

**Федеральное государственное бюджетное образовательное
учреждение высшего образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

Фокин Н.Д.

**VAR-LASSO модель для российской экономики на
большом массиве данных**

Москва 2019

Аннотация. В работе строится большая векторная авторегрессия с L_1 регуляризацией на месячных данных российских макроэкономических показателей с учетом зависимости отечественной экономики от цен на нефть. Цель данной работы – продемонстрировать возможность и преимущества использования описанного подхода для прогнозирования российских макропараметров с помощью большого набора регрессоров, что, с теоретической точки зрения, должно улучшить прогнозы относительно моделей меньшей размерности. Использовались данные по индексам промышленного производства, ценам производителей, инвестициям, экспорту, импорту, процентным ставкам, показателям консолидированного и федерального бюджетов и др. Итоговая база данных состоит из 45 переменных протяженностью в полных 15 лет, на периоде 2002M01-2016M12 – 180 точек, из которых 44 регрессора являются эндогенными, а также один экзогенный – реальная цена на нефть. Используемый в данной работе подход L_1 регуляризации позволяет оценить модель на таком большом объеме данных даже в том случае, если наблюдений меньше чем число оцениваемых параметров. На основе оцененной модели в работе были построены псевдовневыборочные прогнозы индексов промышленного производства, и качество полученных прогнозов сопоставлено с качеством прогнозов по классической модели ARIMA. Результаты оцененной модели свидетельствуют в пользу превосходства оцененной модели над всеми рассмотренными бенчмарками.

Ключевые слова: индексы промышленного производства, ARIMA модель, VAR модель, VAR-LASSO модель, прогнозирование, импульсные отклики, долгосрочные мультипликаторы, цены на нефть

Abstract: This paper contains the construction of the large vector autoregression with L_1 regularization on monthly data of Russian macroeconomic indicators taking into account the high dependence of the domestic economy on oil prices. The point of this work is to demonstrate the possibility and advantages of using the described approach to forecast Russian macroparameters using a large set of regressors, which from the theoretical point of view should improve the forecasts in comparison with models with a smaller dimension. Data on indices of industrial production, producer prices, investments, exports, imports, interest rates, indicators of consolidated and federal budgets, etc. were used. The final database consists of 45 variables with a total length of 15 years, for the period 2002M01-2016M12 - 180 points, from 44 regressors are endogenous, as well as one exogenous - the real oil price. The L_1 regularization approach used in this paper allows us to estimate the model on such a large amount of data even if the observations are less than the number of estimated parameters. Based on the estimated model, we evaluate pseudo out of sample forecasts of indices of industrial production and the quality of the obtained forecasts was compared with the quality of the forecasts for the classical ARIMA model. The results of the evaluated model testify to the superiority of the evaluated model over all the benchmarks considered.

Key words: indices of industrial production, ARIMA model, VAR model, VAR-LASSO model, forecasting, impulse responses, long-run multipliers, oil prices

Фокин Н.Д. младший научный сотрудник лаборатории математического моделирования экономических процессов, Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации

Данная работа подготовлена на основе материалов научно-исследовательской работы, выполненной в соответствии с Государственным заданием РАНХиГС при Президенте Российской Федерации на 2018г.

Оглавление

1 Введение	5
2 LASSO и VAR-LASSO модели	6
3 Описание используемых данных и построение прогнозов	10
4 Заключение	19
Список использованных источников	20
Приложение	22

1 Введение

С учетом возросших объемов доступной экономической статистики за последние годы, большую актуальность приобретают эконометрические модели, позволяющие использовать большое число временных рядов, а также модели, в которых разделение на существенные¹ и несущественные переменные производится одновременно с оценкой параметров модели. В таком случае хорошим инструментом для моделирования является VAR-LASSO модель, которая позволяет использовать любое число регрессоров и их лагов в правой части, в том числе использование данной модели возможно в том случае, если число оцениваемых в правой части параметров превышает число наблюдений.

Основной акцент мы делаем на индексах промышленного производства, в частности, на основе качества прогнозов индексов мы делаем выводы о прогнозной силе модели и противопоставляем VAR-LASSO классическую ARIMA модель. Учитывая недавний переход Росстата к новому общероссийскому классификатору видов экономической деятельности (ОКВЭД 2), данные по индексам промышленного производства в старом ОКВЭД доступны за достаточно большой период времени, но ограничиваются 2016-ым годом. Данные по индексам промышленного производства на основе ОКВЭД 2 публикуются лишь с 2013-го года по настоящее время. В связи с данной ситуацией, мы жертвуем последними двумя годами статистики в пользу более длинных временных рядов, поскольку основной целью работы является анализ применимости рассматриваемого модельного класса для моделирования и прогнозирования российских макропоказателей, для чего необходимо использовать как можно больше статистических данных. Тем не менее, результаты работы могут найти практическую ценность и для анализа текущей макроэкономической ситуации. В частности, в работе строятся функции импульсного отклика ключевых индексов промышленного производства на изменение нефтяных цен, что позволяет понять степень зависимости отечественной экономики от данного внешнеэкономического фактора. Также возможно провести некоторую сцепку временных рядов в старой и новой методологии и получить тем самым единые продолжительные временные ряды

¹ В данной работе термин существенный регрессор используется не в смысле статистической значимости, а в том смысле, что коэффициент при переменной в результате оценки параметров LASSO регрессии не зануляется.

и использовать данную статистическую информацию для прогнозирования, но данная задача выходит за рамки настоящей работы, и мы оставляем её для будущих исследований.

При выборе спецификации кроме индексов промышленного производства в модель было включено большое число других переменных, в качестве ключевой детерминанты российской экономики в модель была экзогенно включена реальная цена на нефть, важность учета которой была продемонстрирована в большом числе российских работ (Казакова, 2009; Идрисов, Казакова, Полбин, 2014; Синельников-Мурылев, Дробышевский, Казакова, 2014; Полбин, 2017б).

