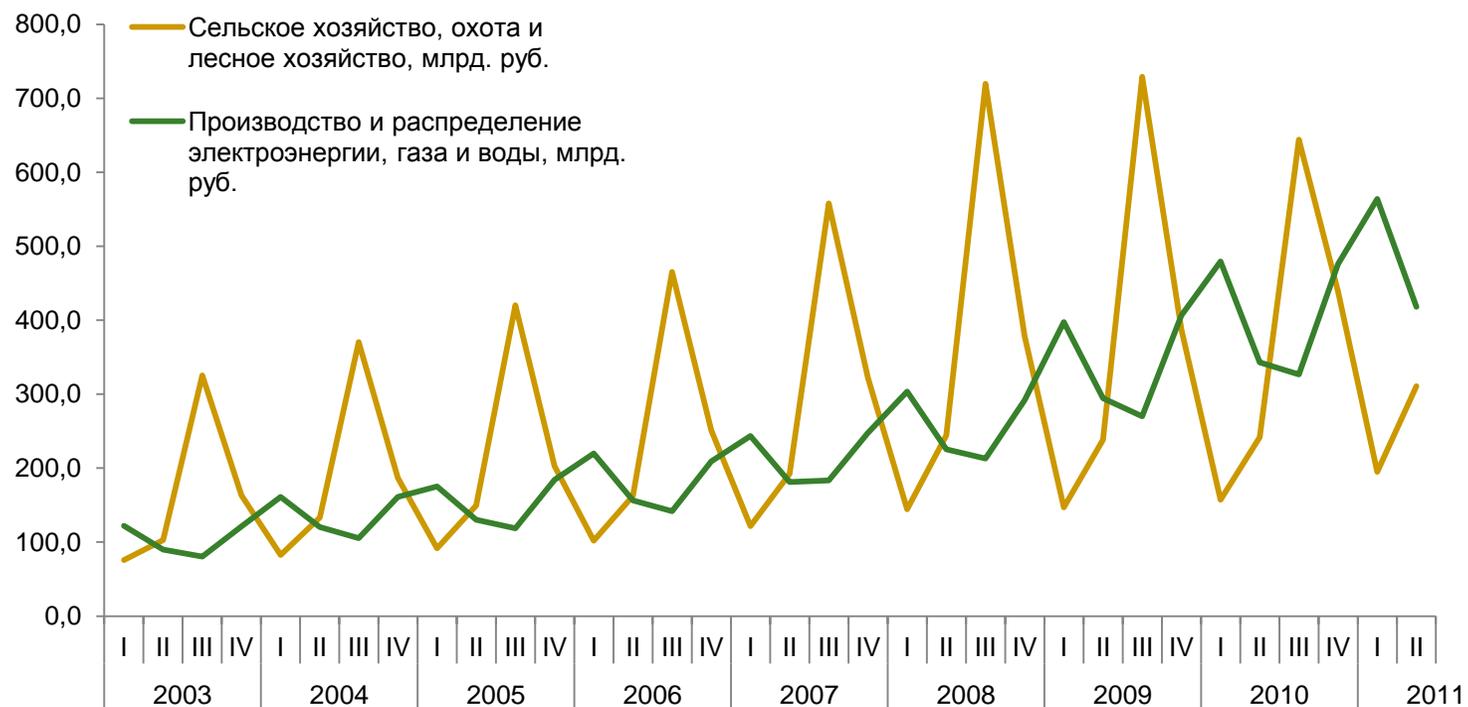


«Сгладить нельзя не сгладить!»  
или пара слов о том, как строить свои  
отношения с сезонностью в данных

# Сезонность

- Регулярно повторяющиеся колебания
- Проявляется в сфере производства

- Основные причины:
  - Климатические факторы
  - Институциональные факторы
  - Ожидания сезонности



# Зачем нужно учитывать?

- Точность прогнозирования
- Дополнительная дисперсия
- Смещение в оценках коэффициентов при исследовании взаимосвязи между несколькими рядами (кажущиеся регрессии)
- Невозможность наблюдать изломы трендов и более длинных циклов

# SARIMA-модели

- Коротко:

$$\text{ARIMA } \underbrace{(p, d, q)}_{\substack{\uparrow \\ \text{(Non-seasonal part} \\ \text{of the model)}}} \underbrace{(P, D, Q)_m}_{\substack{\uparrow \\ \text{(Seasonal part} \\ \text{of the model)}}$$

- Полностью:

$$\underbrace{(1 - \phi_1 B)}_{\substack{\uparrow \\ \text{(Non-seasonal} \\ \text{AR(1)}}}} \underbrace{(1 - \Phi_1 B^4)}_{\substack{\uparrow \\ \text{(Seasonal} \\ \text{AR(1)}}}} \underbrace{(1 - B)}_{\substack{\uparrow \\ \text{(Non-seasonal} \\ \text{difference)}}}} \underbrace{(1 - B^4)}_{\substack{\uparrow \\ \text{(Seasonal} \\ \text{difference)}}}} y_t = \underbrace{(1 + \theta_1 B)}_{\substack{\uparrow \\ \text{(Non-seasonal} \\ \text{MA(1)}}}} \underbrace{(1 + \Theta_1 B^4)}_{\substack{\uparrow \\ \text{(Seasonal} \\ \text{MA(1))}}}} e_t.$$

## О чём вообще выделение сезонности

- Модель сезонности: либо аддитивная:

$$Y_t = U_t + S_t + E_t$$

- Либо мультипликативная

$$Y_t = U_t S_t E_t$$

- Как-то (как – зависит от метода) оценивается сезонная компонента и удаляется из данных

# Методы семейства X11

- US Bureau of Census
- Разрабатываются с 1950-ых годов

Процедура сглаживания итерационная:

1 этап: Сглаживание исходных данных с помощью скользящей средней (13 точек). Выделение первой оценки тренда; очистка данных от тренда:

$$\frac{Y_t}{\hat{U}_t} = \frac{U_t * S_t * E_t}{\hat{U}_t} \approx S_t * E_t$$

2 этап: сглаживание (скользящая средняя 5 или 7 точек) полученного ряда. Получаем оценку сезонности

3 этап: Удаление первой оценки сезонности из данных:

$$\frac{Y_t}{\hat{S}_t} = \frac{U_t * S_t * E_t}{\hat{S}_t} \approx U_t * E_t$$

4 этап: Новая оценка тренда (скользящая средняя: 9, 13 или 23 точки); данные очищаются от нового тренда

5 этап: итоговая оценка сезонности (аналогично этапу 2)

6 этап: очистка данных от сезонности (итоговой)

7 этап: выделение итогового тренда. Получаем данные, очищенные от сезонного компонента

Выбор окна осуществляется на основе соотношения  $\frac{\bar{I}}{\bar{C}}$  ratio

$$\bar{C} = \frac{1}{n-1} \sum_{t=2}^n |C_t - C_{t-1}|$$

$$\bar{I} = \frac{1}{n-1} \sum_{t=2}^n |I_t - I_{t-1}|$$

Если оно меньше 1, то применяется 9-точечная скользящая средняя

От 1 до 3,5 – 13-точечная

Больше 3,5 – 23-точечная

# Технические нюансы

- Использует скользящие средние Хендерсона (1916):
  - Симметричные:  
(-0.01935, -0.02786, 0, 0.06549, 0.14736, 0.21434, 0.24006, 0.21434, 0.14736, 0.06549, 0, -0.02786, -0.01935)
  - Либо (на краях ряда) асимметричные:  
(-0.09186,-0.05811,0.01202,0.11977,0.24390,0.35315,0.42113)

# Технические нюансы

- Использует скользящие средние Хендерсона (1916):
  - Симметричные:  
(-0.01935, -0.02786, 0, 0.06549, 0.14736, 0.21434, 0.24006, 0.21434, 0.14736, 0.06549, 0, -0.02786, -0.01935)
  - Либо (на краях ряда) асимметричные:  
(-0.09186,-0.05811,0.01202,0.11977,0.24390,0.35315,0.42113)
- Смысл их в том, что пропущенный через такой фильтр кубический полином останется неизменным!

# Технические нюансы

- Считается, что кубические полиномы учитывают типичные для экономических временных рядов формы циклов и трендов
- Устраняют практически все высокочастотные (для месячных данных: с частотой выше, чем полугодовые) шумы
- Позволяет учитывать торговые дни (разное количество рабочих дней в разные месяцы), эффекты начала года и прочие календарные эффекты

# X11-ARIMA

- Одна из значимых проблем исходного метода – разное сглаживание середины и концов ряда (на концах приходится использовать несимметричные фильтры). Для решения этой проблемы был разработан метод X11-ARIMA (Statistics Canada, 1980)
- Метод оценивает (автоматически) вспомогательную ARIMA модель, при помощи которой ряд дополняется с краёв, что позволяет использовать обычные симметричные фильтры X11
- При этом данные сначала очищаются от календарных эффектов (при необходимости – при помощи X11), затем дополняются ARIMA-моделью (одной из пяти, выбираемой на основе критерия MAPE) и затем опять сглаживаются

**Table 21.1.** Five Predefined Models

<b>Model #</b>	<b>Specification</b>	<b>Multiplicative</b>	<b>Additive</b>
1	(0,1,1)(0,1,1)s	log transform	no transform
2	(0,1,2)(0,1,1)s	log transform	no transform
3	(2,1,0)(0,1,1)s	log transform	no transform
4	(0,2,2)(0,1,1)s	log transform	no transform
5	(2,1,2)(0,1,1)s	no transform	no transform

