (1.20) |

а автокорреляционная функция вычисляется как:

|  |  |
| --- | --- |
|  | (1.21) |

Уравнение (1.21) показывает, что функция автокорреляции разлагается экспоненциально с лагом и имеет ту же форму, что и уравнение Юла-Уолкера в AR(1). При =1, получаем, что параметр представляет собой корреляцию между последовательными моментами времени. Однако, в отличие от функции автокорреляции для AR(l), является всегда положительной для модели INAR(1).

Относительно свойств распределения процесс INAR(1), отметим, что этот процесс имеет распределительные свойства, аналогичные свойствам AR(1).

Другие свойста модели INAR(1) описаны в [18], [19], [20].

Определяя распределения инновационного процесса , возникает большое количество разнообразных моделей.

# Модель PSINAR(1)



## Варианты распределений

Достаточно много работ посвящено использованию распределения Пуассона как интегральной особенности модели INAR(1). Однако распределение Пуассона не всегда подходит для моделирования, поскольку его среднее значение и дисперсия одинаковы, и это свойство может быть неприемлемым для реальных данных. Кроме того, во многих реальных ситуациях существуют ряды, которые не содержат нулей за большой промежуток времени или даже всегда положительны. В этих ситуациях распределение Пуассона также не подходит для моделирования. Поэтому необходимо ввести различные целочисленные значения временных рядов для решения различных конкретных реальных ситуаций, таких как избыточная дисперсия или нулевая инфляция. Идея рассмотрения распределения для инновационной последовательности, такая, что предельное распределение наблюдений будет удовлетворять данному свойству, широко рассмотрена в [23], где показаны подходы к получению, например, сверхдисперсионного отрицательного биномиального или обобщенного Распределение Пуассона.

Исходя из вышесказанного можно предложить новый INAR(1) процесс с инновациями на основе степенных рядов (power series(PS)). Использование инноваций, происходящих из семейств распределений PS, имеет много преимуществ, так как это семейство распределений, представляющих собой гибкую структуру для статистического моделирования дискретных данных во многих реальных ситуациях [11].

Рассмотрим последовательность дискретных независимых и одинаково распределенных случайные величины , причем распределение каждого индексируется параметром θ и определяется функцией вероятности:

|  |  |
| --- | --- |
|  | (2.1) |

где S является подмножеством неотрицательных целых чисел, зависит только от x и существует такое, что конечно для всех (s может быть равно ). Хотя мы всегда будем рассматривать θ как значение из (0, s), мы также можем предположить, что степенной ряд для C (θ) сходится, к конечному значению для θ ∈ (-s, s). Если это так, то C (θ) имеет производные всех порядков в (-s, s) и эти производные можно получить, дифференцируя степенной ряд почленно. Кроме того, поскольку для всех , C (θ) и всех его производных будут положительными в (0, s) [17].

**Таблица 2.1 – Некоторые распределения, удовлетворяющие (2.1).**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Распределение |  |  |  |  |
| Бернулли | 1 |  |  |  |
| Биномиальное |  |  |  |  |
| Геометрическое | 1 |  | 1 |  |
| Пуассона |  |  |  |  |
| Отрицательное Биномиальное |  |  | 1 |  |
| Логарифмическое |  | ) | 1 |  |

В Таблица 2.1 представлены функции и параметр θ, соответствующие некоторым особым случаям распределений PS, удовлетворяющие (2.1), таким как Бернулли, Биномиальное (с n - целочисленным числом повторений), Геометрическое, Пуассона, Отрицательное Биномиальное и Логарифмическое распределения. Для Отрицательного Биномиального могут существовать ситуации, для которых мы хотим, чтобы r было целочисленным. В этом случае x можно рассматривать как случайное число отказов, пока точно не будут достигнуты успехи в последовательности независимых испытаний, где вероятность отказа равна θ. При использовании Биномиального распределения значение n может быть известно заранее или может быть оценено; то же самое справедливо для значения r при использовании Отрицательного Биномиального распределения.

## Модель PSINAR(1)

Модель (1.9), в которой функция вероятности последовательности удовлетворяет (2.1), называется моделью PSINAR(1)(power series INAR(1)).

Вопрос о том, какое распределение использовать для последовательности субъективный, так как это зависит от конкретной ситуации, с которой мы имеем дело. Например, в разделе «Эпидемиология» предположим, что исследователь контролирует количество индивидуумов в данной популяции, которые не контактировали с конкретным заболеванием, то есть предположим, что представляет собой число здоровых людей в популяции в момент времени t. Пусть α - вероятность того, что здоровый индивидуум останется здоровым, то есть не получит заболевание в следующий момент времени. Предположим также, что с очень небольшой вероятностью больной человек может излечиться от этого заболевания, так что в следующий раз у нас будет не более одного излечившегося человека в популяции. Тогда эволюция может быть описана в (1.9), используя распределение Бернулли для . Предположим теперь, что тот же самый исследователь, желая наблюдать эволюцию лечения в очень специфической группе, предписывает данное лекарство для n больных людей этой группы, так что некоторые из лиц, принимающих это лекарство, могут излечиться. Теперь эволюция может быть описана в (1.9), используя биномиальное распределение для . При лечении серьезного заболевания мы можем считать, что данный человек вылечится или что этот самый индивидуум в конечном итоге умирает. Способ контроля эффективности данного лечения заключается в том, чтобы наблюдать, сколько людей вылечится до того, как человек умрет. Тогда эволюцию можно описать формулой (1.9) с геометрическим распределением для . Если мы заметим, сколько человек вылечится до того, как умрут люди, тогда мы можем использовать отрицательное биномиальное распределение для . Разумный выбор для распределения также может следовать из статистических соображений. Если представляется разумным, что среднее и дисперсия распределения наблюдений равны, то простая модель Пуассона может быть применима. Если дисперсия, по-предположению, меньше среднего, мы должны отбросить Геометрическое и Пуассоновское распределения. Гипотеза может использоваться для выбора между геометрическим и отрицательным биномиальным распределениями.

Из (1.9) следует, что - марковский процесс.

**Теорема 2.1**

При фиксированном , матрица переходных вероятностей этого процесса имеет вид:

|  |  |
| --- | --- |
|  | (2.2) |

для всех .

**Доказательство.**

Для процесса PSINAR (1) с биномиальной оператором прореживания условное распределение , заданное , представляет собой свертку биномиального распределения результата операции прореживания с распределением PS инновационного процесса [21].Таким образом, пусть • обозначает свертку и

и

Тогда,

Если для фиксированного , тогда

и ⇒

таким образом

**Теорема 2.2**

Марковский процесс, определяемый приведенной выше матрицей переходных вероятностей, допускает единственное стационарное распределение.

Маргинальная вероятностная функция задается формулой:

|  |  |
| --- | --- |
|  | (2.3) |

которая представляет собой смешанное распределение.

**Доказательство**

Поскольку , для всех процесс является неприводимым процессом в том смысле, что любое k ∈ S может быть получено из любого l ∈ S. Оно также имеет стационарную матрицу переходных вероятностей, в том смысле, что эти вероятности перехода не связаны с t. Пусть . Согласно [9], существование стационарного распределения эквивалентно

Из [9], если вышеприведенное неравенство справедливо для конкретного y ∈ S, оно так же будет, так как наш процесс неприводим, справедливо для всех x ∈ S. Заметим, без потери общности, что наименьший элемент S равен нулю, в том смысле, что a (0)> 0 в разложении степенных рядов C (θ). Тогда можно доказать, что

Это будет доказано, если существует и положителен. Эквивалентно, можно показать, что существует и конечен.

По индукции

Для утверждение выше уменьшается до

,

что тривиально верно. Полагаем, что это также верно для заданного m. Тогда

Тогда, получаем

Заметим, что при m = 0 получаем , что тривиально верно.

Тогда

Пусть . Из теоремы о промежуточном значении для каждого i можно получить значение ( такое, что

Теперь

Так как положительно, то сумма выше представляет собой сумму положительных членов. Таким образом, остается показать, что бесконечный ряд сходится, если допускаем, что m → ∞. Заметим, что (. Поскольку G имеет производную второго порядка , должна быть непрерывна. Пусть максимальное значение на . Тогда

Хотя мы знаем, что существует уникальное стационарное распределение, получение этого стационарного распределения из приведенных выше уравнений является в общем сложной задачей. Альтернативно, если Ψ - производящая функция (probability generating function (p.g.f.)) для этого стационарного распределения, , нетрудно проверить, что Ψ удовлетворяет для всех . Однако этот подход все еще непростой. В самой простой ситуации, когда , которая соответствует классической модели Пуассона INAR (1), мы легко получим, что стационарное распределение - это распределение Пуассона и его математическое ожидание составляет . С другой стороны, общая проблема получения стационарного распределения наблюдений, учитывая особую , в большинстве ситуаций довольно сложная.

## Свойства модели PSINAR(1)

Моменты случайной величины легко получить из p.g.f. . Тогда математическое ожидание и дисперсия соответственно равны

|  |  |
| --- | --- |
|  | (2.4) |

и

|  |  |
| --- | --- |
|  | (2.5) |

где , и .

Из (1.16), (1.17), (2.4) и (2.5) следует, что математическое ожидание и дисперсия процесса равны:

|  |  |
| --- | --- |
|  | (2.6) |

и

|  |  |
| --- | --- |
|  | (2.7) |

Заметим, что дисперсия может быть либо больше либо меньше математического ожидания, в зависимости от знака . Индекс дисперсии будет равен

|  |  |
| --- | --- |
|  | (2.8) |

Кроме того, среднее и дисперсия будут равны, когда G является линейной функцией, как в случае распределения Пуассона.

Также можно получить из (1.12), (1.13) условное среднее

|  |  |
| --- | --- |
|  | (2.9) |

и вариацию

|  |  |
| --- | --- |
|  | (2.10) |

## Частные случаи модели PSINAR(1)

### Процесс Пуассона PoINAR(1)

и в (2.1) соответствует PoINAR(1).

Матрица переходных вероятностей

Математическое ожидание и дисперсия данного процесса равны соответственно

и

Среднее и дисперсия данного процесса равны и возрастают как в зависимости от , так и от .

Условное среднее и вариация равны соответственно

и

Индекс дисперсии будет равен

В данном случае , и .

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

**Рисунок 2.1 – Графики PoINAR с различными значениями параметров**

### Геометрическая GINAR(1) модель

и в (2.1) соответствует GINAR(1).

Матрица переходных вероятностей

Среднее и дисперсия равны

и

Условное среднее и вариация равны соответственно

и

Очевидно, что и возрастающие функции для и .

Индекс дисперсии будет равен

Таким образом, геометрический процесс INAR (1) можно использовать в качестве модели для сверхдисперсионного неотрицательные целочисленные значения временных рядов. Из приведенного выше выражения мы можем легко заключить, что индекс дисперсии является возрастающей функцией для θ, но убывающей для α.

В данном случае

,

и

.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

**Рисунок 2.2 - Графики GINAR с различными значениями параметров**

# Оценка параметров модели

## Оценка Юле-Уокера

Для выборки размером n процесса автокорреляционная функция равна

где где –средняя величина.

Оценка Юле-Уокера (Yule-Walker estimator(YW)) для , основывается на том факте, что (из (1.21)):

|  |  |
| --- | --- |
|  | (3.1) |

Первый момент задается, как . Используя это, оценку можно получить следующим образом:

,

где берется из (3.1). Оценка параметра может быть получена из уравнения

.

Оценка θ может иметь замкнутую форму, в зависимости от того, какое распределение используется.

Таким образом можно получить оценку параметров

* для PoINAR(1):

следовательно

* для GINAR(1):

тогда

## Условная оценка функции методом наименьших квадратов

В качестве альтернативы метода Юла-Уокера можно рассматривать метод оценки условных наименьших квадратов (Conditional least squares estimation (CLS)).

Условная оценка наименьших квадратов для выглядит следующим образом:

где и . Таким образом, согласно [15], условная оценка наименьших квадратов для может быть представлена, как

|  |  |
| --- | --- |
|  | (3.2) |

и

,

где из (3.2). Аналогично предыдущей оценке, оценку параметра получаем из уравнения .

Тогда можно получить оценку параметров

* для PoINAR(1):

следовательно

* для GINAR(1):

тогда

## Оценка максимального правдоподобия

Полагая фиксированной величиной, условная функция логарифмического правдоподобия для PSINAR(1) выглядит следующим образом:

|  |  |
| --- | --- |
|  | (3.3) |

с в виде (2.2).

Оценки методом максимального правдоподобия (conditional maximum likelihood (CML)) и для и θ определяются как значения и θ, которые максимизируют условную функцию логарифмического правдоподобия в (3.3). В общем случае не будет закрытой формы для оценок CML, и их практическое применение, на практике, потребует численных методов.

## Смещение и вариация оценок

Для изучения свойств построенных оценок будем рассматривать такие характеристики оценки, как смещение и вариация.

**Определение**

Смещением статистической оценки называется уклонение математического ожидания этой оценки от истинного значения параметра :

**Определение**

Вариацией (среднеквадратической ошибкой) статистической оценки называется следующее математическое ожидание:

Заметим, что вариация – всегда скалярная величина.

## Прогнозирование модели

### Идентификация модели

Получив все необходимые оценки параметров модели можно перейти к ее идентификации. Однако, применять полученные оценки можно лишь в том случае, если последовательность действительно является моделью процесса INAR(1).

Для того, чтобы убедится в этом сначала необходимо рассмотреть структуру ряда данных. Автокорреляционная функция для INAR(1), полученная по формуле (1.21) схожа с автокорреляционной функцией для AR(1). Что, в свою очередь, подразумевает, что частная автокорреляционная функция удовлетворяет следующим выражениям

и .

Следовательно, чтобы проверить, подходит ли вообще модель INAR(1) для последовательности данного временного ряда, нужно вычислить сначала автокорреляционную функцию и проверить, является ли она экспоненциально затухающей и положительной, затем вычислить частную автокорреляционную функцию, для нее также проверить, удовлетворяет ли она перечисленным выше свойствам.

Определив AR(1)-подобную структуру автокорреляции, можно анализировать маргинальное распределение ряда и проводить оценку параметров.

Если происходит так, что для ряда данных применимо несколько моделей-кандидатов, то одним из способов выбора наиболее подходящей является использование информационных критериев, AIC и BIC, вычисленных одновременно с оценками максимального правдоподобия.

Другие критерии отбора включают условную сумму квадратов, вычисленную при оценке, или критерии, связанные с прогнозированием.

### Диагностика модели

После определения наилучшей из моделей-кандидатов остается проверить, они действительно адекватны для анализируемых данных, то есть, представляют ли заданные временные ряды собой реализацию рассматриваемой модели.

Очевидный подход к проверке адекватности модели заключается в сопоставлении некоторых характеристик установленной модели с их образцами-аналогами, рассчитанными на основе имеющихся временных рядов. Такое сравнение должно включать структуру автокорреляции, а также маргинальные характеристики, такие как математическое ожидание и дисперсия.

В [14] и [4] были представлены более сложные инструменты, основанные на условных распределениях, которые, следовательно, проверяют прогностическую эффективность. В качестве этого подхода следует проанализировать стандартизированные остатки Пирсона [8], то есть ряд

Для того, чтобы модель была адекватной необходимо, чтобы эти остатки были некоррелированными со средним значением около 0 и вариацией около 1. Вариация больше либо меньше 1 указывает на то, что данные имеют большую или соответственно меньшую дисперсию, чем предлагаемая модель.

