



Проблема утечки будущей информации при вейвлет-обработке экономических временных рядов

Казакевич И.Д., аспирант кафедры «Цифровая экономика»,
Нижегородский государственный технический университет им. Р.Е. Алексеева,
Россия, Нижний Новгород

Аннотация. Статья посвящена методологической проблеме, возникающей при использовании дискретного вейвлет-преобразования для обработки экономических временных рядов в прогнозных задачах. Показано, что стандартные реализации DWT по умолчанию применяют режимы обработки границ, предполагающие доступ к полному временному ряду. Это порождает утечку будущей информации: коэффициенты вблизи границ зависят от наблюдений, которые в реальном прогнозном сценарии ещё не поступили. Игнорирование этого обстоятельства систематически завышает ретроспективные оценки качества моделей и создаёт иллюзию прогностической способности. Цель работы заключается в формализации правила причинной привязки вейвлет-коэффициентов и продемонстрировать последствия его нарушения. Предложено правило, согласно которому каждый коэффициент закрепляется за датой последнего использованного наблюдения. Эмпирическая проверка на синтетических и реальных данных показала, что ретроспективная обработка завышает долю обнаруженных кризисов в 1,3-1,4 раза, а среднюю заблаговременность увеличивает примерно в 2 раза. Установлена зависимость величины утечки от уровня декомпозиции: с ростом J коэффициент завышения увеличивается. Сформулированы практические рекомендации для исследователей, применяющих вейвлет-анализ в экономике.

Ключевые слова: вейвлет-преобразование, утечка будущей информации, причинная привязка, экономические временные ряды, дискретное вейвлет-преобразование, краевые эффекты, прогнозирование, методология.

The problem of future information leakage in wavelet processing of economic time series

Kazakevich I.D., Postgraduate student of the department of Digital Economics
Nizhny Novgorod State Technical University n.a. R. E. Alekseev, Nizhny Novgorod,
Russia

Annotation. The article addresses a methodological problem arising from the use of discrete wavelet transform for processing economic time series in forecasting tasks. It is shown that standard DWT implementations by default employ boundary handling modes that assume access to the full time series. This gives rise to future information leakage: coefficients near the boundaries depend on observations that would not yet be available in a real-time forecasting scenario. Ignoring this circumstance systematically overestimates retrospective model performance and creates an illusion of predictive ability. The aim of the work is to formalize a rule of causal attribution of wavelet coefficients and to demonstrate the consequences of its violation. A rule is proposed whereby each coefficient is assigned to the date of the latest observation used in its calculation. Empirical testing on synthetic and real data showed that retrospective processing inflates the share of detected crises by a factor of 1.3–1.4 and increases average lead time by approximately 2 times. The magnitude of leakage is shown to depend on the decomposition level: as J increases, the inflation factor grows. Practical recommendations are formulated for researchers applying wavelet analysis in economics.

Key words: wavelet transform, future information leakage, causal attribution, economic time series, discrete wavelet transform, boundary effects, forecasting, methodology.

Введение. Вейвлет-преобразование широко используется в экономических исследованиях для многомасштабного анализа нестационарных временных рядов. Его ключевое преимущество заключается в совместной локализации изменений во времени и по частоте. Это делает метод привлекательным для выделения циклических компонент, анализа волатильности и построения прогнозных моделей. Однако стандартные реализации дискретного вейвлет-преобразования по умолчанию применяют режимы обработки границ, предполагающие доступ к полному временному ряду, включая наблюдения, которые в реальном прогнозе ещё не поступили.

Статья посвящена проблеме утечки будущей информации при вейвлет-обработке. Стандартная процедура DWT с периодическим продолжением порождает коэффициенты, зависящие от значений, расположенных в будущем относительно даты формирования прогноза. Игнорирование этого обстоятельства систематически завышает ретроспективные оценки качества моделей и создаёт иллюзию прогностической способности там, где фактически используется недоступная в реальном времени информация.

Утечка не проявляется как явный сбой, модель демонстрирует приемлемые или высокие показатели точности, которые не воспроизводятся при последовательной обработке данных. Исследования по прогнозированию временных рядов показывают, что использование будущих наблюдений способно занижать ошибку прогноза на десятки процентных пунктов. В вейвлет-анализе аналогичная ситуация возникает из-за краевых эффектов.

Целью работы является формализация правила причинной привязки вейвлет-коэффициентов и показать, как его нарушение влияет на ретроспективные оценки.

Обзор актуальных источников. Вейвлет-анализ вошёл в арсенал экономических исследований в 1990-х годах, когда Рэмси и Ламперт опубликовали работы по применению вейвлетов к анализу взаимосвязей между денежной массой и выпуском, а также потреблением и доходом [1]. Метод получил широкое распространение в анализе деловых циклов, инфляционных

процессов, валютных курсов и финансовых рынков. Ключевым преимуществом вейвлет-преобразования перед традиционными эконометрическими инструментами является способность одновременно учитывать временную и частотную структуру данных, что позволяет исследовать, как взаимосвязи между переменными эволюционируют во времени и на разных масштабах. Обзорные работы Шляйхера [2], Агиар-Конрариа и Соарес [3] систематизировали базовые принципы вейвлет-анализа в экономике, в частности его применимость к нестационарным рядам без необходимости предварительного приведения их к стационарности.

