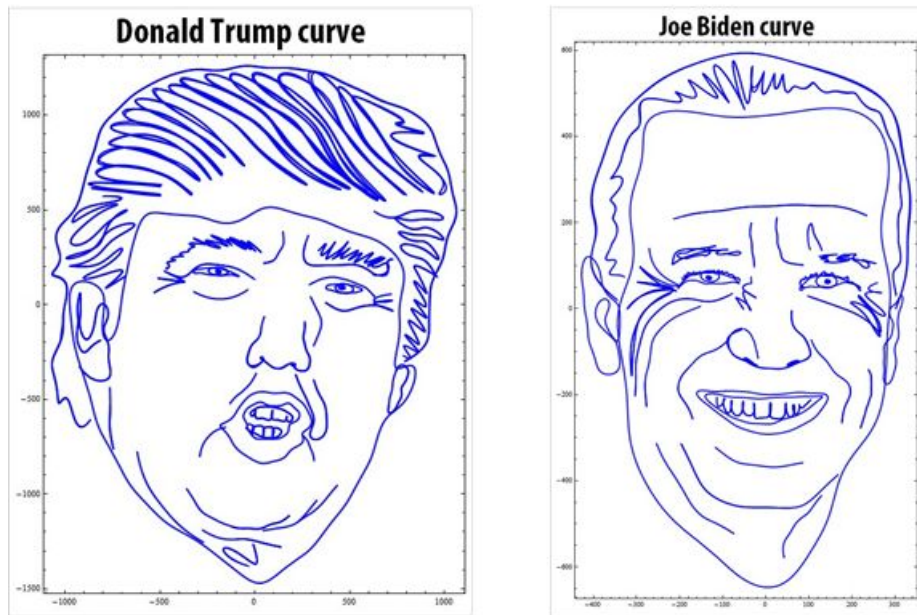


## Фурье-события: использование в задачах бизнеса, прогноз, анализ cash-flow.



**РИСУНКИ ПО ФУРЬЕ-ФОРМУЛАМ**

Использование в бизнесе методов, связанных с дискретным преобразованием Фурье, имеет значительный потенциал. Сдерживающим фактором реализации этого потенциала является высокий входной методологический барьер.

### Основные акценты работы:

- требования к данным для корректной Фурье-аппроксимации временных рядов;
- обоснованность ожиданий от прогнозов;
- небольшой набор гармоник достаточен для аппроксимации сложного ряда;
- что такое Фурье-событие;
- чем и как Фурье-события могут помочь бизнесу;
- Фурье-события в анализе cash-flow.

### **1. Прогноз.**

Задача была поставлена крупным морским перевозчиком и касалась прогнозов на цены фрахта по типам судов. У перевозчика имелась серия подписок на прогнозы международных аналитических компаний, но качество прогнозов его не устраивало. Аналитические компании использовали множественную регрессию, обладали продолжительной статистикой и постоянно увеличивали размерность моделей. При этом они сами признавали достаточно большой процент ошибок в своих прогнозах.

Критерий оценки успешности нового прогноза был следующим: дается фрагмент из исторических данных, формируется прогноз и просчитывается точность прогноза уже свершившегося будущего. Событийность стала сразу явной методологической проблемой. Если США до 2017 года вообще не продавали нефть, а потом сразу перешли в лидеры, то как это может влиять на выводы по историческим данным. Другие события: войны, кризисы, - с точки зрения прогноза являются по сути такими

же событиями, но ситуация с экспортом нефти из США чрезвычайно показательна, чтобы отмахнуться от фактора событийности в методологии прогноза (весовые коэффициенты продуцируют линейность, а события - разрыв и сингулярность).

Были перепробованы многие методы. Самым интересным оказалась аппроксимация рядом Фурье (Фурье-аппроксимация) временных рядов и ее исследование с точки зрения прогноза для бизнеса. При этом имелась техническая проблема - все время появлялся сдвиг аппроксимации от исходного ряда.

## 2. Формирование данных для Фурье-преобразования.

### Необходимые предварительные пояснения.

Дискретное преобразование Фурье применяется к векторам, состоящим из действительных значений. Если рассматривается временной ряд как набор точек <время-значение>, то преобразование Фурье применяется к вектору из последовательности значений временного ряда.

Существуют тонкости применения преобразования Фурье, которые связаны с количеством значений и характеристик промежутков между ними. Например, в исходных временных рядах могут быть неравные промежутки или отсутствовать значения для определенных временных позиций (выходные, праздники).

Во многих случаях полезной является следующая процедура. Исходный временной ряд сначала интерполируется, а затем из интерполированной функции берется нужное количество значений в нужных временных позициях. Таким образом, исходный временной ряд подменяется более частым регулярным рядом с требуемым количеством интерполированных значений.

Далее излагается подход, описанный [A. Dieckmann](http://www-elsa.physik.uni-bonn.de/~dieckman/DFT/DFT.html): <http://www-elsa.physik.uni-bonn.de/~dieckman/DFT/DFT.html> .

### Дискретное преобразование Фурье.

Вектор из действительных значений  $u=u(r)$  преобразуется в вектор комплексных значений  $f(s)$  по следующей формуле (существует несколько вариантов формул для  $F(s,r)$ , дающих эквивалентные результаты):  $f(s)=u(r)*F(s,r)$ , где

$$F(s, r) = \frac{e^{\frac{2i\pi(r-1)(s-1)}{n}}}{\sqrt{n}}$$

, а значения  $s, r$  меняются от 1 до  $n$ .

### Необходимые данные для получения Фурье-спектра.

Полученный вектор  $f(s)$  можно интерпретировать как Фурье-спектр, так как он содержит информацию об амплитудах, частотах и фазах основных гармоник.

При этом существуют требования к  $u(r)$ . Значения  $u(r)$  должны быть определены в точках разбиения интервала, когда интервал имеет длину целого числа шагов одинакового размера. Значение  $r$  соответствует позиции (индексу) в векторе. В общем случае  $r$  определяет позицию во времени (временной ряд) или пространстве (при других измерениях).

Предположим, что вектор  $u(r)$  должен быть определен на интервале  $[t_{\text{Min}}, t_{\text{Max}}]$ , длина которого составляет  $tt=(t_{\text{Max}}-t_{\text{Min}})$ . Пусть  $\text{delta}=tt/n$  соответствует расстоянию между соседними точками интервала, на которых вычисляется  $u(r)$ .

Рассмотрим какой процесс технически должен происходить.

Комплексная экспонента в матрице  $F(s,r)$  может быть интерпретирована как вектор-зонд (зависит от  $s$ ), который вращается в комплексной плоскости с частотой  $(s-1)/tt$  и последовательно продвигается (во времени или пространстве) по  $(r-1)*tt/n$ . Во время умножения матрицы вектор-зонд, соответствующий  $r$ , умножается на конкретное  $u(r)$ , а векторная сумма вычисляется по всем  $r$ , давая комплексное число  $f(s)$ . И так повторяется для всех  $s$  от 1 до  $n$ . Каждый  $f(s)$  показывает наличие или отсутствие составляющей, колеблющейся с частотой, связанной с  $s$ .

Как надо формировать  $u(r)$ ?

К этому моменту исходный временной ряд должен быть интерполирован и отображен на выбранный интервал, кратный целому числу шагов. Достаточное количество точек для аккуратной аппроксимации подбирается эмпирически.

На данном этапе главное, сколько точек брать и каких. Значение  $n$  зафиксировано в  $\text{delta}$ . В этом случае имеем комплект из  $n+1$  точки для всех значений разбиения интервала.