# X12-ARIMA и regARIMA

- В 1990 году был представлен метод X12-ARIMA, более активно использующий ARIMA модели (в частности, regARIMA).
- regARIMA – модель вида

$$\log\left(\frac{Y_t}{D_t}\right) = \beta' X_t + Z_t$$

- $D_t$  – «субъективные априорные» преобразования (если нужны)
- $X_t$  – регрессоры на торговые дни, праздники и т.д.
- $Z_t$  – ARIMA процесс
- Позволяет не только дополнять данные с концов, но и находить выбросы, заполнять пропуски в данных, явно оценивать календарные эффекты
- В более поздних версиях программы выбор модели либо осуществляется из списка (как в X11-ARIMA), либо начинается с самой простой с постепенным усложнением

# Достоинства и недостатки семейства X11

- Непараметрические (следовательно, более гибкие)
- Много дополнительных настроек (у современных версий метода можно руками добавить большое количество календарных эффектов, отдельных событий и прочих вещей, нужных для повышения точности)
- С версии X12-ARIMA – большой набор средств для диагностики сезонности (спектральный анализ, средства для анализа стабильности и качества сезонной корректировки и т.д.)
- [Считается, что лучше работает с короткими (до 5-6 лет) временными рядами]
- Под вопросом остаётся качество работы фильтра X11 (который остаётся в основе всех вариаций метода)
- Другие проблемы (ложная сезонность вокруг точек разрыва, смещения в статистических тестах и др. – подробности ниже)

# TRAMO/SEATS

- Bank of Spain (1996)
- Основывает процедуру сглаживания сезонности на построении ARIMA модели
- Состоит из двух программ:
  - TRAMO – Time Series Regression with ARIMA Noise, Missing Observations and Outliers
  - SEATS – Signal Extraction in ARIMA Times Series

# TRAMO/SEATS

- TRAMO аналогичен regARIMA
- SEATS подбирает оптимальный фильтр в зависимости от данных
- Зная модель для исходного ряда, рассматривая аддитивную сезонность и делая ряд дополнительных предположений получает в явном виде модели для сезонной и трендовой компоненты

# Предпосылки TRAMO/SEATS

- 4 компоненты ряда:
  - Trend-Cycle (низкочастотная компонента)
  - Seasonal (сезонные частоты)
  - Irregular (белый шум)
  - Transitory (всё остальное)
- Все эти компоненты ортогональны
- Все описываются ARIMA моделями
- В сумме дают модель, описывающую поведение ряда в целом (из TRAMO)
- Все компоненты, кроме нерегулярного, не содержат в себе белого шума.

# Встроенные модели в TRAMO/SEATS

Name	Settings
RSA0	Level, Airline model
RSA1	Log/level, outliers detection, Airline model
RSA2	Log/level, working days, Easter, outliers detection, Airline model
RSA3	Log/level, outliers detection, automatic model identification
RSA4	Log/level, working days, Easter, outliers detection, automatic model identification
RSA5	Log/level, trading days, Easter, outliers detection, automatic model identification


$$\Delta\Delta_s x_t = (1 + \theta_1 B)(1 + \theta_s B^s) a_t,$$

# Достоинства и недостатки TRAMO/SEATS

- Параметрический, форма сезонности зависит от данных и является «оптимальной» по каким-то критериям
- Можно очень гибко варьировать модели и параметры
- На практике может давать более гладкие результаты, которые меньше меняются при добавлении новых точек, чем в случае X11
- Меньше диагностики
- Меньше возможностей с точки зрения добавления пользователем дополнительных календарных эффектов и переменных, отвечающих за отдельные события (скорее особенность пакета, чем устройства процедуры)

# X13-ARIMA-SEATS

- Новый метод
- В теории – объединяет достоинства семейства X11 (особенно дополнительные возможности пакета с точки зрения диагностики) и TRAMO/SEATS (форма сезонности определяется данными). При этом есть возможность вместо SEATS использовать обычный фильтр X11
- В рамках одного пакета позволяет применять и сравнивать разные подходы

Сравнение семейства X-11 и TRAMO/SEATS

Hood C. C., Ashley J. D., Findley D. F. An empirical evaluation of the performance of TRAMO/SEATS on simulated series //Proceedings of the Business and Economics Section. – 2000.

- Генерировали ряды с разными формами тренда, сезонности и разным масштабом шума

**Table 2. Results for Default Adjustments**

	12 years	4 years
<b>All 54 Series</b>		
SEATS closer	25 series	11 series
X-12 closer	25 series	39 series
No preference	4 series	4 series
Total	54 series	54 series
<b>Series with Large Irregular (18 series)</b>		
SEATS closer	16 series	0 series
X-12 closer	2 series	17 series
No preference	0 series	1 series
Total	18 series	18 series

**Table 3. Overall Average RRMSQD and RMAD for 12 Years of Data for All 54 Series**

	12 years	4 years
<b>RRMSQD</b>		
SEATS	0.023	0.037
X-12	0.025	0.032
<b>RMAD</b>		
SEATS	0.018	0.030
X-12	0.020	0.027

- Помимо этого, сравнили X-12-ARIMA с X-11-ARIMA. X-12-ARIMA оказался лучше во всех ситуациях

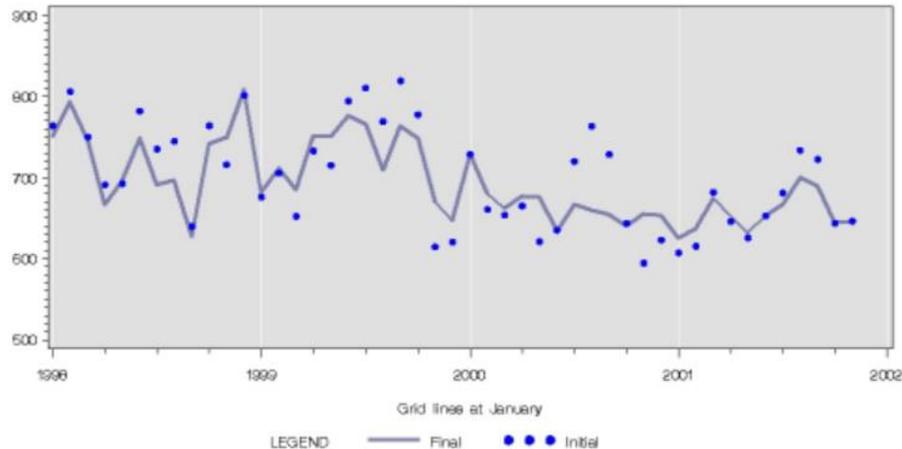
Hood C. C. Comparison of time series characteristics for seasonal adjustments from seats and x-12-arima //ASA proceedings, business and economic statistics section, Alexandria, VA: ASA. – 2002

- Показывается, что для TRAMO/SEATS оценка скорректированного ряда может значительно меняться при добавлении новых данных

**Figure 3. Seasonal Adjustment Revision Graph (Final and Initial) from a Default SEATS Adjustment**

Seasonal Adjustment Values for SEATS

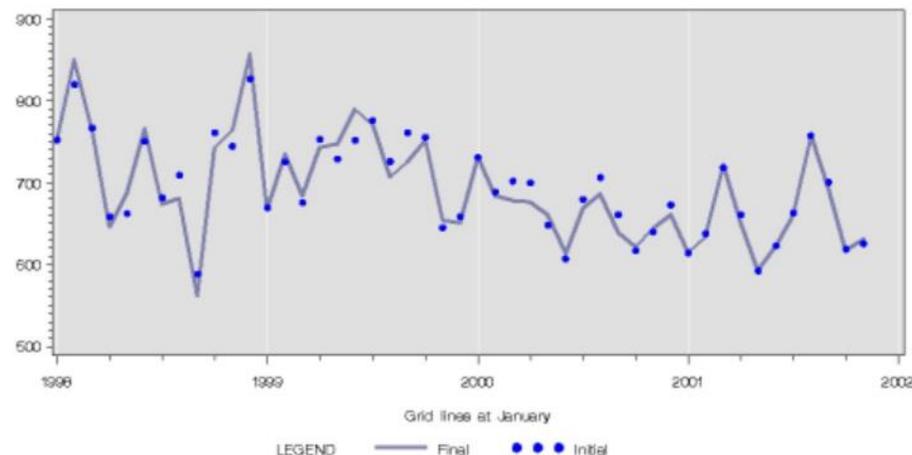
X00gm



**Figure 5. Seasonal Adjustment Revision Graph (Final and Initial) from a Default X-11/X-12-type Adjustment**

Seasonal Adjustment Values for X-12

X00gm



- Помимо этого, SEATS иногда порождает ложную сезонность (находит сезонность там, где её нет)

Wildi M., Schips B. Signal extraction: How (in) efficient are model-based approaches? An empirical study based on TRAMO/SEATS and Census X-12-ARIMA. – Arbeitspapiere//Konjunkturforschungsstelle, Eidgenössische Technische Hochschule Zürich, 2004. – №. 96.