2 LASSO и VAR-LASSO модели

Рассмотрим классическую линейную регрессию с p объясняющими регрессорами и T наблюдениями, удовлетворяющую условиям теоремы Гаусса-Маркова:

$$y = X\beta + \varepsilon, \quad (1)$$

где y – вектор наблюдений объясняемой переменной размерности $T \times 1$, X – матрица наблюдений объясняющих переменных размерности $T \times p$, β – вектор параметров при объясняющих переменных размерности $p \times 1$, ε – вектор случайных ошибок (шоков) размерности $T \times 1$, который пусть кроме удовлетворения условий теоремы Гаусса-Маркова также является нормально распределенной случайной величиной с нулевым математическим ожиданием и некоторой постоянной дисперсией σ_ε^2 , то есть $\varepsilon \sim N(0, \sigma_\varepsilon^2)$.

В такой модели оценка вектора параметров $\hat{\beta}$ имеет вид:

$$\hat{\beta} = (X'X)^{-1}X'y, \quad (2)$$

при условии, что матрица $X'X$ имеет полный столбцовый ранг, то есть определитель матрицы $X'X$ отличен от нуля, $\det(X'X) \neq 0$, что означает, что в пространстве регрессоров нет двух или более переменных, корреляция между которыми по модулю равна единице. Наличие таких переменных в прикладном исследовании означает то, что исследователи не внимательно отнеслись к подбору данных, например, попали в дамми-ловушку, однако такая ситуация в экономике при должном уровне внимания к данным возникает редко. Проблемой аналогичного характера и достаточно часто встречающейся на практике является проблема частичной (несовершенной) мультиколлинеарности, в отличие от вышеописанного примера полной (совершенной) мультиколлинеарности. Частичная мультиколлинеарность подразумевает корреляцию по модулю близкую к единице двух или более регрессоров, включенных в правую часть уравнения. В таком случае определитель матрицы $X'X$ близок к нулю, хоть и не равен ему, за счет чего в обратной матрице к $X'X$ могут получиться очень большие собственные числа, так как в оригинальной матрице они близки к нулю, а если собственный вектор некоторой матрице A равен z , то в матрице A^{-1} он равен $\frac{1}{z}$.

L_1 – регуляризация или LASSO-регрессия, представленная в работе (Tibshirani, 1996), на основе более ранней работы (Breiman, 1995) подразумевает добавление L_1 нормы вектора параметров к штрафной функции метода наименьших квадратов, для регрессии (1) LASSO-преобразование будет иметь вид:

$$Q(\beta) = \sum_{i=1}^T (y_i - X\beta)^2 + \lambda \|\beta\|_1 \rightarrow \min_{\beta} \quad (3)$$

Параметр $\lambda > 0$ является параметром тюннинга, выбор значений для которого будет рассмотрен несколько позже. В данной постановке минимизируемая функция является негладкой и вектор параметр не может быть найден в явном виде путем операций с матрицами данных y и X , в отличие от RIDGE-регрессии, которая является ничем иным, как регуляризацией Тихонова (Тихонов, 1965). Суть регуляризации заключается в том, что мы штрафует функцию потерь $Q(\beta)$ за большие значения параметров из вектора β . Тем самым сводим задачу безусловной минимизации к

задаче условной минимизации на некотором невыпуклом множестве значений вектора параметров β , за счет чего, в отличие от уже упомянутой RIDGE-регрессии, оценки параметров при несущественных регрессорах в LASSO-регрессии зануляются при оценивании модели, что является хорошим свойством L_1 – регуляризации. Таким образом, при оценивании вектора параметров β мы получим разреженный вектор из нулей и отличных от нуля чисел. Регрессоры при ненулевых параметрах будут являться существенными для рассматриваемой объясняемой переменной y .

Параметр λ обычно выбирается на основе кросс-валидации. В данной работе, так как модель строится на данных во времени, используется скользящая кросс-валидация, которая является обобщением классической K-блочной кросс-валидации на случай временных рядов. Ее механизм заключается в следующем: выборка разбивается на тренировочную и тестовую, фиксируется горизонт построения прогноза h , для одного из допустимых значений параметра λ оценивается модель и строится прогноз на h наблюдений вперед, затем для использованного значения параметра λ записывается значение среднеквадратичной ошибки прогноза (MSFE) или корня среднеквадратичной ошибки прогноза (RMSFE). На последнем шаге при оценивании модели мы используем значение параметра λ , минимизирующее ошибку прогноза.

Заметим, что в данной модели обязательной процедурой является стандартизация рассматриваемых переменных, то есть их центрирование и нормирование. В случае, если стандартизация проведена не будет, может возникнуть ситуация, в которой зануляются существенные регрессоры или наоборот. Пусть мы рассматриваем зависимость некоторой переменной, например, ВВП, измеряемого в миллиардах рублей, от инвестиций, измеряемых также в миллиардах рублей, и процентной ставки, измеряемой в процентах. В таком случае, если мы предположим, что оба регрессора значимо влияют на ВВП, при оценивании модели мы получим очень большие значения коэффициента при процентной ставке по сравнению с коэффициентом при инвестициях, так как изменение инвестиций на 1 рубль не должно существенно повлиять на изменения ВВП, а изменение процентной ставки на 1 процентный пункт представляется весьма существенным изменением и может вполне серьезно повлиять на динамику выпуска. Как было упомянуто ранее, в задаче оценки LASSO-регрессии (3) производится штрафование функции потерь при

слишком больших значениях коэффициентов из вектора β , таким образом, мы будем излишне штрафовать функцию потерь за слишком высокие значения параметра при процентной ставке. Кроме того, учитывая, что в модели есть множество эндогенных переменных, при выборе лучшей модели на стадии кросс-валидации нам нужно каким-то образом подобрать модель, которая лучше всего прогнозирует все переменные, а не конкретную переменную. В таком случае вполне разумным кажется складывать MSFE или RMSFE по всем полученным прогнозам, однако данные показатели зависят от дисперсии рассматриваемой переменной, и, чтобы мы могли их сложить и получить оценку качества всей модели в целом, нам необходимо привести переменные в стандартизированный вид.