В  $u=u(r)$  необходимо включать точки только с первой до предпоследней, но не последнюю: всего  $n$ .

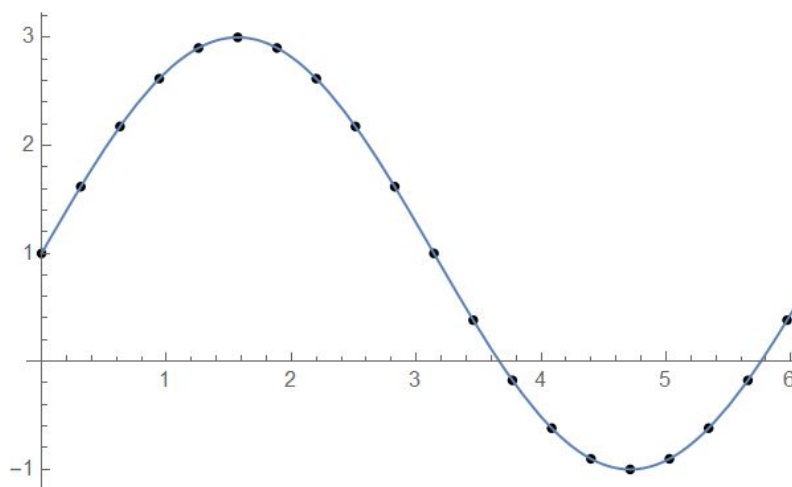
В противном случае при Фурье-аппроксимация будет немного сдвинута относительно исходного временного ряда.

### 3. Наглядная интерпретация Фурье-преобразований.

Для широкого применения Фурье-преобразования на практике необходимо прочувствовать, что оно дает помимо сложных формул, и правильно формировать исходные данные.

Рассмотрим, как Фурье-преобразование действует на синусоидальную функцию. Для этого полезно совместить на одном графике поведение функции и характеристики, которые дает Фурье-преобразование в конкретных точках и в целом на исследуемой функции.

Рассмотрим функцию  $1+\text{Sin}(2\pi x)$  на отрезке  $[0,\pi]$ .



Амплитуда этой функции соответствует 1Hz, так как она повторяет свое движение через  $2\pi$ .

Пусть  $n=20$ , тогда при разбиении интервала на равные части можно получить 21 значение в соответствующих точках разбиения. Но, следуя приведенному выше объяснению, будем оперировать только 20 точками — без последней (только черные на рисунке выше).

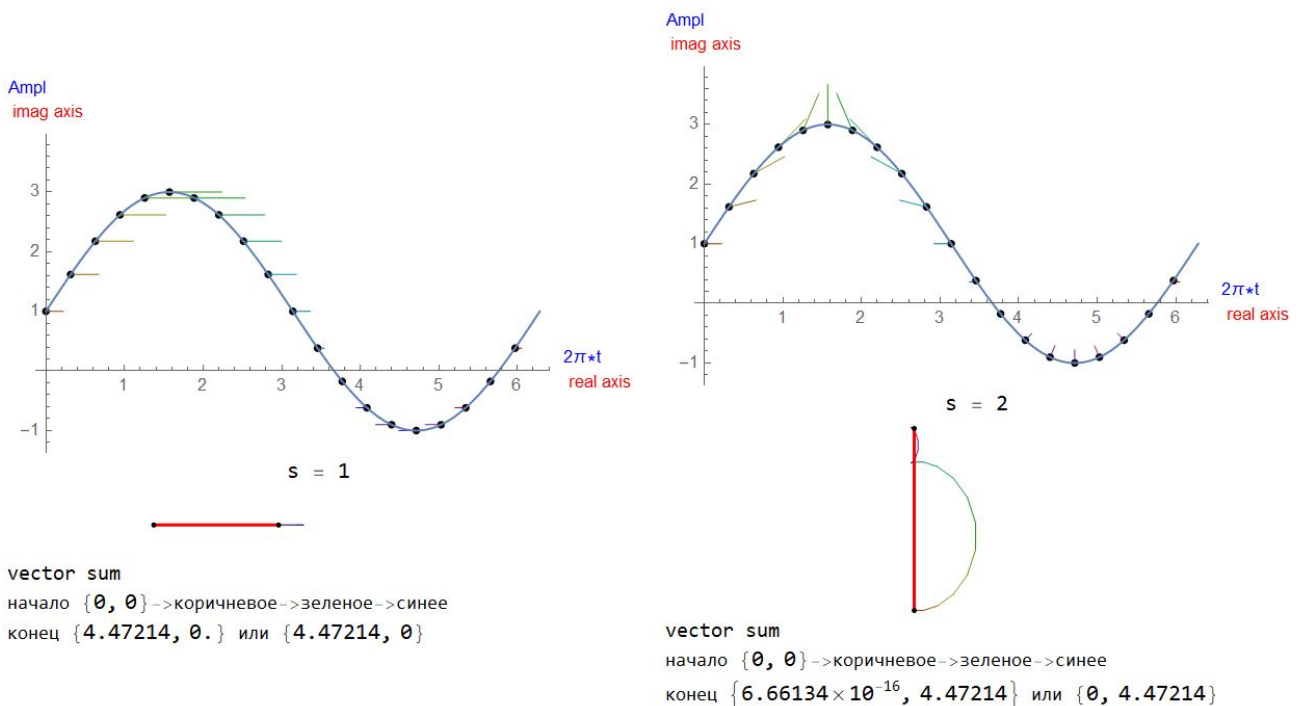
Параметр  $r$  продвигается по оси абсцисс и имеет 20 значений. Параметр  $s$  определяет частоту вращения в  $(s-1)$ Hz.

Далее на рисунках показано вращение вектора-зонда. Каждый вектор-зонд начинается в точке  $u(r)$ , для которой рассчитано значение  $F(s,r)$ . Параметры конца вектора-зонда получаются следующим образом: абсцисса — это произведение  $u(r)*\text{Re}(F(s,r))$ , ордината —  $u(r)*\text{Im}(F(s,r))$ .

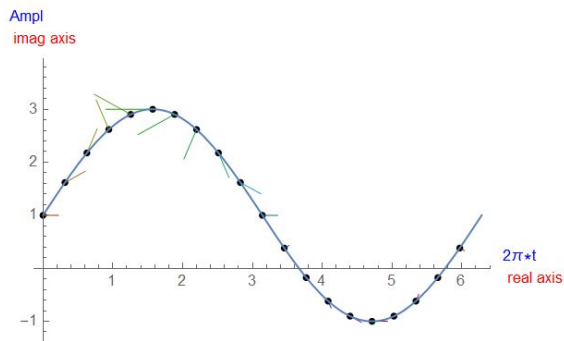
Для наглядности выбрана палитра продвижения векторов-зондов от начала к концу. Начинается с коричневого, далее через зеленый к синему.



На рисунках ниже изображено вращение вектора-зонда, сведенного с точкой на графике, для которой рассчитано Фурье-преобразование, а также путь (vector sum) при непосредственном примыкании соседних векторов-зондов.



Ось ординат отображает амплитуду исходной функции и мнимую часть преобразования Фурье. Ось абсцисс — положение точки на отрезках времени исходной функции и действительную часть преобразования Фурье.



