Т. Андерсон

Статистический анализ временных рядов

Москва
«Книга по Требованию»

УДК 51
ББК 22.1
Т11

Т. Андерсон
Т11 Статистический анализ временных рядов / Т. Андерсон – М.: Книга по Требованию, 2024. – 752 с.

ISBN 978-5-458-26140-1

Монография известного американского специалиста по математической статистике содержит обстоятельное изложение теории статистических выводов для различных вероятностных моделей. Излагаются методы представления временных рядов, оценивания параметров соответствующих вероятностных моделей, проверки гипотез относительно их структуры.

ISBN 978-5-458-26140-1

Эта книга является репринтом оригинала, который мы создали специально для Вас, используя запатентованные технологии производства репринтных книг и печати по требованию.

Сначала мы отсканировали каждую страницу оригинала этой редкой книги на профессиональном оборудовании. Затем с помощью специально разработанных программ мы произвели очистку изображения от пятен, клякс, перегибов и попытались отбелить и выровнять каждую страницу книги. К сожалению, некоторые страницы нельзя вернуть в изначальное состояние, и если их было трудно читать в оригинале, то даже при цифровой реставрации их невозможно улучшить.

Разумеется, автоматизированная программная обработка репринтных книг – не самое лучшее решение для восстановления текста в его первозданном виде, однако, наша цель – вернуть читателю точную копию книги, которой может быть несколько веков.

Поэтому мы предупреждаем о возможных погрешностях восстановленного репринтного издания. В издании могут отсутствовать одна или несколько страниц текста, могут встретиться невыводимые пятна и кляксы, надписи на полях или подчеркивания в тексте, нечитаемые фрагменты текста или загибы страниц. Покупать или не покупать подобные издания – решать Вам, мы же делаем все возможное, чтобы редкие и ценные книги, еще недавно утраченные и несправедливо забытые, вновь стали доступными для всех читателей.

Серия Книжный Ренессанс
www.samizday.ru/reprint

ИЗ ПРЕДИСЛОВИЯ АВТОРА

Авторы книг по статистическому анализу временных рядов по-разному могут подходить к их написанию. Я посвятил книгу математической теории статистических выводов относительно вероятностных моделей, лежащих (по предположению) в основе тех или иных временных рядов. Исходная вероятностная модель может включать детерминированный тренд и случайную составляющую, образующую стационарный случайный процесс. При этом естественно возникает целый ряд статистических задач, относящихся к характеристикам тренда и процесса. В книге всюду, где это возможно, отыскиваются оптимальные процедуры. Изучаются статистические свойства различных методов, причем во многих случаях свойства метода можно описать только для больших выборок, т. е. на основании совокупности наблюдений значений ряда в течение длительного периода времени. Как правило, дается математически строгий вывод соответствующих свойств.

В то время как теория развивается при строгих математических предположениях, приведенные в книге методы могут быть использованы и в тех случаях, когда указанные предположения выполнены не полностью. Часто бывает так, что свойства процедур меняются при этом несущественно. Как бы то ни было, точные результаты теорем служат руководством по использованию соответствующих процедур. В книге приведены примеры приложений развиваемых методов, обсуждаются вычислительные аспекты, вопросы использования и интерпретации этих методов. В то же время вопросы представления процедур в виде программ для вычислительных машин не затрагиваются.

Основой настоящей монографии явился курс лекций, который я читал в течение многих лет в Колумбийском университете обычно в течение одного, а иногда и двух семестров. Однако весь материал, в том виде как он представлен в книге, нельзя изложить в двухсеместровом курсе — лектор, использующий эту книгу как учебник, должен выбрать из нее разделы, представляющиеся ему наиболее интересными и важными. В книге приведено значительное число упражнений. Одни из них являются приложениями описанных методов, другие представляют собой разработку частных случаев общей теории. Часть упражнений связана с уточнением деталей сложных доказательств, остальные посвящены развитию теории.

Я надеюсь, что книга послужит не только учебником для начинающих, но будет полезна статистикам и специалистам, желающим познакомиться с анализом временных рядов, не обращаясь к формальным курсам. Чтение этой книги и выполнение упражнений поможет существенно пополнить запас статистических методов, полезных для анализа временных рядов. Книгу можно использовать также и как справочник. В ней достаточно систематизированно представлен обширный материал, ранее разбросанный по различным источникам. Приводятся также некоторые новые теоремы и методы. В ряде случаев за счет ослабления предположений известных ранее утверждений получены более сильные результаты.