- И TRAMO/SEATS, и X-12-ARIMA в большинстве случаев неверно определяют порядок интеграции ряда в случае, если исходный ряд  $I(0)$
- X-12-ARIMA чаще определяет ряды как  $I(2)$
- X-12-ARIMA и TRAMO/SEATS дают сопоставимые по качеству результаты

# Что применяет Росстат

	Календарь	Период	База
ВВП	Количество рабочих дней и високосные годы	?	Условный средний квартал 2008 г.
Промышленное производство	Праздники, торговые дни (рабочие и субботние) и високосные годы	С 1999 г.	Среднемесячное значение 2010 г.
Сельское и лесное хозяйство	Нет	С 2007 г.	Среднемесячное значение 2010 г.
Строительство	Нет	С 2003 г.	Среднемесячное значение 2010 г.
Транспорт и связь	Праздники, торговые дни (рабочие и субботние) и високосные годы	Вообще с 1992 г., но «При построении текущей модели для наилучшего выполнения сглаживания динамического ряда используются данные с 2003 года»	Среднемесячное значение 2010 г.
Розничная торговля	Праздники, торговые дни (рабочие и субботние) и високосные годы	Вообще с 1992 г., но «При построении текущей модели для наилучшего выполнения сглаживания динамического ряда используются данные с 2008 года»	Среднемесячное значение 2010 г.
Рынок платных услуг населению	Праздники, торговые дни (рабочие и субботние) и високосные годы	Вообще с 2000 г., но «При построении текущей модели для наилучшего выполнения сглаживания динамического ряда используются данные с 2006 года»	Среднемесячное значение 2010 г.
Оптовая торговля	Праздники, торговые дни (рабочие и субботние) и високосные годы	С 2008 г.	Среднемесячное значение 2010 г.
Инвестиции в основной капитал	Нет	С 2004 г.	Среднемесячное значение 2010 г.
Реальные располагаемые денежные доходы	Праздники	С 1992 г.	Среднемесячное значение 2010 г.
Безработица	Нет	С 2011 г.	-

Почему иногда (/часто/всегда) лучше не удалять сезонность из данных?

## Смещение в тестах на единичные корни

- Ghysels, Perron (1993)\*
- Показали (аналитически), что в случае наличия единичного корня в данных асимптотические распределения тестов на единичные корни не меняются
- В случае отсутствия единичного корня, есть смещение в сторону неотвержения гипотезы о наличии единичного корня

\*Ghysels E., Perron P. (1993), The effect of seasonal adjustment filters on tests for a unit root, *Journal of Econometrics*

# Asymptotic bias

- Для процессов без единичного корня:

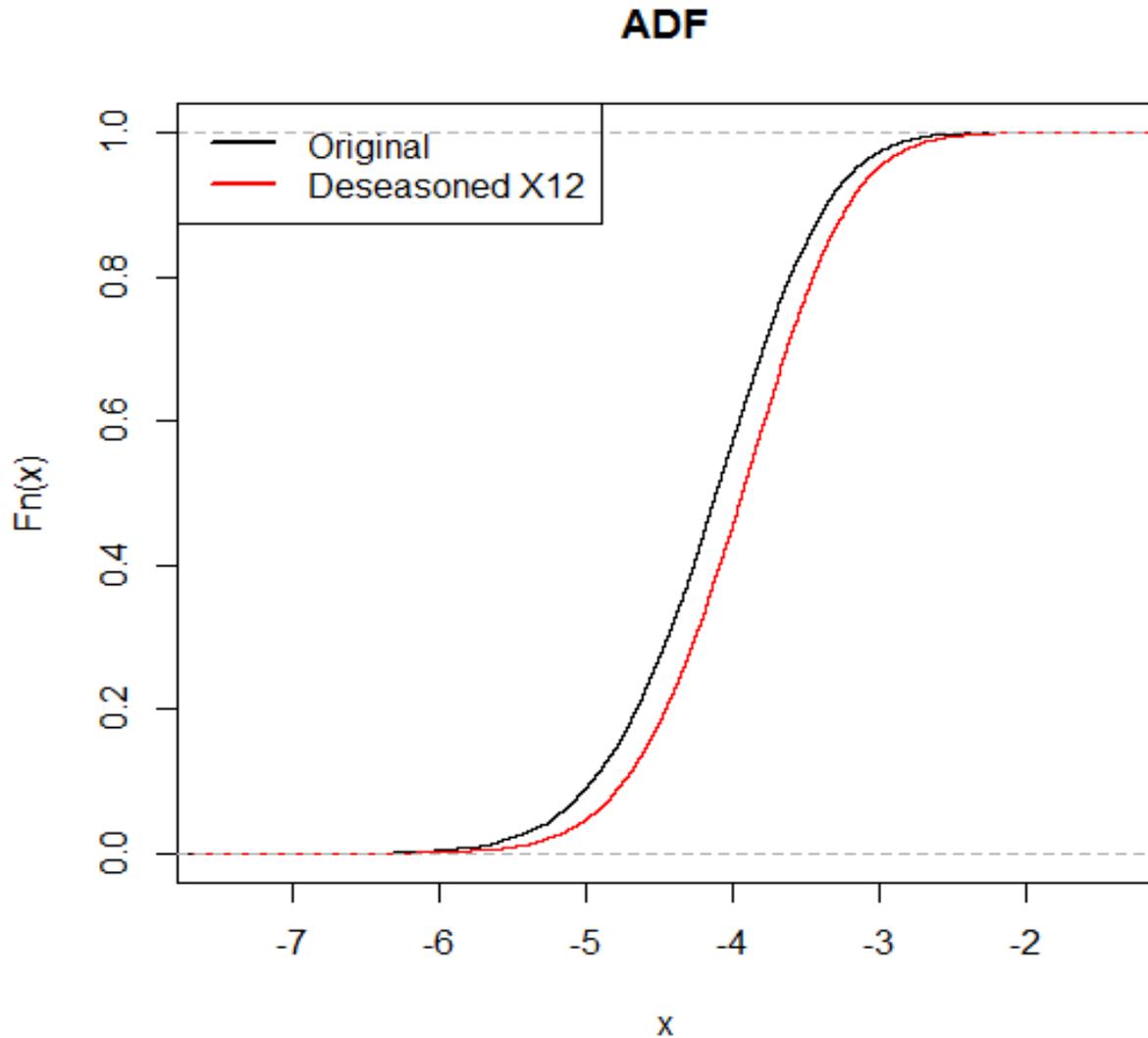
*Proposition 1. Let  $\{y_t\}$  be generated by (4.4). Let  $\hat{\theta}$  be the OLS estimator of  $\theta$  obtained from regression (4.12) and  $\alpha^* = e' \hat{\theta}$ . Similarly let  $\{y_t^F\}$  be generated by  $y_t^F = \psi(L)y_t$  and  $\alpha_F^* = e' \hat{\theta}_F$ , where  $\hat{\theta}_F$  is the vector of least-squares estimators of  $\theta$  in (4.12) when the filtered data,  $y_t^F$ , are used. Then*

$$b(\alpha_F^*, k) \equiv \text{plim}_{T \rightarrow \infty} \alpha_F^* - \text{plim}_{T \rightarrow \infty} \alpha^* = e' A^{-1} V - e' A_F^{-1} V_F, \quad (4.13)$$

*where  $A$  is a  $p \times p$  matrix with elements  $a_{ij} = \gamma_y(i - j)$  ( $i, j = 1, \dots, p$ ) and  $V$  is a  $p \times 1$  vector with elements  $v_i = \gamma_y(i)$  ( $i = 1, \dots, p$ );  $A_F$  and  $V_F$  are defined similarly with elements  $a_{ij}^F = \gamma_y^F(i - j)$  and  $v_i^F = \gamma_y^F(i)$ , with  $\gamma_y^F(s)$  defined by (4.10).*

- Для процессов с единичным корнем  $\text{plim } \alpha = \text{plim } \alpha_F = 1$ , сдвига нет