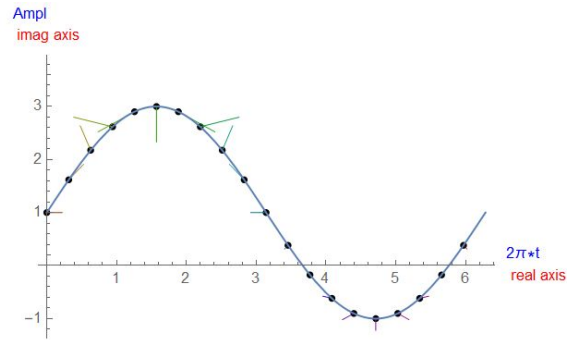
s = 3



vector sum

начало  $\{0, 0\}$  -> коричневое -> зеленое -> синее

конец  $\{6.66134 \times 10^{-16}, 5.55112 \times 10^{-17}\}$  или  $\{0, 0\}$



s = 4

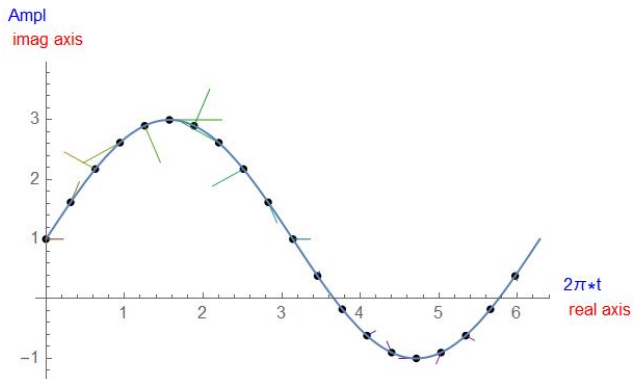


vector sum

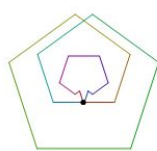
начало  $\{0, 0\}$  -> коричневое -> зеленое -> синее

конец  $\{-4.44089 \times 10^{-16}, -7.77156 \times 10^{-16}\}$  или  $\{0, 0\}$

Сумма векторов показывает конфигурацию движения векторов-зондов. Черная точка обозначает начало движения и конец движения (другая черная точка, если они не совпадают). Для  $s=3$  начало и конец совпадают. Для  $s=1$  и  $s=2$  начало и конец не совпадают.



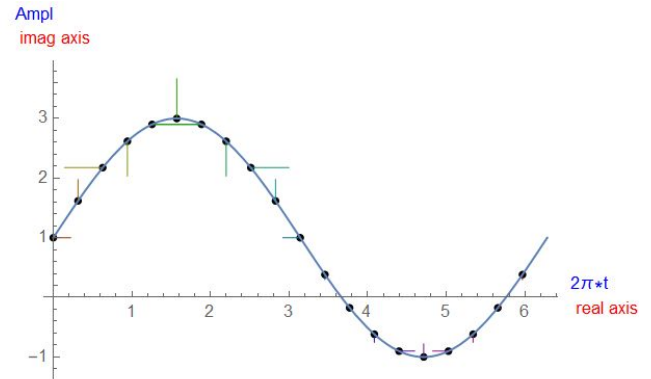
s = 5



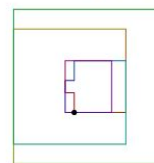
vector sum

начало  $\{0, 0\}$  -> коричневое -> зеленое -> синее

конец  $\{-7.77156 \times 10^{-16}, -2.77556 \times 10^{-17}\}$  или  $\{0, 0\}$



s = 6

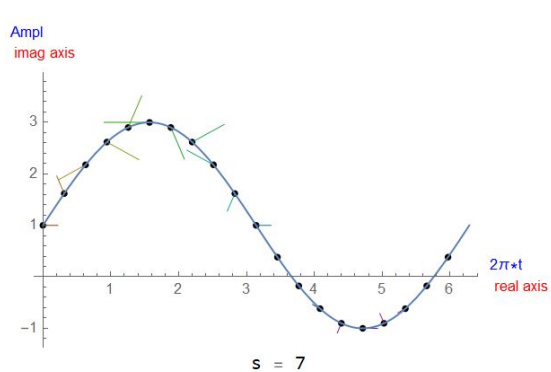


vector sum

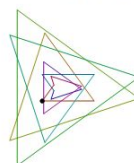
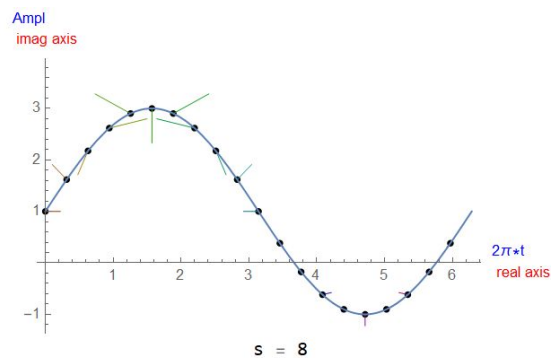
начало  $\{0, 0\}$  -> коричневое -> зеленое -> синее

конец  $\{-3.33067 \times 10^{-16}, 5.55112 \times 10^{-16}\}$  или  $\{0, 0\}$

Отдельно приведены начальные и конечные координаты, а также округленные значения (очень близкие к нулю).

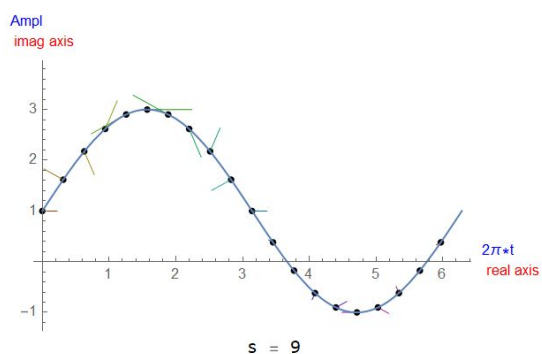


vector sum  
 начало  $\{0, 0\}$  -> коричневое -> зеленое -> синее  
 конец  $\{1.11022 \times 10^{-15}, 0.\}$  или  $\{0, 0\}$

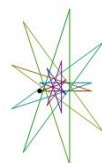
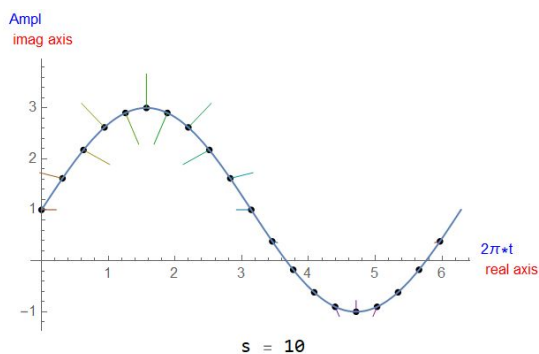


vector sum  
 начало  $\{0, 0\}$  -> коричневое -> зеленое -> синее  
 конец  $\{-9.4369 \times 10^{-16}, -3.33067 \times 10^{-16}\}$  или  $\{0, 0\}$

Значение  $s$  характеризует тестируемую частоту.