Поскольку область анализа временных рядов весьма обширна, передо мной встала задача выбора разделов, которые следует включить в книгу. Материал, помещенный в этой книге, вместе с указанием соответствующих ограничений, при которых развивается теория, описан мною во введении (гл. 1). Определенную информацию дает также предметный указатель. Хочется надеяться, что основные и наиболее важные разделы анализа временных рядов оказались представленными, хотя полнота охвата тематики является делом вкуса. Постоянно создаются новые методы, меняются точки зрения, так что приведенные здесь результаты вряд ли можно считать окончательными. Фактически часть включенного материала уже в настоящее время представляет скорее исторический интерес.

Ввиду большого объема книги будет, по-видимому, полезно дать несколько советов читателям и преподавателям относительно выбора материала для самостоятельного изучения и для целей обучения. Глава 2 содержит краткое замкнутое изложение метода наименьших квадратов. Главы 3 и 4 имеют дело с моделями, в которых случайные составляющие независимы (о них иногда говорят как об «ошибках измерений»). Некоторые идеи и элементы анализа, приведенные в этих главах, используются в дальнейшем. Однако читатель, интересующийся именно последующими главами, может довольно многое пропустить (включая бо́льшую часть § 3.4, 4.3 и 4.4). Процессы авторегрессии, которые оказываются полезными в приложениях и служат примерами стационарных случайных процессов, исследуются в гл. 5. Посвященные теории больших выборок § 5.5 и 5.6 содержат результаты, доказательства которых включают значительное количество деталей и могут быть опущены. Статистические выводы для этих моделей служат основой анализа стационарных процессов «во временно́й области».

Глава 6 представляет собой обширное исследование сериальной корреляции и критериев для проверки независимости. Параграфы 6.3 и 6.4 имеют по преимуществу теоретической интерес. В § 6.5 изучается алгебра квадратичных форм и их отношений. Распределения, моменты и аппроксимирующие распределения получены в § 6.7 и 6.8. Там же даны таблицы процентных точек для соответствующих критериев. Первые пять параграфов гл. 7 служат введением в теорию стационарных случайных процессов и спектральные свойства этих процессов. В гл. 8 развивается теория статистик, относящихся к стационарным случайным процессам. Глава 9 посвящена оцениванию спектральной плотности, являющемуся основой анализа стационарных процессов «в частотной области». В § 10.2 регрессионный анализ (гл. 2) распространяется на стационарные случайные составляющие. Результаты гл. 8 и 9 обобщаются на этот случай в § 10.3. В § 10.4 результаты гл. 6 обобщаются на случай, когда рассматриваются остатки от подобранного тренда. Можно выделить следующие части книги, которые образуют единое целое и которые можно читать независимо от других частей: (I) — гл. 2; (II) — гл. 3 и 4; (III) — гл. 5; (IV) — гл. 6; (V) — гл. 7; (VI) — гл. 8 и 9.

В практических приложениях статистический анализ временных рядов, кроме того, включает и менее формальную технику (которую теперь иногда называют «анализом данных»). Так, пониманию явления способствует простое графическое представление наблюдаемого временного ряда. Могут оказаться полезными те или иные преобразования измерений и связь последних с другими данными. Точно сформулированные процедуры, которые изучаются в этой книге, как правило, используются не в буквальном виде, а изменяются в зависимости от конкретных задач. Тем не менее для исследования статистических методов в рамках математической теории некоторые аспекты анализа необходимо формализовать. Например, задача об определении того, будет ли некоторый эффект достаточно большим, чтобы его можно было считать заметным, иногда формализуется как задача проверки гипотезы о том, что соответствующий параметр просто равен нулю.