VAR-LASSO модель впервые была рассмотрена в работе (Hsu, Hung, Chang, 2014), и чаще всего именно L_1 регуляризация используется в работах по построению больших векторных авторегрессий (Wilms, Croux, 2014, Kock, Callot, 2015). Дальнейшее изложение данной модели мы проведем на основе работы (Nicholson, et al., 2017), авторы которой написали пакет для языка R, в котором будет производиться оценивание модели настоящей работы. Запишем VARX (VAR с экзогенными переменными) модель для $t = 1 \dots T$ наблюдений, с k эндогенными переменными и m экзогенными переменными, для общности допустив разную глубину запаздываний – p для эндогенных и s для экзогенных:

$$\mathbf{y}_t = \mathbf{v} + \sum_{l=1}^p \Phi^{(l)} \mathbf{y}_{t-l} + \sum_{j=1}^s \beta^{(j)} \mathbf{x}_{t-j} + \mathbf{u}_t, \quad (4)$$

где \mathbf{y}_t – вектор эндогенных переменных,

\mathbf{v} – вектор констант,

$\Phi^{(l)}$ – матрица параметров при лаге порядка l эндогенных переменных размером $k \times k$,

\mathbf{x}_t – вектор экзогенных переменных,

$\beta^{(j)}$ – матрица параметров при лаге порядка j экзогенных переменных размером $k \times m$,

u_t – i.i.d. k -мерный белый шум с нулевым вектором математических ожиданий и некоторой ковариационной матрицей Σ_u .

В работе (Nicholson, et al., 2017) рассматривается несколько различных видов штрафных функций для VARX модели, которые позволяют выделять как стандартные группировки, так и другие весьма специфичные варианты штрафных функций. В данной работе мы остановимся на базовой штрафной функции, для которой задачи условной минимизации имеет вид:

$$Q(v, \Phi, \beta) = \sum_{t=1}^T \left(y_t - v - \sum_{l=1}^p \Phi^{(l)} y_{t-l} - \sum_{j=1}^s \beta^{(j)} x_{t-j} \right)^2 + \lambda (||\Phi||_1 + ||\beta||_1) \rightarrow \min_{v, \Phi, \beta} \quad (5)$$

3 Описание используемых данных и построение прогнозов

В приложении, в таблице А1 приведены участвующие в модели регрессоры, из них, в качестве переменных основного интереса данной работы выступают ИПП, индексы промышленного производства по большим отраслям: полезным ископаемым, обрабатываемым производствам, индексу электроэнергии, газа и воды; также в модели участвуют еще более мелкие индексы, а именно: добыча каменного угля, бурого угля и торфа, добыча сырой нефти и нефтяного (попутного) газа, добыча природного газа и газового конденсата, добыча металлических руд, добыча прочих полезных ископаемых, производство пищевых продуктов, включая напитки, и табака, производство кожи, изделий из кожи и производство обуви, производство прочих неметаллических минеральных продуктов, производство машин и оборудования. Кроме того, в таблице 1 присутствуют данные о зарплате, доходах, реальному курсу, показатели платежеспособности частного сектора, данные по экспорту и импорту, валютные резервы ЦБР, денежные агрегаты М0, М2, различные индексы цен и объемов работ, показатели торговли и федерального и консолидированного

бюджетов, уровень безработицы, процентная ставка МІАСР и реальная цена на нефть в качестве экзогенной переменной.

Также в таблице А1 приведена дополнительная информация об обработке переменной, а именно способах очищения от сезонности, источники данных и информация о дефлировании переменной на уровень цен. Кроме того, все ряды были протестированы тестом Филлипса-Перрона на наличие единичного корня, в последнем столбце таблицы приведена информация о порядке их интегрированности. В результате все переменные кроме ставки МІАСР включаются в модель в первых разностях логарифмов, сама переменная МІАСР включается в уровнях, а переменная уровня безработицы в разностях уровней. Все переменные центрированы на свои средние и нормированы на стандартные отклонения.

Перейдем к оцениванию модели, на первых двух третях выборки производится подбор параметров модели, а на последней трети производится скользящая кросс-валидация. Глубина запаздывания по эндогенным переменным составляет четыре лага, по цене на нефть - три лага и текущую цену на нефть (оценивание моделей с большим числом лагов показало отсутствие значимых лагов после четвертого для эндогенных регрессоров и после трех для цены на нефть). В результате в модели при 175 наблюдениях, непосредственно участвующих при оценивании и оставшихся после взятия разностей и лагов переменных, участвует 180 регрессоров. На рисунке 1 изображена матрица разреженности, полученная после оценивания модели. Горизонтально расположены уравнения для переменных по порядку из Таблицы А1.

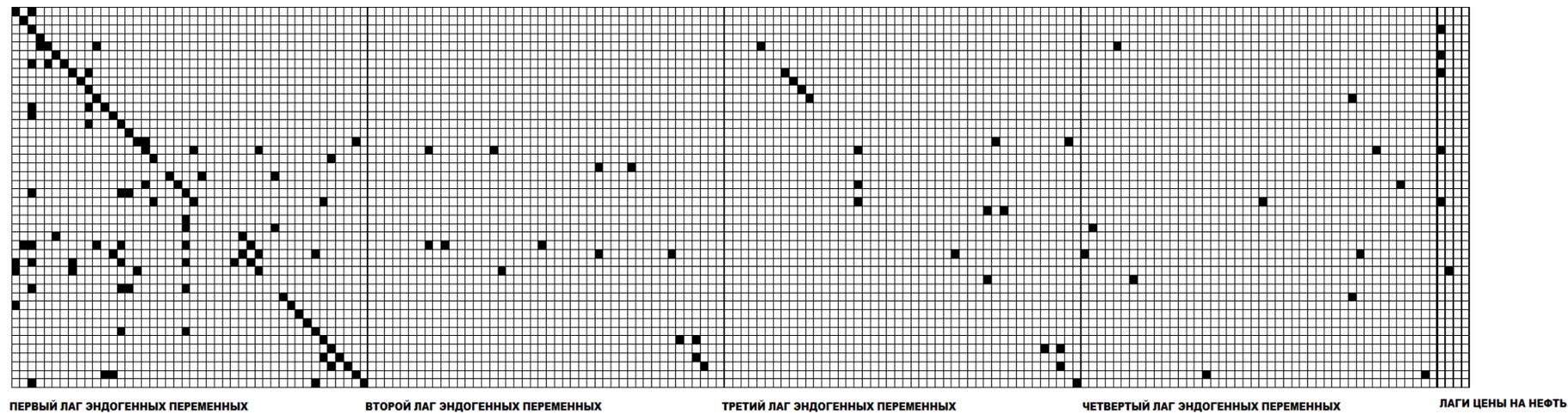


Рисунок 1 – Матрица разреженности, 4 лага каждой переменной. (Отличные от нуля оценки параметров помечены черным независимо от их величины, порядок переменных соответствует таблице А1).