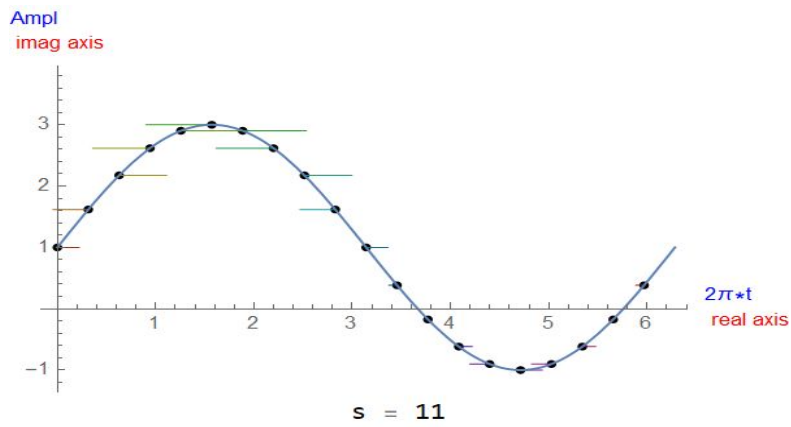
vector sum  
 начало  $\{0, 0\}$  -> коричневое -> зеленое -> синее  
 конец  $\{8.88178 \times 10^{-16}, 1.11022 \times 10^{-16}\}$  или  $\{0, 0\}$



vector sum  
 начало  $\{0, 0\}$  -> коричневое -> зеленое -> синее  
 конец  $\{-2.22045 \times 10^{-16}, 3.60822 \times 10^{-16}\}$  или  $\{0, 0\}$

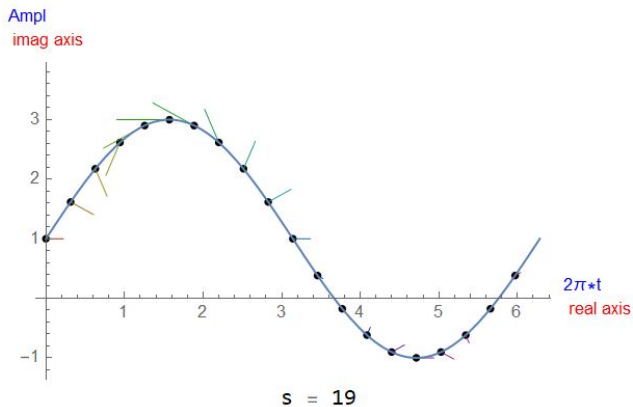
Имеет место симметрия в поведении.  
 Центром является  $s=11$ .



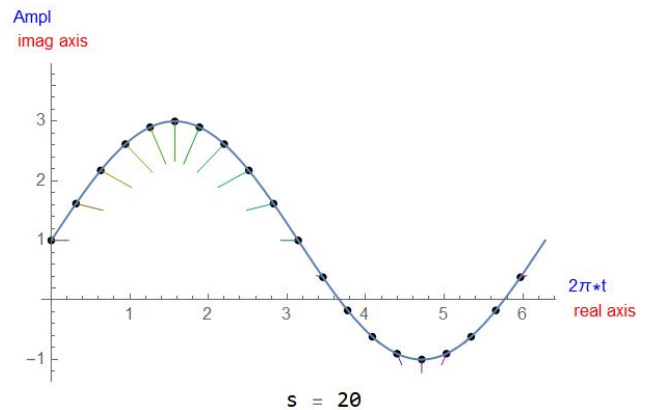


vector sum  
 начало  $\{0, 0\}$  -> коричневое -> зеленое -> синее  
 конец  $\{2.10942 \times 10^{-15}, 0.\}$  или  $\{0, 0\}$

Для примера симметрии приведем рисунки для  $s=19$  и  $s=20$ , которые симметричны  $s=3$  и  $s=2$ .



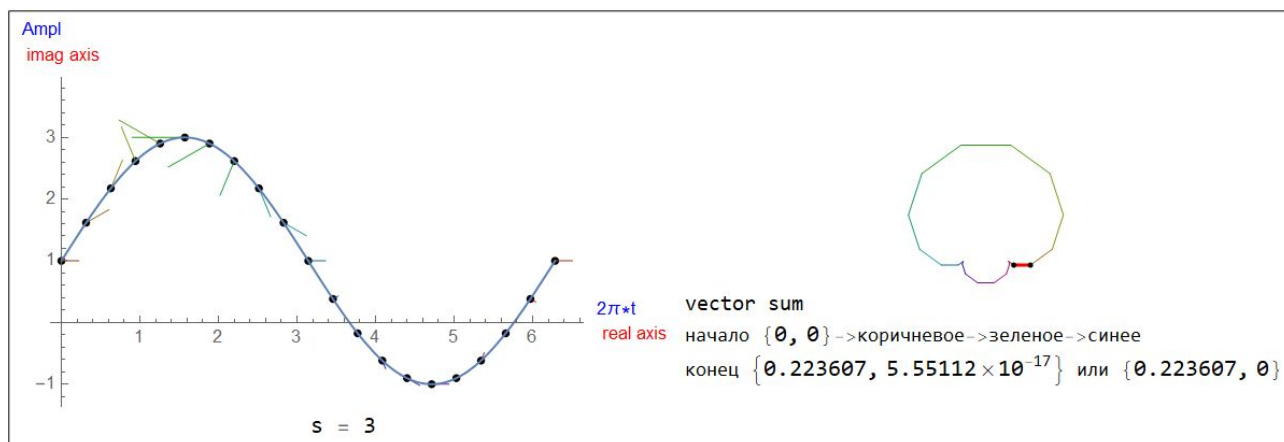
vector sum  
 начало  $\{0, 0\}$  -> коричневое -> зеленое -> синее  
 конец  $\{6.66134 \times 10^{-16}, 5.55112 \times 10^{-17}\}$  или  $\{0, 0\}$



vector sum  
 начало  $\{0, 0\}$  -> коричневое -> зеленое -> синее  
 конец  $\{6.66134 \times 10^{-16}, -4.47214\}$  или  $\{0, -4.47214\}$

Что будет, если взять не 20 точек, а 21, включая последнюю. Пример для  $s=3$ . Он показывает наличие составляющей, колеблющейся с частотой, связанной с  $s=3$ , в

то время как в исходной функции такие колебания отсутствуют. В исходной функции есть колебания только в 1Hz.



Все вышеприведенные графики имеют целью показать важность правильного разбиения интервалов и выборки данных по этим интервалам без последнего значения. Только в этом случае будет иметь место корректная Фурье-аппроксимация исходного ряда и возможность его периодического продолжения.

Остальные аспекты Фурье-аппроксимации достаточно полно представлены в справочной литературе.

#### 4. Анализ реальных временных рядов.

Возвратимся к задаче, которая была описана в самом начале.

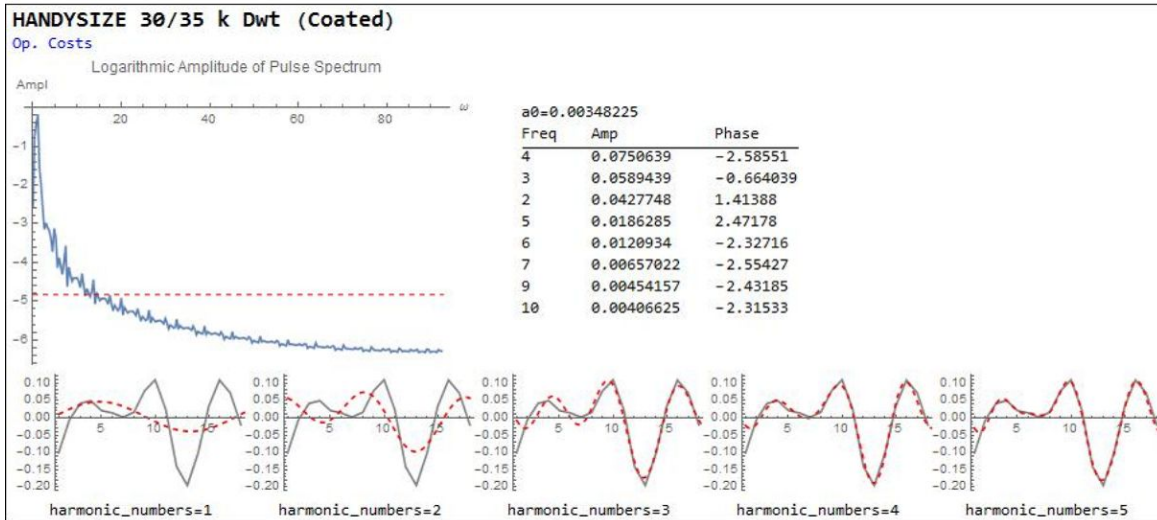
Далее приведена Фурье-аппроксимация исторических данных по ценам на суда для морских перевозок.