По степени сложности эта книга примерно такая же, как и моя более ранняя книга «Введение в многомерный статистический анализ»; требуется лишь некоторое знание алгебры матриц. (Необходимый материал представлен в приложении В к упомянутой книге; дополнительные результаты приводятся в тексте и в упражнениях настоящей книги.) Полезно общее знакомство со статистическими методами. В частности, предполагается, что читатель знаком со стандартным материалом статистического анализа одной случайной переменной, включающим t-критерии и F-статистики, с многомерным нормальным распределением, с элементами оценивания и проверки гипотез. Более тонкие вопросы оценивания, проверки гипотез и теории решений, на которые имеются ссылки, излагаются в упражнениях. [Детальное и строгое изложение вопросов проверки гипотез читатель может найти у Лемана (1959).] Предполагается также знание основ курса высшей математики. Хотя в книге рассматриваются только действительные временные ряды, тем не менее иногда удобно записывать некоторые выражения с привлечением комплексных переменных. Однако фактически теория функций комплексного переменного в книге не применяется, поскольку (за исключением одной задачи) дело ограничивается формулой $e^{i\theta} = \cos\theta + i\sin\theta$. Теория вероятностей использована в объеме, включающем характеристические функции и некоторые основные предельные теоремы. Теория случайных процессов развивается в книге только в необходимых пределах.

Как было отмечено выше, в конце каждой главы приводится большое число задач. Исключением служит лишь первая глава, являющаяся введением. Решения задач были подготовлены Полем Шаманом. Решения, на которые имеются ссылки в тексте или которые демонстрируют некоторые наиболее важные моменты, помещены в приложении В к этой книге.

Я весьма обязан и признателен П. Шаману за его большое содействие в подготовке этой книги. Он принимал участие в отборе материала и его редактировании, в подборе ссылок и задач, в усовершенствовании доказательств и изложения, в исправлении ошибок различного характера. Он прочел рукопись в нескольких вариантах. Мне всегда представлялось излишним говорить (как это обычно принято), что компетентный рецензент рукописи не несет ответственности за ошибки в публикации, так как очевидно, что всякий человек, любезно согласившийся просмотреть рукопись, не может брать на себя такой ответственности. Здесь же такой отказ от ответственности вполне уместен, поскольку П. Шаман исправил так много ошибок, что трудно поверить, чтобы какие-нибудь еще остались. Тем не менее я допускаю, что в подобном материале очень легко ошибиться, и читатель должен порицать за любую найденную ошибку автора (и информировать его об этом).

Я благодарен Д. Хинкли, Т. Сава и Дж. Стаяну, прочитавшим значительную часть рукописи и доказательств и принимавшим участие в подготовке библиографии и указателя. Я выражаю также признательность многим моим коллегам и ученикам за помощь различного рода. Дж. М. Крэддоку, С. У. Дж. Гренджеру, М. Дж. Кендаллу, А. Стьюарту и Герману Вольду я весьма благодарен за то, что они разрешили мне использовать некоторые материалы.

В процессе работы над книгой весьма важным фактором была для меня длительная финансовая поддержка Военно-морского ведомства, оказывавшаяся в течение примерно десяти лет.

Отделения логики и математической статистики постоянно обеспечивали мне чрезвычайно благоприятные условия для работы.

Т. Андерсон

Станфордский университет,
Февраль 1970

Глава 1

ВВЕДЕНИЕ

Временным рядом называют последовательность наблюдений, обычно упорядоченную во времени, хотя возможно упорядочение и по какому-то другому параметру. Основной чертой, выделяющей анализ временных рядов среди других видов статистического анализа, является существенность порядка, в котором призводятся наблюдения. Если во многих задачах наблюдения статистически независимы, то во временных рядах они, как правило, зависимы и характер этой зависимости может определяться положением наблюдений в последовательности. Природа ряда и структура порождающего ряд процесса могут предопределять порядок образования последовательности.

Почти в каждой области встречаются явления, которые интересно и важно изучать в их развитии и изменении во времени. В повседневной жизни могут представлять интерес, например, метеорологические условия, цены на тот или иной товар, те или иные характеристики состояния здоровья индивидуума и т. п. Все они изменяются во времени. Существуют также различные характеристики, относящиеся к целой нации и зависящие уже от совокупности характеристик отдельных индивидуумов, например экономические условия и народонаселение, которые эволюционируют и флуктуируют во времени. С течением времени изменяются деловая активность, режим протекания того или иного производственного процесса, глубина сна человека, восприятие телевизионной программы. Совокупность измерений какой-либо одной характеристики подобного рода в течение некоторого периода времени и представляет собой временной ряд. Это может быть, например, почасовая запись температуры в том или ином месте или ежегодное количество осадков, фиксируемое метеорологической станцией. Это могут быть также поквартальные данные о валовом национальном продукте.

