

УДК 314.748

ДОХОЛЯН ВЛАДИМИР СЕРГЕЕВИЧ

младший научный сотрудник лаборатории математического
моделирования экономических процессов РАНХиГС,
e-mail: doholyan.vs@phystech.edu

ПОЛБИН АНДРЕЙ ВЛАДИМИРОВИЧ

к.э.н., заведующий лабораторией математического моделирования экономических
процессов РАНХиГС, заведующий лабораторией макроэкономического
моделирования ИЭП им.Гайдара,
e-mail: apolbin@ranepa.ru

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ЦИКЛИЧЕСКОЙ БЕЗРАБОТИЦЫ

Аннотация. *Целью работы* является исследование зависимости между циклическими компонентами ВВП и безработицы для прогнозирования последней с помощью методов машинного обучения. В данной работе используются различные методы машинного обучения, в том числе нейронные сети. Также были рассмотрены традиционные эконометрические модели. Модели, основанные на нейронных сетях, показали качество, превосходящее традиционные модели, однако, предсказательная способность моделей, основанных на бинарных решающих деревьях, оказалась ниже, чем линейной регрессии. Циклическая безработица является одной из главных характеристик макроэкономической нестабильности, свидетельством неполной занятости ресурсов. Прогнозирование циклической безработицы представляет практическую ценность с точки зрения использования при построении краткосрочных прогнозов безработицы и реакции безработицы на изменение ВВП. Главным выводом этой работы можно считать подтверждение возможности использования методов машинного обучения для получения прогнозов циклической безработицы, их конкурентоспособность по сравнению с более традиционными линейными регрессиями.

Ключевые слова: прогноз безработицы, закон Оукена, циклическая безработица, машинное обучение, нейронные сети.

DOKHOLYAN VLADIMIR SERGEEVICH

Junior researcher of the laboratory of mathematical
simulation of economic processes of Ranepa,
e-mail: doholyan.vs@phystech.edu

POLBIN ANDREY VLADIMIROVICH

Ph. D. in Economics, head of the laboratory of mathematical modeling of economic
processes Ranepa, head of the laboratory of macroeconomic
simulations of the IEP.Gaidar's,
e-mail: apolbin@ranepa.ru

THE APPLICATION OF MACHINE LEARNING METHODS FOR PREDICTING CYCLICAL UNEMPLOYMENT

Abstract. *The aim* of the work is to study the relationship between the cyclic components of GDP and unemployment to predict the latter using machine learning methods. In this paper, we use a variety of machine learning methods, including neural networks. Traditional econometric

models were also considered. Models based on neural networks showed quality superior to traditional models, however, the predictive ability of models based on binary decision trees turned out to be lower than linear regression. Cyclical unemployment is one of the main characteristics of macroeconomic instability, indicating underemployment of resources. Forecasting cyclical unemployment is of practical value in terms of the use of short-term unemployment forecasts and the response of unemployment to changes in GDP. The main conclusion of this work is the confirmation of the possibility of using machine learning methods to obtain forecasts of cyclical unemployment, their competitiveness in comparison with more traditional linear regressions.

Keywords: *unemployment forecast, oaken's law, cyclic unemployment, machine learning, neural networks.*

1. Введение

Уровень безработицы является одним из главных макроэкономических показателей отражающий уровень экономического развития страны. Особое внимание заслуживает циклическая безработица, представляющая собой разницу между фактическим и естественным уровнями безработицы. Под естественным уровнем безработицы понимается уровень безработицы в долгосрочной перспективе. Циклическая безработица возникает в тех случаях, когда падение совокупного спроса на выпускаемую продукцию вызывает падение совокупного спроса на труд.

В тоже время, зависимость между производством и безработицей представляет как теоретический, так и практический интерес [3,9,20]. Известный закон Оукена связывает линейно циклическую безработицу и циклическую компоненту выпуска, представляющую собой отклонение фактического выпуска от потенциального. Под потенциальным выпуском понимается некоторый долгосрочный уровень выпуска, который определяется производительностью экономики и растет с ростом уровня технологий.