На каждой картинке в первом блоке слева приведен график, показывающий на каком уровне значений амплитуд (красный пунктир) отсекаются гармоники, вносящие незначительный вклад. В первом блоке справа приведены характеристики первых 10 гармоник аппроксимирующего ряда по убыванию амплитуды.

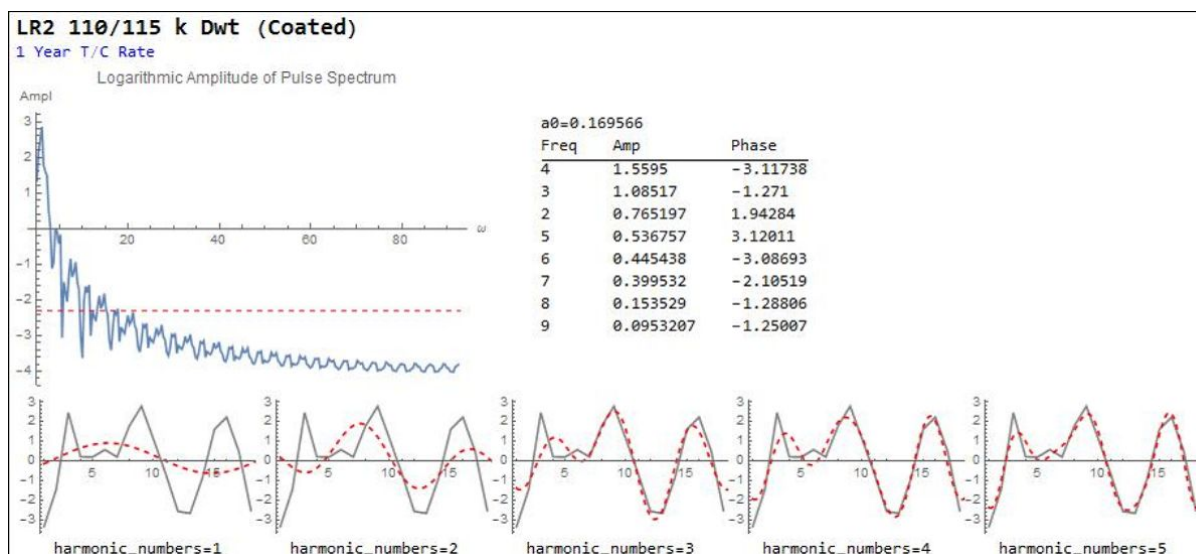
Второй блок состоит из графиков с наращиванием числа гармоник (в порядке от наибольшей амплитуды), используемых для аппроксимации. Итог аппроксимации — красный пунктир.

Для данного временного ряда достаточно 5 гармоник.

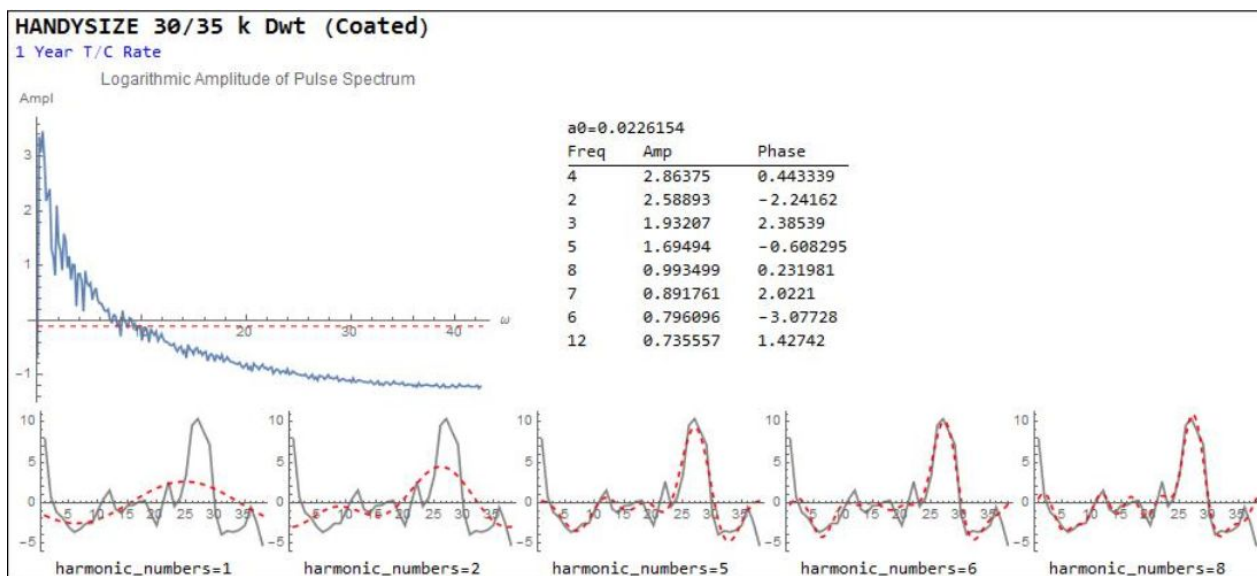




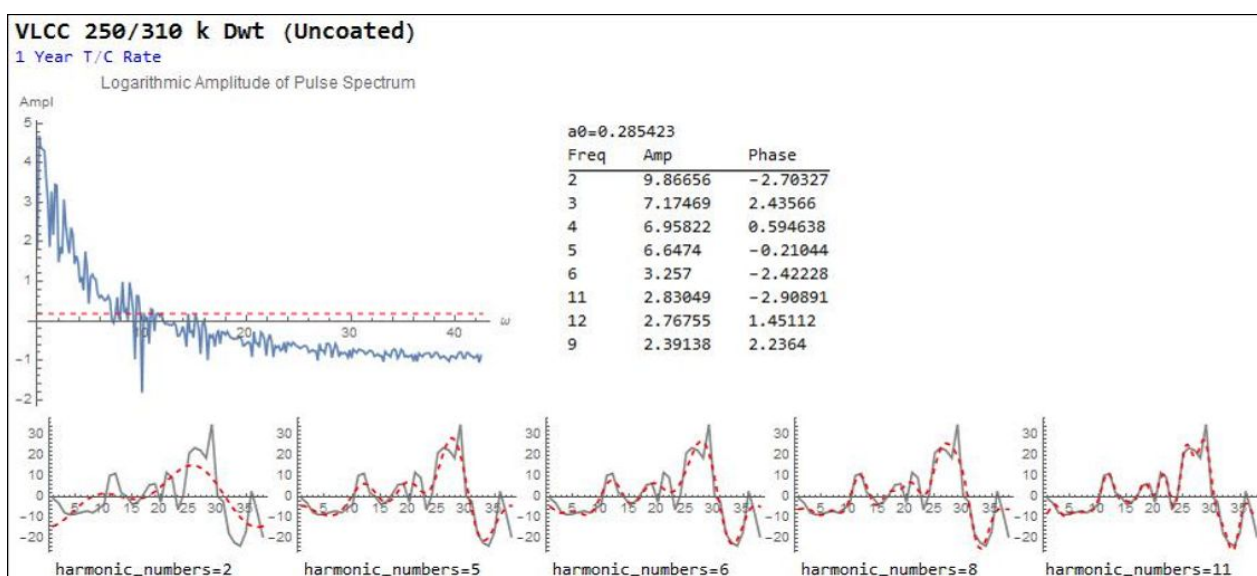
Для данного временного ряда можно ограничиться 5-тью гармониками, если не считать слишком важными совсем старые данные.