К нескольким временным рядам приводит запись электрокардиограммы.

Цели изучения временных рядов могут быть различными. Можно, например, стремиться предсказать будущее на основании знания прошлого, управлять процессом, порождающим ряд, выяснить механизм, порождающий ряд, или просто сжато описать характерные особенности ряда. Как статистики мы будем интересоваться задачами статистических выводов. Именно, на основании ограниченного количества информации, временного ряда конечной длины, мы хотим делать выводы о вероятностном механизме, порождающем этот ряд, анализировать структуру, лежащую в его основе.

В принципе измерение многих величин, таких, как температура и напряжение, может производиться непрерывно. При этом наблюдения можно фиксировать в виде графика. Однако на практике измерения часто производятся все же в дискретные моменты времени. В других случаях, как, например, при подсчете урожайности зерновых культур, измерения вообще могут производиться только в определенные интервалы времени. Как бы то ни было, даже в том случае, когда изучаемые величины регистрируются непрерывно, при обработке их на цифровых вычислительных машинах реально используются только те значения, которые соответствуют дискретному множеству моментов времени. В настоящей книге мы ограничимся только временными рядами, представляющими собой дискретную последовательность наблюдений, производимых через регулярные промежутки времени, такими, как, например, почасовая запись атмосферного давления. Хотя часто имеет значение влияние одной характеристики на другую и взаимодействие во времени сразу нескольких величин, тем не менее во многих исследованиях существенных результатов можно достигнуть и при изучении только одного, отдельно взятого временного ряда. Настоящая книга почти целиком (за исключением того, что касается систем авторегрессии) посвящена статистическим методам анализа одномерных временных рядов. Иными словами, предполагается, что (один и тот же) объект или индивидуум подвергается повторным измерениям только одного типа. Мы будем предполагать, кроме того, что результатом измерения является действительное число, например температура, и что множество исходов не обязательно конечное (или счетное). О результате таких измерений часто говорят как о непрерывной переменной. Мы будем проводить математическое исследование результатов измерений подобного рода, обращаясь с ними так, как если бы сами измерения были непрерывными во времени. Например, ежегодный национальный доход можно измерить, в лучшем случае, с точностью до пенни. Однако его размеры могут быть столь велики, что не произойдет сколько-нибудь серьезной ошибки, если мы будем рассматривать эту величину как непрерывную переменную. Более того, мы будем рассматривать такие временные ряды, которые ведут себя достаточно устойчивым образом, т. е. имеют тенденцию оставаться в определенных границах или по крайней мере меняются медленно, без резких, взрывных изменений. Так, мы рассматривали бы многие метеорологические переменные, но при этом исключили бы ударные волны.

Пусть наблюдаемым временным рядом является $y_1, y_2, \ldots, y_T$. Мы будем понимать эту запись следующим образом. Имеется T чисел, представляющих собой наблюдение некоторой переменной в T равностоящих моментов времени. Эти моменты для удобства перенумерованы целыми числами 1, 2, ..., T. Достаточно общей математической (статистической, или вероятностной) моделью служит модель вида

$$(1) \qquad y_t \quad f(t) + u_t, \quad t = 1, 2, \ldots, T.$$

В этой модели наблюдаемый ряд рассматривается как сумма некоторой полностью детерминированной последовательности $\{f(t)\}$, которую можно назвать систематической составляющей, и случайной последовательности $\{u_t\}$, подчиняющейся некоторому вероятностному закону. (Иногда для этих двух составляющих используются соответственно термины сигнал и шум). Эти компоненты наблюдаемого ряда ненаблюдаемы; они являются теоретическими величинами. Например, если производится измерение количества ежедневно выпадаемых осадков, то $f(t)$ может представлять собой климатическую норму, получающуюся долговременным усреднением за большой период, а u_t те капризы и нерегулярности в погоде, которые характеризуют отклонения от климатической нормы. Точный смысл указанного разложения зависит не только от самих данных, но частично и от того, что понимается под повторением эксперимента, результатом которого являются эти данные. Мы используем здесь так называемую «частотную» интерпретацию. Мы полагаем, что по крайней мере принципиально можно повторять всю ситуацию целиком, получая новые совокупности наблюдений. При таком повторении эксперимента функция $f(t)$ должна была бы оставаться одной и той же, а случайные составляющие оказывались бы различными как различные реализации случайного процесса. Случайные составляющие, кроме всего прочего, могут включать в себя и ошибки наблюдений. (При этом [1] $f(t) = \mathcal{E} y_t$.)

