

# Мульти-масштабный адаптивный регрессионный анализ

Предпосылки, методология и экспериментальная проверка

**Афанасьев Д.О.<sup>a</sup>   Федорова Е.А.<sup>b,a</sup>   Гиленко Е.В.<sup>c</sup>**

<sup>a</sup>Финансовый университет при Правительстве РФ, Москва

<sup>b</sup>Высшая школа экономики, Москва

<sup>c</sup>Санкт-Петербургский государственный университет, Высшая Школа Менеджмента, Санкт-Петербург

21 апреля 2016

XVII Апрельская международная научная конференция ВШЭ



# Содержание

- 1** Мотивация
- 2 Методология
- 3 Эксперимент
- 4 Результаты

# Анализ временных рядов на исходном уровне данных

Задача восстановления зависимости от экзогенных переменных – классическая задача анализа временных рядов:

$$y(t) = f(\vec{x}_k(t); \varepsilon(t); \vec{\beta}_k(t) | \Omega_T), \Omega_T = \{y_t, x_t\}_{t \in [1, T]}$$

Но, что мы можем сказать о внутренней структуре зависимости на характерных временных масштабах?

- какова реакция  $y$  на долгосрочное изменение  $x_k$ ?
- связана ли цикличность  $y$  с соответствующими колебаниями  $x_k$ ?
- значимо ли влияние экзогенного фактора  $x_k$  в краткосрочном периоде?
- что можно сказать о произвольных временных масштабах?

# Мульти-масштабный аналитический подход

Позволяет выявлять взаимосвязи экономических показателей на различных временных масштабах и определять факторы, вносящие основной вклад в комплексный механизмы взаимодействия.

Состоит из 2х основных шагов:

- 1 Декомпозиция исходных данных на наборы мод с различными периодами колебаний  $T$  (идентифицируют экономический смысл).
- 2 Применение тех или иных методов эконометрического анализа к полученным модам.

Моды должны быть независимы и слабо связаны на кросс-масштабах  $\Rightarrow$  отражают влияние различных внутренних факторов.

## Шаг 1. Декомпозиция на моды

Из наблюдаемого временного ряда выделяются ненаблюдаемые структурные компоненты, соответствующие различным временным масштабам (тренд, сезонность, цикличность, шум и т.д.).

### ИНСТРУМЕНТЫ:

- Фурье-разложение (De Jong, 2006; Weron и др., 2010)
- Вейвлета-декомпозиция (Janczuga и др., 2013; Nowotarski и др., 2013)
- Декомпозиция на эмпирические моды, EMD (Huang и др., 1998; Wu и Huang, 2009; Torres и др., 2011)

### ОГРАНИЧЕНИЯ:

Метод	Нестационарность	Нелинейность	Априорный базис
Fourier	Нет	Нет	Да
Wavelet	Да	Нет	Да
EMD	Да	Да	Нет

## Шаг 2. Эконометрический анализ на уровне мод

К полученным модам применяется те или иные инструменты эконометрического анализа для ответа на вопросы о взаимосвязи исследуемых явлений на различных временных масштабах.

ПРИМЕРЫ:

- Yu и др. (2015) – исследование причинно-следственных взаимосвязей между европейским рынком квот на выбросы  $CO_2$  и рынком сырой нефти марки Brent. Инструмент: линейный и нелинейный тесты на причинность по Грейнджеру.
- Afanasyev и др. (2015) – анализ взаимосвязи цены и спроса на электроэнергию для 2х оптовых российских рынков на сутки вперед. Инструмент: зависящая от времени внутренняя корреляция TDIC (Chen и др., 2010).

Мульти-масштабный аналитический подхода + Адаптивная к данным методика декомпозиции EMD = Мульти-масштабный адаптивный анализ.

# Зависимость коэффициентов от времени

Стандартная практика – постулируем неизменность коэффициентов эконометрической модели на всем исследуемом периоде. Но соответствует ли это реальной ситуации?

ИНСТРУМЕНТЫ:

- 1 Модель пространства состояний SSM (Karakatsani и Bunn, 2008) и фильтр Калмана (Kim и Nelson, 1999).
- 2 Модели с переключениями режимов – TAR, STAR, MRS (Hamilton, 1990; Zachmann, 2013).