Данный временной ряд вполне хорошо аппроксимируется 8-мью гармониками.



В этом случае желательно учитывать 11 гармоник.



Таким образом, исторические данные в достаточно динамичной сфере деятельности (цены на суда для морских перевозок) удастся хорошо аппроксимировать в среднем 10-тью гармониками.

В целом задачу прогноза, когда по фрагменту исторических данных с некоторой погрешностью восстанавливается уже имевшее место будущее, можно считать решенной, если известны основные (некоторые) гармоники аппроксимации.

При этом понятно, что прогноз на будущее, который даст Фурье-аппроксимация, будет по сути совсем неправильный: это становится ясным из-за прозрачного механизма построения Фурье-аппроксимации.

При множественной регрессии, когда говорится о 70% достоверности прогноза, имеет место все то же самое, но непрозрачный механизм построения позволяет необоснованно надеяться, что в целом (на 70%) прогноз будет правильным.

## **5. Фурье-события.**

Фурье-события появляются при допущении, что имеют место базовые циклические процессы (гармоники), которые накладываются и совмещаются с важными событиями, также представленными гармониками.

Таким образом, все гармоники Фурье-аппроксимации разделяются на две части: базовые гармоники процесса и гармоники событий. При этом важно помнить, что в сумме базовые и событийные гармоники дают адекватную аппроксимацию исходного ряда.

В этом случае для хорошего прогноза достаточно знать базовые циклы и иметь перечень событий и обстоятельств, по которым следует формировать скользящий прогноз по ожидаемым или уже произошедшим событиям или их цепочкам. Но это несколько другая, не традиционная технология формирования прогноза.

Методологически оправданы следующие два метода для фиксации Фурье-событий.

Первый метод связан с вычитанием из полного временного ряда всевозможных сочетаний гармоник, которые его аппроксимируют, и сопоставлением известных событий получающимся экстремумам или устойчивым отклонениям. Так как почти во всех отраслях есть аналитические компании, которые собирают статистику и делают обзоры по трендам (в каких-то отраслях даже еженедельно), то поиск важных событий на дату не является достаточно сложной проблемой.

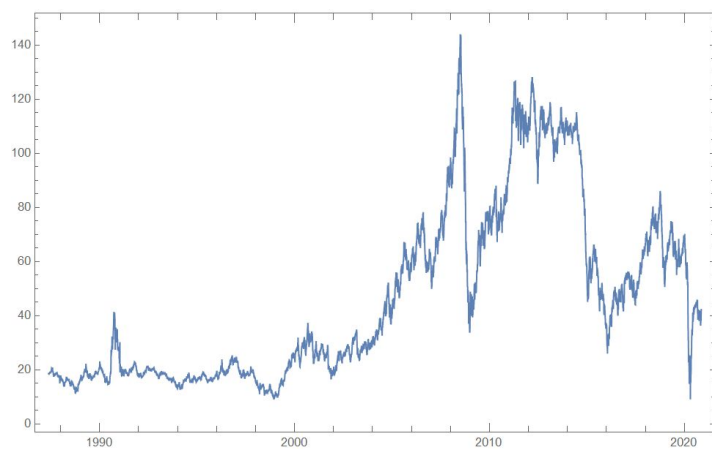
Второй, назовем его методом сплитования, связан с разбиением полного временного ряда на разные по длине периоды и поиском «похожих» периодов по сопоставимым гармоникам. При описанном подходе к Фурье-аппроксимации такая задача может быть полностью автоматизирована.

Метод сплитования качественно отличается от первого метода, так как имеется нелинейная для всего разбиения первоначальная операция выделения своего тренда (линейной регрессии) из каждой составляющей выбранного разбиения.

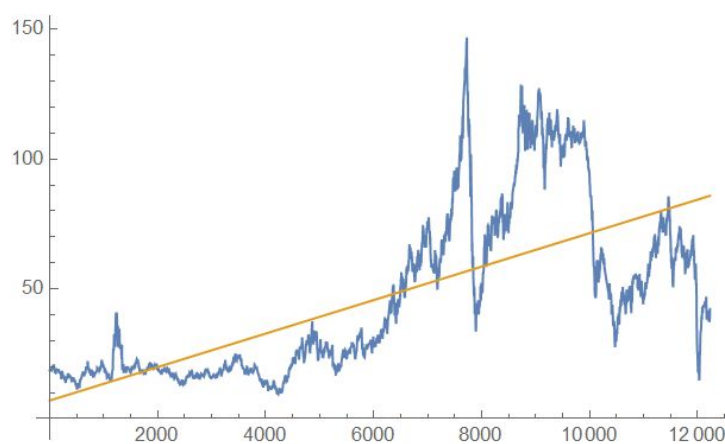
## **6. Аналитика данных для цен на нефть через Фурье-события.**

Для примера рассмотрим цены на нефть Europe Brent Spot Price FOB. Источник: <https://www.eia.gov/dnav/pet/hist/RBRTED.htm>, Thomson Reuters. Данные ежедневные в долларах США с 20 мая 1987 по 10 ноября 2020.

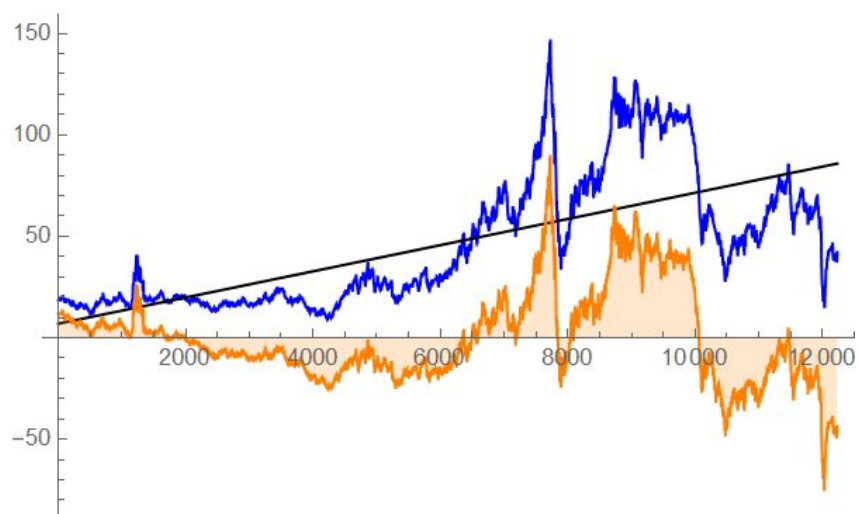
Исходный временной ряд



Выделяем тренд — линейную регрессию.