### Прогнозирование

В наиболее распространенной процедуре построения прогнозов в моделях временных рядов используется условное математическое ожидание.

Модель INAR (1) удовлетворяет следующим свойствам:

Таким образом, для

Прогнозирование на h шаг вперед, основанное на условном ожидании INAR(1)

было получено [3] и [5], но вряд ли будет давать целые оценки. Для получения конгерентных предсказаний для [7] предлагают использовать значение, которое минимизирует ожидаемую абсолютную погрешность, то есть вместо условного ожидания можно использовать условную медиану, которая минимизирует ожидаемую абсолютную ошибку.

Для этой цели для модели INAR(1) должно быть рассчитано условное распределение на h шаг вперед, то есть нужно расчитать вероятности перехода на h шаг вперед

Как только это распределение доступно, соответствующие условная медиана и мода могут быть использованы в качестве когерентного точечного прогноза. Фактически, условная медиана также удовлетворяет свойству оптимальности, поскольку она минимизирует среднюю абсолютную ошибку.

Остается единственный вопрос о том, как вычислить . Имея то, что для INAR(1)

Так как распределение известно, то с помощью (2.2):

и

## Экспериментальные исследования

Чтобы понять относительные достоинства каждого из трех методов оценки, были проведены экспериментальные исследования. Начальное значение процесса было принято как ожидаемое значение процесса. Эксперимент по моделированию проводился для рядов со значениями n = 50, 75, 100 и 200 и для разных значений параметров . Значения параметра θ также брались различными, но для PoINAR(1) представим результаты только для , а вот для GINAR(1)

CMLбыла найдена путем максимизации условной функции логарифмического правдоподобия в (3.3). 200 повторений проводились по каждому образцу для каждой возможной комбинации параметров и были рассчитаны смещение и вариация оценок параметров.

Результаты моделирования для PoINAR(1), когда , который является типичным, представлены в таблицах 1 и 2, которые показывают соответственно смещение и вариацию при различных значениях и различных размерах выборки, которые были рассмотрены.

**Таблица 3.1 – Смещение результатов различных методов оценки для модели PoINAR (1) (θ = 1,0)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Смещение (α) | | | Смещение (θ) | | |
| n | α | Y-W | CLS | CML | Y-W | CLS | CML |
| 50 | 0,1 | -0,0576 | -0,0341 | -0,0172 | 0,0962 | 0,0501 | 0,0502 |
|  | 0,3 | -0,0923 | -0,0620 | -0,0787 | 0,2007 | 0,1429 | 0,1781 |
|  | 0,5 | -0,1105 | -0,0726 | -0,0687 | 0,2672 | 0,1834 | 0,1706 |
|  | 0,7 | -0,1047 | -0,0617 | -0,0420 | 0,2863 | 0,1762 | 0,0678 |
|  | 0,9 | -0,0877 | -0,0359 | -0,0217 | 0,4869 | 0,2335 | -0,0733 |
|  |  |  |  |  |  |  |  |
| 75 | 0,1 | -0,0269 | -0,0112 | 0,0493 | 0,0591 | 0,0283 | 0,0299 |
|  | 0,3 | -0,1122 | -0,0905 | 0,1779 | 0,2475 | 0,2043 | 0,1772 |
|  | 0,5 | -0,1125 | -0,0865 | 0,1804 | 0,2647 | 0,2100 | 0,0336 |
|  | 0,7 | -0,0925 | -0,0638 | 0,0123 | 0,2565 | 0,1775 | 0,0079 |
|  | 0,9 | -0,0576 | -0,0260 | -0,0130 | 0,3819 | 0,2036 | -0,0659 |
|  |  |  |  |  |  |  |  |
| 100 | 0,1 | -0,0457 | -0,0339 | 0,1127 | 0,0960 | 0,0708 | 0,0548 |
|  | 0,3 | -0,0852 | -0,0700 | 0,1595 | 0,1788 | 0,1490 | 0,0747 |
|  | 0,5 | -0,0889 | -0,0705 | 0,1691 | 0,2534 | 0,2142 | -0,0572 |
|  | 0,7 | -0,0635 | -0,0432 | 0,0010 | 0,2083 | 0,1527 | 0,0213 |
|  | 0,9 | -0,0418 | -0,0217 | -0,0073 | 0,2886 | 0,1828 | -0,0530 |
|  |  |  |  |  |  |  |  |
| 200 | 0,1 | -0,0462 | -0,0405 | 0,0957 | 0,0774 | 0,0654 | 0,0357 |
|  | 0,3 | -0,0727 | -0,0651 | 0,2022 | 0,1707 | 0,1572 | 0,0290 |
|  | 0,5 | -0,0808 | -0,0721 | 0,0350 | 0,2196 | 0,2010 | 0,0453 |
|  | 0,7 | -0,0392 | -0,0299 | -0,0064 | 0,1269 | 0,1022 | 0,0420 |
|  | 0,9 | -0,0209 | -0,0106 | -0,0040 | 0,1603 | 0,1020 | -0,0335 |

Из Таблица 3.1 видно, что:

* Из 20 случаев, которые представляют собой различные комбинации значений и размера выборки, смещен вниз и смещен вверх в 20 случаях CLS и 20 случаях оценки Юла-Уокера. Это обратное соотношение следует ожидать из уравнений, определяющих оценки. Однако вероятность того, что две оценки и имеют один и тот же знак смещения имеет место только, если значение превышает среднее значение процесса во большинстве случаев. Это видно из следующего соотношения:
* Величина смещений и увеличивается с увеличением как в оценках Юла-Уокера так и CLS. Увеличение смещения намного больше (примерно в восемь раз), чем у . Напротив, смещение CML не показывает такой тенденции к увеличению смещения. Это наблюдение иллюстрируется графически на Рисунок 3.1 и Рисунок 3.2, которые показывают соответственно смещения в и для трех методов оценки на основе выборки значения 75 и значений

|  |
| --- |
| **Рисунок 3.1 - Результаты смещения (α) при размере выборки 75** |
| **Рисунок 3.2 - Результаты смещения () при размере выборки 75** |

* Как в оценках Юла-Уокера, так и CLS и, в меньшей степени, в оценках CML, величина смещения взаимно связана с размером выборки. Можно видеть, что, удвоив размер выборки, величина смещения каждого из и будет уменьшена наполовину как в оценках Юла-Уокера, так и CLS.

**Таблица 3.2– Вариация результатов различных методов оценки для модели PoINAR (1) (θ = 1,0)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Вариация (α) | | | Вариация (θ) | | |
| n | α | Y-W | CLS | CML | Y-W | CLS | CML |
| 50 | 0,1 | 0,0237 | 0,0210 | 0,0110 | 0,0522 | 0,0442 | 0,0339 |
|  | 0,3 | 0,0372 | 0,0312 | 0,0322 | 0,0959 | 0,0728 | 0,0852 |
|  | 0,5 | 0,0332 | 0,0238 | 0,0198 | 0,1354 | 0,0900 | 0,0829 |
|  | 0,7 | 0,0271 | 0,0174 | 0,0097 | 0,1860 | 0,1108 | 0,0637 |
|  | 0,9 | 0,0087 | 0,0028 | 0,0011 | 0,4975 | 0,2252 | 0,0586 |
|  |  |  |  |  |  |  |  |
| 75 | 0,1 | 0,0174 | 0,0157 | 0,0725 | 0,0328 | 0,0271 | 0,0293 |
|  | 0,3 | 0,0228 | 0,0191 | 0,1659 | 0,0699 | 0,0544 | 0,1092 |
|  | 0,5 | 0,0238 | 0,0187 | 0,1141 | 0,1062 | 0,0812 | 0,1463 |
|  | 0,7 | 0,0156 | 0,0103 | 0,0129 | 0,1490 | 0,1017 | 0,0735 |
|  | 0,9 | 0,0049 | 0,0020 | 0,0008 | 0,2341 | 0,1208 | 0,0383 |
|  |  |  |  |  |  |  |  |
| 100 | 0,1 | 0,0108 | 0,0104 | 0,1196 | 0,0264 | 0,0236 | 0,0366 |
|  | 0,3 | 0,0153 | 0,0133 | 0,1636 | 0,0537 | 0,0436 | 0,0867 |
|  | 0,5 | 0,0172 | 0,0140 | 0,0657 | 0,0993 | 0,0790 | 0,0957 |
|  | 0,7 | 0,0088 | 0,0066 | 0,0091 | 0,0807 | 0,0601 | 0,1364 |
|  | 0,9 | 0,0041 | 0,0022 | 0,0005 | 0,2870 | 0,1716 | 0,0334 |
|  |  |  |  |  |  |  |  |
| 200 | 0,1 | 0,0066 | 0,0063 | 0,1577 | 0,0151 | 0,0136 | 0,0253 |
|  | 0,3 | 0,0104 | 0,0094 | 0,1308 | 0,0464 | 0,0415 | 0,1014 |
|  | 0,5 | 0,0107 | 0,0093 | 0,0692 | 0,0648 | 0,0561 | 0,0852 |
|  | 0,7 | 0,0051 | 0,0043 | 0,0040 | 0,0547 | 0,0456 | 0,0303 |
|  | 0,9 | 0,0012 | 0,0008 | 0,0001 | 0,0997 | 0,0723 | 0,0152 |

Аналогично были получены оценки параметров для GINAR(1) модели. Таблицы с результатами экспериментов приведены в приложении.

# Модели искажений

## 4.1. Модели искажений авторегрессионных временных рядов

Пусть в качестве исходного ряда имеется временной ряд авторегрессии порядка p. Эту исходную модель, задаваемую уравнением (1.1) будем обозначать ,

Параметрами модели являются порядок авторегрессии , вектор коэффициентов авторегрессии , «белый шум» , .

В зависимости от того, имеются ли искажения в самих параметрах модели, в канале наблюдений или нарушаются ли первоначальные предположения относительно модели (или ее параметров), различают несколько типов искажений [27]:

* Классификация наблюдений (или частный случай классификации – округление)

Вместо исходных значений временного ряда регистрируются классифицированные (округленные) наблюдения – временной ряд ,

где

* Выбросы

Значение исходного ряда в некоторые моменты времени может быть заменено дополнительной величиной:

где – последовательность независимых в совокупности случайных величин Бернулли, - последовательность случайных величин, описывающих «выбросы». Тогда – вероятность появления «выброса».

Выбросы бывают нескольких видов:

1. аддитивные выбросы:
2. заменяющие выбросы:
3. инновационные выбросы (выбросы в значениях ).

* Пропуски

В какие-то моменты времени t величина ряда по каким-либо причинам не регистрировалась.

Различают два вида пропусков: случайные (в повторении пропусков нет закономерности) и неслучайные (в частоте или времени появления пропусков имеется какая-то система).

Пропуски часто регистрируются тогда, когда значения исходного ряда настолько нетипично, что его выгоднее считать неизвестным.

* Неоднородность

Это искажение наблюдается, если не являются одинаково распределенными случайными величинами.

Неоднородность возможна в:

1. дисперсии: ;
2. математическом ожидании: ;
3. функции распределения:

Этот тип искажения можно описать и так: в каждый регистрируемый момент времени имеет разные вероятностные характеристики, однако эти характеристики различаются не более, чем на . Поэтому такие искажения еще называют –неоднородностью.

* Зависимость

Это искажение проявляется тогда, когда нарушается первоначальное предположение о независимости случайных величин – «белого шума», то есть

* Случайный величины не являются гауссовскими

Этот случай отличается от –неоднородности тем, что распределение всегда одно и то же, но оно не является нормальным.

* Параметры модели зависят от времени

То есть в модели

где – уровень искажения, – неслучайная действительная функция.

* Функциональные искажения

В данной работе более подробно рассмотрены искажения типа «выбросы».

## 4.2. Модели искажений для INAR(1).

Пусть искажение происходит в момент времени и исходная выборка размера ( из удовлетворяет (1.9).

Рассмотрим три вида искажений:

* сдвиг уровня (level shift (LS));

В данном случае будет рассматриваться новая выборка размера , где

где .

* выброс (spiky outlier (SO));

В данном случае также будет рассматриваться новая выборка размера , где

где .

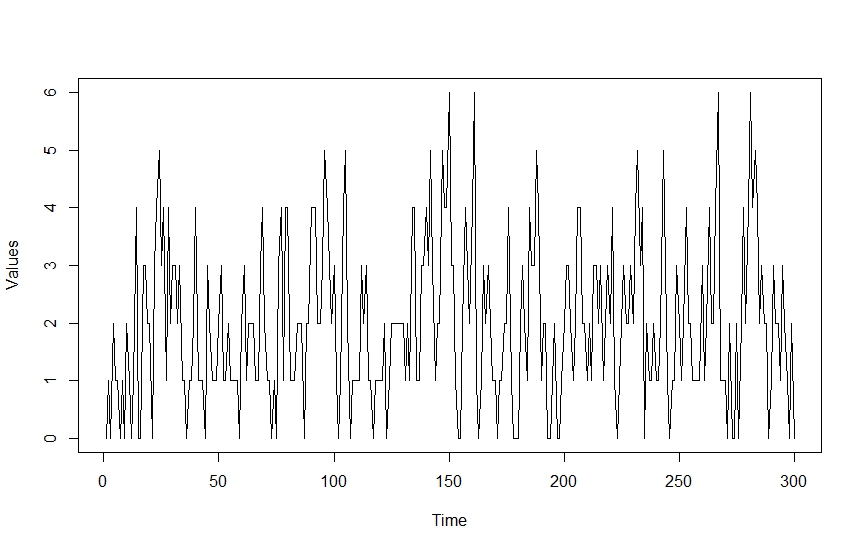
* временный сдвиг (transient shift (TS));

В первых двух случаях искажения происходят в самих данный и рассматривается новый ряд, а в данном случае изменения происходят в самой инновационной последовательности, что влияет на последующие значения, то есть модель примет вид:

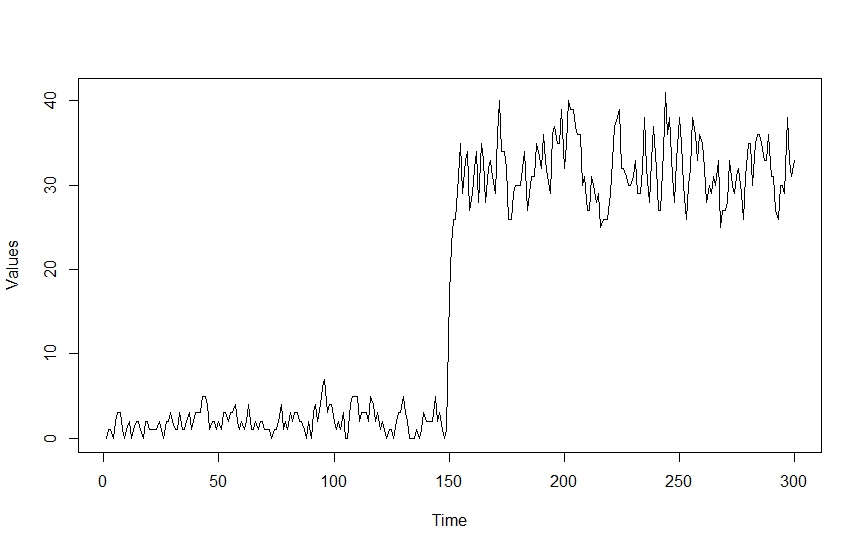
где

где .