ОГРАНИЧЕНИЯ:

- SSM – ограничение на процесс генерации данных DGP для коэффициентов (обычно случайное блуждание).
- TAR, STAR – ограничения на пороги и функции переключения.
- MRS – ограничение на тип латентного процесса.

*Как отказаться от этих ограничений в контексте мульти-масштабного аналитического подхода?*

# Мульти-масштабный адаптивный регрессионный анализ

Зависящая от времени внутренняя корреляция TDIC (Chen и др., 2010) и мульти-масштабный адаптивный корреляционный анализ (Afanasyev и др., 2015):

- рассчитывает корреляцию на уровне внутренних мод, т.е. различных временных масштабов
- использует скользящее окно для учета изменения корреляции во времени
- размер окна адаптивен к данным, т.к. представляет собой мгновенный период колебаний (преобразование Гильберта)

НАШ ВКЛАД:

- Переход от корреляционного к регрессионному анализу  $\Rightarrow$  Зависящая от времени внутренняя регрессия (TDIR) и Мульти-масштабный адаптивный регрессионный анализа (MSARA).
- Методология новая и ранее не апробировалась  $\Rightarrow$  необходимо выполнить проверку на симуляционных экспериментах.



# Содержание

- 1 Мотивация
- 2 Методология**
- 3 Эксперимент
- 4 Результаты

# Декомпозиция на эмпирические моды

## Идея

Исходный сигнал декомпозируется на сумму внутренних модовых функций IMF (Huang и др., 1998) с помощью алгоритма “просеивания” (стоп-критерии: в форме Коши, S-критерий).

## Внутренняя модовая функция IMF

- 1 число экстремумов и число переходов через нулевую точку должны быть либо равны, либо отличаться не более чем на единицу
- 2 в любой момент времени среднее значение огибающей, построенной на локальных максимумах, и огибающей, построенной на локальных минимумах, равняется нулю

Исходный сигнал  $x(t)_{t \in [1, T]}$  может быть восстановлен в виде суммы  $I$  полученных IMF (эмпирический базис разложения) и остатка  $r(t)$ :

$$x(t) = \sum_{i=1}^I IMF_i(t) + r(t).$$

# Декомпозиция на эмпирические моды

## Хронология развития

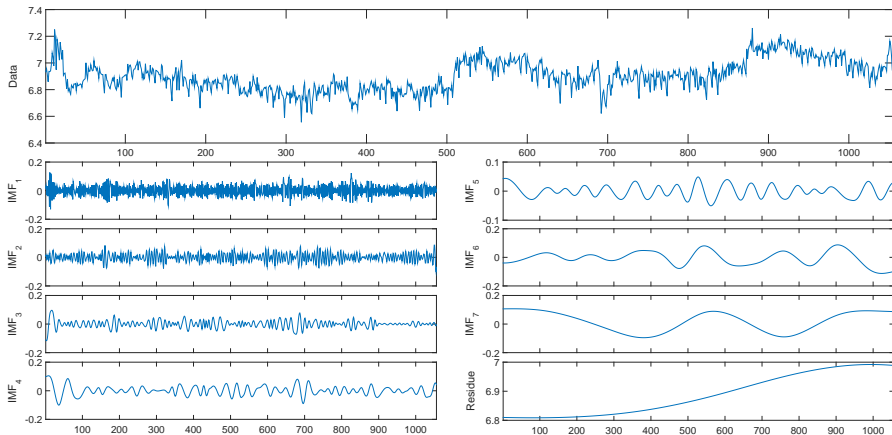
Метод	Авторы	Основные проблемы
Декомпозиция на эмпирические моды, EMD	Huang и др. (1998)	Перемешивание мод ("mode mixing")
Ансамблевая декомпозиция на эмпирические моды, EEMD	Wu и Huang (2009)	Остаточный шум, неустойчивое число мод, размер ансамбля, вычислительные ресурсы
Декомпозиция по полной системе эмпирических мод с адаптивным шумом, CEEMDAN	Torres и др. (2011); Colominas и др. (2012)	Размер ансамбля

## CEEMDAN vs EEMD:

- Устойчива к изменению амплитуды добавляемого шума
- Точность восстановления исходного сигнала практически не зависит от соотношения сигнал-шум (SNR)
- Менее требовательна к вычислительным ресурсам