Закон Оукена имеет важную макроэкономическую интерпретацию: для того, чтобы уменьшить безработицу необходимо простимулировать спрос. Сдвиг совокупного спроса вызывают изменения в производстве, что, в свою очередь, приводит к тому, что фирмы нанимают новых работников. Прогнозирование циклической безработицы несёт в себе неоспоримую практическую ценность как для построения безусловных прогнозов безработицы, так и сценарных прогнозов условно на заданной траектории снижения или роста агрегированного выпуска.

Однако закон Оукена в своей линейно форме может не выполняться (особенно в периоды кризисов) по ряду объективных причин:

- издержки обучения новых работников могут быть весьма высоки, что удерживает фирмы от увольнения уже имеющихся работников при ВВП ниже нормального уровня
- фирмы могут изменять не численность работников, а количество рабочих часов, т.е. переводить работников на сверхурочную работу если ВВП превышает нормальный уровень
- увольнения работников не является непрерывным процессом, часто увольнения имеют массовый характер.

Данные факторы могут рационализировать разработку различного рода нелинейных моделей для прогнозирования безработицы, таких как нейронные сети, которые, согласно работе [5], могут приблизить любую гладкую функцию с заданной точностью. Однако небольшое количество наблюдений в макроэкономических временных рядах обуславливает наличие проблемы проклятия размерности, когда информации в имеющейся выборке данных попросту недостаточно для качественной оценки большого количества параметров, что приводит к проблеме переобучения. Переобучение представляет собой явление, когда модель хорошо прогнозирует на объектах из обучающей выборки, но крайне плохо предсказывает для объектов, не учувствовавших в обучении (тестовой выборке). В последнее время набирают популярность методы машинного обучения, которые позволяют моделировать более сложные, нелинейные зависимости, чем традиционные модели, и стараются решить проблему переобучения за счёт регуляризации. В качестве примеров успешного применения моделей машинного обучения для зарубежных экономик, среди многих, можно привести [8, 11, 17, 21]. В работе [17] нейронные сети используются для прогнозирования инфляции в США, Японии и Еврозоны, в

работе [21] была предложена нейронная сеть для прогнозирования ИПЦ Греции, в работе [8] с помощью моделей случайного леса прогнозировалось ВВП, в работе [11] нейронные сети применялись для выделения сезонной компоненты.

Методы машинного обучения нашли интересные применения и для анализа российских экономических временных рядов. И.Ю. Золотова и В.В. Дворкин в своей работе [4] использовали однослойный перцептрон для прогнозирования цен на электроэнергию. Полученная модель краткосрочного прогнозирования цен на электроэнергию смогла дать достаточно точные предсказания значения цен в условиях волатильности и неопределенности в разные сезоны года.

Работа [1] представляет собой обзор и сравнение различных моделей машинного обучения для прогнозирования инфляции России. В частности, автор рассмотрел модели с регуляризацией, а также ансамблевые методы (случайный лес и бустинг). Результаты показали, что модели, основанные на решающих деревьях (случайный лес бустинг) могут, как минимум, не хуже предсказывать инфляцию, чем традиционные эконометрические модели, такие как случайное блуждание и авторегрессия. Как итог, данное исследование подтверждает возможность более точного прогнозирования инфляции в России с помощью методов машинного обучения.

В работе [2] авторы использовали нейронную сеть для решения задачи классификации платежеспособности предприятия по финансовым показателям их публичных отчетностях. В работе [7] демонстрируется превосходство VAR-LASSO-модели над более традиционными моделями, такими как VAR-модель с подбором лагов на основе критериев Акаике и Шварца, а также ARIMA для задачи прогнозирования индексов промышленного производства.

Однако, методы машинного обучения для моделирования безработицы РФ, насколько нам известно, ещё не применялись, что вносит элемент новизны настоящего исследования.