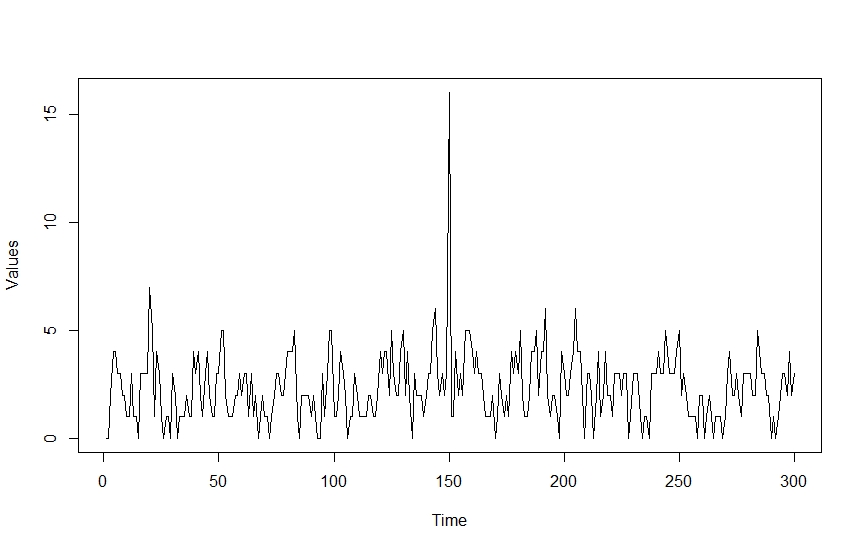
На следующих рисунках изображены ряд без искажений, ряд с искажением типа LS, SO и TS соответственно. Исходная модель представляет модель PoInar(1) с параметрами для всех случаев.



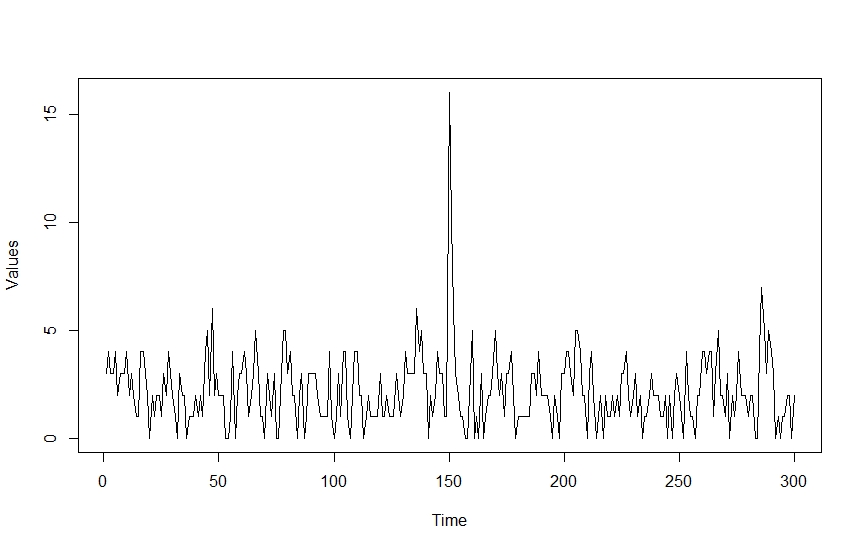
**Рисунок 4.1 - Ряд PoInar(1) без искажений.**



**Рисунок 4.2 - Ряд PoInar(1) с искажением типа LS.**



**Рисунок 4.3- Ряд PoInar(1) с искажением типа SO.**



**Рисунок 4.4- Ряд PoInar(1) с искажением типа TS.**

## 4.3. Оценка параметров модели при наличии искажений

Вопрос оценки параметров модели при наличии искажений отличается от аналогичного для стандартных моделей. Проведем сравнение параметров оценок параметров для модели PoINAR(1) и модели PoINAR(1) с искажениями различных типов. Для этого рассмотрим оценку параметра модели α оценками Y-W и CLS, приведенными в главе 3. Рассматриваем значения параметра и . Значение ε для искажений принимаем равным 10, 50 или 100.

**Таблица 4.1 – Смещение оценки параметра α при различных видах выбросов.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Смещение (α) | | | | | | | |
|  |  | Без искажений | | SO | | TS | | LS | |
| ε | α | Y-W | CLS | Y-W | CLS | Y-W | CLS | Y-W | CLS |
| 10 | 0,1 | -0,0420 | -0,0308 | -0,0682 | -0,0606 | -0,0439 | -0,0369 | 0,8300 | 0,8397 |
|  | 0,3 | -0,0935 | -0,0777 | -0,1714 | -0,1610 | -0,0579 | -0,0484 | 0,6371 | 0,6485 |
|  | 0,5 | -0,0969 | -0,0799 | -0,2316 | -0,2169 | -0,0698 | -0,0563 | 0,4546 | 0,4655 |
|  | 0,7 | -0,0504 | -0,0284 | -0,2425 | -0,2221 | -0,0418 | -0,0280 | 0,2733 | 0,2856 |
|  | 0,9 | -0,0492 | -0,0281 | -0,1355 | -0,1060 | -0,0363 | -0,0225 | 0,0831 | 0,1000 |
|  |  |  |  |  |  |  |  |  |  |
| 50 | 0,1 | -0,0468 | -0,0350 | -0,1093 | -0,1083 | -0,0169 | -0,0158 | 0,8694 | 0,8799 |
|  | 0,3 | -0,1033 | -0,0888 | -0,2933 | -0,2917 | -0,0405 | -0,0391 | 0,6734 | 0,6838 |
|  | 0,5 | -0,1042 | -0,0857 | -0,4782 | -0,4758 | -0,0403 | -0,0384 | 0,4792 | 0,4898 |
|  | 0,7 | -0,0603 | -0,0393 | -0,6369 | -0,6320 | -0,0226 | -0,0202 | 0,2841 | 0,2952 |
|  | 0,9 | -0,0374 | -0,0186 | -0,6340 | -0,6137 | -0,0175 | -0,0142 | 0,0855 | 0,1024 |
|  |  |  |  |  |  |  |  |  |  |
| 100 | 0,1 | -0,0575 | -0,0455 | -0,1115 | -0,1110 | -0,0135 | -0,0130 | 0,8718 | 0,8819 |
|  | 0,3 | -0,0890 | -0,0753 | -0,3071 | -0,3065 | -0,0186 | -0,0180 | 0,6763 | 0,6869 |
|  | 0,5 | -0,0839 | -0,0658 | -0,4999 | -0,4990 | -0,0184 | -0,0176 | 0,4811 | 0,4916 |
|  | 0,7 | -0,0588 | -0,0366 | -0,6909 | -0,6890 | -0,0228 | -0,0215 | 0,2848 | 0,2964 |
|  | 0,9 | -0,0382 | -0,0193 | -0,8110 | -0,8026 | -0,0227 | -0,0202 | 0,0856 | 0,1028 |

**Таблица 4.2 – Вариация оценки параметра α при различных видах выбросов.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Вариация (α) | | | | | | | |
|  |  | Без искажений | | SO | | TS | | LS | |
| ε | α | Y-W | CLS | Y-W | CLS | Y-W | CLS | Y-W | CLS |
| 10 | 0,1 | 0,0125 | 0,0120 | 0,0099 | 0,0090 | 0,0104 | 0,0099 | 0,6881 | 0,7055 |
|  | 0,3 | 0,0227 | 0,0196 | 0,0390 | 0,0354 | 0,0174 | 0,0160 | 0,4026 | 0,4187 |
|  | 0,5 | 0,0200 | 0,0164 | 0,0672 | 0,0604 | 0,0137 | 0,0122 | 0,2080 | 0,2182 |
|  | 0,7 | 0,0111 | 0,0078 | 0,0601 | 0,0508 | 0,0097 | 0,0083 | 0,0745 | 0,0814 |
|  | 0,9 | 0,0055 | 0,0029 | 0,0207 | 0,0139 | 0,0029 | 0,0017 | 0,0068 | 0,0101 |
|  |  |  |  |  |  |  |  |  |  |
| 50 | 0,1 | 0,0076 | 0,0073 | 0,0121 | 0,0119 | 0,0019 | 0,0019 | 0,7550 | 0,7738 |
|  | 0,3 | 0,0202 | 0,0170 | 0,0892 | 0,0883 | 0,0033 | 0,0033 | 0,4539 | 0,4676 |
|  | 0,5 | 0,0217 | 0,0178 | 0,2321 | 0,2298 | 0,0062 | 0,0061 | 0,2296 | 0,2399 |
|  | 0,7 | 0,0088 | 0,0060 | 0,4160 | 0,4094 | 0,0034 | 0,0032 | 0,0804 | 0,0874 |
|  | 0,9 | 0,0023 | 0,0012 | 0,4161 | 0,3909 | 0,0011 | 0,0009 | 0,0072 | 0,0106 |
|  |  |  |  |  |  |  |  |  |  |
| 100 | 0,1 | 0,0117 | 0,0104 | 0,0126 | 0,0125 | 0,0016 | 0,0015 | 0,7600 | 0,7776 |
|  | 0,3 | 0,0210 | 0,0182 | 0,0970 | 0,0966 | 0,0029 | 0,0029 | 0,4581 | 0,4716 |
|  | 0,5 | 0,0176 | 0,0146 | 0,2557 | 0,2547 | 0,0021 | 0,0021 | 0,2314 | 0,2417 |
|  | 0,7 | 0,0117 | 0,0080 | 0,4703 | 0,4678 | 0,0015 | 0,0014 | 0,0811 | 0,0879 |
|  | 0,9 | 0,0035 | 0,0018 | 0,6468 | 0,6338 | 0,0008 | 0,0006 | 0,0073 | 0,0106 |

Из Таблица 4.1 и Таблица 4.2 видно, что и смещение и вариация растут при наличии выбросов, но от размера ε не имеют сильной зависимости. Причем, при выбросах типа TS значения смещения и вариации меняются меньше всего (в среднем в 2,5 раза). В тоже время, при выбросах типа LS эти показатели увеличиваются в более чем 20 раз, что можно видеть из Рисунок 4.5, Рисунок 4.6, Рисунок 4.7, Рисунок 4.8. Из всего вышесказанного вытекает необходимость построения новых оценок, которые в меньшей степени будут зависеть от искажений.

**Рисунок 4.5 – Смещение оценки Y-W для α**

**Рисунок 4.6** **– Смещение оценки CLS для α**

**Рисунок 4.7 -** **Вариация оценки Y-W для α**

**Рисунок 4.8 -** **Вариация оценки CLS для α**

Наблюдения с выбросами требуют для анализа робастной модификации метода наименьших квадратов. В работе [2] для регрессионной модели вместо оценки , получаемой в результате минимизации

|  |  |
| --- | --- |
|  | (4.1) |

рассматривается M-оценка , как решение задачи минимизации

|  |  |
| --- | --- |
|  | (4.2) |

где ρ(t) – функция, которая растет медленнее, чем . В работе [5] предлагается обобщенная M-оценка, получаемая как решение системы

|  |  |
| --- | --- |
|  | (4.3) |

В данной работе система уравнений (4.3) используется для оценивания параметров авторегрессионной модели, причем оценки записываются в рекуррентной форме

|  |  |
| --- | --- |
|  | (4.4) |

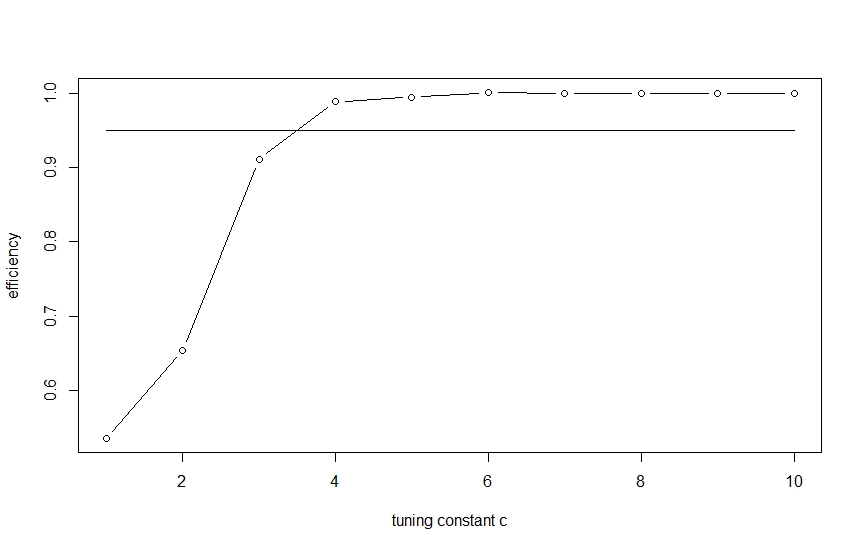
где

c – параметр. Этот метод оценки параметров в дальнейшем именуем MCLS.

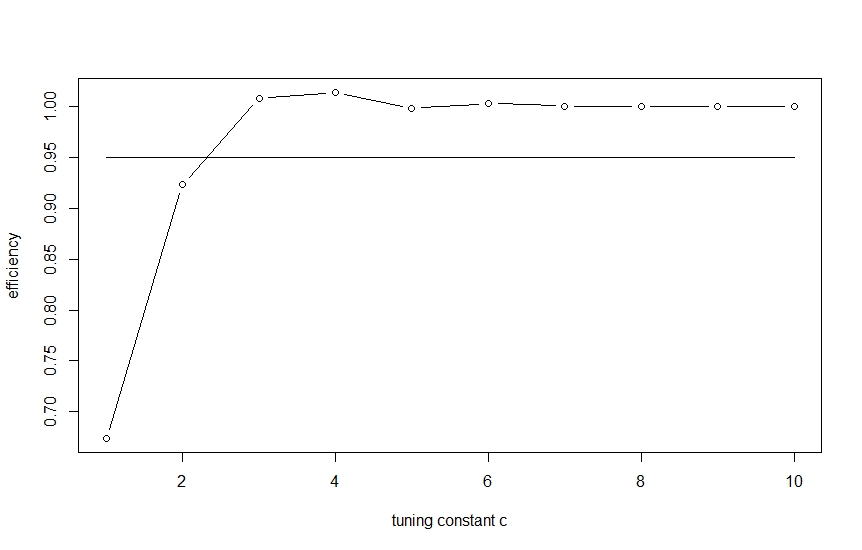
## Выбор константы усечения с

Основным условием надежности метода является вычисление правильного значения с для достижения определенных критериев оптимальности. Константа усечения с выбирается так, чтобы обеспечить заранее определенный уровень относительной эффективности. Таким образом, на данных, не содержащих искажения, MSE MCLS сравнивается с MSE другой оценки (не требующей выбора параметров (например: CLS), и это должно привести к следующему равенству:

В следующем примере моделирования 1000 выборок чистых данных, то есть данных без искажений, генерируются в соответствии с моделью (1.9). Параметры модели оцениваются для численных значений с, изменяющихся от 1 до 10, и вычисляется относительная эффективность.