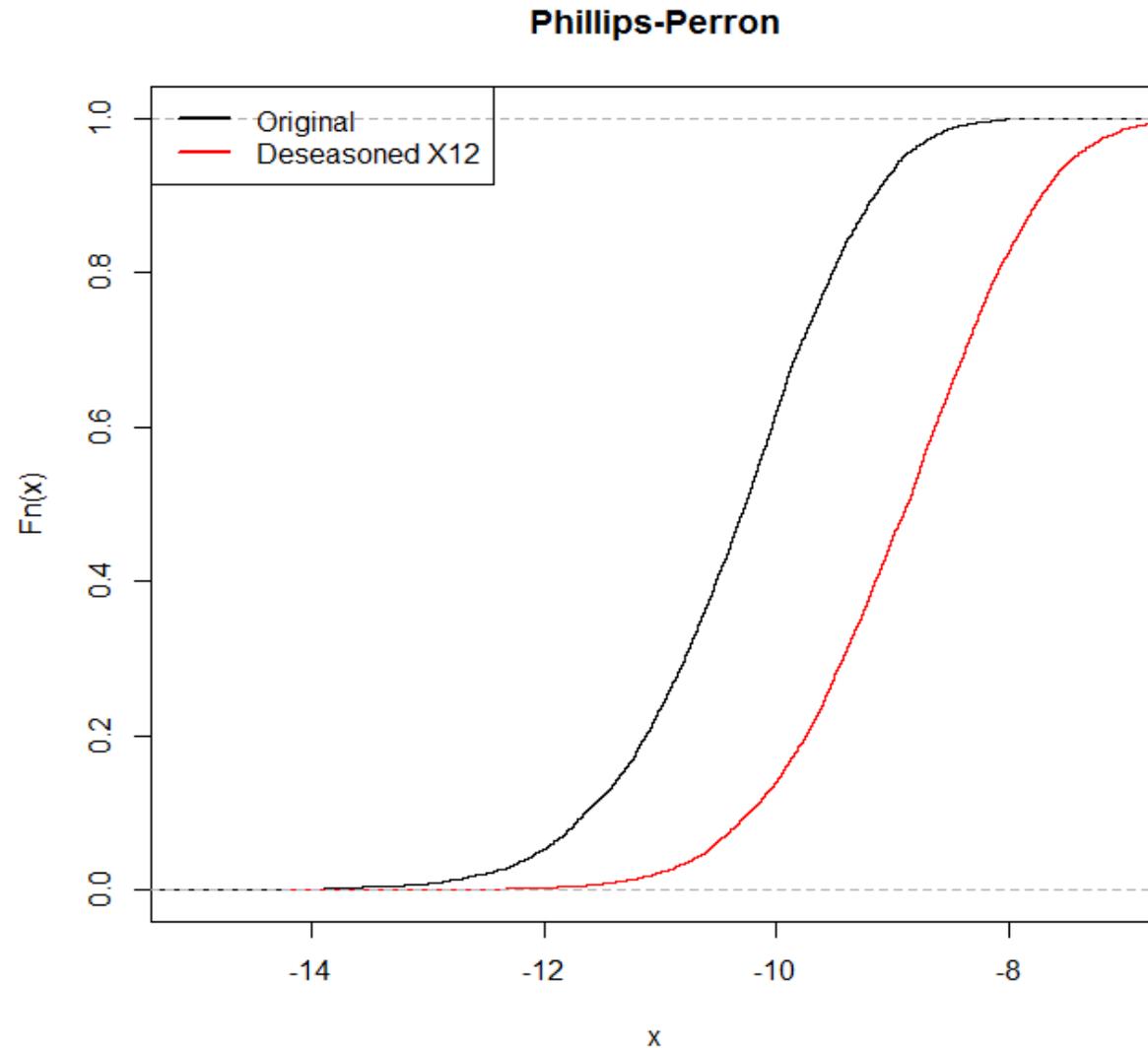
# Монте-Карло, истинный ряд - TS



Доля неверно  
определённых рядов

Значимость	Исходный ряд	Сглаженный ряд
10%	0,0488	0,0831
5%	0,1354	0,2018
1%	0,4593	0,5754