# Декомпозиция на эмпирические моды

## Пример



# Зависящая от времени внутренняя регрессия (TDIR)

## Начало

- Декомпозируем  $y$  и  $x_k$  с помощью CEEMDAN:

$$y(t) = \sum_{i=1}^{I_y} IMF_i^y(t), \quad x_k(t) = \sum_{i=1}^{I_{x_k}} IMF_i^{x_k}(t)$$

- Исключаем из анализа:

- Моды, статистически не отличимые от дробного гауссового шума (Flandrin и др., 2004).
- Моды, вносящие незначительный вклад в общую наблюдаемую дисперсию временного ряда.

- Иерархическая кластеризация мод  $IMF_i^y$  и  $IMF_i^{x_k}$  (Chih-Yu и др., 2013), полное связывание + корреляционная дистанция  $\Rightarrow$  набор кластерных мод  $CIMF_i^y$  и  $CIMF_i^{x_k}$ .

# Зависящая от времени внутренняя регрессия (TDIR)

## Продолжение 1

- Вычисляем мгновенные периоды  $T_i^y(t)$  и  $T_i^{x_k}(t)$  для  $CIMF_i^y$  и  $CIMF_i^{x_k}$  (преобразование Гильберта). Размер адаптивного окна (Chen и др., 2010):

$$\tau_i^n(t) = [t - nt_i^m(t)/2 : t + nt_i^m(t)/2]$$

$t_i^m(t) = \max(T_i^y(t), T_i^{x_1}(t), \dots, T_i^{x_K}(t))$ ,  $n \geq 1$  – число циклов в окне.

- Восстанавливаем линейную регрессию  $CIMF_i^y$  на  $CIMF_i^{x_k}$  в адаптивном скользящем окне  $\tau_i^n(t)$  ( $\beta_0^i = 0 | i \neq I$ ):

$$CIMF_i^y(t) = \beta_0^i + \sum_{k=1}^K \beta_k^i \cdot CIMF_i^{x_k}(t) + \varepsilon_i(t), \varepsilon_i(t) \sim i.i.d. N(0, \sigma_i^2), t \in \tau_i^n(t)$$

### Результат:

- $I \times K$  временных рядов  $\hat{\beta}_k^i(t)$  и  $SE_k^i(t)$  коэффициентов  $\beta_k^i(t)$
- $I \times K$  временных рядов статистик и  $p$ -значений  $t$ -теста
- $I$  временных рядов статистик и  $p$ -значений  $F$ -теста

# Зависящая от времени внутренняя регрессия (TDIR)

## Продолжение 2

- Рассчитываем характерные для временного масштаба оценки  $\overline{\beta}_k^i$  и их стандартные ошибки  $\overline{SE}_k^i$ :

$$\overline{\beta}_k^i = MeB(\hat{\beta}_k^i(t) \cdot \mathbb{I}(p \leq \alpha)), \overline{SE}_k^i = MeB(SE_k^i(t))$$

$MeB$  – оператор бустрапированной медианы

$\mathbb{I}(p \leq \alpha)$  – индикаторная функция

$p$  –  $p$ -значение  $t$ -теста значимости коэффициента  $\hat{\beta}_k^i(t)$

$\alpha$  – уровень значимости.

# Содержание

- 1 Мотивация
- 2 Методология
- 3 Эксперимент**
- 4 Результаты



## Схема генерации данных

Структура объясняемой переменной  $x$  ( $t = 1, \dots, 1095$ ):

- долгосрочный тренд  $x^T = \exp(0.0001 \cdot t)$
- циклические колебания  $x^C = \sin((2\pi/365) \cdot t)$
- сезонные осцилляции  $x^S = \cos((4\pi/365) \cdot t)$

Схема генерации компонент  $i = \{T, C, S\}$  объясняемой переменной  $y$ :

$$y^i = \beta_0^i + \beta_1^i x^i + \varepsilon^i, \beta_0^i \neq 0 | i = T, \varepsilon^i \sim i.i.d. N(0, \sigma_i^2)$$

Итоговая объясняемая переменная:

$$y = y^S + y^C + y^T$$

# Параметры эксперимента

Рассмотренные ситуации:

- 1 Постоянные коэффициенты  $\beta_1^i$  модели для всех компонент
- 2 Зависящий от времени коэффициент  $\beta_1^S$  для сезонной составляющей
- 3 Гетероскедастичность в виде монотонного роста дисперсии  $\sigma_S^2$  для сезонной компоненты

Компонента	Постоянные коэфф.		Динамические коэфф.		Гетероскедастичность	
	$\beta_1$	$\sigma$	$\beta_1$	$\sigma$	$\beta_1$	$\sigma$
Сезон, S	3	1	$3 \cdot (1 + 0.001t)$	1	3	$1 \cdot \sqrt{1 + 0.005t}$
Цикл, C	4	1.2	4	1.2	4	1.2
Тренд, T	2	0.2	2	0.2	2	0.2

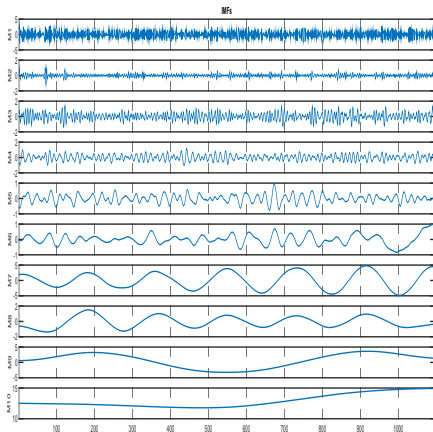
Пояснения: для всех рассмотренных случаев константа модели  $\beta_0^i = 9|i = T$  и отлична от нуля только для трендовой составляющей, т.е.  $\beta_0^i = 0|i \neq T$ .

# Содержание

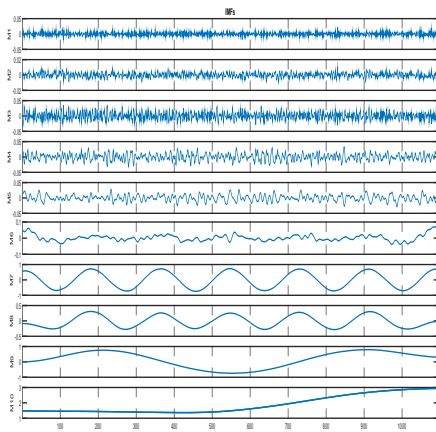
- 1 Мотивация
- 2 Методология
- 3 Эксперимент
- 4 Результаты**