В таблице 1 представлены суммы MSFE на периоде кросс-валидации для VAR-LASSO модели, а также для четырех бенчмарков: выборочного среднего, VAR модели с подбором лагов на основе критерия Акаике (Akaike, 1974) и Шварца (Schwarz, 1978), а также ARIMA модели. В результате VAR-LASSO модель оказывается наилучшей на последних двух третях выборки, худшей оказывается VAR модель с выбором лагов на основе информационного критерия, что скорее всего связано с тем, что критерий Акаике завышает число включаемых в модель лагов, за счет чего падает точность оценивания, а, следовательно, и прогноза, так как в классической VAR модели отсутствует регуляризация и на оценивание параметров необходимо некоторое разумное число наблюдений.

Таблица 1 – Сумма вневыборочных MSFE по всем стандартизированным переменным на периоде кросс-валидации (последние 2/3 выборки)

	VAR-LASSO	MEAN	AIC VAR	BIC VAR	ARIMA BIC
Сумма MSFE по всем стандартизированным переменным	44.754	47.107	133.431	46.105	46.905

Для наглядной визуализации качества прогнозов на рисунках 2 и 3 приведены псевдовневыборочные прогнозы аналогично работе (Полбин, Фокин, 2017) темпа роста индекса промышленного производства. Визуализация прогнозов по индексам обрабатывающих производств, добычи полезных ископаемых и электроэнергии, газа и воды не приводится с целью ограничиться самым хорошим примером, так как качество прогнозов по VAR-LASSO модели относительно ARIMA для общего ИПП является максимальным из всех. Прогнозы строятся по принципу разделения выборки на тренировочную и тестовую. Изначально на периоде 2002M01-2014M12 оценивается VAR-LASSO модель, пропорции для подбора параметров и кросс-валидации остаются теми же – две трети и одна треть соответственно. Затем строится вневыборочный прогноз на двенадцать месяцев (для первой итерации это весь 2015-ый год) и процедура повторяется пока последней точкой в тренировочной выборке не окажется последняя доступная нам точка, а именно декабрь 2016-го года.

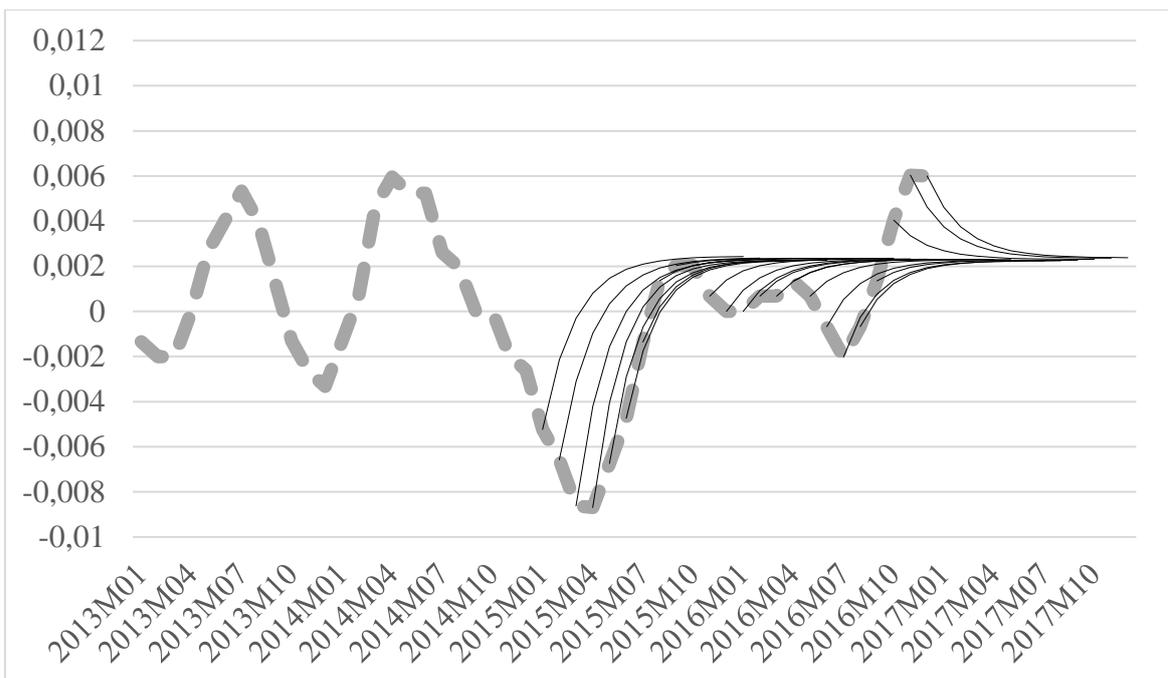


Рисунок 2 – Псевдовневыборочные прогнозы темпа роста ИПП VAR-LASSO
(пунктирная линия – истинные значения ряда)