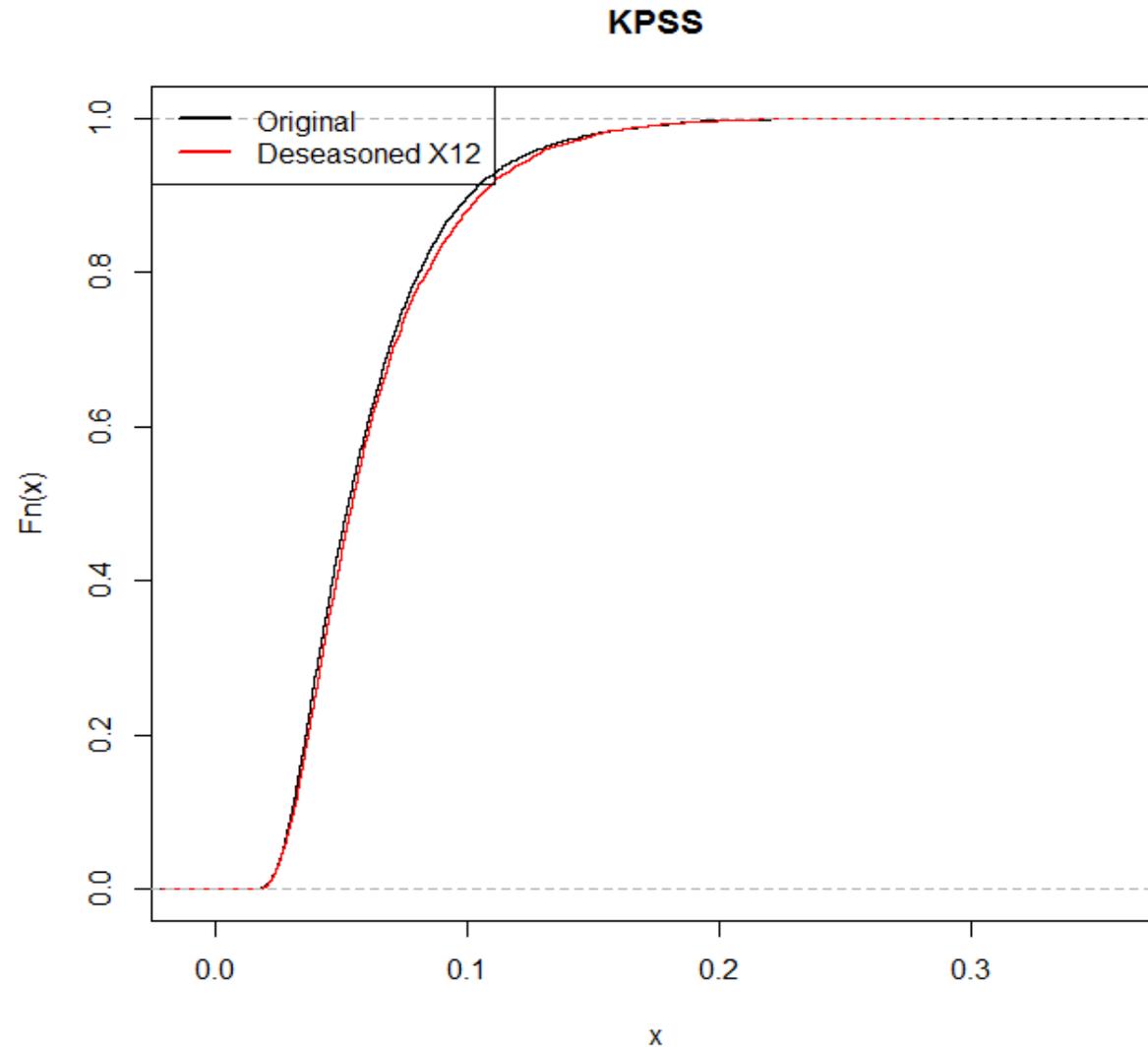
# Монте-Карло, истинный ряд - TS



Доля неверно  
определённых рядов

Значим ость	Исходн ый ряд	Сглаже нный ряд
10%	0	0
5%	0	0
1%	0	0

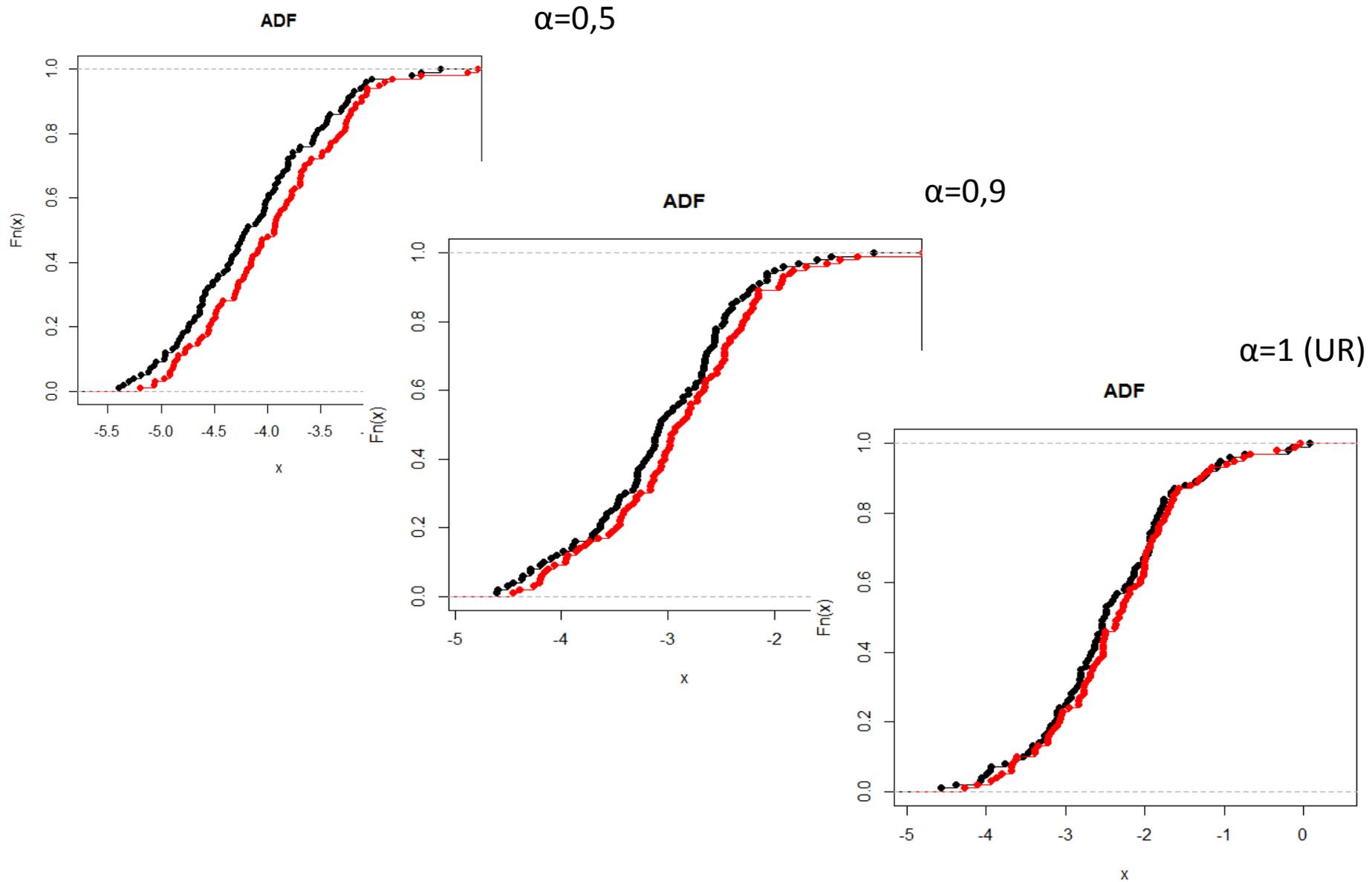
# Монте-Карло, истинный ряд - TS



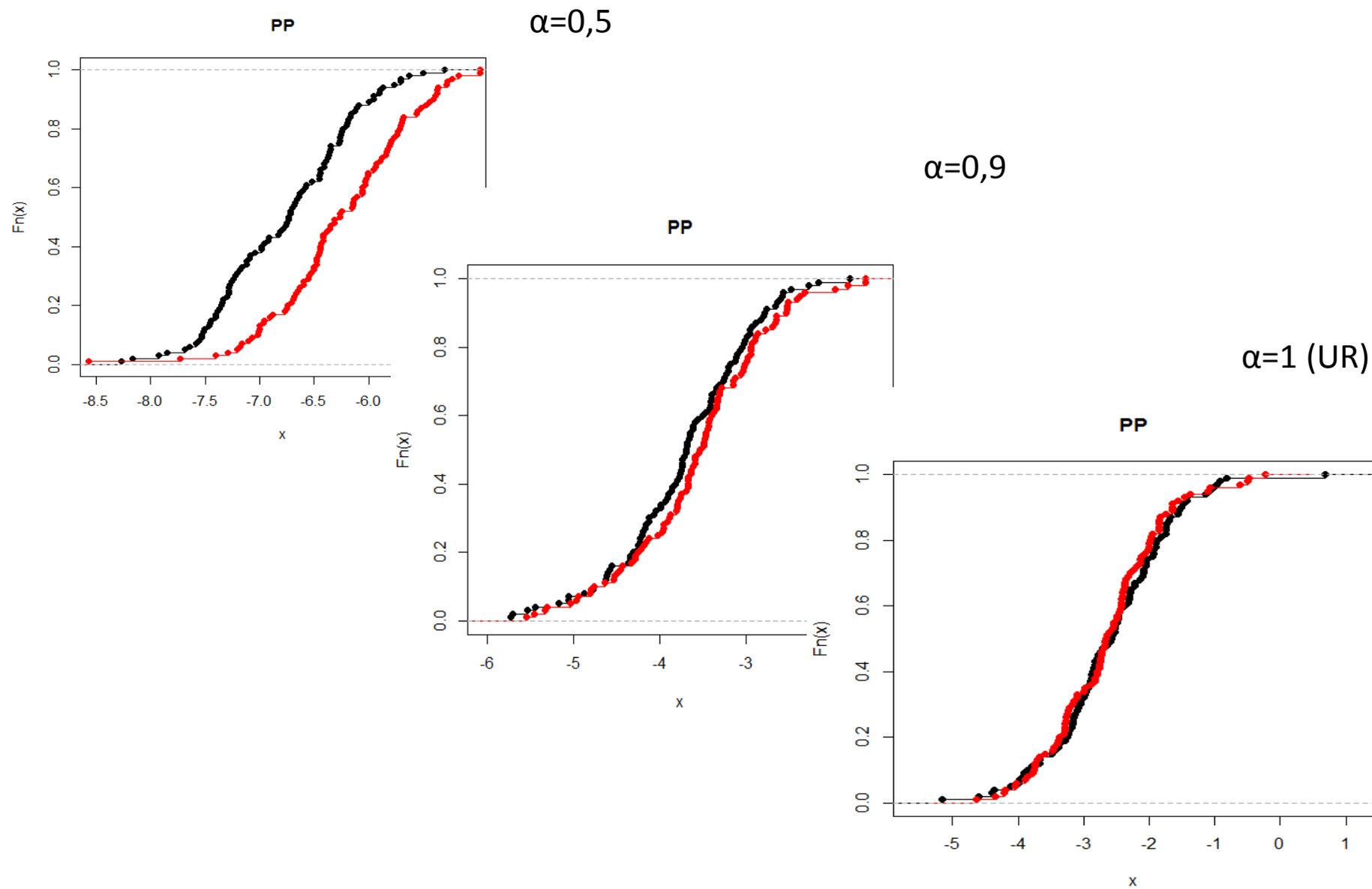
Доля неверно  
определённых рядов

Значим ость	Исходн ый ряд	Сглаже нный ряд
10%	0,055	0,063
5%	0,0234	0,027
1%	0	0

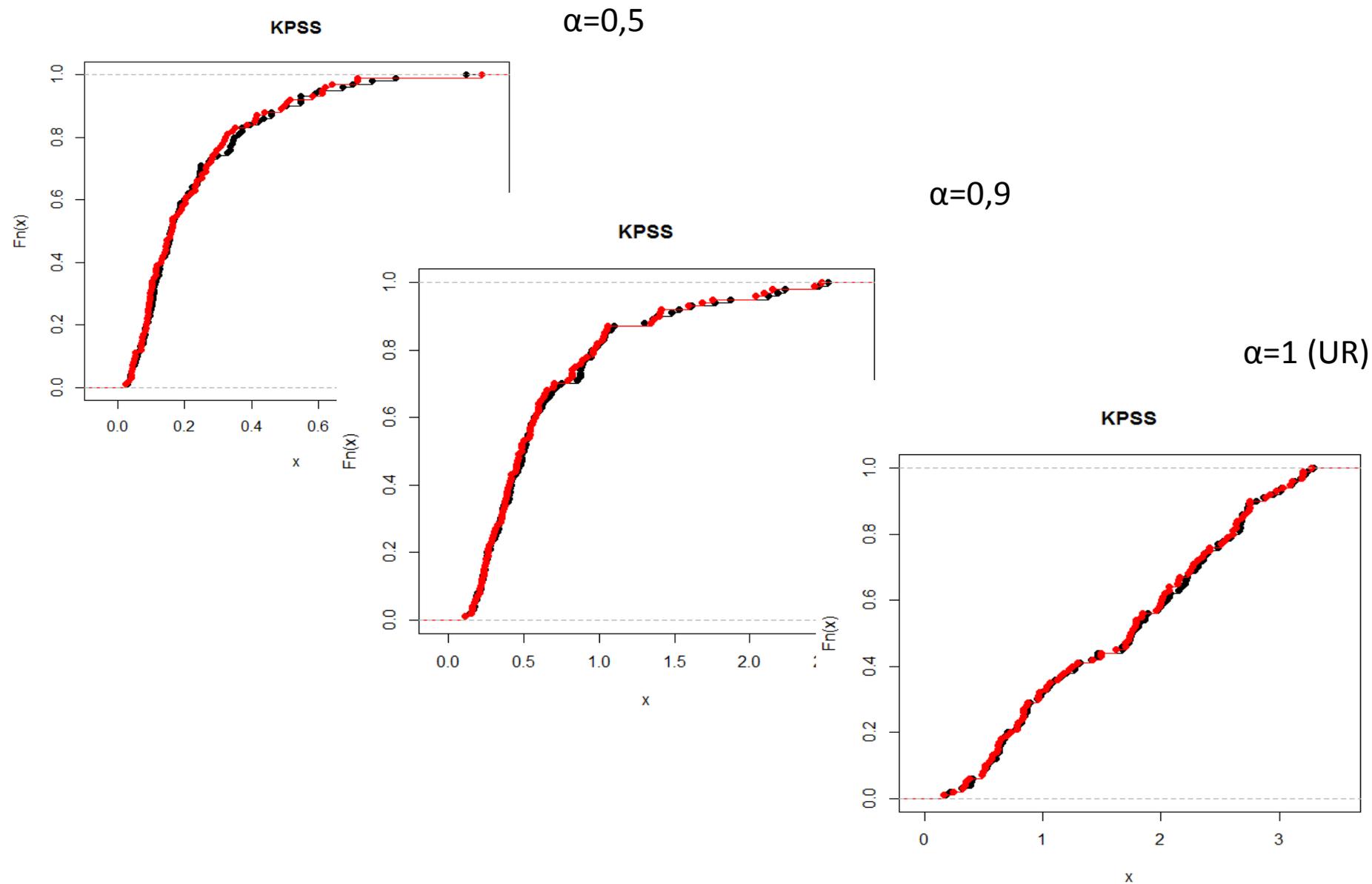
# Монте-Карло, истинный ряд AR(1)