# Эмпирические моды $y$ и $x$

Объясняемая переменная  $y$

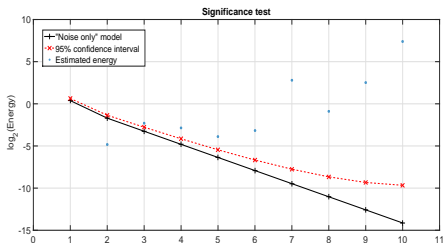


Объясняющая переменная  $x$

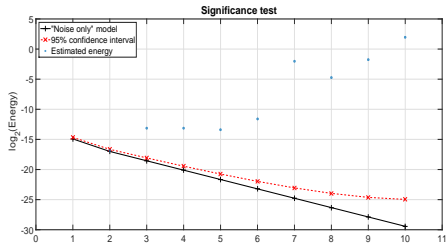


# Отбор эмпирических мод

Объясняемая переменная  $y$



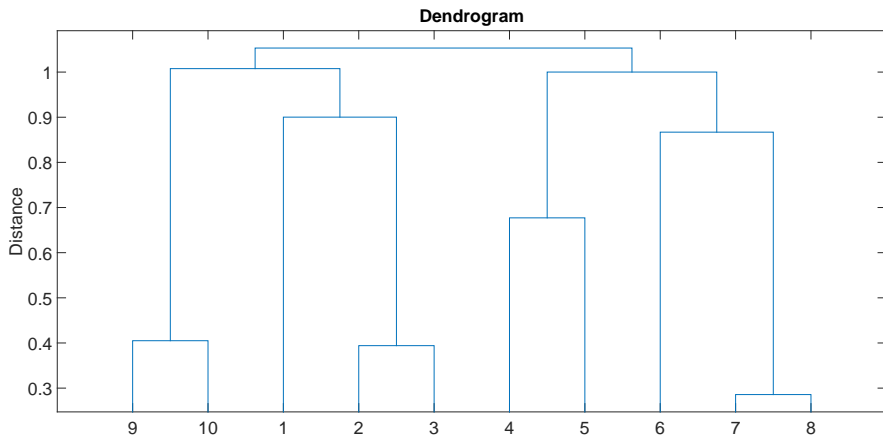
Объясняющая переменная  $x$



## Характеристики мод

$IMF_i$	Средний период $\overline{T}_i$		Дисперсия $\sigma_i$		Дисперсия $\sigma_i$ (%)	
	$y$	$x$	$y$	$x$	$y$	$x$
1	1	2	1.30	0.00	5.7	0.0
2	3	2	0.04	0.00	0.2	0.0
3	4	2	0.20	0.00	0.9	0.0
4	7	4	0.14	0.00	0.6	0.0
5	14	8	0.07	0.00	0.3	0.0
6	34	18	0.11	0.00	0.5	0.0
7	118	118	6.91	0.25	30.1	18.9
8	141	133	0.53	0.04	2.3	3.0
9	626	487	5.20	0.27	22.7	20.5
10	-	-	1.14	0.30	5.0	22.7

## Иерархическая кластеризация эмпирических мод



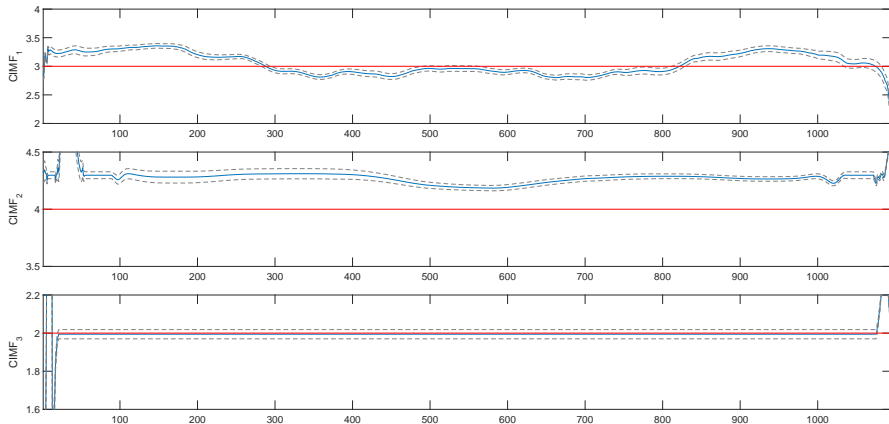
**1**  $CIMF_1 = IMF_7 + IMF_8$  – сезонные колебания