**Рисунок 4.9 – Выбор константы усечения с для модели PoINAR(1)**



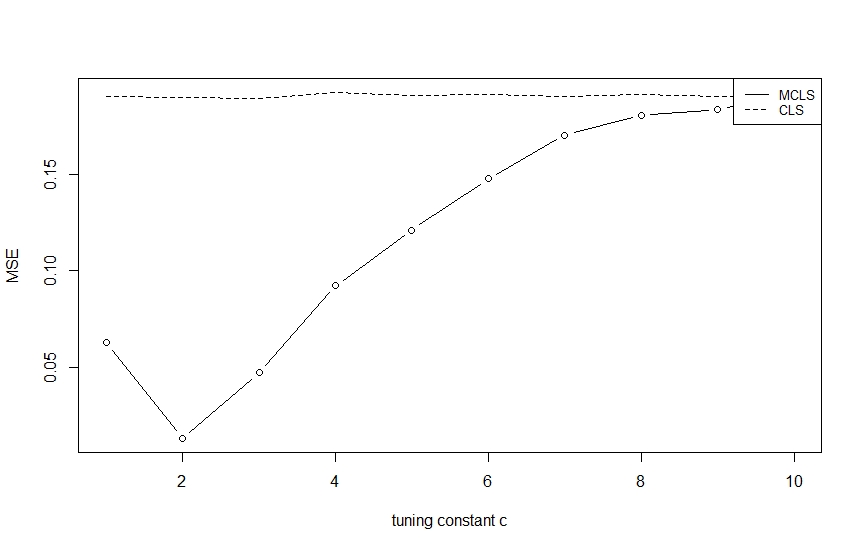
**Рисунок 4.10 - Выбор константы усечения с для модели GINAR(1)**

На Рисунок 4.9 и Рисунок 4.10 показаны относительные эффективности метода оценки в двух случаях, соответствующих модели PoINAR(1) и GINAR(1). Во всех случаях MCLS обладает чрезвычайно высоким уровнем эффективности, который фактически достигает эффективности CLS при увеличении значения c. Рекомендуется, чтобы наиболее подходящая величина c, обеспечивающая 95% -ный уровень эффективности, составляла от 3 до 4 для модели PoINAR(1) и от 2 до 3 для модели GINAR(1).

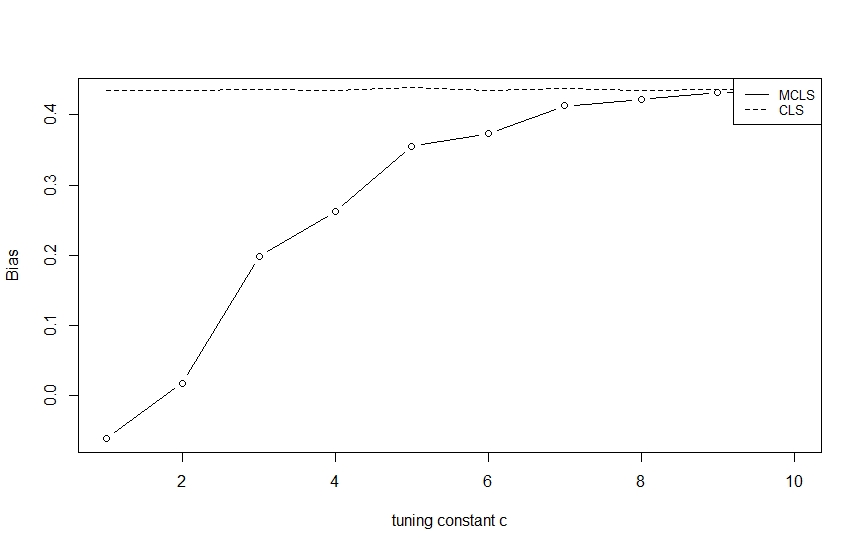
## Экспериментальные исследования

### Сдвиг уровня и временной сдвиг

В данном разделе рассмотрим оценки параметров моделей с искажениями типа сдвиг уровня и временной сдвиг, которые происходят в момент времени . Рассмотрена модель с искажениями PoINAR(1) c параметрами и .

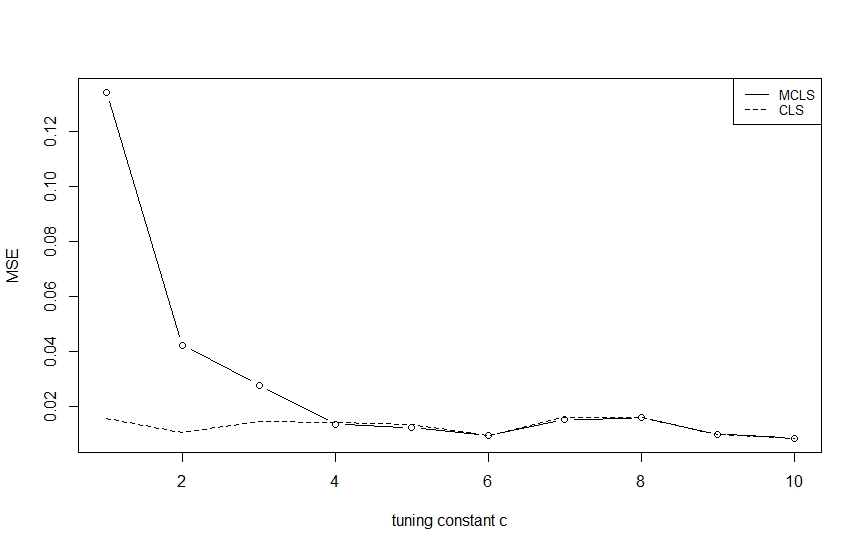


**Рисунок 4.11 – Вариация для LS модели PoINAR(1)**

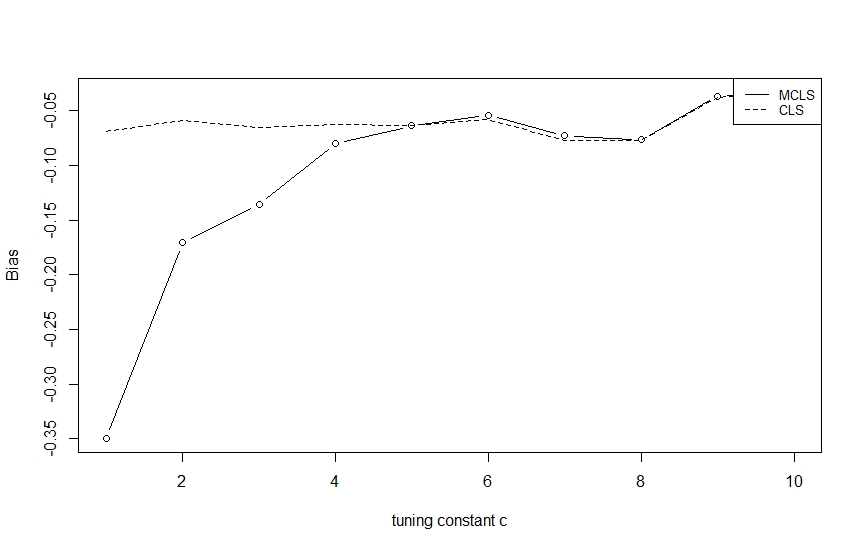


**Рисунок 4.12 – Смещение для LS модели PoINAR(1)**

Экспериментальное исследование по эффекту вмешательства LS или TS показывает, что в этих случаях применение м-оценки гораздо эффективнее.



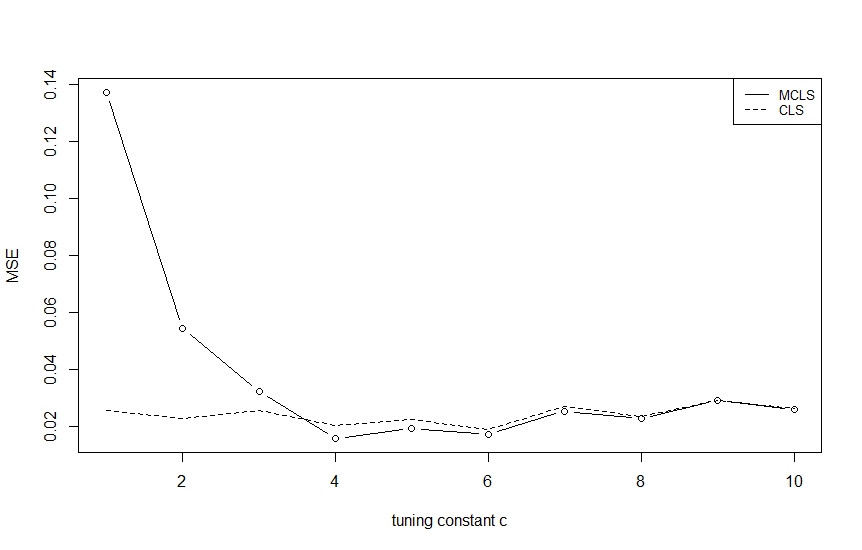
**Рисунок 4.13** **– Вариация для TS модели PoINAR(1)**



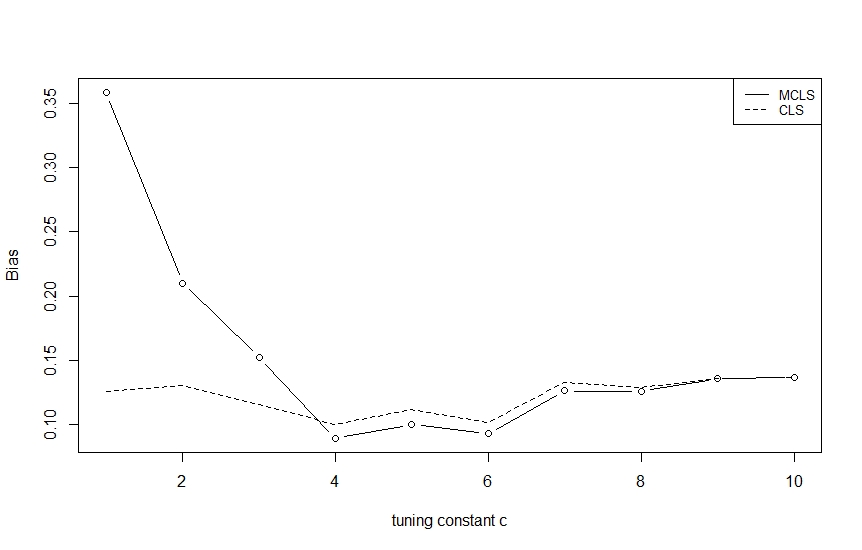
**Рисунок 4.14** **– Смещение для TS модели PoINAR(1)**

### Выбросы

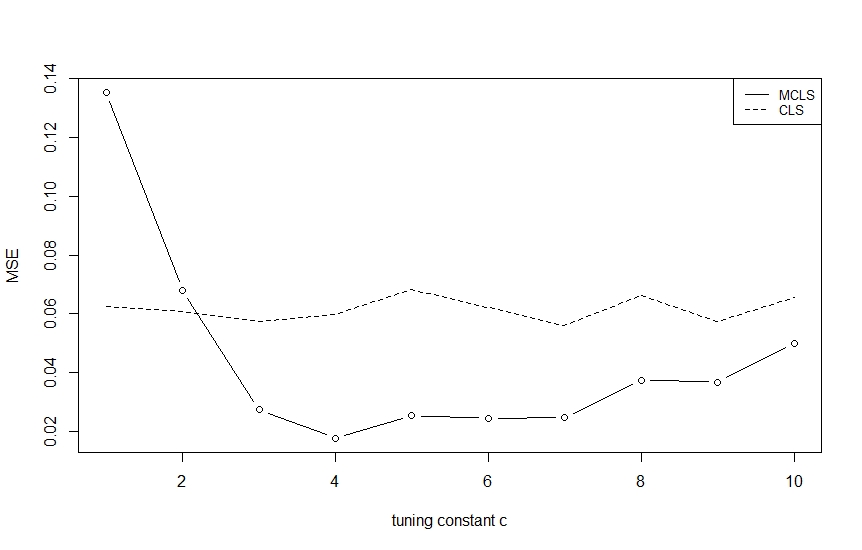
**Единичный выброс** в момент врмени размером 10 или 30. Это имеет влияние на наблюдаемую серию в момент времени и несколько лагов после этой временной точки. По рисункам, приведенным ниже видно, что при константе усечения большей 2 качество MCLS оценки сравнимо с качестов оценки CLS, а в некоторых случаях даже лучше.



**Рисунок 4.15** **– Вариация для SO модели PoINAR(1) при**



**Рисунок 4.16** **– Смещение для SO модели PoINAR(1) при**



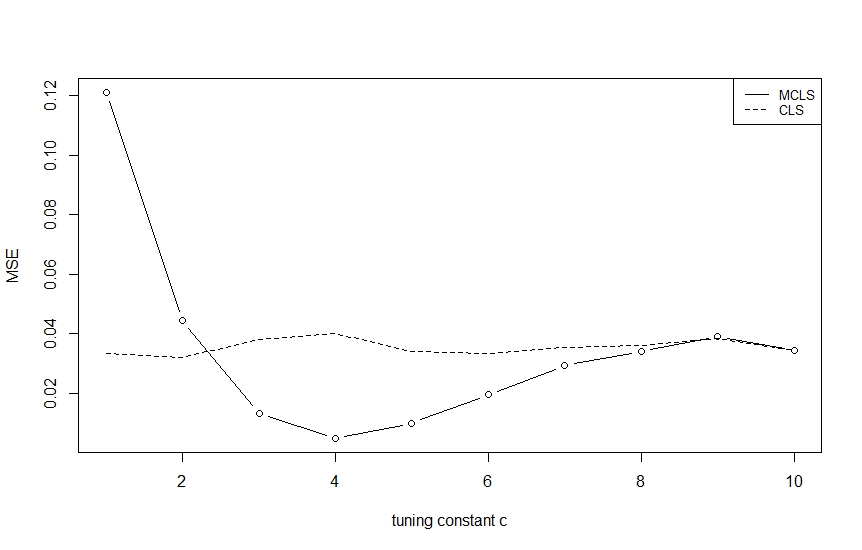
**Рисунок 4.17 – Вариация для SO модели PoINAR(1) при**



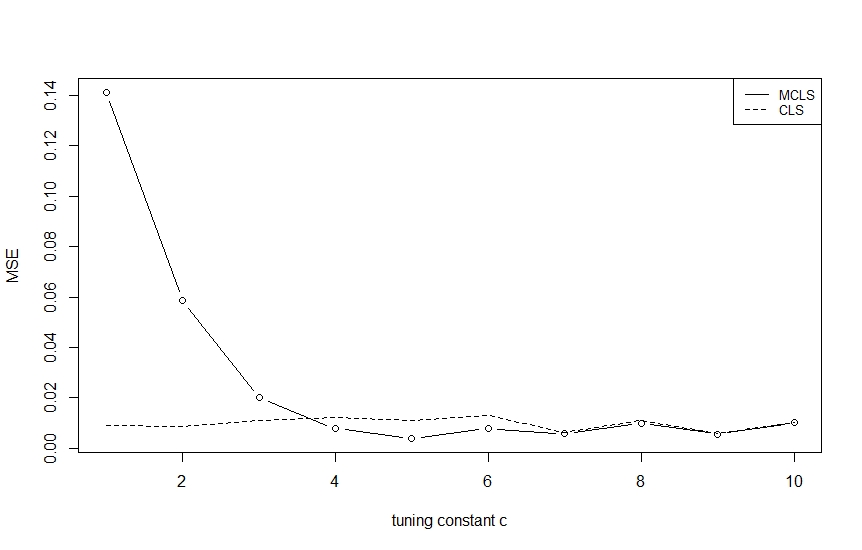
**Рисунок 4.18** **– Вариация для SO модели PoINAR(1) при**

**Набор последовательных выбросов**

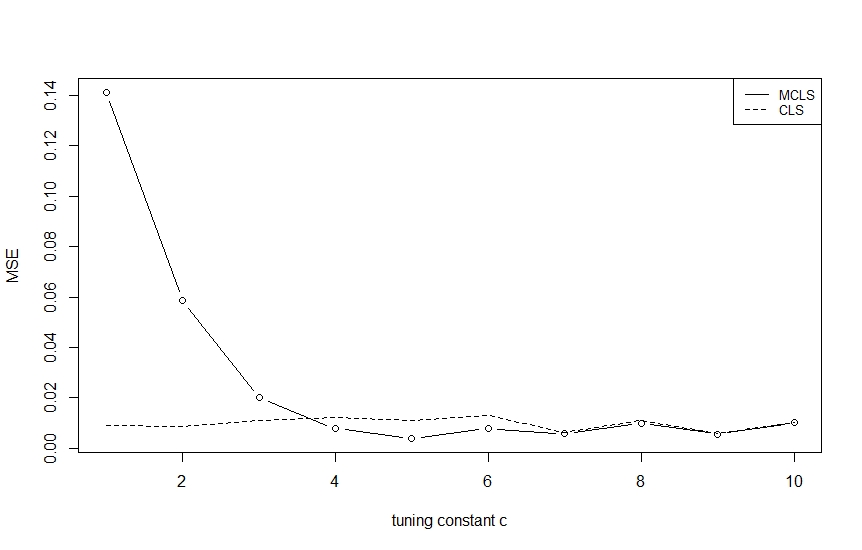
В данном случае промежутки времени смежны. Следовательно, если, например, мы имеем временные ряды размером n = 100, k = 15, и ε = 10, то мы будем наблюдать пятнадцать последовательных выбросов, возникающих в моменты времени t = 25, 26, ..., 39. Ясно, что наилучшей оценкой в данном случае является м-оценка MCLS при оптимальном выборе константы усечения. Также из нижеприведенных рисунков видно, что от количества выбросов оценки не зависят.