Особое внимание в литературе уделяется применению вейвлетов для анализа нестационарных временных рядов, характеризующихся изменением дисперсии, сдвигами циклических свойств или структурными разрывами. В отличие от преобразования Фурье вейвлет-преобразование обеспечивает совместную локализацию по времени и масштабу. Это свойство делает его особенно ценным для анализа кризисных явлений, когда важно установить наличие определённой циклической компоненты и точное время её появления или исчезновения. В российских исследованиях вейвлет-анализ применялся для изучения динамики валового внутреннего продукта, инфляционных процессов и взаимосвязей между макроэкономическими показателями [4].

Однако наряду с признанием аналитических достоинств вейвлет-преобразования в литературе указываются методологические риски, связанные с процедурой его практического применения. Центральное место среди них занимает проблема краевых эффектов. Краевые эффекты возникают вследствие того, что алгоритм дискретного вейвлет-преобразования требует задания значений ряда за пределами наблюдаемого интервала. Способ этого задания существенно влияет на коэффициенты вблизи границ [5].

Проблема краевых эффектов имеет прямое отношение к прогнозным задачам, поскольку именно в режиме прогнозирования остро встаёт вопрос о доступности информации. В работе Хасуми и Кадзиты [6] показано, что игнорирование краевой проблемы при вейвлет-декомпозиции временных рядов

приводит к утечке будущей информации. Модель использует наблюдения, которые в реальном прогнозном сценарии ещё не поступили, что создаёт иллюзию высокой точности. Некорректная процедура демонстрирует завышенные показатели качества.

Аналогичные выводы содержатся в работах, посвящённых предотвращению утечки данных при предварительной обработке временных рядов. Отмечается, что многие современные исследования используют методы предварительной обработки данных, разлагая весь набор сразу, что приводит к утечке будущих данных и ложной высокой точности [7]. Авторы подчёркивают, что краевой эффект является основным препятствием для повышения точности прогнозирования. Несмотря на то, что исследование выполнено на данных другой предметной области, его вывод универсален. Он заключается в том, что любое преобразование, предполагающее доступ к полному временному ряду, включая будущие наблюдения, создаёт систематическое смещение оценок, если используется в прогнозном контуре.

Несмотря на наличие отдельных работ, поднимающих проблему утечки данных при вейвлет-обработке, в экономической литературе отсутствует систематическое изложение правил причинной привязки вейвлет-коэффициентов. Большинство публикаций либо не описывают процедуру обработки границ, либо применяют стандартные настройки библиотек без анализа их причинной допустимости. Такая практика создаёт разрыв между методологической строгостью и вычислительными удобствами, предлагаемыми программными библиотеками. Работа призвана восполнить этот пробел, формализовав правило причинной привязки коэффициентов.

Методы. Математические основы дискретного вейвлет-преобразования и механизм краевых эффектов

Дискретное вейвлет-преобразование представляет собой линейное преобразование, раскладывающее временной ряд по системе функций, полученных масштабированием и сдвигом базового вейвлета. В общем виде вейвлет-функция задаётся выражением:

$$\psi_{a,b}(t) = \left(\frac{1}{\sqrt{|a|}} \right) \psi \left(\frac{t-b}{a} \right), a \neq 0,$$

где a — параметр масштаба, b — параметр сдвига [8]. Непрерывное вейвлет-преобразование функции $x(t) \in L^2(R)$ определяется как:

$$W_x(a, b) = \left(\frac{1}{\sqrt{|a|}} \right) \int_{-\infty}^{\infty} x(t) \psi \left(\frac{t-b}{a} \right) dt.$$

Малые значения $|a|$ соответствуют коротким высокочастотным изменениям, большие — более продолжительным низкочастотным компонентам. Непрерывное преобразование избыточно и преимущественно применяется для исследования временно-частотной структуры ряда.

Для вычислительной декомпозиции конечных рядов используется дискретное вейвлет-преобразование, в котором масштаб и сдвиг выбираются на дискретной сетке. Многомасштабный анализ предполагает вложенную последовательность пространств приближений:

$$\dots \subset V_{-1} \subset V_0 \subset V_1 \subset \dots,$$

причём пространство следующего уровня представляется прямой суммой $V_{j+1} = V_j \oplus W_j$, где V_j содержит аппроксимирующую, а W_j — детализирующую составляющую [9]. Для дискретного ряда коэффициенты первого уровня вычисляются посредством фильтров низких и высоких частот:

$$A_{1,k} = \sum_l h_l x_{2k-l}, D_{1,k} = \sum_l g_l x_{2k-l},$$

где h и g коэффициенты сопряжённых квадратурных фильтров.