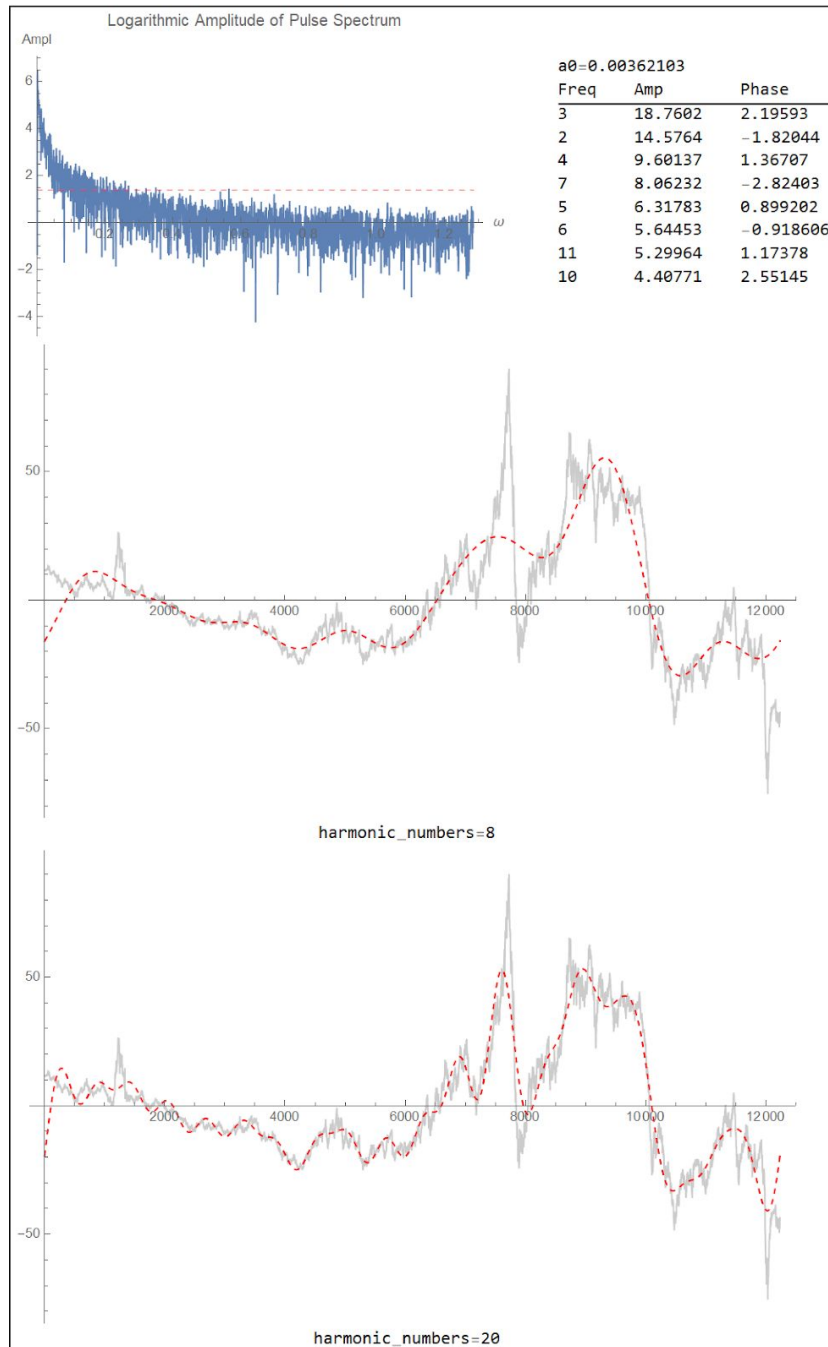
Очищаем исходные данные от тренда (линейный тренд всегда можно восстановить).



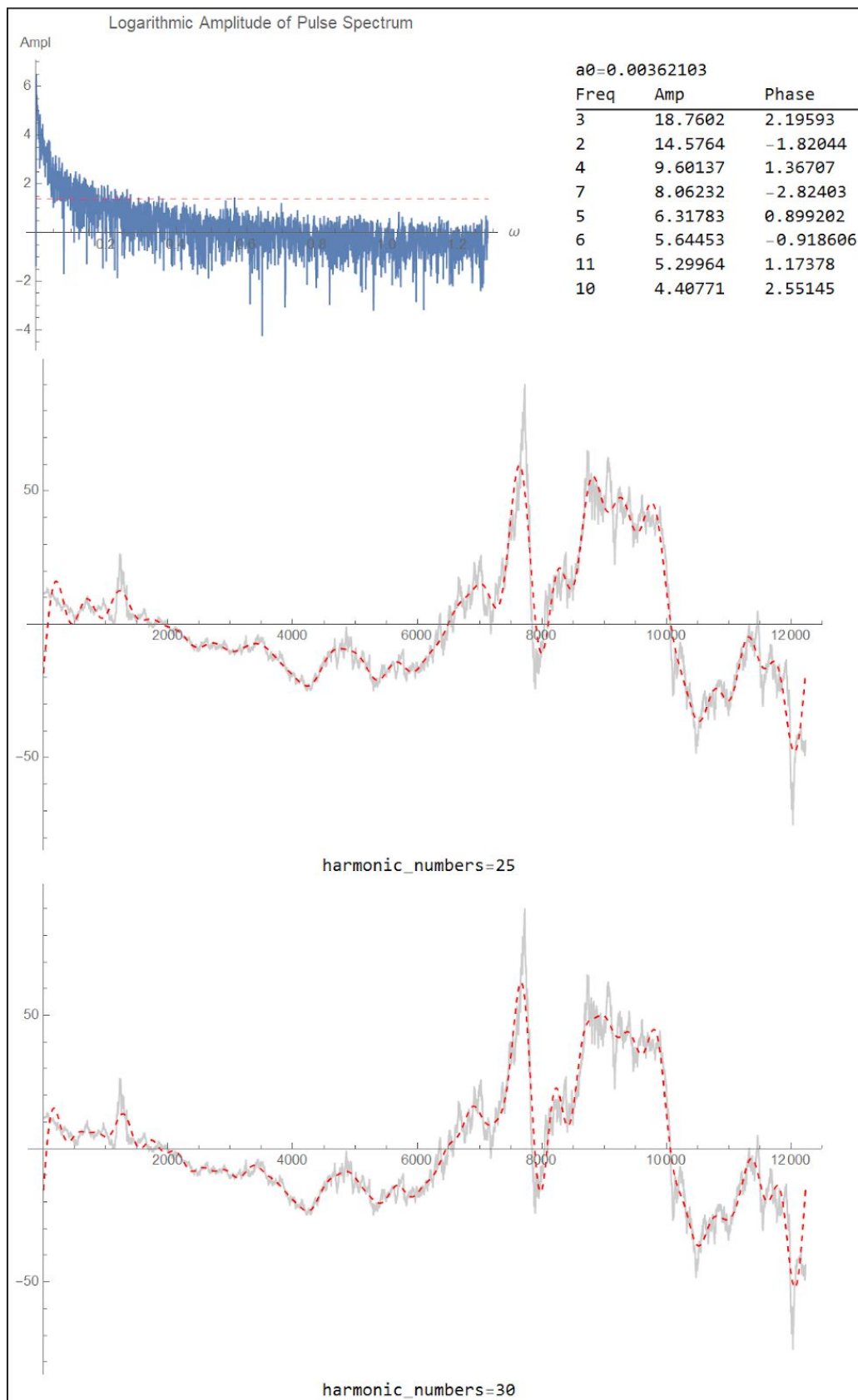
Синий график — исходные данные. Черный — тренд. Оранжевый - нормализованные данные (без тренда).

Находим аппроксимацию.

Пока все не очень хорошо: 8 и 20 гармоник для такого ряда будет мало.



Для 30 гармоник результат вполне приемлемый.