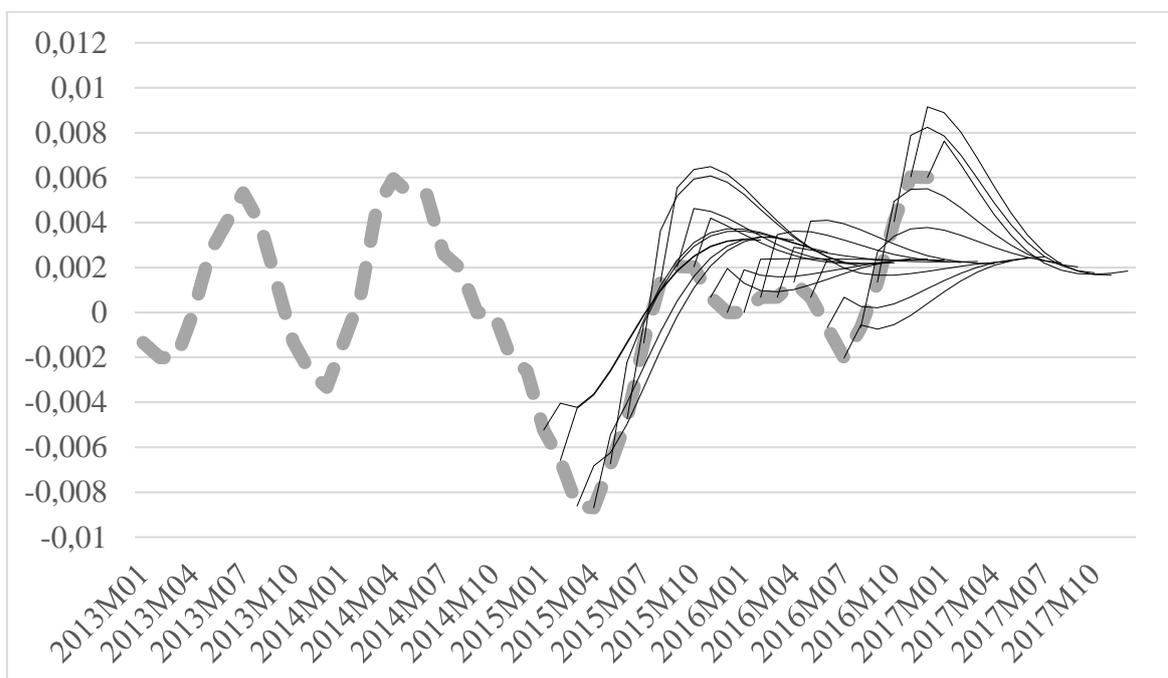


Рисунок 3 – Псевдовневыборочные прогнозы темпа роста ИПП ARIMA (пунктирная линия – истинные значения ряда)

В результате мы видим, что прогнозы по VAR-LASSO модели достаточно быстро стягиваются к среднему. При этом заметим, что цена на нефть не входит в уравнение для индекса промышленного производства, чем и может быть обусловлено

отсутствие видимой динамики прогнозов. Прогнозы по ARIMA модели оказываются весьма волатильными, однако сходимость к среднему также прослеживается. Количественно измерить качество прогнозов по общему и крупным индексам промышленного производства мы можем с помощью RMSFE и их отношений, представленных в таблицах 2 и 3 соответственно.

Таблица 2 – Вневыборочные RMSFE на периоде 2015-2016, нестандартизированные переменные (жирным шрифтом выделено наименьшее из двух RMSFE).

Шаг прогноза	ИПП VAR-LASSO	ИПП ARIMA	ИПП полез. ископ. VAR-LASSO	ИПП полез. ископ. ARIMA	ИПП обр. произв. VAR-LASSO	ИПП обр. произв. ARIMA	ИПП электр., газ, вода VAR-LASSO	ИПП электр., газ, вода. ARIMA
1	0.0022	0.0025	0.0033	0.0031	0.0052	0.0037	0.0032	0.0026
2	0.0033	0.0028	0.0033	0.0032	0.006	0.0047	0.0030	0.0037
3	0.0035	0.0032	0.0029	0.0031	0.0053	0.0052	0.0028	0.0038
4	0.0031	0.0035	0.0026	0.0028	0.005	0.0052	0.0027	0.0034
5	0.0027	0.0035	0.0026	0.0028	0.0039	0.0043	0.0027	0.0030
6	0.0023	0.0030	0.0021	0.0025	0.0029	0.0033	0.0027	0.0027
7	0.0023	0.0027	0.0022	0.0024	0.0023	0.0025	0.0027	0.0027
8	0.0023	0.0026	0.0022	0.0022	0.0021	0.0023	0.0026	0.0029
9	0.0024	0.0027	0.0018	0.0020	0.0019	0.0024	0.0027	0.0030
10	0.0024	0.0029	0.0018	0.0019	0.0017	0.0024	0.0028	0.0029
11	0.0024	0.0029	0.0018	0.0020	0.0017	0.0025	0.0029	0.0029
12	0.0023	0.0029	0.0018	0.0020	0.0018	0.0025	0.0029	0.0030
Среднее за все периоды	0.0022	0.0029	0.0024	0.0025	0.0033	0.0034	0.0028	0.0030

Таблица 3 – Отношение вневыборочных RMSFE BIGVAR к ARIMA на периоде 2015-2016, нестандартизированные переменные.

Шаг прогноза	ИПП VAR-LASSO к ARIMA	ИПП полез. ископ. VAR-LASSO к ARIMA	ИПП обр. произв. VAR-LASSO к ARIMA	ИПП электр., газ, вода VAR-LASSO к ARIMA
1	0.88	1.06	1.41	1.23
2	1.18	1.03	1.28	0.81
3	1.09	0.94	1.02	0.74
4	0.89	0.93	0.96	0.79

5	0.77	0.93	0.91	0.90
6	0.77	0.84	0.88	1.00
7	0.85	0.92	0.92	1.00
8	0.88	1.00	0.91	0.90
9	0.89	0.90	0.79	0.90
10	0.83	0.95	0.71	0.97
11	0.83	0.90	0.68	1.00
12	0.79	0.90	0.72	0.97
Среднее за все периоды	0.76	0.95	0.97	0.93

Исходя из результатов в таблицах выше, можно заключить, что даже в том случае (добыча полезных ископаемых и электроэнергия, газ и вода), когда в уравнение для индекса входят только собственные лаги (случай ARIMA), и не входят лаги других переменных, VAR-LASSO модель имеет более высокую прогнозную силу, нежели обычная ARIMA. Такой результат обусловлен тем фактом, что в VAR-LASSO модели при оценивании подбирается параметр тюнинга, и оценивание модели производится не только согласно некоторому методу, например, OLS или ML, а с учетом кросс-валидации, которая улучшает прогноз на будущее. В случае общего ИПП и обрабатывающих производств, где кроме авторегрессионной части в уравнении участвуют также лаги других регрессоров, преимущество над ARIMA формируется как за счет кросс-валидации, так и за счет большого числа информации используемой при построении прогноза. Самое большое преимущество в случае общего ИПП - в среднем на всех периодах прогноза VAR-LASSO превышает качество прогнозов относительно ARIMA почти на одну четверть – 24%. Также стоит учитывать, что при добавлении новой точки в обучающую выборку число значимых коэффициентов в том или ином уравнении может увеличиваться, что также потенциально может улучшать прогноз относительно базовой модели ARIMA.