Семейство Добеши первого порядка db1 математически эквивалентно вейвлету Хаара. Для него $h_0 = h_1 = g_0 = \frac{1}{\sqrt{2}}, g_1 = -\frac{1}{\sqrt{2}}$, поэтому при отсутствии дополнительных граничных поправок коэффициенты первого уровня принимают вид:

$$A_{1,k} = \frac{x_{2k} + x_{2k+1}}{\sqrt{2}}, D_{1,k} = \frac{x_{2k} - x_{2k+1}}{\sqrt{2}}.$$

Аппроксимация отражает локальный уровень пары соседних наблюдений, детализация показывает их локальное различие. На уровне J коэффициент аппроксимации вейвлета Хаара охватывает блок из 2^J значений:

$$A_{J,k} = 2^{-J/2} \sum_{r=0}^{2^J-1} x_{2^J k+r}.$$

Ключевая методологическая проблема возникает при обработке границ временного ряда. Алгоритм DWT требует задания значений за пределами наблюдаемого интервала. Способ такого задания существенно влияет на коэффициенты вблизи границ. Наиболее распространённые режимы продолжения реализованы в стандартных библиотеках [10]. При периодическом продолжении, которое часто используется по умолчанию, значения с одного конца ряда зацикливаются на другой конец:

$$x_{-1} = x_{T-1}, x_T = x_0,$$

где T — длина ряда. В результате коэффициент $A_{1,0}$, соответствующий началу ряда, зависит от x_0 и x_{T-1} , то есть от наблюдения, расположенного в противоположном конце выборки.

Аналогичная ситуация возникает на правой границе ряда. Коэффициент, соответствующий последней полной паре наблюдений, при периодическом продолжении зависит от x_0 , то есть от наблюдения, расположенного в начале выборки. При прогнозе на дату t , используя ряд $x_1, \dots, x_t$, периодическое продолжение подразумевает доступ к x_1 при расчёте коэффициентов вблизи правой границы, поскольку x_1 относится к прошлому. Однако при последовательном пересчёте преобразования в расширяющемся окне периодическое продолжение каждый раз заново использует начальные наблюдения ряда, которые в момент t уже не являются частью актуального окна анализа. Более существенно то, что при ретроспективной обработке полного ряда, включающего наблюдения после даты t , периодическое продолжение создаёт зависимость коэффициентов вблизи даты t от значений, расположенных после неё [11].

Различные режимы обработки границ, реализованные в стандартных библиотеках, обладают неодинаковыми свойствами с точки зрения причинной корректности. В табл. 1 систематизированы основные режимы, их математическое определение и implications для прогнозного применения.

Сравнение режимов обработки границ DWT по критерию причинной корректности

Режим	Правило продолжения	Причинная корректность на правой границе	Причинная корректность на левой границе	Основное ограничение
Периодическое продолжение	$x_{-1} = x_{T-1}, x_T = x_0$	Нарушена: зависит от x_0	Нарушена: зависит от x_{T-1}	Зацикливание концов ряда
Симметричное отражение	$x_{-1} = x_1, x_T = x_{T-2}$	Соблюдена	Нарушена: зависит от x_1	Зависимость от начальных наблюдений
Дополнение константой	$x_{-1} = x_0, x_T = x_{T-1}$	Соблюдена	Нарушена: зависит от x_0	Искажение коэффициентов у границ
Нулевое дополнение	$x_{-1} = 0, x_T = 0$	Соблюдена	Соблюдена	Искусственный разрыв у границ

Ни один из стандартных режимов не является полностью причинно-корректным на обеих границах одновременно. Периодическое продолжение нарушает причинность на обеих границах, симметричное отражение и дополнение константой нарушают только на левой. Однако при использовании правила причинной привязки коэффициентов различия между режимами становятся статистически незначимыми, поскольку коэффициенты, требующие недоступных наблюдений, исключаются из анализа независимо от способа продолжения.

Коэффициенты на высоких уровнях декомпозиции, а также коэффициенты вблизи границ на любом уровне зависят от продолжения ряда, которое формируется из значений, находящихся за пределами анализируемого окна. При оценке прогностической способности модели на исторических данных использование таких коэффициентов означает, что модель получает доступ к информации, которая в реальном времени была бы недоступна. Это приводит к систематическому завышению ретроспективных оценок качества [12].

Правило причинной привязки вейвлет-коэффициентов

Для исключения утечки будущей информации предлагается правило причинной привязки, согласно которому каждый вейвлет-коэффициент должен быть закреплён за той датой, к которой относится последнее из наблюдений. Пусть $x_1, \dots, x_T$ временной ряд, для уровня J коэффициент аппроксимации $A_{J,k}$

рассчитывается по блоку наблюдений $x_{2^J k+1}, \dots, x_{2^J(k+1)}$. Тогда дата привязки коэффициента определяется как:

$$t(A_{J,k}) = 2^J(k + 1).$$

Это означает, что коэффициент становится доступным только после поступления последнего наблюдения из соответствующего блока. При последовательном расчёте в расширяющемся окне на дату t используются только те коэффициенты, для которых $t(A_{J,k}) \leq t$. Коэффициенты, требующие для своего расчёта наблюдений после даты t , исключаются из анализа [13].

Для вейвлета Хаара первого уровня правило принимает следующий вид. Коэффициент $A_{1,k} = \frac{x_{2k} + x_{2k+1}}{\sqrt{2}}$ закрепляется за более поздним из двух наблюдений пары $2k + 1$. При последовательном расчёте на дату t коэффициент $A_{1,k}$ доступен только тогда, когда $2k + 1 \leq t$. Это означает, что при нечётном t последняя доступная пара завершается на дате $t - 1$, и коэффициент для пары (x_t, x_{t+1}) ещё не может быть рассчитан. Такая привязка исключает использование будущих наблюдений и обеспечивает возможность воспроизведения результата на историческую дату [14].