Мы все имеем определенные интуитивные представления о том, что следовало бы понимать под временным параметром в подобных моделях или процессах. Одно из таких представлений состоит в том, что время течет в одном направлении. Другое — что события, близкие по времени, должны быть сравнительно сильно связаны, а события, разделенные большими промежутками времени, не должны

[1] Символом $\mathcal{E}X$ в этой книге всюду обозначается математическое ожидание случайной величины X.— *Прим. перев.*

иметь сильной связи. Можно рассматривать различные варианты математической модели (1), в которых влияние времени может сказываться либо только на функции или последовательности $f(t)$, либо только на вероятностном процессе, определяющем случайную составляющую u_t, либо, наконец, на обеих этих компонентах. Первая часть книги посвящена анализу временных рядов, соответствующих так называемым моделям ошибок, в которых наблюдения рассматриваются как результат независимых случайных отклонений от некоторой функции, представляющей тренд. Во второй части книги мы будем иметь дело уже с последовательностями зависимых случайных величин, обычно со стационарными случайными процессами, выделяя при этом процессы авторегрессии. В конце книги будут рассмотрены модели, в которых на тренд накладывается случайная составляющая, образующая стационарный случайный процесс. Необходимые сведения о стационарных случайных процессах приводятся в гл. 7.

Во многих случаях модель можно определить с точностью до конечного числа параметров. Задачи статистических выводов будут связаны тогда именно с этими параметрами. В других ситуациях модель оказывается более неопределенной и приходится использовать непараметрические методы. Разумеется, модель должна достаточно хорошо представлять механизм образования соответствующего ряда. Однако, будучи математической абстракцией, она является лишь только приближением к реальному явлению. Сколь же точно можно определить модель, зависит от уровня знаний об исследуемом процессе и соответственно от той информации, которую мы можем получить с использованием статистического анализа, зависящего от характера этих знаний. В данной книге будет описано много методов и их свойств. Делается это для того, чтобы иметь возможность выбрать приемлемый метод, приводящий к полезным результатам. При этом внимание уделяется как самому статистическому выводу, так и его математическому обоснованию.

Первоначально анализ временных рядов базировался на моделях, в которых влияние временного параметра проявлялось только в систематической составляющей. Эту ситуацию можно было бы назвать классической, поскольку в известной степени она восходит к тем временам, когда Гаусс и другие развивали теорию и метод наименьших квадратов с целью применения их в астрономии и физике. В таких моделях мы предполагаем, что течение времени никак не отражается на случайной составляющей. Точнее говоря, мы предполагаем, что математическое ожидание (т. е. среднее значение) случайной составляющей тождественно равно нулю, дисперсия равна некоторой постоянной и что значения u_t в различные моменты времени некоррелированы. Такое определение приводит к тому, что всякую зависимость от времени приходится включать в систематическую составляющую $f(t)$. Последовательность $f(t)$ может

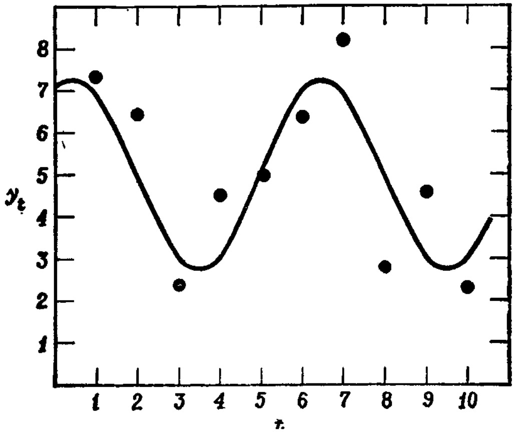

Рис. 1.1.
Ряд с тригонометрическим трендом.

зависеть от некоторых неизвестных коэффициентов и от известных величин, меняющихся со временем. В этом случае ее называют «функцией регрессии». Методы статистических выводов для коэффициентов функции регрессии оказываются полезными во многих областях статистики. Своеобразие же методов, относящихся именно к временным рядам, состоит в том, что здесь исследуются те модели, в которых упомянутые выше величины, меняющиеся со временем, являются известными функциями t.