**2**  $CIMF_2 = IMF_9$  – циклические колебания

**3**  $CIMF_3 = IMF_{10}$  – тренд

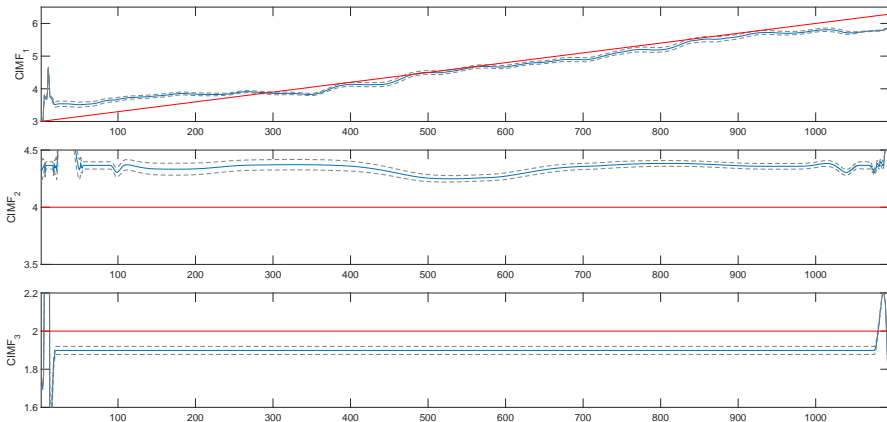
# Динамика оценки $\hat{\beta}_1^i(t)$

Постоянные коэффициенты



# Динамика оценки $\hat{\beta}_1^i(t)$

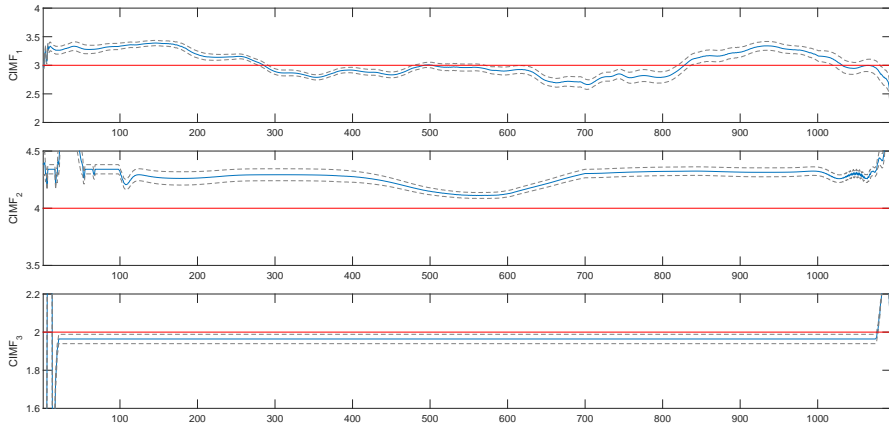
Переменный коэффициент





# Динамика оценки $\hat{\beta}_1^i(t)$

Гетероскедастичность



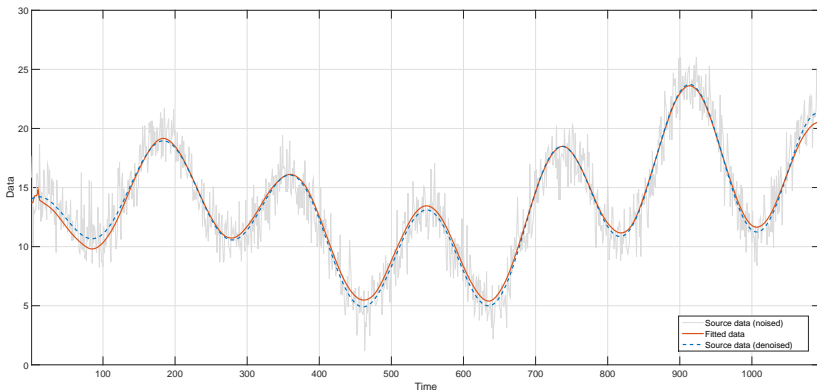
# Усредненная на временном масштабе оценка $\overline{\beta}_k^i$

$CIMF_i$	Постоянный коэффициент		Переменный коэффициент		Гетероскедастичность	
	$\overline{\beta}_1^i$	$MAPE$	$\overline{\beta}_1^i$	$MAPE$	$\overline{\beta}_1^i$	$MAPE$
1 (сезон)	2.985 <sup>***</sup> (0.025)	5.5%	4.607 <sup>***</sup> (0.027)	4.5%	2.973 <sup>***</sup> (0.035)	6.1%
2 (цикл)	4.282 <sup>***</sup> (0.014)	7.4%	4.359 <sup>***</sup> (0.015)	9.1%	4.291 <sup>***</sup> (0.020)	7.5%
3 (тренд)	1.994 <sup>***</sup> (0.012)	1.6%	1.899 <sup>***</sup> (0.011)	5.9%	1.964 <sup>***</sup> (0.013)	3.2%
	$\overline{\beta}_0^i$		$\overline{\beta}_0^i$		$\overline{\beta}_0^i$	
3 (тренд)	9.147 <sup>***</sup> (0.024)	2.0%	9.323 <sup>***</sup> (0.021)	3.7%	9.106 <sup>***</sup> (0.025)	1.6%

Примечания: \*\*\* – значим на 1% уровне, \*\* – значим на 5% уровне, \* – значим на 10% уровне. В скобках приведена стандартная ошибка коэффициента.