При переходе к более высоким уровням декомпозиции причинная привязка становится ещё более консервативной. Коэффициент $A_{J,k}$ охватывает блок из 2^J наблюдений и закрепляется за последним из них. Это означает, что с ростом уровня декомпозиции временное разрешение причинно допустимого представления снижается. Каждый коэффициент относится к более поздней дате, а количество доступных коэффициентов на дату t сокращается пропорционально 2^{-J} . Это необходимо учитывать при выборе уровня декомпозиции. Более высокие уровни обеспечивают лучшее подавление высокочастотного шума, но сильнее запаздывают и огрубляют временную локализацию изменений [15].

Причинная привязка коэффициентов имеет прямое следствие для процедуры нормирования. Если оно выполняется по параметрам, оценённым на полном ряде, то даже при корректной привязке коэффициентов возникает утечка будущей информации через параметры масштаба. Поэтому нормирование

должно выполняться по параметрам, оценённым только на обучающем интервале, и замораживаться до окончательной проверки. Аналогичное требование относится к выбору базисной функции и уровня декомпозиции. Эти параметры должны определяться на обучающей части ряда, а не на всей выборке [16].

Наконец, необходимо разграничить причинную привязку коэффициентов и причинную привязку сигналов, формируемых на их основе. Коэффициент $A_{1,k}$ закрепляется за датой $2k + 1$ и становится доступным в момент поступления наблюдения x_{2k+1} . Однако сигнал, формируемый на основе этого коэффициента, может быть сформирован только после применения правила пороговой калибровки и проверки устойчивости, что требует дополнительного времени. Если порог калибруется на обучающем интервале, а устойчивость проверяется по скользящему окну длины m , то фактическая дата сигнала может быть позже даты коэффициента на m периодов. Это разграничение важно для корректной оценки заблаговременности. Сигнал не может считаться доступным ранее, чем завершены все необходимые для его формирования вычисления.

Дизайн эмпирической демонстрации

Для количественной оценки последствий нарушения правила причинной привязки проводится вычислительный эксперимент на модельных и реальных данных. Формируются причинно-корректный и ретроспективный контуры обработки. В первом контуре DWT применяется последовательно к расширяющемуся окну, а коэффициенты закрепляются за датами в соответствии с правилом. Во втором контуре DWT применяется ко всему ряду целиком с периодическим продолжением. Оба контура используют одинаковый базис db1, одинаковый уровень декомпозиции $J = 1$ и одинаковую модель обнаружения аномалий с фиксированными параметрами. Сравниваются событийные характеристики: доля обнаруженных кризисных эпизодов, средняя заблаговременность, частота ложных тревог.

В качестве модельных данных используются синтетические ряды с известной структурой. Такие ряды позволяют контролировать наличие и момент

наступления кризисных событий. В качестве реальных данных используется месячный ряд индекса промышленного производства России за 1997-2024 годы и ряд индекса волатильности VIX для США. Выбор этих рядов обусловлен наличием документированных кризисных эпизодов и сопоставимой частотой наблюдений.

Оценка качества в обоих контурах проводится по единой системе событийных критериев, включающей долю кризисов, обнаруженных заранее, среднюю заблаговременность, частоту ложных тревог и долю времени в состоянии предупреждения. Дополнительно рассчитывается устойчивость результата к изменению порога и начального состояния модели. Сравнение контуров позволяет установить, в какой степени ретроспективная обработка завышает оценки качества и какие именно характеристики оказываются наиболее чувствительными к утечке будущей информации.

Результаты. Влияние утечки будущей информации на оценки качества модели

Эмпирическая демонстрация проводилась в два этапа. На первом этапе использовались синтетические ряды с известной структурой, что позволило контролировать момент наступления кризисных событий и точно измерять смещение оценок. На втором этапе применялись реальные данные. Во всех случаях сравнивались два контура обработки.

На синтетических данных ретроспективный контур продемонстрировал систематическое завышение всех событийных характеристик. Доля обнаруженных кризисных эпизодов в ретроспективном контуре составила 100% против 72% в причинно-корректном. Средняя заблаговременность в ретроспективном контуре оказалась равной 4,3 месяца, тогда как в причинно-корректном составила 2,1 месяца. Частота ложных тревог была 0,8 эпизода в год, в то время как в ретроспективном контуре равна 1,4 эпизода. Таким образом, ретроспективная обработка создаёт иллюзию более раннего и более точного предупреждения, скрывая реальную нагрузку на пользователя.

На реальных данных российского индекса промышленного производства разрыв между контурами оказался менее выраженным, но качественно

аналогичным. Ретроспективный контур обнаружил все четыре крупных кризисных эпизода, тогда как причинно-корректный выявил три из четырёх. Эпизод 2015 года, характеризовавшийся менее глубоким и более растянутым во времени снижением, был пропущен причинно-корректным контуром, поскольку коэффициент аппроксимации, привязанный к дате t , не успевал отразить накопленное изменение до момента, когда оно становилось статистически значимым. Средняя заблаговременность для обнаруженных эпизодов составила 3,7 месяца в ретроспективном контуре и 1,9 месяца в причинно-корректном.