С помощью импульсных откликов мы сможем давать прогноз для индексов промышленного производства в ОКВЭД 2 и других переменных на период после 2016-го года для тех рядов, в уравнении которых значим хотя бы один лаг нефтяных цен. Согласно матрице разреженности, изображенной на рисунке 1, цена на нефть является значимой для трех индексов промышленного производства, а именно для индекса обрабатывающих производств, индекса добычи сырой нефти и нефтяного (попутного газа) и индекса добычи металлических руд, кроме того цена на нефть

оказывается значимой для реального обменного курса, импорта и индекса цен промышленных товаров. В связи с этим перейдем к построению импульсных откликов данных показателей на перманентный 10% шок цен на нефть (рисунок 4) и расчету долгосрочных мультипликаторов (таблица 4).

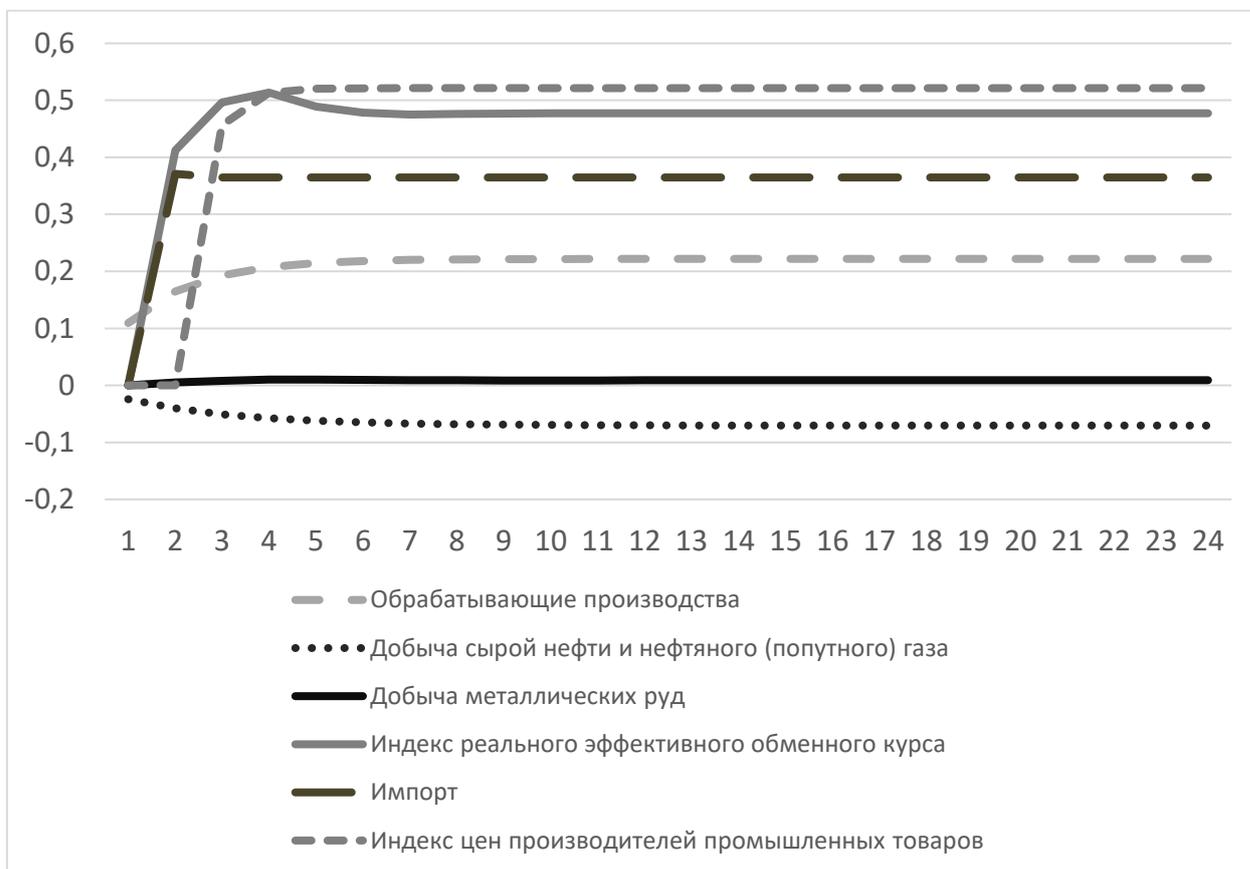


Рисунок 4 – Импульсные отклики рассмотренных показателей на 10% перманентный шок цен на нефть.

Таблица 4 – Таблица долгосрочных мультипликаторов рассмотренных переменных в ответ на 10% шок нефтяных цен.

Переменная	Величина долгосрочного мультипликатора по ценам на нефть
Обработывающие производства	2.22%
Добыча сырой нефти и нефтяного (попутного) газа	-0.71%
Добыча металлических руд	0.09%

Индекс реального эффективного обменного курса	4.77%
Импорт	3.65%
Индекс цен производителей промышленных товаров	5.21%

Сходимость для всех откликов наступает достаточно быстро – в течение полугода почти все отклики выходят на долгосрочный уровень. Таким образом, при построении прогнозов используя новый ОКВЭД полугодовой прогноз является уже долгосрочным и сценарные прогнозы при разных ценах на нефть релевантны именно на этот период. Отдельно отметим достаточно близкую оценку долгосрочной эластичности (4.4%) в работе (Скроботов, Фокин, 2018), в которой на достаточно схожем периоде (1999M1-2014M10) оценивалась долгосрочная эластичность реального курса рубля по ценам на нефть с помощью пороговой модели коррекции ошибок. В еще одной работе по оценке долгосрочной эластичности реального курса по ценам на нефть (Полбин, 2017а) также на достаточно близком к рассматриваемому в данной работе периоде (1999M1-2016M11), автор получает долгосрочный мультипликатор в размере 3.3%, что несколько ниже, чем оцененный мультипликатор с помощью VAR-LASSO модели.