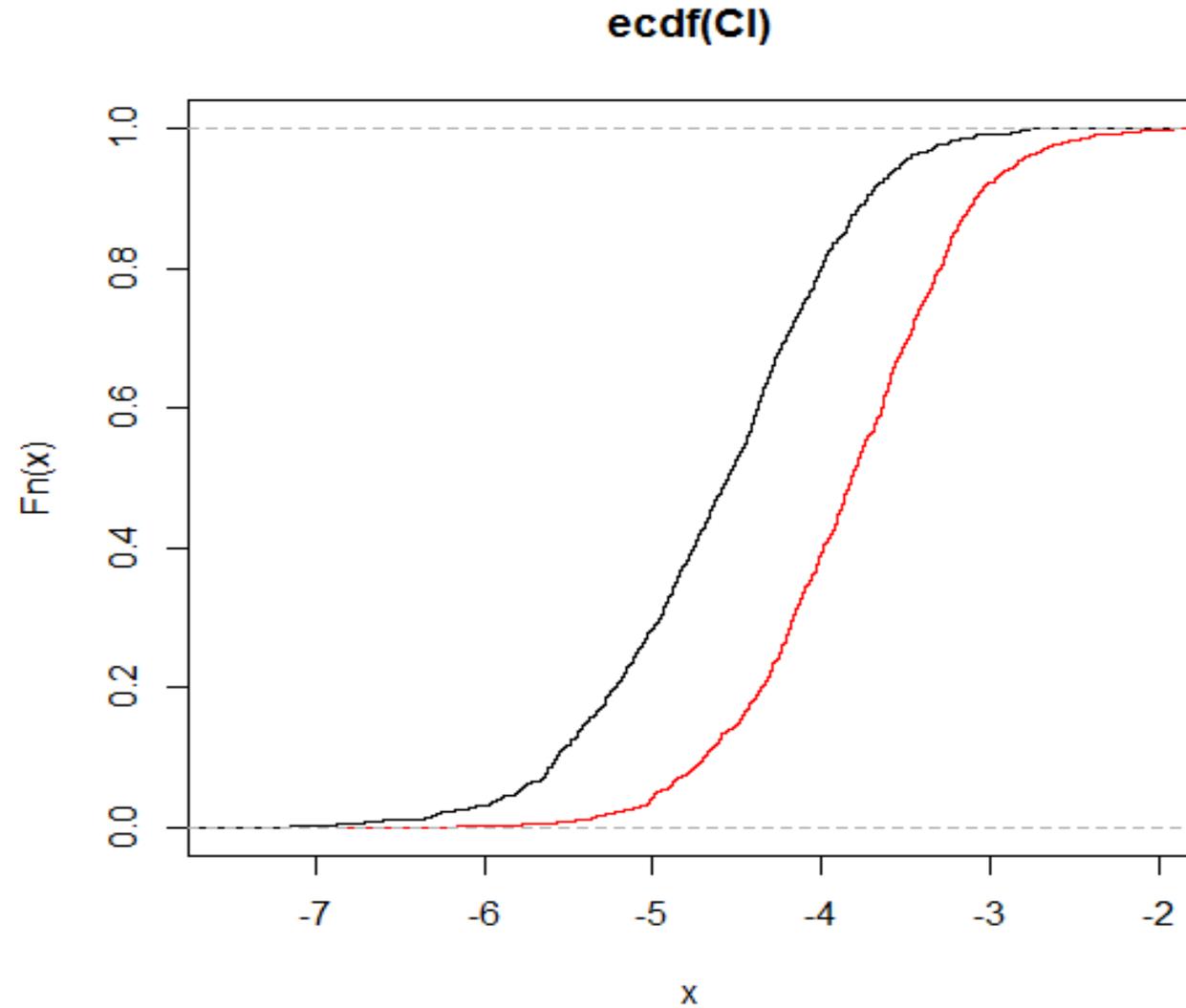
# Монте-Карло, истинный ряд AR(1)



# Монте-Карло, истинный ряд AR(1)



# Монте-Карло: коинтеграция при помощи теста Энгла-Грейнджера



# Точки перелома

- Matas-Mir, Osborn, Lombardi (2008)\*
- Показали (преимущественно методом Монте-Карло), что сглаживание:
  - Уменьшает величину шока
  - Увеличивает его продолжительность
  - Сдвигает точку окончания спада

\* Matas-Mir A., Osborn D.R., Lombardi M.J. (2008), The Effect of Seasonal Adjustment on the Properties of Business Cycle Regimes, *Journal of Applied Econometrics*

# На реальных данных: вероятность рецессии

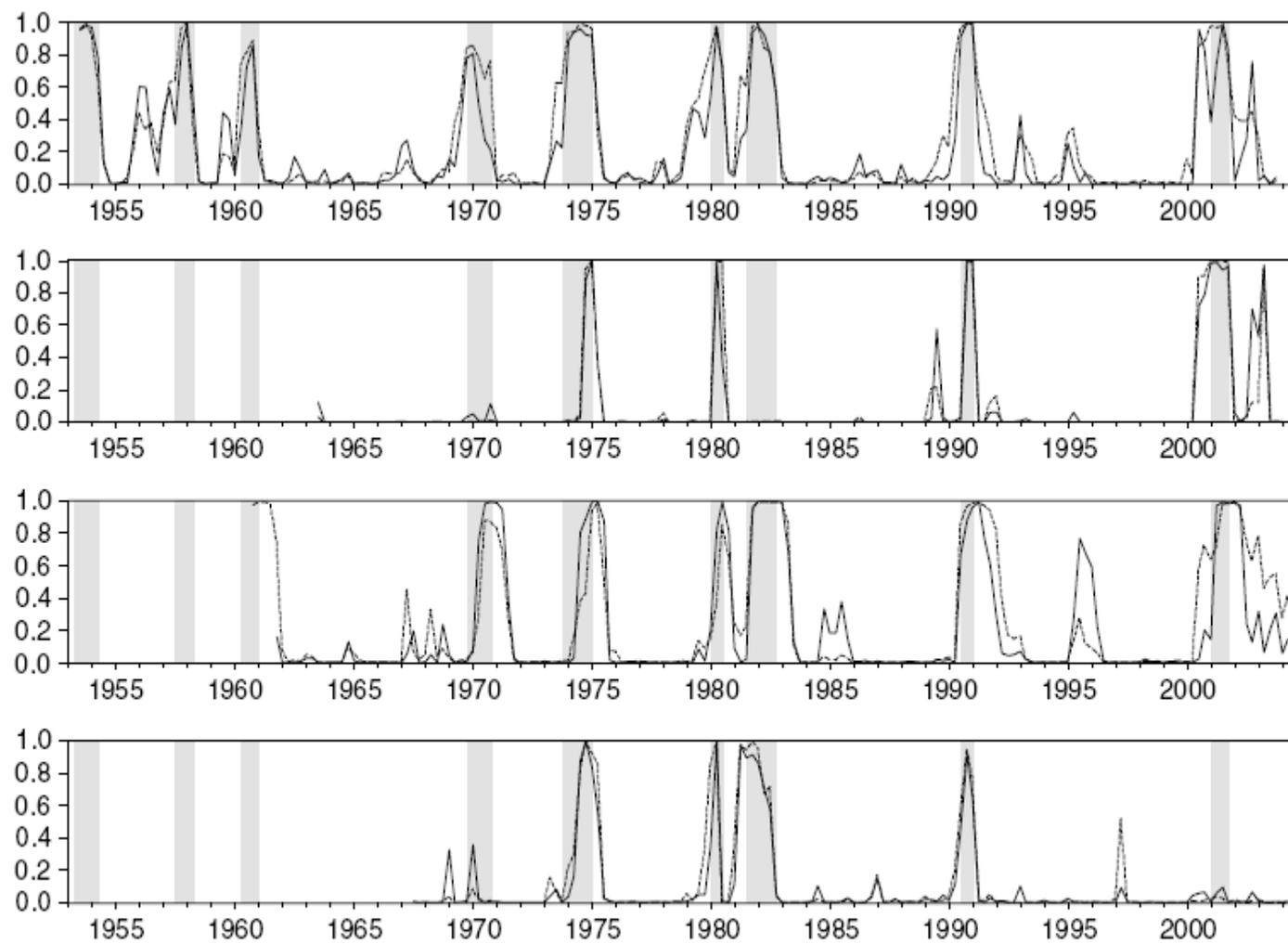


Figure 3. Estimated recession probabilities for US business cycle indicators. *Note:* Estimated recession probabilities for seasonally adjusted data are indicated by the dashed line and for unadjusted data by the solid line. Series analysed are GDP, industrial production, employment and sales (respectively, from top to bottom). Shading indicates NBER recession observations

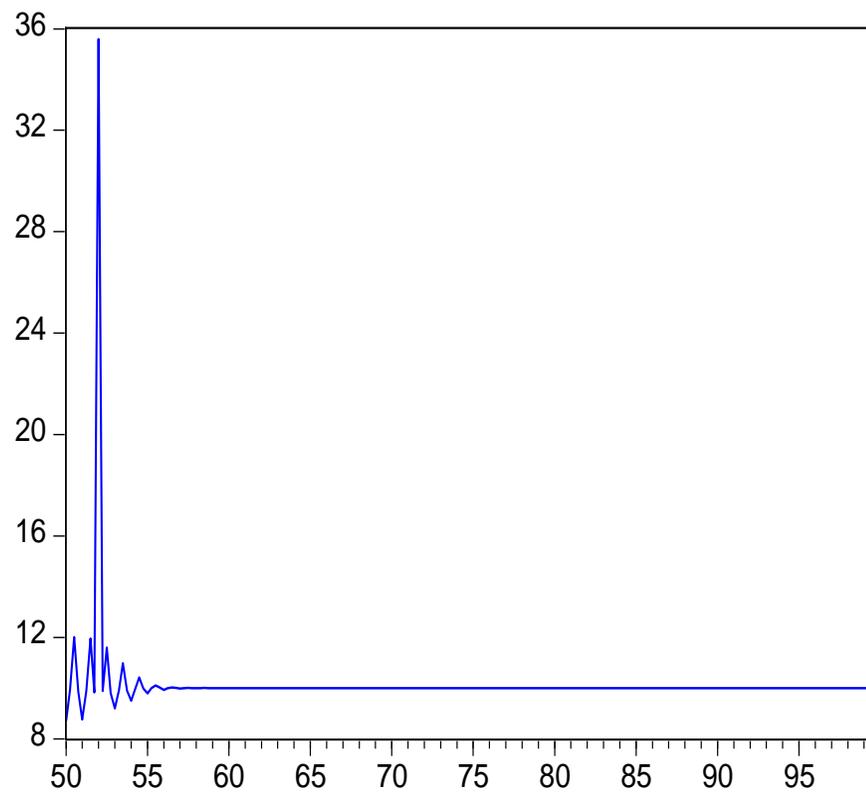
# Неустойчивость к выбросам и сдвигам

- Bruce, Jurke (1992)\*
- Проверяли на реальных данных, сравнивали с другими процедурами
- Обнаружили неустойчивость к выбросам и структурным сдвигам, но отметили большую гибкость самой процедуры