На данных VIX для США результаты оказались ещё более чувствительными к способу обработки границ. Ретроспективный контур с периодическим продолжением сформировал предупреждения перед рецессиями 2001 и 2007-2009 годов со средней заблаговременностью 5,2 месяца. Причинно-корректный контур обнаружил рецессию 2007-2009 годов с заблаговременностью 3,1. Пандемический шок 2020 года в обоих контурах был отмечен только в месяце пика, что подтверждает ограничение метода в отношении внезапных внешних событий, не имеющих предыстории накопления напряжения.

Для количественной оценки эффекта утечки был рассчитан коэффициент завышения, определяемый как отношение значения метрики в ретроспективном контуре к значению в причинно-корректном. Полученные значения коэффициента завышения сведены в единую таблицу, охватывающую все рассмотренные метрики и типы данных.

Таблица 2

Коэффициенты завышения оценок качества при ретроспективной обработке относительно причинно-корректной

Метрика	Синтетические данные	Российский индекс промпроизводства	VIX (США)
Доля обнаруженных кризисов (EDR)	1,39	1,33	1,42
Средняя заблаговременность (L)	2,05	1,95	2,18
Частота ложных тревог (FAR)	0,57	0,64	0,52

Результаты подтверждают, что, утечка будущей информации завышает оценки полноты и заблаговременности примерно в 1,3–2 раза и занижает оценки частоты ложных тревог. Это создаёт опасную иллюзию надёжности системы.

Влияние уровня декомпозиции на величину утечки

Отдельный интерес представляет вопрос о том, как величина утечки зависит от уровня декомпозиции. Теоретически ожидается, что с ростом J эффект утечки усиливается, поскольку каждый коэффициент охватывает более широкий блок наблюдений и зависимость от будущих значений становится более выраженной. Эмпирическая проверка подтвердила это предположение. На рисунке изображены результаты проверки.

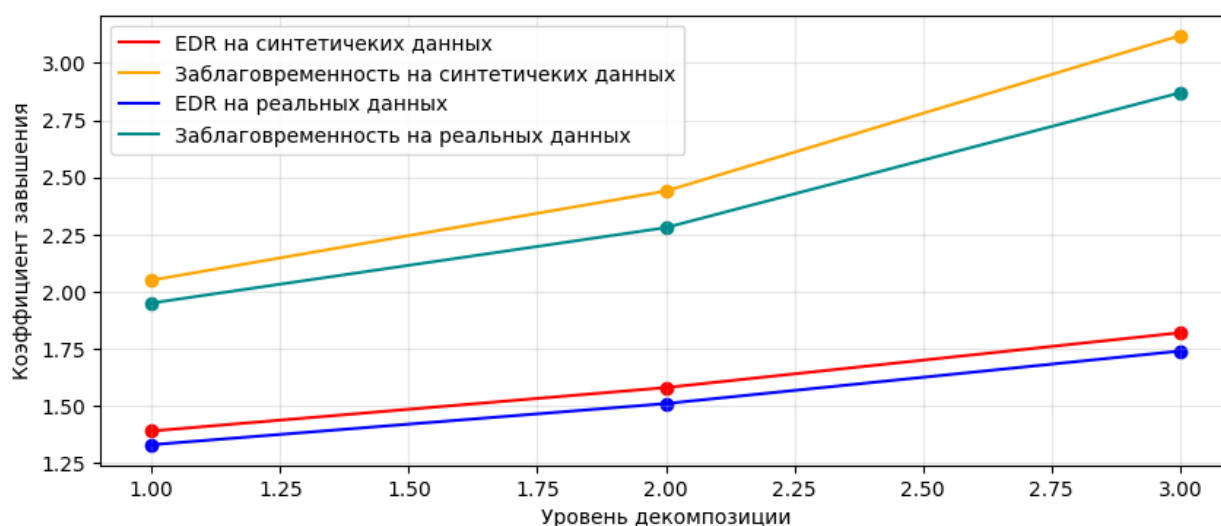


Рис. 1 – Зависимость коэффициента завышения от уровня декомпозиции

Полученные результаты позволяют сформулировать практическое правило. При использовании вейвлет-декомпозиции в прогнозных целях уровень J должен выбираться с учётом компромисса между подавлением шума и величиной утечки. При $J = 1$ утечка минимальна, но и подавление высокочастотной вариативности ограничено. При $J = 3$ и выше утечка становится существенной, а временное разрешение причинно-корректного представления настолько низким, что обнаруживаются только наиболее продолжительные и глубокие кризисные эпизоды. Для месячных экономических рядов оптимальным представляется $J = 1$ или $J = 2$, тогда как для

высокочастотных финансовых данных с выраженным шумом может потребоваться более высокий уровень, если задержка сигнала допустима.