Данная работа имеет следующую структуру: во втором разделе кратко описаны методы, используемые в работе, в третьем описаны данные и методология, в четвертом показаны результаты работы всех моделей, а пятый раздел содержит заключение данной работы.

2. Краткий обзор методов

Линейные модели

Все линейные модели имеют следующий вид:

$$y_t = \alpha + \bar{b} \times \bar{u}_t + \varepsilon_t, \text{ где}$$

y_t - целевая переменная с момент времени t ;

α - константа;

\bar{b} - вектор весов;

\bar{u}_t - вектор признаков для периода t ;

ε_t - случайная ошибка в момент времени t .

Разница между различными моделями заключается в способе получения коэффициентов. В классическом методе наименьших квадратов минимизируется среднеквадратическая ошибка:

$$Q = \sum_{t=0}^n (y_t - \hat{y}_t)^2 = \sum_{t=0}^n (y_t - \alpha - \bar{b} \times \bar{u}_t)^2$$

Одним из главных преимуществ этого метода является то, что решение этой оптимизационной задачи можно выписать аналитически.

$$b = (X^T X)^{-1} X^T y; \text{ где}$$

b - вектор коэффициентов;

X - матрица признаков;

y - вектор наблюдений объясняемой переменной.

Однако, при абсолютной мультиколлинеарности признаков матрица может оказаться необратимой и решения не будет существовать. И даже если мультиколлинеарность не будет абсолютной, то это может привести к большим стандартным ошибкам, а оценки параметров будут крайне неустойчивым. Поэтому появление новых данных может привести к полному пересмотру значений коэффициентов при регрессорах. Штраф за большие значения коэффициентов, т.е. регуляризация, может решить эту проблему. Одним из вариантов является L2 регуляризация, в рамках которой функция потерь будет иметь следующий вид:

$$Q = \sum_{t=0}^n (y_t - \alpha - \bar{b} \times \bar{u}_t)^2 + \lambda \sum_{i=1}^p b_i^2,$$

Где λ определяет силу регуляризации. Этот параметр необходимо определить заранее, как правило, с помощью кросс-валидации. При больших λ модель сводится в константу. Важным достоинством этой регуляризации является единственность решения, причем это решение можно найти аналитически:

$$b = (X^T X + \lambda I)^{-1} X^T y, \text{ где}$$

I – единичная матрица соответствующего размера

Другим вариантом является L1 регуляризация и функция потерь принимает вид:

$$Q = \sum_{t=0}^n (y_t - \alpha - \bar{b} \times \bar{u}_t)^2 + \lambda \sum_{i=1}^p |b_i|$$

В этом случае аналитического решения не существует. Важно, что L1 регуляризация обнуляет коэффициенты при незначительных или избыточных признаках. Поэтому ее можно использовать для предварительного отбора признаков для других моделей.

Можно использовать сразу оба регуляризатора, в этом случаи функция потерь примет вид:

$$Q = \sum_{t=0}^n (y_t - \alpha - \bar{b} \times \bar{u}_t)^2 + \lambda_1 \sum_{i=1}^p |b_i| + \lambda_2 \sum_{i=1}^p b_i^2$$

Линейная модель с такой функцией ошибки называется эластичной сетью (elastic net)

Ансамбли

Следующие две модели представляют собой композицию более простых базовых моделей – решающих деревьев. Решающее дерево представляет собой граф в виде бинарного (как правило) дерева, состоящего из корневых(внутренних) вершин и листьев. В каждой внутренней вершине записано условие вида $x^j > t$ (где x^j – это j -ый признак, а t – некоторый порог), а в каждом листе прогноз. В качестве прогноза в листе берется среднее значение целевой переменной в этом листе. Построение модели проходит последовательно, от корня к листьям. На каждом этапе выборка в текущем узле разбивается на 2, которые попадают в правое и левое поддерево согласно условию в узле. Процесс продолжается до критерия остановки. Критерием остановки может служить, например, глубина дерева или количество объектов при котором узел уже считается листом.