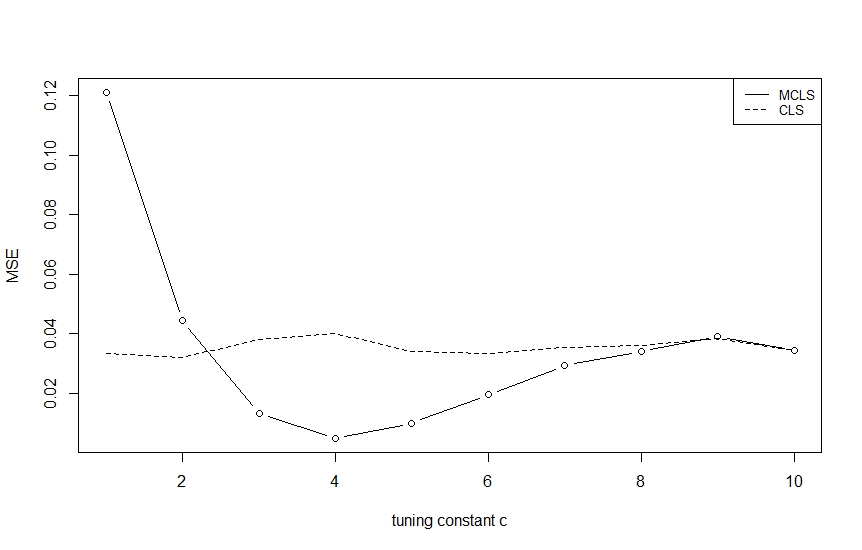
**Рисунок 4.19** **– Вариация для набора SO модели PoINAR(1) при**



**Рисунок 4.20** **– Вариация для набора SO модели PoINAR(1) при**



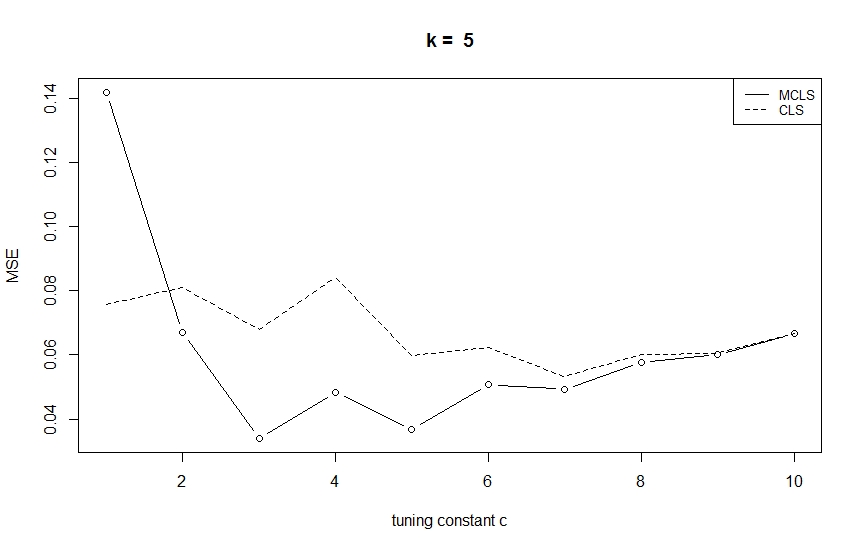
**Рисунок 4.21** **– Вариация для набора SO модели PoINAR(1) при**



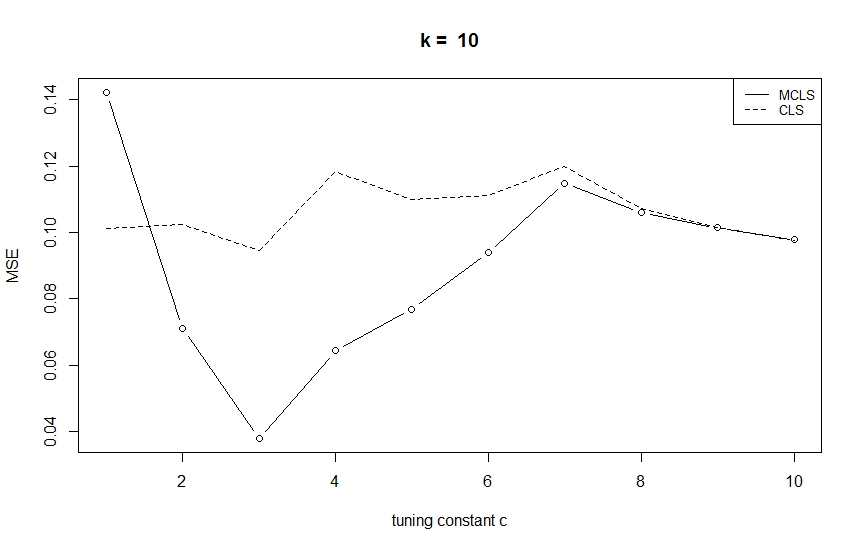
**Рисунок 4.22** **– Вариация для набора SO модели PoINAR(1) при**

**Изолированные выбросы**

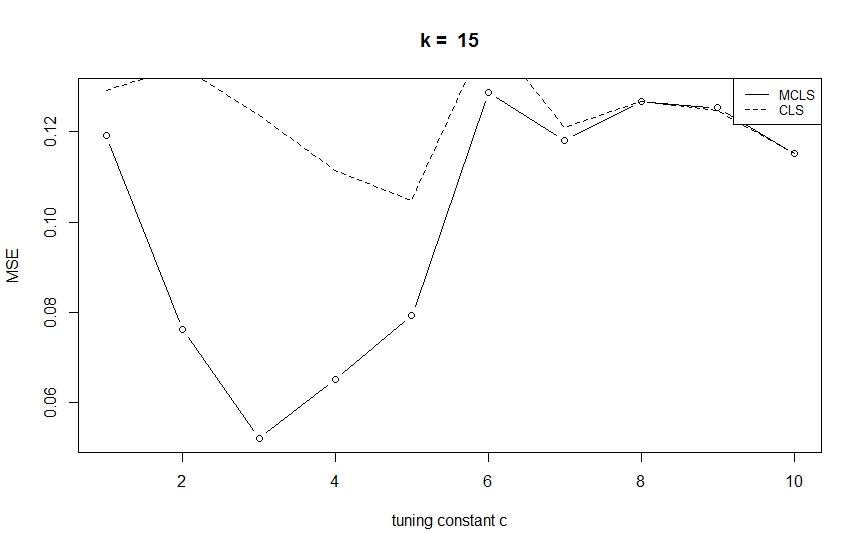
Далее мы рассмотрим случай нескольких изолированных больших аддитивных выбросов. Пусть k = 5, 10, 15, 20 и предположим, что размер ε равен 10, 20, 30. Выбросы добавляются к наблюдаемым данным в случайно выбранных позициях. Из рисунков ниже видно, что чем больше колличество изолированных выбросов, тем точнее (при правильном выборе константы с) становится оценка параметров.



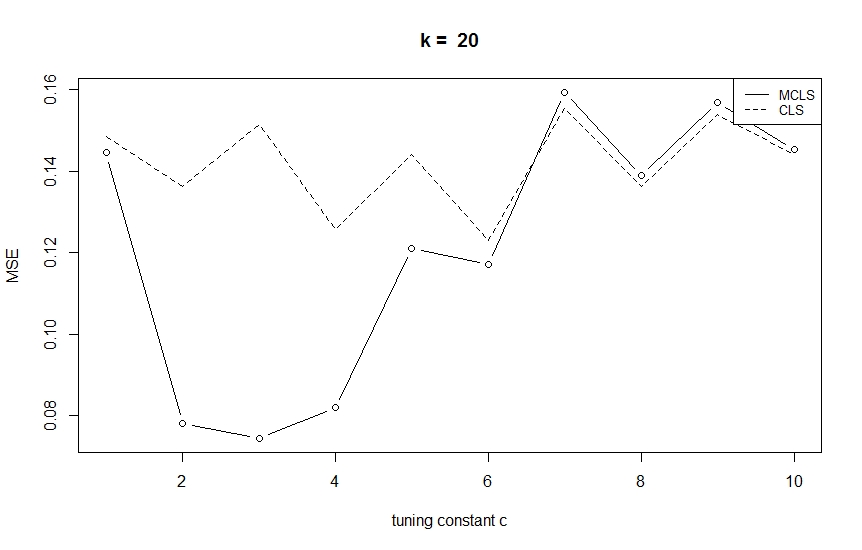
**Рисунок 4.23** **– Вариация для изолированных SO модели PoINAR(1) при**



**Рисунок 4.24 – Вариация для изолированных SO модели PoINAR(1) при**



**Рисунок 4.25** **– Вариация для изолированных SO модели PoINAR(1) при**



**Рисунок 4.26** **– Вариация для изолированных SO модели PoINAR(1) при**

# Заключение

В работе рассмотрена модель INAR(1) и ее спецификация, в случае когда инновационная последовательность удовлетворяет (2.1) – модель PSINAR(1), а также частные случаи - PoINAR(1) и GINAR(1).

Основные результаты работы:

1. Рассмотрены свойства авторегрессионных временных рядов и аналогичные свойства для модели INAR(1).
2. Описана модель PSINAR(1) и ее основные характеристики.
3. Рассмотрены частные случаи модели PSINAR(1) и их свойства.
4. Построены оценки методами Юле-Уокера, наименьших квадратов и максимального правдоподобия.
5. Для построенных методов проведены экспериментальные исследования и анализ с помощью смещения и вариации.
6. Построены модели с искажениями типа: временной сдвиг, сдвиг уровня и выброс.
7. С помощью экспериментального исследования получено, что построенные оценки являются смещенными для модели с искажениями, поэтому для таких моделей с искажениями построены м-оценки.
8. Предложена процедура прогнозирования.
9. Разработано ПО для анализа модели INAR(1), ее оценок, а также для прогнозирования модели.

# Список литературы

1. Al–Osh M.A. ,Alzaid A.A. (1987). First–order integer–valued autoregressive (INAR(1)) process. Journal of Time Series Analysis, 261-275.
2. Andrews D.F. (1974). A robust method for multiple linear regression. Technometrics, 523–531.
3. Brannas K. . (1994). Estimation and testing in integer-valued AR(1) models. Umea Economic Studies, 335.
4. Christou V., Fokianos K. (2015). Estimation and testing linearity for non-linear mixed poisson autoregressions. Electronic Journal of Statistics, 1357-1377.
5. Denby L., Laren W.A. (1977). Robust regression estimators compared via Monte – Carlo. Comm. Statist. Assoc., 335–362.
6. Freeland R. K., McCabe B. P. M. (2005). Asymptotic properties of CLS estimators in the Poisson AR(1) model. Statistcs & Probability Letters, 147-153.
7. Freeland R.K., McCabe B.P.M. (2003). Forecasting discrete valued low count series. International Journal of Forecasting, 427-434.
8. Harvey Andrew C., Fernandes Cristiano. (1989). Time Series Models for Count or Qualitative Observations: Reply. Journal of Business and Economic Statistics, 422.
9. Hoel P. G., Port S. C., Stone C. J. (1972). Introduction to Stochastic Processes. Long Grove, IL: Waveland Press.
10. Jazi M. A., Jones G., Lai C. D. (2012). First-order integer valued AR processes with zero inflated Poisson innovations. . Journal of Time Series Analysis, 954-963.
11. Johnson N. L., Kemp A. K., Kotz, S. . (2005). Univariate Discrete Distributions 3rd ed. . Hoboken, NJ: Wiley.
12. Jung R. C., Ronning G., Tremayne, A. R. (2005). Estimation in conditional first order autoregressiongression with discrete support. Statistical Papers, 195-224.
13. Jung R. C., Tremayne, A. R. (2003). Testing for serial dependence in time series models of counts. Journal ofTime Series Analysis, 65-84.
14. Jung Robert C., Tremayne A. R. (2011). Convolution‐closed models for count time series with applications. Journal of Time Series Analysis, 268-280.
15. Klimko L., Nelson P. (1978). On conditional least squares estimation for stochastic processes. The Annals of Statistic, 629-642.
16. McKenzie E. (1985). Some simple models for discrete variate time series. Water Resources Bulletin, 645–650.
17. Noack A. (1950). A class of random variables with discrete distributions. Annals of Mathematical Statistics, 127-132.
18. Park Y. ,Oh C. W. (1997). Some asymptotic properties in INAR(1) processes with Poisson marginals. Statistical Papers, 287-302.
19. Pavlopoulos H., Karlis D. (2008). INAR(1) modeling of overdispersed count series with an enviromental application. Environmetrics, 369-393.
20. Silva M. E.,Oliveira V. L. (2004). Difference equations for the higher-order moments and cumulants of the INAR(1) model. Journal of Time Series Analysis, 317-333.
21. Sprott D. A. (1983). Estimating the parameters of a convolution by maximum likelihood. Journal of the American Statistical Association, 457-460.
22. Stuetal F. W., Van Harn K. (1979). Discrete analogues of self-decomposability and stability. The Annals of Probability, 893-899.
23. Weiß C.H. (2008). Thinning operations for modelling time series of counts—a survey. Advances in Statistical Analysis, 319-341.
24. Дженкинс Г. Бокс Дж. (1974). Анализ временных рядов, прогноз и управление. М.: Мир.
25. Жук Е.Е. Харин Ю.С. Зуев Н.М. (2011). Теория вероятностей математическая и прикладная статистика. Минск: БГУ.
26. Ибрагимов Н.М. и др. Суслов В.И. (2005). Эконометрика. Новосибирск: Новосибирский государственный университет.
27. Харин Ю.С. (2008). Оптимальность и робастность в статистическом прогнозировании. Минск: БГУ.