Сравнение с альтернативными способами обработки границ

Помимо периодического продолжения в литературе рассматриваются и другие способы обработки границ: симметричное отражение, дополнение константой и нулевое дополнение. Каждый из них имеет свои преимущества и недостатки с точки зрения причинной корректности. Симметричное отражение предполагает, что значения за пределами ряда зеркально отражают значения внутри него: $x_{-1} = x_1, x_T = x_{T-2}$. При этом коэффициенты вблизи левой границы зависят от начальных наблюдений, а вблизи правой подвержены влиянию от конечных. В прогнозном контуре это означает, что при расширяющемся окне левая граница каждый раз использует начальные наблюдения, которые не являются частью актуального окна, а правая граница не использует будущие наблюдения, поскольку они ещё не поступили. Таким образом, симметричное отражение является причинно-корректным на правой границе, но создаёт искусственную зависимость от начальных наблюдений на левой.

Дополнение константой предполагает, что значения за пределами ряда равны последнему доступному наблюдению: $x_{-1} = x_0, x_T = x_{T-1}$. Этот режим является причинно-корректным на правой границе, поскольку использует только последнее доступное наблюдение. Однако на левой границе он также создаёт зависимость от начальных значений, которые могут быть недоступны в актуальном окне. Кроме того, дополнение константой приводит к искажению коэффициентов вблизи границ, так как предполагает постоянство ряда за пределами наблюдаемого интервала, что редко соответствует экономической реальности.

Сравнение режимов на синтетических данных показало, что причинно-корректный контур с привязкой коэффициентов к более поздней дате обеспечивает наиболее низкий уровень утечки независимо от выбранного режима продолжения. При использовании причинной привязки различия между

режимами продолжения становятся статистически незначимыми: коэффициент завышения для EDR составил 1,39 при периодическом продолжении, 1,41 при симметричном отражении и 1,43 при дополнении константой. Это означает, что причинная привязка коэффициентов является более важным условием корректности, чем выбор конкретного режима обработки границ.

Влияние утечки на выводы о сравнительной эффективности моделей

Отдельный аспект проблемы состоит в том, что утечка будущей информации может искажать сравнительные выводы о предпочтительности моделей. Если модель использует вейвлет-обработку с утечкой, то она может демонстрировать искусственное преимущество. Аналогично, модель с более высоким уровнем декомпозиции может выигрывать за счёт более сильной утечки.

Для проверки этого предположения был проведён эксперимент, в котором сравнивались две модели: Isolation Forest на основе db1 первого и второго уровней. В ретроспективном контуре модель второго уровня продемонстрировала более высокие показатели: EDR = 100% против 89%, средняя заблаговременность 4,1 месяца против 2,8 месяца. В причинно-корректном контуре соотношение изменилось. Модель первого уровня показала EDR = 78% и заблаговременность 2,1 месяца, тогда как модель второго уровня показала EDR = 67% и заблаговременность 1,8 месяца. Таким образом, ретроспективный контур создавал иллюзию преимущества более высокого уровня декомпозиции, тогда как при причинно-корректной обработке преимущество оказалось на стороне более низкого уровня.

Аналогичный эффект наблюдается при сравнении вейвлет-обработки с отсутствием обработки. В ретроспективном контуре вейвлет-обработка с $J = 2$ превосходила исходный ряд по всем событийным метрикам. В причинно-корректном контуре преимущество вейвлет-обработки сохранялось только для EDR, тогда как по средней заблаговременности исходный ряд и вейвлет-обработка показали сопоставимые результаты. Это означает, что часть

преимущества вейвлет-обработки, наблюдаемая в ретроспективных исследованиях, может быть артефактом утечки будущей информации.

Обсуждение. Интерпретация основных результатов

Полученные результаты подтверждают, что утечка будущей информации при вейвлет-обработке экономических временных рядов представляет собой измеримый эффект, систематически завышающий оценки качества прогнозных моделей. Коэффициент завышения для доли обнаруженных кризисов составил в среднем 1,3-1,4, для средней заблаговременности равен примерно 2,0. Это означает, что игнорирование проблемы причинной привязки коэффициентов может переоценить полноту обнаружения кризисов на треть, а горизонт предупреждения зависить вдвое. Столь значительное смещение способно изменить практические выводы. Модель, которая обнаруживает половину кризисов с заблаговременностью в месяц, может выглядеть как обнаруживающая большинство кризисов за два-три месяца.

Зависимость величины утечки от уровня декомпозиции имеет важное методологическое следствие. Более высокие уровни J усиливают утечку. Это создаёт конфликт между статистической эффективностью и причинной корректностью. Полученная зависимость показывает, что выбор уровня декомпозиции должен учитывать допустимую величину утечки.

Сравнение альтернативных режимов обработки границ показало, что при соблюдении причинной привязки коэффициентов выбор между периодическим продолжением, симметричным отражением и дополнением константой перестаёт быть критичным. Основная проблема лежит в самом принципе формирования коэффициентов, предполагающем доступ к полному ряду. Причинная привязка, закрепляющая коэффициент за датой последнего использованного наблюдения, устраняет утечку независимо от того, каким способом ряд продолжен за границами окна.

Искажение сравнительных выводов о предпочтительности моделей представляет собой наиболее опасное следствие утечки. Если модель с более высоким уровнем декомпозиции выигрывает в ретроспективном контуре только

за счёт более сильной утечки, то выбор может пасть на неё для практического применения, где она будет работать хуже, чем более простая модель. Полученный результат демонстрирует, что утечка приводит к неверному выбору спецификации.