Пусть в вершину m попало множество X_m объектов из обучающей выборки. Тогда параметры j и t подбираются так, чтобы минимизировать критерий ошибки:

$$Q(X_m, j, t) \rightarrow \min$$

Параметры j и t подбираются перебором. После чего получаются 2 поддерева:

$$X_l = \{x \in X_m | [x^j \leq t]\}, X_r = \{x \in X_m | [x^j > t]\}$$

Критерий ошибки можно записать следующим образом:

$$Q(X_m, j, t) = \frac{|X_l|}{|X_m|} H(X_l) + \frac{|X_r|}{|X_m|} H(X_r),$$

где $H(X)$ – критерий информативности.

Он характеризует разброс ответов (однородность) объектов. Для задачи регрессии в качестве критерия информативности служит дисперсия:

$$H(X) = \frac{1}{N} \sum_{i \in X} (y_i - \bar{y}(X))^2$$



Схема 1: Пример решающего дерева

Случайный лес

Случайный лес (RF) есть композиция решающих деревьев. Предсказание RF получается путем усреднения предсказания всех решающих деревьев.

$$RF(\bar{u}_t) = \frac{1}{N} \sum_{i=1}^N DT(\bar{u}_t), \text{ где}$$

DT – модель бинарного дерева

Идея подобных алгоритмов заключается в уменьшении разброса ошибки при предсказании. Ошибку модели на новых данных можно представить как сумму 3 компонент:

- шум – это характеристика данных и проявляется для любой модели;
- смещение – это среднее отклонение (для различных обучающих выборок) от прогноза идеальной модели;
- разброс – это дисперсия ответов моделей, обученных на различных обучающих выборках.

Смещение композиции деревьев совпадает со смещением отдельного решающего дерева, но разброс уже вычисляется следующим способом:

$$(\text{разброс композиции}) = \frac{1}{N} (\text{разброс отдельного } DT) + (\text{корреляция между } DF)$$

Если деревья не зависимы, то разброс композиции будет в N раз меньше. Для того, чтобы уменьшать зависимость между отдельными деревьями, каждый из них обучается на подмножестве обучающей выборки. Также можно при построении решающих деревьев в каждом узле проверять не все признаки, а некоторое их подмножество. Это делает деревья более независимыми и соответственно уменьшает разброс модели случайного леса.

Градиентный Бустинг

Случайный лес – композиция глубоких деревьев, которые строятся независимо друг от друга, то есть процесс построения дерева ненаправленный и поэтому для решения задач требуется огромное количество деревьев. Эту проблему способен решить метод градиентного бустинга. Идея бустинга (обозначим как α) заключается в том, что деревья строятся последовательно и каждое следующее дерево стремится исправить ошибку уже построенной композиции.

Процесс построения этой модели можно описать следующим способом. Сначала строится базовый алгоритм, как правило, он крайне прост:

$$h_0(x) = \frac{1}{l} \sum_{i=1}^l y_i$$

На этом шаге $\alpha_0 = b_0$, затем каждый следующий алгоритм обучается так, чтобы уменьшить ошибку композиции:

$$\sum_{i=1}^l L(y_i, a_{N-1}(x_i) + b_N(x_i)) \rightarrow \min$$

После чего новый алгоритм добавляется к композиции с некоторым весом $\mu < 1$:

$$a_N(x) = a_{N-1}(x) + \mu * b_N(x)$$

Нейронная сеть

Главная причина роста популярности искусственных нейронных сетей состоит в том, что они могут аппроксимировать практически любую нелинейную функцию с любой степенью точности. Следовательно, при применении к временному ряду, который характеризуется действительно нелинейной зависимостью, модели ИНН могут показать лучший результат по сравнению с линейными моделями. Одним из главных недостатков ИНН является их сложность и неинтерпретируемость коэффициентов. По этой причине ИНН часто рассматривают как модель «Черного ящика», в основном для целей прогнозирования.