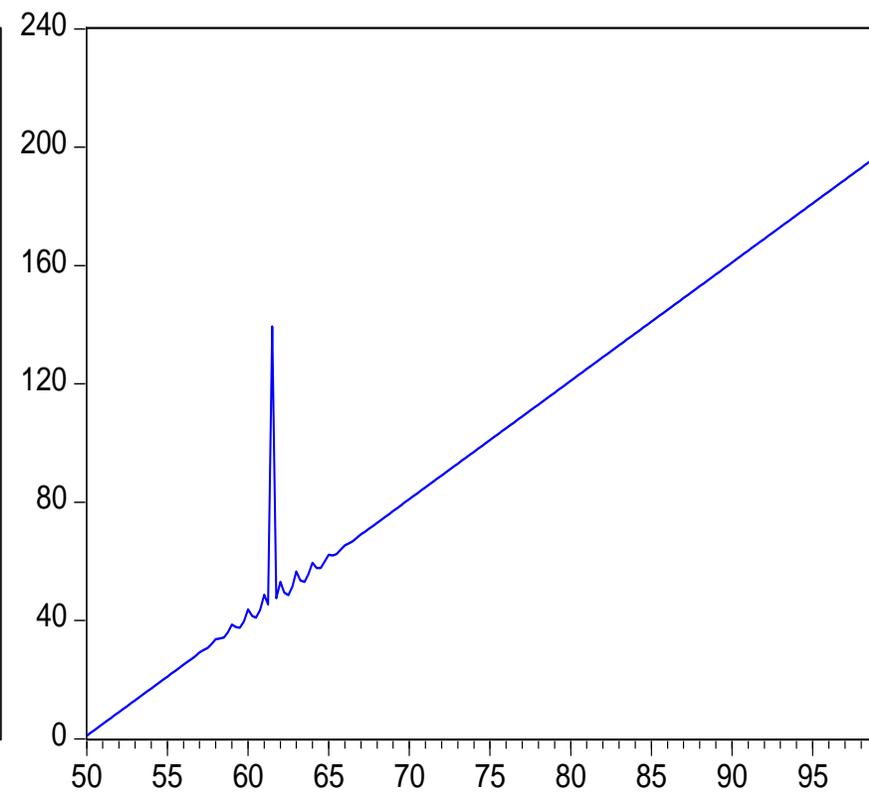
\*Bruce A. G., Jurke S. R. (1992), Non-Gaussian Seasonal Adjustment: X-12-ARIMA Versus Robust Structural Models, *BUREAU OF THE CENSUS, STATISTICAL RESEARCH DIVISION, Statistical Research Report Series*

# Выбросы и X12

N2\_SA



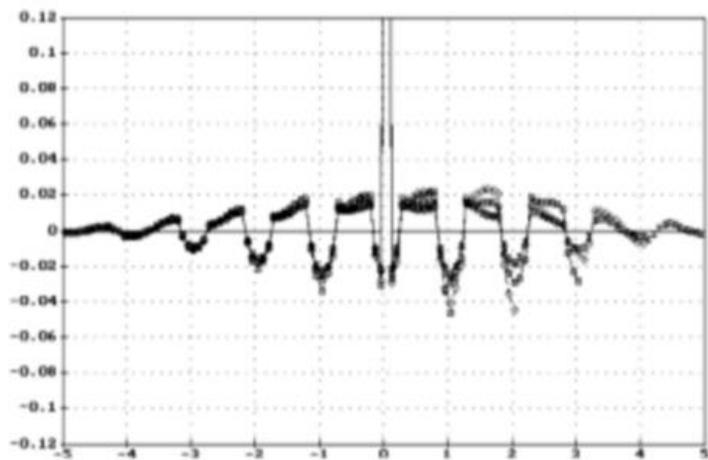
LIN\_SA\_X12



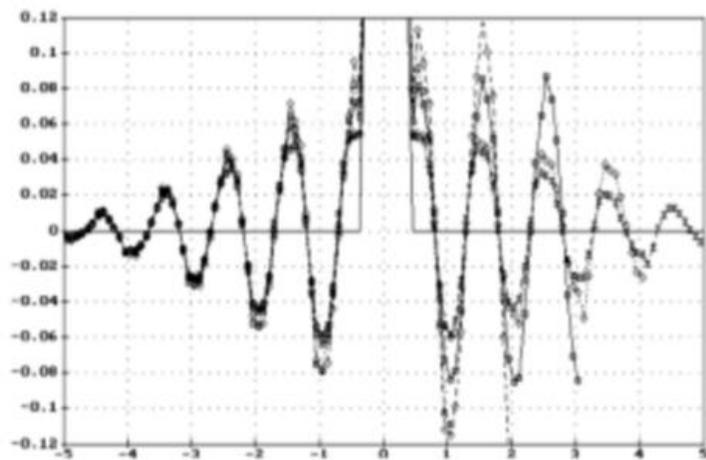
# Ложные сигналы

- Бессонов, Петроневи́ч (2013)\*
- Сравнивали влияние разных типов выбросов на результирующий сезонно скорректированный ряд
- Показывается, что в результате сезонной корректировки отдельные выбросы в данных могут порождать фиктивную сезонность, фиктивные фор- и афтершоки.

\* Анна Р., Петроневи́ч А. В., БЕССОНОВ В. А. Сезонная корректировка как источник ложных сигналов //Экономический журнал Высшей школы экономики. – 2013. – Т. 17. – №. 4.

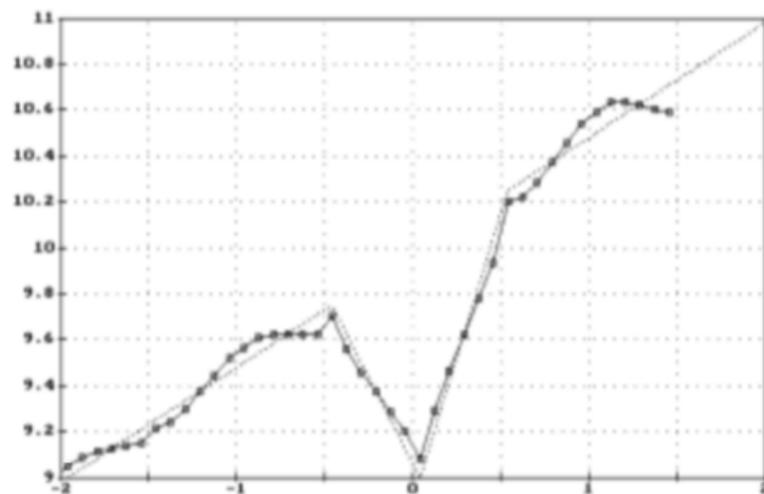


а)  $l = 1$

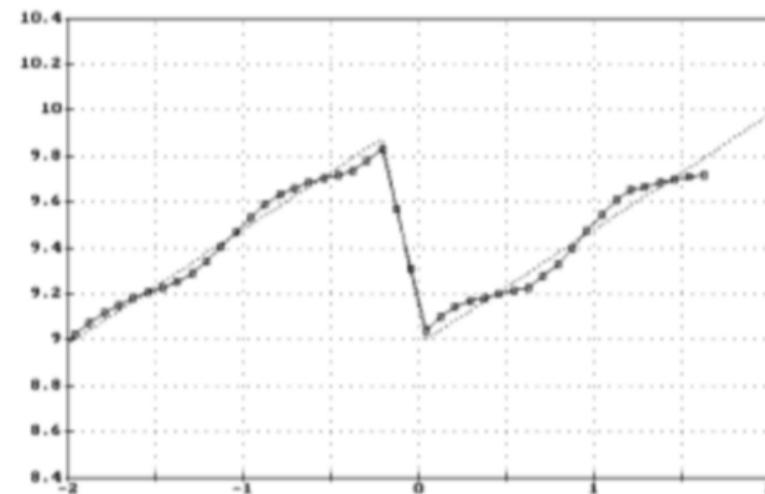


б)  $l = 5$

**Рис. 5.** Фиктивные фор- и афтершоки для флуктуации типа «пик» вблизи правого края временных рядов ( $\tau = 12, 24, 36, 48$  и  $60$  месяцев). Линия – исходный ряд; маркеры – сезонно скорректированные ряды