# Приложения



Таблицы сравнения смещения и вариации оценок параметров модели GINAR(1) при различных значениях параметра .

**Таблица А.1 – Смещение результатов различных методов оценки для модели GINAR (1) (θ = 0.7).**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Смещение (α) | | | Смещение (θ) | | |
| n | α | Y-W | CLS | CML | Y-W | CLS | CML |
| 50 | 0,1 | -0,0515 | -0,0322 | 0,0091 | 0,1199 | 0,1364 | 0,0038 |
|  | 0,3 | -0,0707 | -0,0435 | -0,0003 | 0,1131 | 0,1327 | 0,0027 |
|  | 0,5 | -0,0708 | -0,0370 | 0,0035 | 0,1111 | 0,1356 | 0,0051 |
|  | 0,7 | -0,0929 | -0,0470 | -0,0023 | 0,0815 | 0,1188 | -0,0002 |
|  | 0,9 | -0,0772 | -0,0322 | -0,0027 | 0,0449 | 0,0893 | 0,0024 |
|  |  |  |  |  |  |  |  |
| 75 | 0,1 | -0,0195 | -0,0077 | -0,0007 | 0,1280 | 0,1396 | 0,0002 |
|  | 0,3 | -0,0500 | -0,0309 | -0,0029 | 0,1111 | 0,1264 | -0,0013 |
|  | 0,5 | -0,0457 | -0,0249 | -0,0046 | 0,1124 | 0,1284 | -0,0010 |
|  | 0,7 | -0,0580 | -0,0315 | -0,0001 | 0,1162 | 0,1395 | 0,0092 |
|  | 0,9 | -0,0501 | -0,0197 | -0,0012 | 0,0700 | 0,1132 | 0,0027 |
|  |  |  |  |  |  |  |  |
| 100 | 0,1 | -0,0198 | -0,0110 | -0,0049 | 0,1305 | 0,1384 | 0,0013 |
|  | 0,3 | -0,0366 | -0,0245 | 0,0015 | 0,1123 | 0,1222 | -0,0014 |
|  | 0,5 | -0,0429 | -0,0274 | -0,0029 | 0,1049 | 0,1169 | -0,0035 |
|  | 0,7 | -0,0512 | -0,0307 | 0,0003 | 0,1008 | 0,1202 | 0,0047 |
|  | 0,9 | -0,0381 | -0,0164 | 0,0019 | 0,0854 | 0,1162 | 0,0084 |
|  |  |  |  |  |  |  |  |
| 200 | 0,1 | -0,0224 | -0,0181 | -0,0058 | 0,1190 | 0,1228 | -0,0026 |
|  | 0,3 | -0,0280 | -0,0216 | -0,0036 | 0,1166 | 0,1214 | -0,0011 |
|  | 0,5 | -0,0345 | -0,0272 | -0,0051 | 0,1131 | 0,1191 | 0,0005 |
|  | 0,7 | -0,0246 | -0,0158 | -0,0003 | 0,1169 | 0,1255 | 0,0024 |
|  | 0,9 | -0,0211 | -0,0105 | -0,0004 | 0,1007 | 0,1197 | 0,0025 |

**Таблица А.2 – Вариация результатов различных методов оценки для модели GINAR (1) (θ = 0.7).**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Вариация (α) | | | Вариация (θ) | | |
| n | α | Y-W | CLS | CML | Y-W | CLS | CML |
| 50 | 0,1 | 0,0219 | 0,0211 | 0,0066 | 0,0237 | 0,0291 | 0,0018 |
|  | 0,3 | 0,0246 | 0,0206 | 0,0060 | 0,0245 | 0,0302 | 0,0019 |
|  | 0,5 | 0,0224 | 0,0168 | 0,0040 | 0,0261 | 0,0330 | 0,0016 |
|  | 0,7 | 0,0239 | 0,0144 | 0,0016 | 0,0269 | 0,0397 | 0,0019 |
|  | 0,9 | 0,0103 | 0,0034 | 0,0003 | 0,0203 | 0,0278 | 0,0017 |
| 75 | 0,1 | 0,0111 | 0,0107 | 0,0034 | 0,0229 | 0,0263 | 0,0011 |
|  | 0,3 | 0,0174 | 0,0170 | 0,0034 | 0,0181 | 0,0238 | 0,0009 |
|  | 0,5 | 0,0123 | 0,0101 | 0,0027 | 0,0206 | 0,0248 | 0,0011 |
|  | 0,7 | 0,0127 | 0,0087 | 0,0010 | 0,0295 | 0,0364 | 0,0014 |
|  | 0,9 | 0,0055 | 0,0023 | 0,0002 | 0,0263 | 0,0382 | 0,0053 |
| 100 | 0,1 | 0,0100 | 0,0097 | 0,0025 | 0,0218 | 0,0241 | 0,0007 |
|  | 0,3 | 0,0113 | 0,0104 | 0,0029 | 0,0172 | 0,0198 | 0,0006 |
|  | 0,5 | 0,0108 | 0,0094 | 0,0018 | 0,0170 | 0,0200 | 0,0008 |
|  | 0,7 | 0,0092 | 0,0068 | 0,0009 | 0,0199 | 0,0245 | 0,0009 |
|  | 0,9 | 0,0038 | 0,0017 | 0,0001 | 0,0274 | 0,0346 | 0,0041 |
| 200 | 0,1 | 0,0053 | 0,0051 | 0,0012 | 0,0168 | 0,0178 | 0,0004 |
|  | 0,3 | 0,0055 | 0,0051 | 0,0015 | 0,0163 | 0,0176 | 0,0004 |
|  | 0,5 | 0,0051 | 0,0045 | 0,0009 | 0,0160 | 0,0175 | 0,0004 |
|  | 0,7 | 0,0037 | 0,0032 | 0,0005 | 0,0199 | 0,0220 | 0,0005 |
|  | 0,9 | 0,0014 | 0,0008 | 0,0001 | 0,0218 | 0,0257 | 0,0005 |

**Таблица А.3 – Смещение результатов различных методов оценки для модели GINAR (1) (θ = 0.5).**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Смещение (α) | | | Смещение (θ) | | |
| n | α | Y-W | CLS | CML | Y-W | CLS | CML |
| 50 | 0,1 | -0,0345 | -0,0208 | 0,0133 | 0,4586 | 0,4928 | 0,2264 |
|  | 0,3 | -0,1239 | -0,1030 | -0,0356 | 0,3486 | 0,3841 | 0,1047 |
|  | 0,5 | -0,1210 | -0,0919 | -0,0186 | 0,3404 | 0,3836 | 0,1278 |
|  | 0,7 | -0,1052 | -0,0704 | -0,0146 | 0,3505 | 0,4144 | 0,1746 |
|  | 0,9 | -0,0803 | -0,0409 | -0,0052 | 0,3078 | 0,3761 | 0,1756 |
|  |  |  |  |  |  |  |  |
| 75 | 0,1 | -0,0394 | -0,0301 | 0,0002 | 0,4443 | 0,4685 | 0,2004 |
|  | 0,3 | -0,0839 | -0,0713 | -0,0206 | 0,3737 | 0,3933 | 0,1173 |
|  | 0,5 | -0,0992 | -0,0809 | -0,0147 | 0,3299 | 0,3583 | 0,0898 |
|  | 0,7 | -0,0681 | -0,0457 | -0,0012 | 0,3796 | 0,4211 | 0,1932 |
|  | 0,9 | -0,0585 | -0,0329 | -0,0052 | 0,3395 | 0,4070 | 0,1779 |
|  |  |  |  |  |  |  |  |
| 100 | 0,1 | -0,0373 | -0,0307 | 0,0037 | 0,4500 | 0,4647 | 0,2000 |
|  | 0,3 | -0,0918 | -0,0820 | -0,0280 | 0,3494 | 0,3652 | 0,0677 |
|  | 0,5 | -0,0838 | -0,0715 | -0,0184 | 0,3481 | 0,3662 | 0,0902 |
|  | 0,7 | -0,0651 | -0,0489 | -0,0067 | 0,3777 | 0,4084 | 0,1570 |
|  | 0,9 | -0,0418 | -0,0226 | 0,0006 | 0,3575 | 0,4257 | 0,1955 |
|  |  |  |  |  |  |  |  |
| 200 | 0,1 | -0,0332 | -0,0299 | -0,0048 | 0,4427 | 0,4506 | 0,1562 |
|  | 0,3 | -0,0656 | -0,0605 | -0,0184 | 0,3582 | 0,3662 | 0,0193 |
|  | 0,5 | -0,0715 | -0,0653 | -0,0138 | 0,3355 | 0,3447 | 0,0336 |
|  | 0,7 | -0,0475 | -0,0399 | -0,0041 | 0,3836 | 0,3997 | 0,1131 |
|  | 0,9 | -0,0251 | -0,0160 | 0,0018 | 0,4094 | 0,4367 | 0,2330 |

**Таблица А.4 – Вариация результатов различных методов оценки для модели GINAR (1) (θ = 0.5).**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Вариация (α) | | | Вариация (θ) | | |
| n | α | Y-W | CLS | CML | Y-W | CLS | CML |
| 50 | 0,1 | 0,0212 | 0,0214 | 0,0104 | 0,2610 | 0,2994 | 0,1852 |
|  | 0,3 | 0,0337 | 0,0292 | 0,0109 | 0,1630 | 0,1935 | 0,0934 |
|  | 0,5 | 0,0352 | 0,0282 | 0,0089 | 0,1622 | 0,2032 | 0,1025 |
|  | 0,7 | 0,0247 | 0,0172 | 0,0047 | 0,1947 | 0,2518 | 0,1453 |
|  | 0,9 | 0,0114 | 0,0046 | 0,0006 | 0,2180 | 0,2463 | 0,1711 |
|  |  |  |  |  |  |  |  |
| 75 | 0,1 | 0,0141 | 0,0135 | 0,0066 | 0,2350 | 0,2602 | 0,1559 |
|  | 0,3 | 0,0251 | 0,0232 | 0,0097 | 0,1759 | 0,1927 | 0,1025 |
|  | 0,5 | 0,0234 | 0,0194 | 0,0052 | 0,1502 | 0,1709 | 0,0841 |
|  | 0,7 | 0,0148 | 0,0110 | 0,0026 | 0,2109 | 0,2493 | 0,1575 |
|  | 0,9 | 0,0068 | 0,0034 | 0,0004 | 0,2281 | 0,3073 | 0,1768 |
|  |  |  |  |  |  |  |  |
| 100 | 0,1 | 0,0123 | 0,0119 | 0,0058 | 0,2330 | 0,2482 | 0,1465 |
|  | 0,3 | 0,0207 | 0,0190 | 0,0071 | 0,1454 | 0,1579 | 0,0617 |
|  | 0,5 | 0,0181 | 0,0159 | 0,0043 | 0,1522 | 0,1663 | 0,0770 |
|  | 0,7 | 0,0112 | 0,0088 | 0,0018 | 0,1818 | 0,2072 | 0,1141 |
|  | 0,9 | 0,0039 | 0,0022 | 0,0003 | 0,2097 | 0,2822 | 0,1572 |
|  |  |  |  |  |  |  |  |
| 200 | 0,1 | 0,0065 | 0,0063 | 0,0031 | 0,2088 | 0,2161 | 0,1012 |
|  | 0,3 | 0,0102 | 0,0096 | 0,0030 | 0,1388 | 0,1450 | 0,0276 |
|  | 0,5 | 0,0110 | 0,0100 | 0,0023 | 0,1286 | 0,1349 | 0,0367 |
|  | 0,7 | 0,0053 | 0,0045 | 0,0010 | 0,1755 | 0,1892 | 0,0843 |
|  | 0,9 | 0,0018 | 0,0012 | 0,0001 | 0,2447 | 0,2638 | 0,1879 |

**Таблица А.5 – Смещение результатов различных методов оценки для модели GINAR (1) (θ = 0.3).**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Смещение (α) | | | Смещение (θ) | | |
| n | α | Y-W | CLS | CML | Y-W | CLS | CML |
| 50 | 0,1 | -0,0523 | -0,0435 | -0,0174 | 1,3699 | 1,4295 | 1,3699 |
|  | 0,3 | -0,1645 | -0,1486 | -0,1418 | 0,8265 | 0,8692 | 0,8068 |
|  | 0,5 | -0,2346 | -0,2117 | -0,1871 | 0,6159 | 0,6618 | 0,5635 |
|  | 0,7 | -0,1976 | -0,1634 | -0,1546 | 0,6565 | 0,7279 | 0,6131 |
|  | 0,9 | -0,1204 | -0,0768 | -0,0639 | 0,8134 | 0,9736 | 0,7812 |
|  |  |  |  |  |  |  |  |
| 75 | 0,1 | -0,0695 | -0,0632 | -0,0394 | 1,2503 | 1,2858 | 1,2503 |
|  | 0,3 | -0,1473 | -0,1366 | -0,1395 | 0,8148 | 0,8441 | 0,8129 |
|  | 0,5 | -0,1965 | -0,1810 | -0,1791 | 0,6279 | 0,6619 | 0,5997 |
|  | 0,7 | -0,1647 | -0,1432 | -0,1424 | 0,6793 | 0,7286 | 0,6516 |
|  | 0,9 | -0,0804 | -0,0575 | -0,0641 | 1,0027 | 1,1046 | 0,9815 |
|  |  |  |  |  |  |  |  |
| 100 | 0,1 | -0,0601 | -0,0552 | -0,0378 | 1,3188 | 1,3469 | 1,3188 |
|  | 0,3 | -0,1599 | -0,1523 | -0,1519 | 0,7749 | 0,7976 | 0,7708 |
|  | 0,5 | -0,1864 | -0,1750 | -0,1727 | 0,6317 | 0,6544 | 0,6118 |
|  | 0,7 | -0,1510 | -0,1340 | -0,1373 | 0,6809 | 0,7229 | 0,6639 |
|  | 0,9 | -0,0626 | -0,0450 | -0,0530 | 1,0517 | 1,1279 | 1,0358 |
|  |  |  |  |  |  |  |  |
| 200 | 0,1 | -0,0479 | -0,0457 | -0,0383 | 1,2880 | 1,3001 | 1,2880 |
|  | 0,3 | -0,1334 | -0,1297 | -0,1322 | 0,8310 | 0,8405 | 0,8310 |
|  | 0,5 | -0,1633 | -0,1578 | -0,1633 | 0,6345 | 0,6463 | 0,6345 |
|  | 0,7 | -0,1186 | -0,1121 | -0,1186 | 0,7383 | 0,7543 | 0,7383 |
|  | 0,9 | -0,0316 | -0,0248 | -0,0290 | 1,2366 | 1,2733 | 1,2315 |