На первый взгляд, знак мультипликатора добычи сырой нефти и нефтяного (попутного) газа может показаться странным, так как отрицательный коэффициент означает, что при росте цен на нефть добыча падает, а при падении – растет. Однако если изучить динамику индекса за периоды кризисов 2008-2009 годов и 2014-2015, можно увидеть, что во время колоссальных падений цен на нефть индекс в след за ценой не уменьшался. То есть во время падения цены на нефть нефтедобывающий сектор наращивает добычу с целью продавать больше нефти и получать дополнительную прибыль.

Остальные мультипликаторы интерпретируются вполне тривиально, в случае импорта интерпретация следующая: при росте цен на нефть фактически происходит трансфер богатства в страну-экспортер, за счет чего данная страна может потреблять больший объем товаров и услуг, в том числе и импорта, за счет чего увеличивается его физический объем. Аналогичная ситуация в случае цен производителей промышленных товаров: когда в стране увеличивается объем богатства, которое

потребители могут потратить на покупку товаров, производители увеличивают цены. В случае обрабатывающих производств интерпретация также вполне очевидна, так как обработка сырой нефти непосредственно является одним из видов обрабатывающего производства. Мультипликатор при индексе добычи металлических руд также положителен, но по отношению к остальным его значение очень мало – всего 0,09% при росте цены на нефть на 1%.

4 Заключение

В данной работе была построена большая VAR-LASSO модель с использованием 45 регрессоров месячной частоты. В частности, был рассмотрен случай, когда общее число оцениваемых параметров превышает число доступных наблюдений. На основе оцененной модели было проведено несколько экспериментов по тестированию качества прогнозов как модели в целом, так и на отдельных временных рядах российских индексов промышленного производства. В результате VAR-LASSO модель превосходит рассмотренные бенчмарки во всех случаях. По совокупной ошибке прогноза модели VAR-LASSO превосходит ARIMA, прогноз среднего, VAR на основе AIC и BIC. В случае отдельных индексов промышленного производства модель прогнозирует лучше, чем ARIMA. При переходе на ОКВЭД 2 ряды по индексам промышленного производства стали весьма короткими для полноценного эконометрического анализа. Однако используя построенные в данной работе импульсные отклики по ценам на нефть, можно строить сценарные прогнозы для импорта, реального курса рубля, индекса цен производителей промышленных товаров, а также индекса обрабатывающих производств, индекса добычи сырой нефти и нефтяного (попутного) газа и индекса добычи металлических руд. Результаты данной работы могут найти применение при построении краткосрочных (несколько месяцев) и среднесрочных (полгода) прогнозов различных макроэкономических показателей российской экономики.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Идрисов Г., Казакова М., Полбин А. Теоретическая интерпретация влияния нефтяных цен на экономический рост в современной России //Экономическая политика. – 2014. – Т. 5. – С. 150-171.
2. Казакова М.В. Вклад нефтегазового сектора в динамику экономических показателей в России и в мировой практике //Российский внешнеэкономический вестник. – 2009. – №. 8.
3. Полбин А.В. Моделирование реального курса рубля в условиях изменения режима денежно-кредитной политики //Вопросы экономики. – 2017. – №. 4. – С. 61-78.
4. Полбин А.В. Оценка влияния шоков нефтяных цен на российскую экономику в векторной модели коррекции ошибок //Вопросы экономики. – 2017. – Т. 10. – С. 27-49.
5. Полбин А., Фокин Н.К. вопросу о долгосрочной взаимосвязи реального потребления домохозяйств с реальным доходом в РФ //Экономическое развитие России. – 2017. – Т. 24. – №. 10. – С. 6-16.
6. Синельников-Мурылев С., Дробышевский С., Казакова М. Декомпозиция темпов роста ВВП России в 1999-2014 годах //Экономическая политика. – 2014. – №. 5. – С. 7-37.
7. Скроботов А.А., Фокин Н.Д. Тестирование асимметричной сходимости реального обменного курса к равновесному//Экономическая политика. – 2018. – Т. 13. – №. 3.
8. Тихонов А.Н. О некорректных задачах линейной алгебры и устойчивом методе их решения //ДАН СССР. – 1965. – Т. 163. – №. 3.
9. Akaike H. A new look at the statistical model identification //IEEE transactions on automatic control. – 1974. – Т. 19. – №. 6. – С. 716-723.
10. Breiman L. Better subset regression using the nonnegative garrote //Technometrics. – 1995. – Т. 37. – №. 4. – С. 373-384
11. Hsu N.J., Hung H.L., Chang Y.M. Subset selection for vector autoregressive processes using lasso //Computational Statistics & Data Analysis. – 2008. – Т. 52. – №. 7. – С. 3645-3657.
12. Kock A.B., Callot L. Oracle inequalities for high dimensional vector autoregressions //Journal of Econometrics. – 2015. – Т. 186. – №. 2. – С. 325-344.

13. Nicholson W.B., Matteson D.S., Bien J. VARX-L: Structured regularization for large vector autoregressions with exogenous variables //International Journal of Forecasting. – 2017. – T. 33. – №. 3. – C. 627-651.
14. Schwarz G. et al. Estimating the dimension of a model //The annals of statistics. – 1978. – T. 6. – №. 2. – C. 461-464.
15. Tibshirani R. Regression shrinkage and selection via the lasso //Journal of the Royal Statistical Society. Series B (Methodological). – 1996. – C. 267-288.
16. Wilms I., Croux
17. C. Sparse cointegration. – 2014.