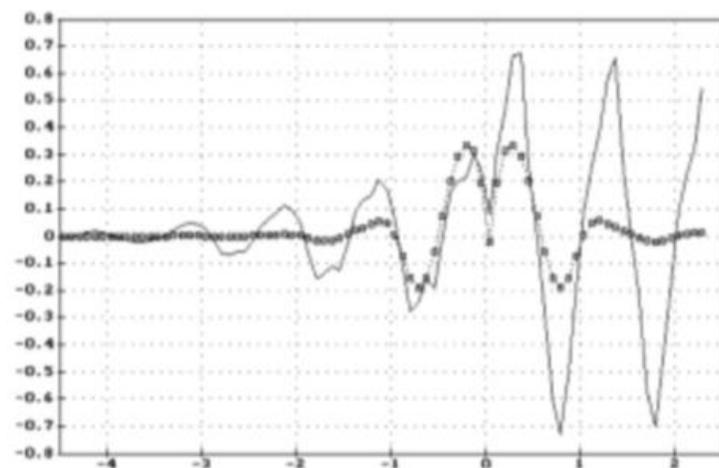
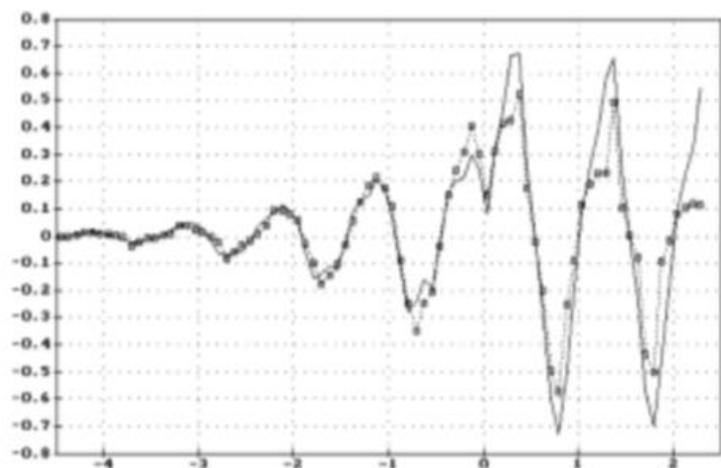


а) флуктуация типа «пик» вблизи края,  $l = 6, \tau = 17$



б) флуктуация типа «ступень» вблизи края,  $l = 3, \tau = 19$

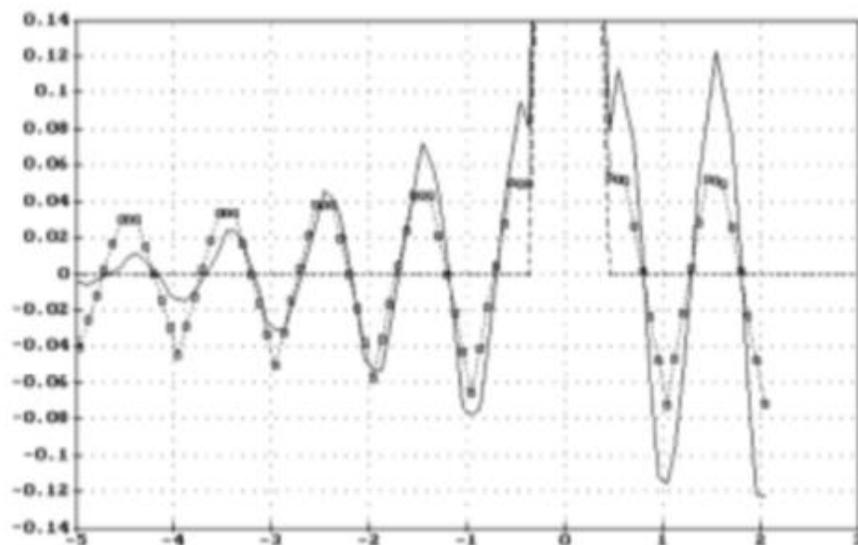
**Рис. 7.** Иллюстрация влияния фиктивных фор- и афтершоков на идентификацию краткосрочных тенденций. Линия – исходный ряд; маркеры – сезонно скорректированный ряд



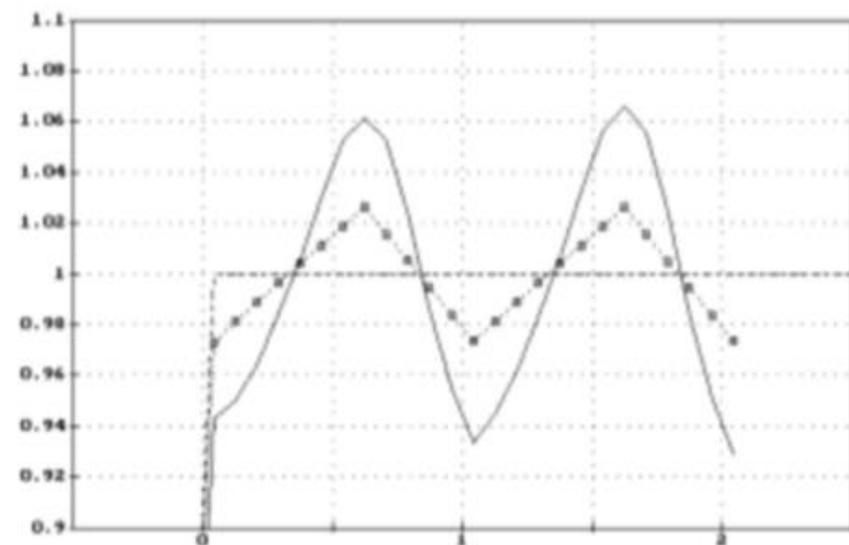
а) X-11 (сплошная линия) и X-12-ARIMA (маркеры)

б) X-11 (сплошная линия) и TRAMO/SEATS (маркеры)

**Рис. 12.** Сравнение ложных сигналов разных алгоритмов декомпозиции при изменении сезонной волны вблизи правого края временного ряда ( $\tau = 27$ )



а) флуктуация типа «пик» вблизи края,  $l = 5$ ,  $\tau = 24$



б) флуктуация типа «ступень» вблизи края,  $l = 5$ ,  $\tau = 24$

**Рис. 11.** Сравнение ложных сигналов для алгоритмов X-11 (сплошная линия) и TRAMO/SEATS (маркеры)

Спасибо за внимание!

- [https://github.com/bdemeshev/epsilon/blob/master/e\\_001/e\\_001.pdf](https://github.com/bdemeshev/epsilon/blob/master/e_001/e_001.pdf)

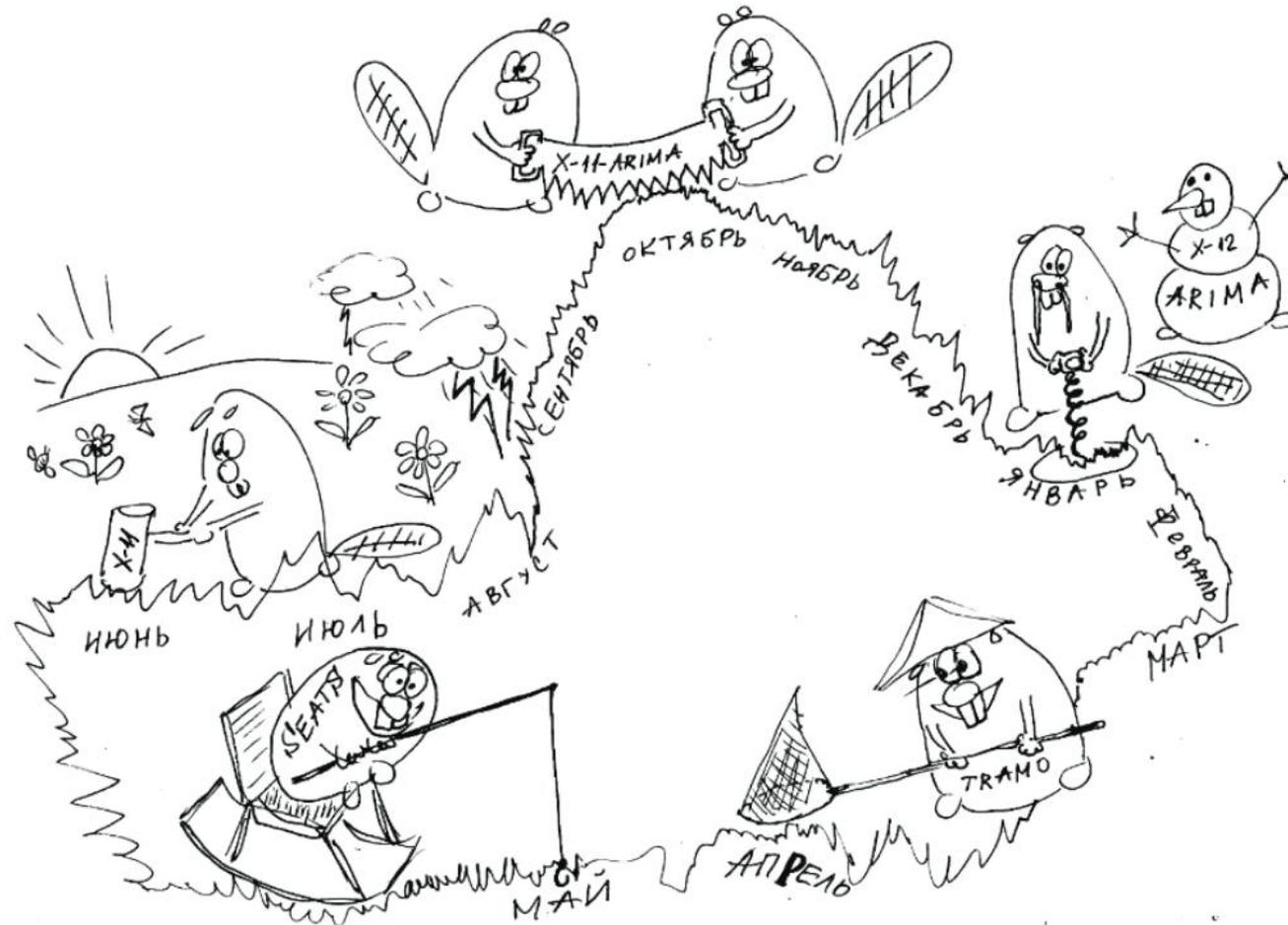


Рис. 1. Бобры выполняют сезонную корректировку