# Внутривыборочное качество

Мера качества	Постоянные коэфф.		Переменный коэфф.		Гетероскедастичность	
	$y_d$	$y$	$y_d$	$y$	$y_d$	$y$
<i>MAE</i>	0.26	1.29	0.29	1.30	0.25	1.84
<i>MAPE</i>	2.4%	10.9%	2.8%	11.9%	2.4%	16.2%
<i>MSE</i>	0.11	2.59	0.13	2.63	0.12	5.33
<i>RMSE</i>	0.34	1.61	0.36	1.62	0.34	2.31



# Выводы

- TDIR позволяет выявить наличие и оценить величину взаимосвязей явлений на различных временных масштабах.
- TDIR способна улавливать изменение данных взаимосвязей во времени.
- Гетероскедастичность не оказывает существенного влияния на результаты, получаемые с помощью TDIR.
- Внутривыборочное качество TDIR на уровне исходных (не декомпозированных) данных является достаточно высоким.
- Мульти-масштабный адаптивный регрессионный анализ потенциально дает возможность более детально исследовать комплексные механизмы взаимосвязи экономических процессов, понимать их внутреннюю структуру, а также выявлять эволюцию данных механизмов во времени.
- Вопрос апробации данной методологии на прикладных задачах пока остается открытым.

# Список литературы

- Afanasyev, D., Fedorova, E., Popov, V., 2015. Fine structure of the price–demand relationship in the electricity market: Multi-scale correlation analysis. *Energy Economics* 51, 215–226.
- Chen, N., Wu, Z., Huang, N., 2010. The time-dependent intrinsic correlation based on the empirical mode decomposition. *Advances in Adaptive Data Analysis* 2 (2), 223–265.
- Chih-Yu, K., Shao-Kuan, W., Pi-Wen, T., 2013. Ensemble empirical mode decomposition with supervised cluster analysis. *Advances in Adaptive Data Analysis* 5 (1).
- Colominas, M., Schlotthauer, G., Torres, M., Flandrin, P., 2012. Noise-assisted EMD methods in action. *Advances in Adaptive Data Analysis* 4 (4).
- De Jong, C., 2006. The nature of power spikes: a regime-switch approach. *Stud. Nonlinear Dynamics and Econometrics* 10 (3), Article 3.
- Flandrin, P., Goncalves, P., Rilling, G., 2004. Detrending and denoising with empirical mode decomposition. в: 12th European Signal Processing Conference. сс. 1581–1584.
- Hamilton, J., 1990. Analysis of time series subject to changes in regime. *Journal of Econometrics* 45, 39–70.
- Huang, N., Shen, Z., Long, S., Wu, M., Shih, H., Zheng, Q., Yen, N., Tung, C., Liu, H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. в: *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*. Том 454. сс. 903–995.
- Janczura, J., Truck, S., Weron, R., Wolff, R., 2013. Identifying spikes and seasonal components in electricity spot price data: A guide to robust modelings. *Energy Economics* 38, 96–100.
- Karakatsani, N., Bunn, D., 2008. Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. *International Journal of Forecasting* 24, 764–785.
- Kim, C., Nelson, C., 1999. *Space models with regime switching*. MIT Press.
- Nowotarski, J., Tomczyk, J., Weron, R., 2013. Robust estimation and forecasting of the long-term seasonal component of electricity spot prices. *Energy Economics* 39, 13–27.
- Torres, M., Colominas, M., Schlotthauer, G., Flandrin, P., 2011. A complete ensemble empirical mode decomposition with adaptive noise. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 4144–4147.
- Weron, R., Bierbrauer, M., Trueck, S., 2010. Modeling electricity prices: jump diffusion and regime switching. *Phys. A* 336, 39–48.
- Wu, Z., Huang, N., 2009. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis* 1 (1), 1–41.
- Yu, L., Li, J., Tang, L., Wang, S., 2015. Linear and nonlinear granger causality investigation between carbon market and crude oil market: A multi-scale approach. *Energy Economics* 51, 300–311.
- Zachmann, G., 2013. A stochastic fuel switching model for electricity prices. *Energy Economics* 35, 5–13.

Спасибо за внимание!  
Вопросы?

Афанасьев Д.О.: [dmafanasyev@gmail.com](mailto:dmafanasyev@gmail.com), <http://dmafanasyev.ru>  
Федорова Е.А.: [ecolena@mail.ru](mailto:ecolena@mail.ru)