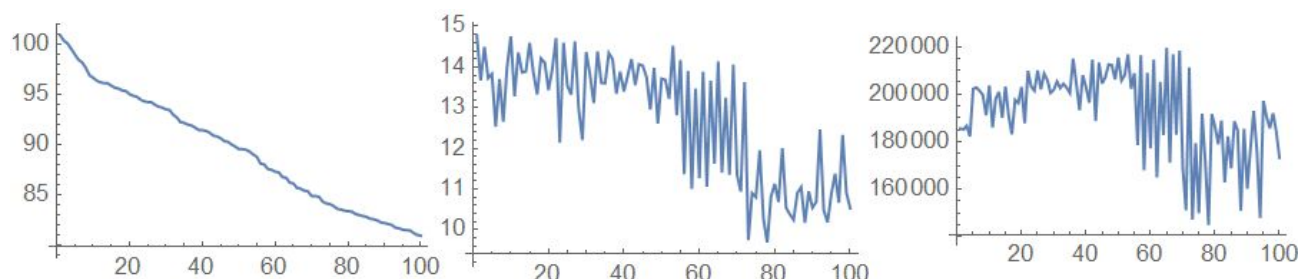
Перейдем к методике выделения Фурье-событий. Проиллюстрируем один из подходов для случая аппроксимации исходного ряда 8-мью гармониками.

Найдем все возможные сочетания гармоник из 8 штук. Их будет 255 штук. Для каждого из 255 сочетаний вычислим абсолютное значение по точечной разнице



между исходным рядом и конструкцией (рядом), порождаемой конкретным сочетанием гармоник.

Для нового ряда вычислим максимум, квадратичное отклонение и общую сумму значений (возможно, что надо вычислять и другие показатели: среднее и т.п.). На рисунках эти показатели изображены последовательно. Они соответствуют первой сотни значений, упорядоченных по убыванию максимума.



Рассмотрим первые 60 из выбранных 100. И далее из них выберем (визуально) интересные. Графики приведены ниже. Номер под картинкой соответствует порядковому номеру сочетания из 255. Серый график — исходный ряд, красный ряд - из сочетания гармоник.

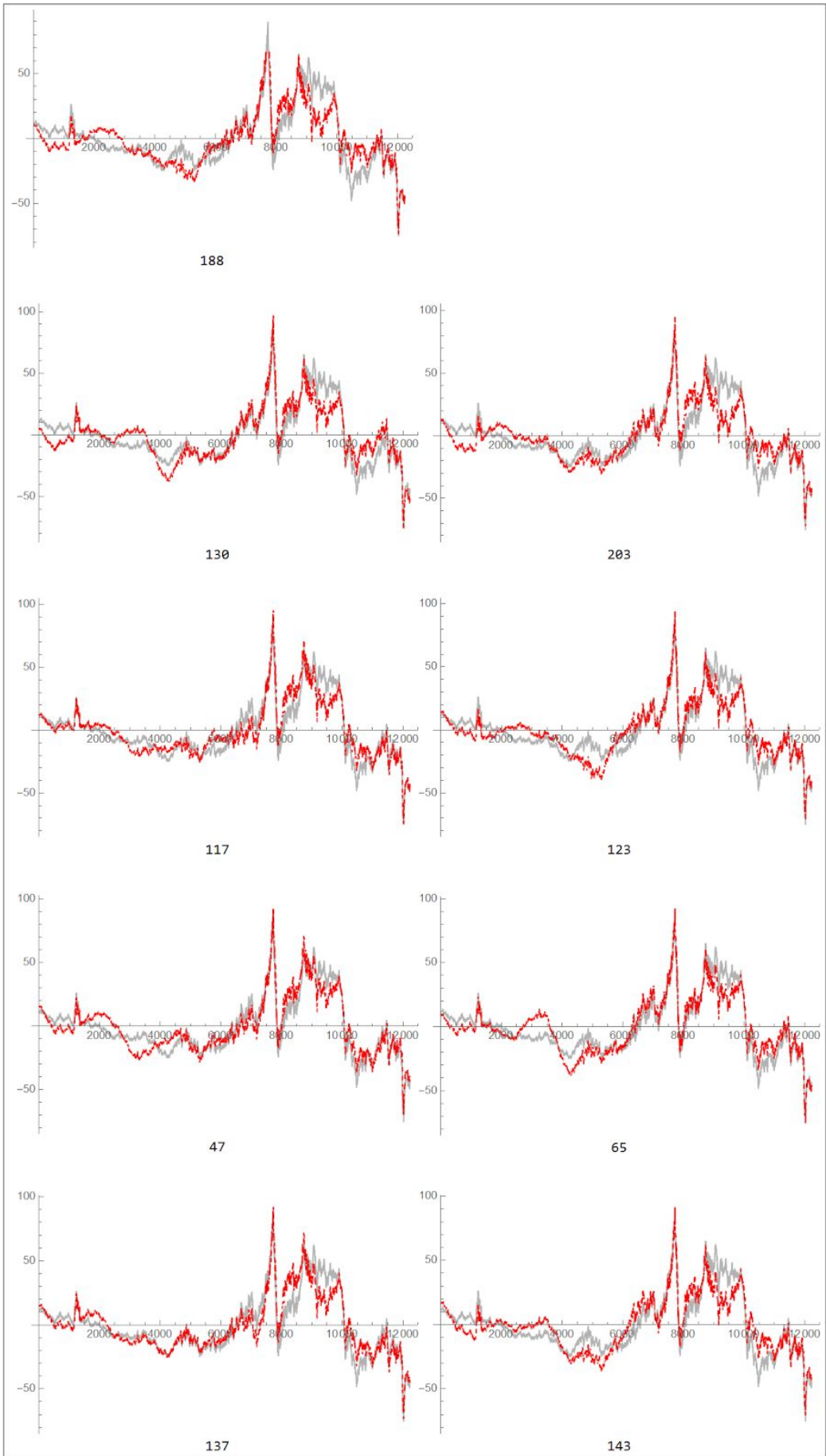
Что считать «интересным» является как раз содержательной задачей для бизнеса. Все, что было до сих пор — просто стандартная техника.

Что в итоге получилось? Из набора гармоник, которые хорошо аппроксимируют исходный ряд, выбраны сочетания, которые на каких-то участках очень хорошо соответствуют исходному временному ряду, а на других показывают явное несоответствие. Как раз последние участки являются кандидатами на анализ событий, имевших место в этот период (все графики дневные с явным соответствием дате).

Кроме того, наличие участков очень хорошо прилегающих графиков дает основание для выведения характеристик «нормы» для динамики отражаемых процессов.

Целевой установкой анализа является выявление гармоник, которые соответствуют событиям. Обратная задача — это выделение базовых циклических процессов.

Метод сплитования важен потому, что процесс, представленный одним временным рядом, может быть по сути составным и зависеть от событий более высокого порядка: глобальный кризис и т.п.



## **7. Фурье-события при анализе cash-flow.**

Процесс анализа временного ряда связан с ожиданием того, что в нем доминируют циклические процессы. В общем случае такие ожидания могут быть не оправданы. Дело даже не в том, что такого доминирования циклическости нет. Просто из-за конкретного способа формирования временного ряда может быть невозможным выявить циклическость в данном конкретном ряду.

Другое дело cash-flow. По сути производственно-коммерческой деятельности большинство процессов изначально циклические по способу формирования. Отклонения от нормы связаны с событиями, которые нарушают эту циклическость. Использование метода Фурье-событий при анализе cash-flow позволяет выявить объективную «норму», а также индикаторы отклонений.

В терминах Фурье-событий задача анализа cash-flow является хорошо алгоритмизируемой для применения методов искусственного интеллекта и нейросетей (машинного обучения).