**Таблица А.6 – Вариация результатов различных методов оценки для модели GINAR (1) (θ = 0.3).**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Вариация (α) | | | Вариация (θ) | | |
| n | α | Y-W | CLS | CML | Y-W | CLS | CML |
| 50 | 0,1 | 0,0235 | 0,0228 | 0,0113 | 2,1704 | 2,3757 | 2,1704 |
|  | 0,3 | 0,0504 | 0,0451 | 0,0360 | 0,8324 | 0,9099 | 0,8308 |
|  | 0,5 | 0,0796 | 0,0680 | 0,0549 | 0,4981 | 0,5673 | 0,4937 |
|  | 0,7 | 0,0618 | 0,0472 | 0,0420 | 0,5639 | 0,6600 | 0,5593 |
|  | 0,9 | 0,0259 | 0,0140 | 0,0090 | 1,2064 | 1,5834 | 1,2016 |
|  |  |  |  |  |  |  |  |
| 75 | 0,1 | 0,0174 | 0,0170 | 0,0078 | 1,7493 | 1,8470 | 1,7493 |
|  | 0,3 | 0,0389 | 0,0360 | 0,0337 | 0,7699 | 0,8236 | 0,7698 |
|  | 0,5 | 0,0565 | 0,0504 | 0,0474 | 0,4620 | 0,5109 | 0,4593 |
|  | 0,7 | 0,0420 | 0,0347 | 0,0341 | 0,5576 | 0,6251 | 0,5530 |
|  | 0,9 | 0,0115 | 0,0076 | 0,0084 | 1,4542 | 1,7159 | 1,4503 |
|  |  |  |  |  |  |  |  |
| 100 | 0,1 | 0,0140 | 0,0137 | 0,0070 | 1,9186 | 1,9986 | 1,9186 |
|  | 0,3 | 0,0386 | 0,0361 | 0,0334 | 0,6731 | 0,7113 | 0,6727 |
|  | 0,5 | 0,0503 | 0,0463 | 0,0431 | 0,4556 | 0,4858 | 0,4538 |
|  | 0,7 | 0,0329 | 0,0278 | 0,0277 | 0,5526 | 0,6287 | 0,5504 |
|  | 0,9 | 0,0070 | 0,0046 | 0,0055 | 1,4948 | 1,6289 | 1,4928 |
|  |  |  |  |  |  |  |  |
| 200 | 0,1 | 0,0074 | 0,0072 | 0,0048 | 1,7320 | 1,7647 | 1,7320 |
|  | 0,3 | 0,0254 | 0,0244 | 0,0245 | 0,7251 | 0,7406 | 0,7251 |
|  | 0,5 | 0,0328 | 0,0310 | 0,0328 | 0,4261 | 0,4416 | 0,4261 |
|  | 0,7 | 0,0195 | 0,0178 | 0,0195 | 0,5966 | 0,6206 | 0,5966 |
|  | 0,9 | 0,0024 | 0,0019 | 0,0021 | 1,8641 | 1,9554 | 1,8630 |

# Приложение В

**Листинг программы**

PoINAR<-function(n,alpha,lambda){

#set.seed(2880)

tt=2:n

X=vector(length = n)

el=rpois(1,lambda = lambda)

X[1]=el

for(t in 2:n){

e=rpois(1,lambda = lambda)

print

nt=X[t-1]

S=vector(length = nt)

for(i in 1:nt){

S[i]=rbinom(1,1,alpha)

}

alpha.oX=sum(S)

X[t]=alpha.oX+e

}

return (X)

}

GINAR<-function(n,alpha,p){

#set.seed(2880)

tt=2:n

X=vector(length = n)

el=rgeom(1,prob = p)

X[1]=el

for(t in 2:n){

e=rgeom(1,prob = p)

print

nt=X[t-1]

S=vector(length = nt)

for(i in 1:nt){

S[i]=rbinom(1,1,alpha)

}

alpha.oX=sum(S)

X[t]=alpha.oX+e

}

return (X)

}

GINAR\_LevelShift<-function(n,alpha,p,eps){

#set.seed(2880)

tt=2:n

X=vector(length = n)

el=rgeom(1,prob = p)

X[1]=el

for(t in 2:n){

e=rgeom(1,prob = p)

print

nt=X[t-1]

S=vector(length = nt)

for(i in 1:nt){

S[i]=rbinom(1,1,alpha)

}

alpha.oX=sum(S)

X[t]=alpha.oX+e

if(t>=n/4)

{

X[t]=X[t]+eps

}

}

return (X)

}

GINAR\_TransientShift<-function(n,alpha,p,eps){

#set.seed(2880)

tt=2:n

X=vector(length = n)

el=rgeom(1,prob = p)

X[1]=el

for(t in 2:n){

e=rgeom(1,prob = p)

print

nt=X[t-1]

S=vector(length = nt)

for(i in 1:nt){

S[i]=rbinom(1,1,alpha)

}

alpha.oX=sum(S)

if(t==n/4)

{

e=e+eps

}

X[t]=alpha.oX+e

}

return (X)

}

GINAR\_Outlier<-function(n,alpha,p,eps){

#set.seed(2880)

tt=2:n

X=vector(length = n)

el=rgeom(1,prob = p)

X[1]=el

for(t in 2:n){

e=rgeom(1,prob = p)

print

nt=X[t-1]

S=vector(length = nt)

for(i in 1:nt){

S[i]=rbinom(1,1,alpha)

}

alpha.oX=sum(S)

X[t]=alpha.oX+e

}

Y=X

Y[n/4]=Y[n/4]+eps

return (Y)

}

PoINAR\_Outlier<-function(n,alpha,lambda,eps){

tt=2:n

X=vector(length = n)

el=rpois(1,lambda = lambda)

X[1]=el

for(t in 2:n){

e=rpois(1,lambda = lambda)

print

nt=X[t-1]

S=vector(length = nt)

for(i in 1:nt){

S[i]=rbinom(1,1,alpha)

}

alpha.oX=sum(S)

X[t]=alpha.oX+e

}

Y=X

Y[n/2]=Y[n/2]+eps

return (Y)

}

PoINAR\_OutlierPatch<-function(n,alpha,lambda,eps,k){

tt=2:n

X=vector(length = n)

el=rpois(1,lambda = lambda)

X[1]=el

for(t in 2:n){

e=rpois(1,lambda = lambda)

print

nt=X[t-1]

S=vector(length = nt)

for(i in 1:nt){

S[i]=rbinom(1,1,alpha)

}

alpha.oX=sum(S)

X[t]=alpha.oX+e

}

Y=X

for(i in 1:k)

{

Y[n/4+i]=Y[n/4+i]+eps

}

return (Y)

}

PoINAR\_OutlierAtTimes<-function(n,alpha,lambda,eps,k){

tt=2:n

X=vector(length = n)

times = sample(100, k, replace = FALSE)

el=rpois(1,lambda = lambda)

X[1]=el

for(t in 2:n){

e=rpois(1,lambda = lambda)

print

nt=X[t-1]

S=vector(length = nt)

for(i in 1:nt){

S[i]=rbinom(1,1,alpha)

}

alpha.oX=sum(S)

X[t]=alpha.oX+e

}

Y=X

for(i in 1:k)

{

Y[times[i]]=Y[times[i]]+eps

}

return (Y)

}

PoINAR\_TransientShift<-function(n,alpha,lambda,eps){

tt=2:n

X=vector(length = n)

el=rpois(1,lambda = lambda)

X[1]=el

for(t in 2:n){

e=rpois(1,lambda = lambda)

print

nt=X[t-1]

S=vector(length = nt)

for(i in 1:nt){

S[i]=rbinom(1,1,alpha)

}

alpha.oX=sum(S)

if(t==n/4)

{

e=e+eps

}

X[t]=alpha.oX+e

}

return (X)

}

PoINAR\_LevelShift<-function(n,alpha,lambda,eps){

tt=2:n

X=vector(length = n)

el=rpois(1,lambda = lambda)

X[1]=el

for(t in 2:n){

e=rpois(1,lambda = lambda)

print

nt=X[t-1]

S=vector(length = nt)

for(i in 1:nt){

S[i]=rbinom(1,1,alpha)

}

alpha.oX=sum(S)

X[t]=alpha.oX+e

if(t>=n/4)

{

X[t]=X[t]+eps

}

}

return (X)

}

INAR\_YWalpha<-function(X){

n=length(X)

m=mean(X)

Y=X-m

Z1=Y[1:n-1]

Z2=Y[2:n]

Z3=Z1\*Z2

Z4=Y^2

result=sum(Z3)/sum(Z4)

return(result)

}

INAR\_YWlambda<-function(X,alpha){

m=mean(X)

result=(1-alpha)\*m

return(result)

}

INAR\_YWp<-function(X,alpha){

m=mean(X)

result=1/((1-alpha)\*m)

return(result)

}

INAR\_CLSEalpha<-function(X){

n=length(X)

z1=X[1:n-1]

z2=X[2:n]

z3=z1\*z2

s1=sum(z3)

s2=sum(z1)\*sum(z2)/n

z4=z1^2

s3=sum(z4)

s4=sum(z1)^2/n

result=(s1-s2)/(s3-s4)

return(result)

}

INAR\_MCLSEalpha<-function(X,c){

n=length(X)

for(i in 1:n)

{

if(X[i]>c)

{

X[i]=c

}

}

z1=X[1:n-1]

z2=X[2:n]

z3=z1\*z2

s1=sum(z3)

s2=sum(z1)\*sum(z2)/n

z4=z1^2

s3=sum(z4)

s4=sum(z1)^2/n

result=(s1-s2)/(s3-s4)

return(result)

}

ChooseConstantCPo<-function(){

est=vector(length = 10)

for(i in 1:10){

e1=vector(length=50)

e2=vector(length=50)

for(j in 1:50)

{

x=PoINAR(100,0.3,1)

e1[j]=INAR\_CLSEalpha(x)

e2[j]=INAR\_MCLSEalpha(x,i)

}

est[i]=local.mse(e1,0.5)/local.mse(e2,0.5)

}

tw=seq(0.95,0.95,length.out=10)

plot(est,xlab="tuning constant c",ylab="efficiency", type = "b")

lines(tw)

return(est)

}

ChooseConstantCG<-function(){

est=vector(length = 10)

for(i in 1:10){

e1=vector(length=50)

e2=vector(length=50)

for(j in 1:50)

{

x=GINAR(100,0.3,0.7)

e1[j]=INAR\_CLSEalpha(x)

e2[j]=INAR\_MCLSEalpha(x,i)

}

est[i]=local.mse(e1,0.5)/local.mse(e2,0.5)

}

tw=seq(0.95,0.95,length.out=10)

plot(est,xlab="tuning constant c",ylab="efficiency", type = "b")

lines(tw)

return(est)

}

INAR\_CLSElambda<-function(X,alpha){

n=length(X)

z1=X[1:n-1]

z2=X[2:n]

result=(sum(z2)-alpha\*sum(z1))/n

return(result)

}

INAR\_CLSEp<-function(X,alpha){

n=length(X)

z1=X[1:n-1]

z2=X[2:n]

result=1/((sum(z2)-alpha\*sum(z1))/n)

return(result)

}

local.bias<-function(X,y){

n=length(X)

m=mean(X)

result=m-y

return(result)

}

local.mse<-function(X,y){

n=length(X)

z1=X-y

z2=z1^2

result=mean(z2)

return(result)

}

INAR\_CMLPo<-function(data){

rho1 <- acf(data, plot=FALSE)[[1]][2]

barX <- mean(data)

mue <- barX\*(1-rho1)

estml <- suppressWarnings(optim(c(mue,rho1), F\_PoINAR,method = "BFGS", lower=c(0.0001,0.0001), upper=c(9999,0.9999),control=list(ndeps=c(1e-4,1e-4)), data=data, hessian=TRUE))

lambdaestml <- estml$par[[1]]

alphaestml <- estml$par[[2]]

return(c(alphaestml, lambdaestml))

}

TP\_PoINAR <- function(k,l,lambda,alpha){

tp <- 0

if(min(k,l)>=0){

for(j in c(0:min(k,l))){

tp <- tp + dbinom(j,l,alpha)\*dpois(k-j,lambda)

}

}

}

F\_PoINAR <- function(par,data){

T <- length(data)

value <- -log(dpois(data[1], par[1]/(1-par[2]))) #full likelihood, otherwise use 0 here

for(t in c(2:T)) {

value <- value-log(TP\_PoINAR(data[t], data[t-1], par[1], par[2]))

}

if(is.na(value)||is.infinite(value)){

value=100

}

value

}

INAR\_CMLG<-function(data){

rho1 <- acf(data, plot=FALSE)[[1]][2]

barX <- mean(data)

mue <- 1-(barX\*(1-rho1))

estml <- suppressWarnings(optim(c(mue,rho1), F\_GINAR, method='BFGS', lower=c(0.0001,0.0001), upper=c(9999,0.9999), control=list(ndeps=c(1e-4,1e-4)), data=data, hessian=TRUE))

p <- estml$par[[1]]

alphaestml <- estml$par[[2]]

return(c(alphaestml, p))

}

TP\_GINAR <- function(k,l,p,alpha){

tp <- 0

if(min(k,l)>=0){

for(j in c(0:min(k,l))){

tp <- tp + dbinom(j,l,alpha)\*dgeom(k-j,p)

}

}

tp

}

F\_GINAR <- function(par,data){

T <- length(data)

value <- -log(dgeom(data[1], par[1])) #full likelihood, otherwise use 0 here

for(t in c(2:T)) {

value <- value-log(TP\_GINAR(data[t], data[t-1], par[1], par[2]))

}

if(is.na(value)||is.infinite(value)){

value=100

}

value

}

testParamPoINAR<-function(t){

n1=c(50,75,100,200)

alpha1=c(0.1,0.3,0.5,0.7,0.9)

lambda=1

estYW=vector(length = t)

estCLS=vector(length = t)

estCML=vector(length = t)

estYW\_l=vector(length = t)

estCLS\_l=vector(length = t)

estCML\_l=vector(length = t)

b=matrix(data=NA,nrow=20,ncol=8)

mse=matrix(data=NA,nrow=20,ncol=8)

for(i in 1:length(n1))

{

for(k in 1:length(alpha1))

{

for(j in 1:t)

{

x=PoINAR(n1[i],alpha1[k],lambda)

estYW[j]=INAR\_YWalpha(x)

estYW\_l[j]=INAR\_YWlambda(x,estYW[j])

estCLS[j]=INAR\_CLSEalpha(x)

estCLS\_l[j]=INAR\_CLSElambda(x, estCLS[j])

cml=INAR\_CMLPo(x)

estCML[j]=cml[1]

estCML\_l[j]=cml[2]

}

mse[5\*(i-1)+k,1]=n1[i]

mse[5\*(i-1)+k,2]=alpha1[k]

mse[5\*(i-1)+k,3]=local.mse(estYW,alpha1[k])

mse[5\*(i-1)+k,4]=local.mse(estCLS,alpha1[k])

mse[5\*(i-1)+k,5]=local.mse(estCML,alpha1[k])

mse[5\*(i-1)+k,6]=local.mse(estYW\_l,lambda)

mse[5\*(i-1)+k,7]=local.mse(estCLS\_l,lambda)

mse[5\*(i-1)+k,8]=local.mse(estCML\_l,lambda)

b[5\*(i-1)+k,1]=n1[i]

b[5\*(i-1)+k,2]=alpha1[k]

b[5\*(i-1)+k,3]=local.bias(estYW,alpha1[k])

b[5\*(i-1)+k,4]=local.bias(estCLS,alpha1[k])

b[5\*(i-1)+k,5]=local.bias(estCML,alpha1[k])

b[5\*(i-1)+k,6]=local.bias(estYW\_l,lambda)

b[5\*(i-1)+k,7]=local.bias(estCLS\_l,lambda)

b[5\*(i-1)+k,8]=local.bias(estCML\_l,lambda)

}

}

result=data.frame(b,mse)

return(result)

}

testEstimMethod<-function(t){

x<-testParamPoINAR(t)

#y<-testParamGINAR(t)

return(x)