Приложение

Таблица А1 - Описание используемых данных

Показатель	Имя переменной	Доп. информация	Интегрированность согласно РР тесту, 5% уровень
ИПП	<i>IPP</i>	Очищение от сезонности - SOPHIST	I(1)
ИПП полезные ископаемые	<i>POLEZ</i>	Очищение от сезонности - SOPHIST	I(1)
ИПП обрабатывающие производства	<i>OBR</i>	Очищение от сезонности - SOPHIST	I(1)
ИПП электроэнергия, газ, вода	<i>ELECTR</i>	Очищение от сезонности - SOPHIST	I(1)
Добыча каменного угля, бурого угля и торфа	<i>MCLP2</i>	Очищение от сезонности - SOPHIST	I(1)
Добыча сырой нефти и нефтяного (попутного) газа	<i>ECOG2</i>	Очищение от сезонности - SOPHIST	I(1)
Добыча природного газа и газового конденсата	<i>ENGC2</i>	Очищение от сезонности - SOPHIST	I(1)
Добыча металлических руд	<i>MMET2</i>	Очищение от сезонности - SOPHIST	I(1)
Добыча прочих полезных ископаемых	<i>OMAQ2</i>	Очищение от сезонности - SOPHIST	I(1)
Производство пищевых продуктов, включая напитки, и табака	<i>MFBT2</i>	Очищение от сезонности - SOPHIST	I(1)
Производство кожи, изделий из кожи и производство обуви	<i>MLPP2</i>	Очищение от сезонности - SOPHIST	I(1)
Производство прочих неметаллических минеральных продуктов	<i>MONMP2</i>	Очищение от сезонности - SOPHIST	I(1)
Производство машин и оборудования	<i>MERP2</i>	Очищение от сезонности - SOPHIST	I(1)

Реальная зарплата, индекс	WAGE	Очищение от сезонности - SOPHIST	I(1)
Реальный доход, индекс	INCOME	Очищение от сезонности - SOPHIST	I(1)
ИПЦ	CPI	Очищение от сезонности - SOPHIST	I(1)
Индекс реального эффективного обменного курса	REER	IMF	I(1)
Задолженность покупателей	DBT_P_M	Очищение от сезонности – X12 ARIMA EViews, поделен на ИПЦ, SOPHIST	I(1)
Дебиторская задолженность	DBT_T_M	Очищение от сезонности – X12 ARIMA EViews, поделен на ИПЦ, SOPHIST	I(1)
Просроченная дебиторская задолженность	DBT_UNP_M	Очищение от сезонности – X12 ARIMA EViews, поделен на ИПЦ, SOPHIST	I(1)
Экспорт	EXPORT	Очищение от сезонности – X12 ARIMA EViews, поделен на ИПЦ, SOPHIST	I(1)
Валютные резервы ЦБР	ICR_M	Очищение от сезонности – X12 ARIMA EViews, SOPHIST	I(1)
Импорт	IMPORT	Очищение от сезонности – X12 ARIMA EViews, поделен на ИПЦ, SOPHIST	I(1)
Задолженность в бюджет	LIAB_B_M	Очищение от сезонности – X12 ARIMA	I(1)

		EIEWS, поделен на ИПЦ, SOPHIST	
Задолженность поставщикам	<i>LIAB_S_M</i>	Очищение от сезонности – X12 ARIMA EIEWS, поделен на ИПЦ, SOPHIST	I(1)
Кредиторская задолженность	<i>LIAB_T_M</i>	Очищение от сезонности – X12 ARIMA EIEWS, поделен на ИПЦ, SOPHIST	I(1)
Просроченная кредиторская задолженность	<i>LIAB_UNP_M</i>	Очищение от сезонности – X12 ARIMA EIEWS, поделен на ИПЦ, SOPHIST	I(1)
Денежный агрегат M0	<i>M0</i>	Очищение от сезонности – X12 ARIMA EIEWS, поделен на ИПЦ, SOPHIST	I(1)
Индекс цен на строительно-монтажные работы	<i>PRICE_BUILD</i>	Очищение от сезонности – X12 ARIMA EIEWS, SOPHIST	I(1)
Денежный агрегат M2	<i>M2_M</i>	Очищение от сезонности – X12 ARIMA EIEWS, поделен на ИПЦ, SOPHIST	I(1)
Индекс цен производителей промышленных товаров	<i>PRICE_PROD</i>	Очищение от сезонности – X12 ARIMA EIEWS, SOPHIST	I(1)
Индекс цен на грузовые перевозки	<i>PRICE_TRANS P</i>	Очищение от сезонности – X12 ARIMA EIEWS, SOPHIST	I(1)
Международные резервы ЦБР	<i>IR_M</i>	Очищение от сезонности – X12	I(1)

		ARIMA EIEWS, поделен на ИПЦ, SOPHIST	
Индекс реального объема сельскохозяйственного производства	<i>AGR_M_DIRI</i>	Очищение от сезонности - SOPHIST	I(1)
Индекс реального объема работ, выполненных по виду деятельности "Строительство"	<i>CNSTR_M_DIRI</i>	Очищение от сезонности - SOPHIST	I(1)
Ввод в действие жилых домов	<i>CONSTR_M_N AT</i>	Очищение от сезонности - SOPHIST	I(1)
Индекс пассажирооборота транспорта общего пользования	<i>TRP_M_PASS_ DIRI</i>	Очищение от сезонности – X12 ARIMA EIEWS, поделен на ИПЦ, SOPHIST	I(1)
Индекс реального оборота розничной торговли	<i>RTRD_M_DIRI</i>	Очищение от сезонности – X12 ARIMA EIEWS, поделен на ИПЦ, SOPHIST	I(1)
Расходы консолидированного бюджета	<i>CB_E</i>	Очищение от сезонности – X12 ARIMA EIEWS, поделен на ИПЦ, SOPHIST	I(1)
Доходы консолидированного бюджета	<i>CB_I</i>	Очищение от сезонности – X12 ARIMA EIEWS, поделен на ИПЦ, SOPHIST	I(1)
Расходы федерального бюджета	<i>FED_E</i>	Очищение от сезонности – X12 ARIMA EIEWS, поделен на ИПЦ, SOPHIST	I(1)
Доходы федерального бюджета	<i>FED_I</i>	Очищение от сезонности – X12 ARIMA	I(1)

		EViews, поделен на ИПЦ, SOPHIST	
Уровень безработицы	<i>UNEMP</i>	-	I(1)
Ставка MIACR	<i>MIACR</i>	Усреднение однодневных ставок, cbr.ru	I(0)
Реальная цена на нефть	<i>POIL</i>	Номинальная цена, поделенная на долларовой ИПЦ. EIA.GOV, FRED	I(1)