В рамках сделанных ограничений можно выделить два типа временных последовательностей $f(t)$, часто называемых трендом. Один тип представляют медленно меняющиеся функции времени, примером которых могут служить полиномы достаточно низкой степени. К другому типу принадлежат циклические последовательности, например, конечные отрезки ряда Фурье, представляющие собой конечные суммы пар синусоидальных и косинусоидальных составляющих. Такой парой может являться $\alpha \cos \lambda t + \beta \sin \lambda t$ $(0 < \lambda < \pi)$. Ее можно записать и с использованием одной только функции косинус, именно $\rho \cos (\lambda t - \theta)$. Период этой функции времени равен $2\pi/\lambda$, т. е. она повторяет свои значения всякий раз по прошествии времени $2\pi/\lambda$. Соответствующая частота, т. е. величина, обратная периоду, равна $\lambda/(2\pi)$. Коэффициент $\rho = \sqrt{\alpha^2 + \beta^2}$ является амплитудой, а θ — фазой указанной функции. Считается, что наблюдаемый ряд представляет собой сумму подобного отрезка ряда Фурье $f(t)$ и случайной составляющей. На рис. 1.1 представлены значения функции $y_t = 5 + 2 \cos 2\pi t/6 + \sin 2\pi t/6 + u_t$, где составляющая u_t нормально распределена с нулевым средним и единичной дисперсией. [Функция $f(t)$ представлена здесь в виде функции от непрерывного аргумента t.] Последовательные значения y_t разбросаны случайным образом по обе стороны от кривой $y = f(t)$. Если даже эта кривая известна и если известен закон распределения ошибки, то информация о значениях $y_1, \ldots, y_{t-1}$ не оказывает в данной модели никакой помощи в предсказании значения y_t. Поведение графика функции $f(s)$ для $s > t - 1$ не зависит от значений $y_1, \ldots, y_{t-1}$.

Подобная модель может оказаться приемлемой в ряде физических и экономических задач. В астрономии, например, $f(t)$ может описывать пространственное положение (по одной из координат) планеты в моменты времени t. Так как телескоп — прибор не идеальный, а состояние атмосферы постоянно изменяется, определение соответствующей координаты планеты производится с некоторой, хотя и достаточно малой, ошибкой. Эта ошибка наблюдения никак не влияет ни на последующие положения планеты, ни на реальные наблюдения этих положений. В случае свободно колеблющегося маятника его смещение (измеренное от нижнего положения) является тригонометрической функцией $\rho \cos(\lambda t - \theta)$.

Одной из общих моделей, в которой влияние временного параметра проявляется в случайной составляющей, является стационарный случайный процесс. Проиллюстрируем это примером процесса авторегрессии. Предположим, что y_1 имеет некоторое распределение с нулевым средним. Пусть y_1 и y_2 имеют совместное распределение, совпадающее с совместным распределением случайных величин y_1 и $\rho y_1 + u_2$, где u_2 не зависит от y_1 и имеет нулевое математическое ожидание. Совместное распределение $y_1, y_2, \ldots, y_{t-1}, y_t$ для $t = 3, 4, \ldots$ будем полагать в свою очередь таким же, как совместное распределение $y_1, y_2, \ldots, y_{t-1}, \rho y_{t-1} + u_t$, причем предполагается, что случайная величина u_t не зависит от $y_1, \ldots, y_{t-1}$ и имеет нулевое математическое ожидание. Если маргинальные распределения u_2, $u_3, \ldots$ совпадают, а распределение y_1 выбрано надлежащим образом, то последовательность $\{y_t\}$ образует стационарный случайный процесс, именно процесс авторегрессии, и

$$y_t = \rho y_{t-1} + u_t \tag{2}$$

является стохастическим разностным уравнением первого порядка. Такое построение для $\rho = 1/2$ иллюстрирует рис. 1.2. В этой модели «возмущение» u_t оказывает влияние и на y_t, и на все последующие y_r. Из указанного построения вытекает, что условное математическое ожидание y_t при заданных значениях $y_1, \ldots, y_{t-1}$ удовлетворяет равенству

$$\mathscr{E}(y_t \mid y_1, \ldots, y_{t-1}) = \rho y_{t-1}. \tag{3}$$

(В действительности для процесса первого порядка значения y_t и $y_{t-2}, \ldots, y_1$ условно независимы при заданном значении y_{t-1}.) Если