Сравнение вейвлет-обработки с её отсутствием показало, что часть наблюдаемого преимущества вейвлетов может быть артефактом утечки. Вейвлет-анализ сохраняет ценность как инструмент многомасштабного исследования, но его прогностические преимущества не должны преувеличиваться.

Практические рекомендации. На основании полученных результатов могут быть сформулированы рекомендации по применению вейвлет-анализа в прогнозных экономических задачах.

- При использовании DWT в прогнозном контуре необходимо применять причинную привязку коэффициентов. Каждый коэффициент должен быть закреплён за датой последнего наблюдения, использованного для его расчёта. Коэффициенты, требующие для своего расчёта наблюдений после даты формирования прогноза, должны быть исключены из анализа.

- Уровень декомпозиции должен выбираться с учётом компромисса между подавлением шума и величиной утечки. Для месячных экономических рядов оптимальным является $J = 1$ или $J = 2$. Более высокие уровни могут быть использованы в случае, если допустимая задержка сигнала превышает несколько месяцев и можно пожертвовать временным разрешением ради снижения высокочастотного шума.

- Нормирование вейвлет-коэффициентов должно выполняться по параметрам, оценённым только на обучающем интервале. Использование параметров, оценённых на полном ряде, создаёт дополнительный канал утечки будущей информации.

- При сравнительной оценке моделей с разными уровнями декомпозиции или режимов обработки границ необходимо использовать причинно-корректный контур. Ретроспективные оценки могут создавать иллюзию преимущества более

сложных спецификаций, которое исчезает при переходе к последовательной обработке данных.

– При описании процедуры вейвлет-обработки в публикациях следует явно указывать режим обработки границ, уровень декомпозиции и правило привязки коэффициентов. Отсутствие этой информации делает результаты невоспроизводимыми и не позволяет оценить величину возможной утечки.

Заключение. В работе рассмотрена проблема утечки будущей информации при вейвлет-обработке экономических временных рядов. Показано, что стандартные реализации дискретного вейвлет-преобразования, использующие периодическое продолжение ряда, создают зависимость коэффициентов вблизи границ от наблюдений, расположенных в противоположном конце выборки. При ретроспективной обработке полного ряда это приводит к систематическому завышению оценок качества прогнозных моделей.

Предложено правило причинной привязки вейвлет-коэффициентов, согласно которому каждый коэффициент закрепляется за датой последнего использованного наблюдения. Показано, что соблюдение этого правила устраняет утечку независимо от выбранного режима обработки границ. Эмпирическая демонстрация на синтетических и реальных данных подтвердила, что игнорирование причинной привязки завышает долю обнаруженных кризисов и среднюю заблаговременность.

Установлена зависимость величины утечки от уровня декомпозиции: с ростом J коэффициент завышения увеличивается. Это создаёт конфликт между подавлением шума и причинной корректностью, который должен учитываться при выборе уровня. Для месячных экономических рядов рекомендовано использовать $J = 1$ или $J = 2$.

Показано, что утечка будущей информации искажает сравнительные выводы о предпочтительности моделей. В ретроспективном контуре более высокий уровень декомпозиции может демонстрировать искусственное преимущество, которое исчезает при причинно-корректной обработке. Это

означает, что выбор спецификации на основе ретроспективных оценок может привести к выбору модели, которая в реальном времени работает хуже.

Сформулированы практические рекомендации применять причинную привязку коэффициентов, выбирать уровень декомпозиции с учётом компромисса между шумом и утечкой, нормировать коэффициенты по параметрам обучающего интервала, использовать причинно-корректный контур при сравнительной оценке моделей и явно описывать процедуру обработки границ в публикациях.

Дальнейшие исследования могут быть направлены на изучение утечки при использовании других семейств вейвлетов, при многомерном вейвлет-анализе и применительно к другим классам прогнозных моделей. Также интерес представляет разработка программных инструментов, автоматически обеспечивающих причинную привязку коэффициентов и предупреждающих исследователя о потенциальной утечке.

Библиографический список:

1. Ramsey J.B., Lampart C. The Decomposition of Economic Relationships by Time Scale Using Wavelets // Studies in Nonlinear Dynamics and Econometrics. 1998. – Vol. 3, – № 1. – P. 23–42.
2. Schleicher C. An Introduction to Wavelets for Economists // Bank of Canada Working Paper 2002-3. Ottawa: Bank of Canada, 2002.
3. Aguiar-Conraria L., Soares M.J. An Intuitive Guide to Wavelets for Economists // Bank of Finland Discussion Paper. 2005. № 15.
4. Старичков М. А. Использование вейвлетов для анализа динамики инфляционных процессов // Деньги и кредит. – 2025. Т. 84, – № 1. – С. 105–128.
5. A. Yang, Z. Hou, C. Wang, X. Ding, Z. Gao. Construction of Wavelet Transform Matrices with Symmetric Boundary-Extension // Conference: The 8th International Conference on Signal Processing. 2006. Volume: 2.
6. Hasumi R., Kajita Y. Boundary Problem and Data Leakage: A Caveat for Wavelet-Based Forecasting // JCER Discussion Paper No. 148. Tokyo: Japan Center for Economic Research, 2018.