}

testPlotPoINAR<-function(t){

n=75

alpha=0.1

lambda=1

estYW=vector(length = t)

estCLS=vector(length = t)

estCML=vector(length = t)

estYW\_l=vector(length = t)

estCLS\_l=vector(length = t)

estCML\_l=vector(length = t)

b=matrix(data=NA,nrow=9,ncol=7)

mse=matrix(data=NA,nrow=9,ncol=7)

for(i in 1:9)

{

for(j in 1:t)

{

x=PoINAR(n,alpha,lambda)

estYW[j]=INAR\_YWalpha(x)

estYW\_l[j]=INAR\_YWlambda(x,estYW[j])

estCLS[j]=INAR\_CLSEalpha(x)

estCLS\_l[j]=INAR\_CLSElambda(x, estCLS[j])

cml=INAR\_CMLPo(x)

estCML[j]=cml[1]

estCML\_l[j]=cml[2]

}

b[i,1]=alpha

b[i,2]=abs(local.bias(estYW,alpha))\*100

b[i,3]=abs(local.bias(estCLS,alpha))\*100

b[i,4]=abs(local.bias(estCML,alpha))\*100

b[i,5]=abs(local.bias(estYW\_l,lambda))\*100

b[i,6]=abs(local.bias(estCLS\_l,lambda))\*100

b[i,7]=abs(local.bias(estCML\_l,lambda))\*100

mse[i,1]=alpha

mse[i,2]=abs(local.mse(estYW,alpha))\*100

mse[i,3]=abs(local.mse(estCLS,alpha))\*100

mse[i,4]=abs(local.mse(estCML,alpha))\*100

mse[i,5]=abs(local.mse(estYW\_l,lambda))\*100

mse[i,6]=abs(local.mse(estCLS\_l,lambda))\*100

mse[i,7]=abs(local.mse(estCML\_l,lambda))\*100

alpha=alpha+0.1

}

matplot(b[,1],b[,2:4],type="l")

matplot(b[,1],b[,5:7],type="l")

matplot(mse[,1],mse[,2:4],type="l")

matplot(mse[,1],mse[,5:7],type="l")

return(b)

}

testParamGINAR<-function(t,p){

n1=c(50,75,100,200)

alpha1=c(0.1,0.3,0.5,0.7,0.9)

estYW=vector(length = t)

estCLS=vector(length = t)

estCML=vector(length = t)

estYW\_p=vector(length = t)

estCLS\_p=vector(length = t)

estCML\_p=vector(length = t)

b=matrix(data=NA,nrow=20,ncol=8)

mse=matrix(data=NA,nrow=20,ncol=8)

for(i in 1:length(n1))

{

for(k in 1:length(alpha1))

{

for(j in 1:t)

{

x=GINAR(n1[i],alpha1[k],p)

estYW[j]=INAR\_YWalpha(x)

estYW\_p[j]=INAR\_YWp(x,estYW[j])

estCLS[j]=INAR\_CLSEalpha(x)

estCLS\_p[j]=INAR\_CLSEp(x, estCLS[j])

cml=INAR\_CMLG(x)

estCML[j]=cml[1]

estCML\_p[j]=cml[2]

}

mse[5\*(i-1)+k,1]=n1[i]

mse[5\*(i-1)+k,2]=alpha1[k]

mse[5\*(i-1)+k,3]=local.mse(estYW,alpha1[k])

mse[5\*(i-1)+k,4]=local.mse(estCLS,alpha1[k])

mse[5\*(i-1)+k,5]=local.mse(estCML,alpha1[k])

mse[5\*(i-1)+k,6]=local.mse(estYW\_p,p)

mse[5\*(i-1)+k,7]=local.mse(estCLS\_p,p)

mse[5\*(i-1)+k,8]=local.mse(estCML\_p,p)

b[5\*(i-1)+k,1]=n1[i]

b[5\*(i-1)+k,2]=alpha1[k]

b[5\*(i-1)+k,3]=local.bias(estYW,alpha1[k])

b[5\*(i-1)+k,4]=local.bias(estCLS,alpha1[k])

b[5\*(i-1)+k,5]=local.bias(estCML,alpha1[k])

b[5\*(i-1)+k,6]=local.bias(estYW\_p,p)

b[5\*(i-1)+k,7]=local.bias(estCLS\_p,p)

b[5\*(i-1)+k,8]=local.bias(estCML\_p,p)

}

}

res=data.frame(b,mse)

return(res)

}

testParamPoINARWithShift<-function(t,eps){

n=100

alpha1=c(0.1,0.3,0.5,0.7,0.9)

lambda=1

estYW=vector(length = t)

estCLS=vector(length = t)

estYW\_AO=vector(length = t)

estCLS\_AO=vector(length = t)

estYW\_TS=vector(length = t)

estCLS\_TS=vector(length = t)

estYW\_LS=vector(length = t)

estCLS\_LS=vector(length = t)

b=matrix(data=NA,nrow=5,ncol=9)

for(k in 1:length(alpha1))

{

for(j in 1:t)

{

x=PoINAR(n,alpha1[k],lambda)

x\_AO=PoINAR\_Outlier(n,alpha1[k],lambda,eps)

x\_TS=PoINAR\_TransientShift(n,alpha1[k],lambda,eps)

x\_LS=PoINAR\_LevelShift(n,alpha1[k],lambda,eps)

estYW[j]=INAR\_YWalpha(x)

estCLS[j]=INAR\_CLSEalpha(x)

estYW\_AO[j]=INAR\_YWalpha(x\_AO)

estCLS\_AO[j]=INAR\_CLSEalpha(x\_AO)

estYW\_TS[j]=INAR\_YWalpha(x\_TS)

estCLS\_TS[j]=INAR\_CLSEalpha(x\_TS)

estYW\_LS[j]=INAR\_YWalpha(x\_LS)

estCLS\_LS[j]=INAR\_CLSEalpha(x\_LS)

}

b[k,1]=alpha1[k]

b[k,2]=local.mse(estYW,alpha1[k])

b[k,3]=local.mse(estCLS,alpha1[k])

b[k,4]=local.mse(estYW\_AO,alpha1[k])

b[k,5]=local.mse(estCLS\_AO,alpha1[k])

b[k,6]=local.mse(estYW\_TS,alpha1[k])

b[k,7]=local.mse(estCLS\_TS,alpha1[k])

b[k,8]=local.mse(estYW\_LS,alpha1[k])

b[k,9]=local.mse(estCLS\_LS,alpha1[k])

}

return(b)

}

testParamPoINARWithShiftLambda<-function(t,eps){

n=100

alpha1=c(0.1,0.3,0.5,0.7,0.9)

lambda=1

estYW=vector(length = t)

estCLS=vector(length = t)

estYW\_AO=vector(length = t)

estCLS\_AO=vector(length = t)

estYW\_TS=vector(length = t)

estCLS\_TS=vector(length = t)

estYW\_LS=vector(length = t)

estCLS\_LS=vector(length = t)

b=matrix(data=NA,nrow=5,ncol=9)

for(k in 1:length(alpha1))

{

for(j in 1:t)

{

x=PoINAR(n,alpha1[k],lambda)

x\_AO=PoINAR\_Outlier(n,alpha1[k],lambda,eps)

x\_TS=PoINAR\_TransientShift(n,alpha1[k],lambda,eps)

x\_LS=PoINAR\_LevelShift(n,alpha1[k],lambda,eps)

estYW[j]=INAR\_YWlambda(x, INAR\_YWalpha(x))

estCLS[j]=INAR\_CLSElambda(x,INAR\_CLSEalpha(x))

estYW\_AO[j]=INAR\_YWlambda(x,INAR\_YWalpha(x\_AO))

estCLS\_AO[j]=INAR\_CLSElambda(x,INAR\_CLSEalpha(x\_AO))

estYW\_TS[j]=INAR\_YWlambda(x,INAR\_YWalpha(x\_TS))

estCLS\_TS[j]=INAR\_CLSElambda(x,INAR\_CLSEalpha(x\_TS))

estYW\_LS[j]=INAR\_YWlambda(x,INAR\_YWalpha(x\_LS))

estCLS\_LS[j]=INAR\_CLSElambda(x,INAR\_CLSEalpha(x\_LS))

}

b[k,1]=alpha1[k]

b[k,2]=local.mse(estYW,alpha1[k])

b[k,3]=local.mse(estCLS,alpha1[k])

b[k,4]=local.mse(estYW\_AO,alpha1[k])

b[k,5]=local.mse(estCLS\_AO,alpha1[k])

b[k,6]=local.mse(estYW\_TS,alpha1[k])

b[k,7]=local.mse(estCLS\_TS,alpha1[k])

b[k,8]=local.mse(estYW\_LS,alpha1[k])

b[k,9]=local.mse(estCLS\_LS,alpha1[k])

}

return(b)

}

testF<-function(){

x=testParamPoINARWithShiftLambda(50,10)

y=testParamPoINARWithShiftLambda(50,50)

z=testParamPoINARWithShiftLambda(50,100)

d=data.frame(x,y,z)

return(d)

}

MEstimator<-function(n,alpha,lambda){

x<-PoINAR(n,alpha,lambda)

resp=x[2:n]

x1=x[1:n-1]

d=as.data.frame(cbind(resp,x1))

estCUBIF=glmrob(resp~x1,family=poisson(),data=d,method = "Mqle")

coefMqle=estCUBIF$coef

print(d)

t=coefMqle[1]+coefMqle[2]\*x1[2]

print(coefMqle)

}

makeGraph1<-function(){

alpha=c(0.3,0.5,0.7)

lambda=c(1,3)

linetype <- c(1:2)

for(i in 1:length(alpha))

{

for(j in 1:length(lambda))

{

x1=PoINAR(100,alpha[i],lambda[j])

x2=seq(lambda[j]/(1-alpha[i]),lambda[j]/(1-alpha[i]), length.out = 100)

a=alpha[i]

b=lambda[j]

plot(x1,xlab="Time",ylab="Value", type = "b",lty=linetype[1], main=paste("alpha: ",a,", ","lambda: ", b))

lines(x2,lty=linetype[2])

legend("topright", legend=c("PoInar", "mu"),

lty=linetype, cex=0.8)

}

}

}

makeGraph2<-function(){

alpha=c(0.3,0.5,0.7)

p=c(0.3,0.5,0.7)

linetype <- c(1:2)

for(i in 1:length(alpha))

{

for(j in 1:length(p))

{

x1=GINAR(100,alpha[i],p[j])

x2=seq((1-p[j])/(p[j]\*(1-alpha[i])),(1-p[j])/(p[j]\*(1-alpha[i])), length.out = 100)

a=alpha[i]

b=p[j]

plot(x1,xlab="Time",ylab="Value", type = "b",lty=linetype[1], main=paste("alpha: ",a,", ","p: ", b))

lines(x2,lty=linetype[2])

legend("topright", legend=c("GINAR", "mu"),

lty=linetype, cex=0.8)

}

}

}

testWithShift<-function(){

mse1=vector(length = 10)

mse2=vector(length = 10)

bias1=vector(length = 10)

bias2=vector(length = 10)

for(c in 1:10)

{

e1=vector(length=50)

e2=vector(length=50)

for(i in 1:50)

{

x<-PoINAR\_TransientShift(100,0.5,1,5)

e1[i]=INAR\_CLSEalpha(x)

e2[i]=INAR\_MCLSEalpha(x,c)

}

mse1[c]=local.mse(e1,0.5)

mse2[c]=local.mse(e2,0.5)

bias1[c]=local.bias(e1,0.5)

bias2[c]=local.bias(e2,0.5)

}

linetype <- c(1:2)

plot(mse2,xlab="tuning constant c",ylab="MSE", type = "b",lty=linetype[1])

lines(mse1,lty=linetype[2])

legend("topright", legend=c("MCLS", "CLS"),

lty=linetype, cex=0.8)

plot(bias2,xlab="tuning constant c",ylab="Bias", type = "b",lty=linetype[1])

lines(bias1,lty=linetype[2])

legend("topright", legend=c("MCLS", "CLS"),

lty=linetype, cex=0.8)

}

testWithOutlier<-function(){

mse1=vector(length = 10)

mse2=vector(length = 10)

bias1=vector(length = 10)

bias2=vector(length = 10)

eps=c(5,10)

for(j in 1:2)

{

for(c in 1:10)

{

e1=vector(length=50)

e2=vector(length=50)

for(i in 1:50)

{

x<-PoINAR\_Outlier(100,0.5,1,eps[j])

e1[i]=INAR\_CLSEalpha(x)

e2[i]=INAR\_MCLSEalpha(x,c)

}

mse1[c]=abs(local.mse(e1,0.5))

mse2[c]=abs(local.mse(e2,0.5))

bias1[c]=abs(local.bias(e1,0.5))

bias2[c]=abs(local.bias(e2,0.5))

}

linetype <- c(1:2)

plot(mse2,xlab="tuning constant c",ylab="MSE", type = "b",lty=linetype[1])

lines(mse1,lty=linetype[2])

legend("topright", legend=c("MCLS", "CLS"),

lty=linetype, cex=0.8)

plot(bias2,xlab="tuning constant c",ylab="Bias", type = "b",lty=linetype[1])

lines(bias1,lty=linetype[2])

legend("topright", legend=c("MCLS", "CLS"),

lty=linetype, cex=0.8)

}

}

testWithOutlierPatch<-function(){

mse1=vector(length = 10)

mse2=vector(length = 10)

bias1=vector(length = 10)

bias2=vector(length = 10)

k=c(5,10,15,20)

for(j in 1:length(k))

{

for(c in 1:10)

{

e1=vector(length=50)

e2=vector(length=50)

for(i in 1:50)

{

x<-PoINAR\_OutlierPatch(100,0.5,1,5,k[j])

e1[i]=INAR\_CLSEalpha(x)

e2[i]=INAR\_MCLSEalpha(x,c)

}

mse1[c]=abs(local.mse(e1,0.5))

mse2[c]=abs(local.mse(e2,0.5))

bias1[c]=abs(local.bias(e1,0.5))

bias2[c]=abs(local.bias(e2,0.5))

}

linetype <- c(1:2)

plot(mse2,xlab="tuning constant c",ylab="MSE", type = "b",lty=linetype[1])

lines(mse1,lty=linetype[2])

legend("topright", legend=c("MCLS", "CLS"),

lty=linetype, cex=0.8)

}

}

testWithOutlierAtTimes<-function(){

mse1=vector(length = 10)

mse2=vector(length = 10)

bias1=vector(length = 10)

bias2=vector(length = 10)

k=c(5,10,15,20)

for(j in 1:length(k))

{

for(c in 1:10)

{

e1=vector(length=50)

e2=vector(length=50)

for(i in 1:50)

{

x<-PoINAR\_OutlierAtTimes(100,0.5,1,5,k[j])

e1[i]=INAR\_CLSEalpha(x)

e2[i]=INAR\_MCLSEalpha(x,c)

}

mse1[c]=abs(local.mse(e1,0.5))

mse2[c]=abs(local.mse(e2,0.5))

bias1[c]=abs(local.bias(e1,0.5))

bias2[c]=abs(local.bias(e2,0.5))

}

linetype <- c(1:2)

plot(mse2,xlab="tuning constant c",ylab="MSE", type = "b",lty=linetype[1],main=paste("k = ",k[j]))

lines(mse1,lty=linetype[2])

legend("topright", legend=c("MCLS", "CLS"),

lty=linetype, cex=0.8)

}

}