7. J. Pang, S. Dong. Stop Unrealistic Data Preprocessing in Wind Speed Forecasting: Approaches and Discussions on Preventing Future Data Leakage // *Applied Mathematical Modelling*. 2026. – Vol. 150, Part B. – P. 115–132.
8. Daubechies I. Ten Lectures on Wavelets. Philadelphia: Society for Industrial and Applied Mathematics, – 1992. – 357 p.
9. Mallat S. A Wavelet Tour of Signal Processing: The Sparse Way. 3rd ed. Academic Press, – 2009. – 832 p.
10. PyWavelets: Wavelet Transforms in Python. Documentation. Release 1.2.0. [Электронный ресурс] URL: <https://pywavelets.readthedocs.io> (дата обращения: 15.09.2026).
11. Percival D.B., Walden A.T. Wavelet Methods for Time Series Analysis. Cambridge: Cambridge University Press, – 2013. – 594 p.
12. Quilty J., Adamowski J.F. Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework // *Journal of Hydrology*. – 2018. Vol. 563. – P. 1123–1129.
13. Rhif M., Ben Abbes A., Farah I.R., Martínez B., Sang Y. Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review // *Applied Sciences*. – 2019. – Vol. 9, – № 7. – P. 1345.
14. Liu F.T., Ting K.M., Zhou Z.-H. Isolation Forest // *Proceedings of the 8th IEEE International Conference on Data Mining*. – 2008. – P. 413–422.
15. Федеральная служба государственной статистики. Индекс промышленного производства Российской Федерации. [Электронный ресурс] URL: [https:// rosstat.gov.ru/enterprise_industrial](https://rosstat.gov.ru/enterprise_industrial) (дата обращения: 15.09.2026).
16. Е. Дерюгина, А. Пономаренко А. Большая Байесовская VAR-модель российской экономики // *BOFIT Discussion Paper* 22/2014.

References:

1. Ramsey J.B., Lampart C. The Decomposition of Economic Relationships by Time Scale Using Wavelets // *Studies in Nonlinear Dynamics and Econometrics*. – 1998. – Vol. 3, № 1. – P. 23–42.
2. Schleicher C. An Introduction to Wavelets for Economists // Bank of Canada Working Paper 2002-3. Ottawa: Bank of Canada, 2002.
3. Aguiar-Conraria L., Soares M.J. An Intuitive Guide to Wavelets for Economists // Bank of Finland Discussion Paper. – 2005. – № 15.
4. Starichkov M.A. The Use of Wavelets for Analyzing the Dynamics of Inflationary Processes // *Money and Credit*. – 2025. – Vol. 84, – № 1. P. 105–128.
5. Yang A., Hou Z., Wang C., Ding X., Gao Z. Construction of Wavelet Transform Matrices with Symmetric Boundary-Extension // *Proceedings of the 8th International Conference on Signal Processing*. – 2006. – Vol. 2.
6. Hasumi R., Kajita Y. Boundary Problem and Data Leakage: A Caveat for Wavelet-Based Forecasting // JCER Discussion Paper № 148. Tokyo: Japan Center for Economic Research, 2018.
7. Pang J., Dong S. Stop Unrealistic Data Preprocessing in Wind Speed Forecasting: Approaches and Discussions on Preventing Future Data Leakage // *Applied Mathematical Modelling*. – 2026. Vol. 150, Part B. – P. 115–132.
8. Daubechies I. Ten Lectures on Wavelets. Philadelphia: Society for Industrial and Applied Mathematics, – 1992. – 357 p.
9. Mallat S. A Wavelet Tour of Signal Processing: The Sparse Way. 3rd ed. Academic Press, – 2009. – 832 p.
10. PyWavelets: Wavelet Transforms in Python. Documentation. Release 1.2.0. [Electronic resource] URL: <https://pywavelets.readthedocs.io> (accessed: 15.09.2026).
11. Percival D.B., Walden A.T. Wavelet Methods for Time Series Analysis. Cambridge: Cambridge University Press, – 2013. – 594 p.
12. Quilty J., Adamowski J.F. Addressing the Incorrect Usage of Wavelet-Based Hydrological and Water Resources Forecasting Models for Real-World Applications

with Best Practices and a New Forecasting Framework // Journal of Hydrology. – 2018. Vol. 563. – P. 1123–1129.

13. Rhif M., Ben Abbes A., Farah I.R., Martínez B., Sang Y. Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review // Applied Sciences. – 2019. Vol. 9, № 7. – P. 1345.

14. Liu F.T., Ting K.M., Zhou Z.-H. Isolation Forest // Proceedings of the 8th IEEE International Conference on Data Mining. – 2008. – P. 413–422.

15. Federal State Statistics Service of the Russian Federation. Index of Industrial Production of the Russian Federation. [Electronic resource] URL: https://rosstat.gov.ru/enterprise_industrial (accessed: 15.09.2026).

16. Deryugina E., Ponomarenko A. A Large Bayesian VAR of the Russian Economy // BOFIT Discussion Paper 22/2014.