Применение полноценных нейронных сетей затруднено вследствие небольшого размера обучающей выборки. Philip Hans Franses, Dick van Dijk в «Nonlinear Time Series Models in Empirical Finance» рассмотрели следующую модель основанную на нейронной сети:

$$y_t = x_t' \phi + \sum_{k=1}^q \beta_k G(x_t' \gamma_k) + \varepsilon_t, \text{ где}$$

x_t – вектор признаков;

G – функция активации;

β_k – вес k -го нейрона;

γ_k – вектор коэффициентов в k -м нейроне.

Первое слагаемое представляет собой обычную линейную регрессию, а второе есть одно-слойная нейронная сеть с q скрытыми нейронами.

Параметры ИНН модели могут быть получены путем минимизации функции потерь. В качестве функции потерь, как правило, используется среднеквадратическая ошибка.

$$Q_n(\theta) = \sum_{t=1}^n [y_t - \hat{y}_t(\theta)]^2, \text{ где}$$

θ – параметры модели

Данный функционал минимизируется с помощью метода обратного распространения ошибки. Однако, как отмечалось выше, превосходный результат на обучающей выборке не является гарантией что ИНН будет хорошо работать на новых данных, так как это может быть вызвано переобучением.

Большие веса у коэффициентов способствуют переобучению. Поэтому штраф за большие веса приводит к повышению качества предсказаний модели. Коэффициенты регуляризации необходимо подбирать заранее. Необходимо заметить, что при использовании регуляризации обязательно надо нормировать входные данные.

$$Q_n(\theta) = \sum_{t=1}^n [y_t - \hat{y}_t(\theta)]^2 + r_\phi \sum_{i=0}^k \phi_i^2 + r_\beta \sum_{j=0}^q \beta_j^2 + r_\gamma \sum_{j=0}^q \sum_{i=0}^k \gamma_{i,j}^2, \text{ где}$$

$r_\phi, r_\beta, r_\gamma$ – коэффициенты регуляризации.

Существует несколько распространённых методов масштабирования. Можно преобразовывать данные к интервалу $[0,1]$, применяя

$$z_t^* = \frac{z_t - \min(z_t)}{\max(z_t) - \min(z_t)}$$

Другой способ состоит в том, чтобы переменные имели нулевое среднее и стандартное отклонение равное 1. Для этого применяется преобразование:

$$z_t^* = \frac{z_t - \bar{z}_t}{\sigma(z_t)}$$

Выбор функции масштабирования зависит сугубо от начальных зависимостей в данных.

Как уже было отмечено выше, если количество нейронов в скрытом слое q станет значительным, то модель может стать слишком гибкой и некоторые нейроны будут подстраиваться под шум. Поэтому крайне важно правильно подобрать количество нейронов в скрытом слое и силу регуляризации. Одним из способов уменьшения вероятности переобучения, который часто применяется, является так называемая кросс-валидация. В этом случае имеющиеся наблюдения делятся на обучение и валидацию. Параметры ANN оцениваются по обучающей выборке, но во время итеративной оптимизации также регистрируется функция потерь (сумма квадратов ошибок) на валидации. За итоговые параметры модели принимаются не те, которые привели к минимальной ошибке на обучающей выборке, а те, что минимизировали функцию потерь на валидации. В этом случае параметры могут описывать общие нелинейные закономерности, которые присутствуют в данных, а любое дальнейшее улучшение подгонки на обучающей выборке, которое может быть достигнуто, приведет к тому, что сеть начнет описывать выбросы или другие нетипичные события. [10, 12]

3. Обработка данных и методология

Данные представляют собой квартальные данные безработицы и ВВП, взятые с сайта Росстата. ВВП приведен к уровню цен 2008 года. Исходные данные представлены на графиках ниже:

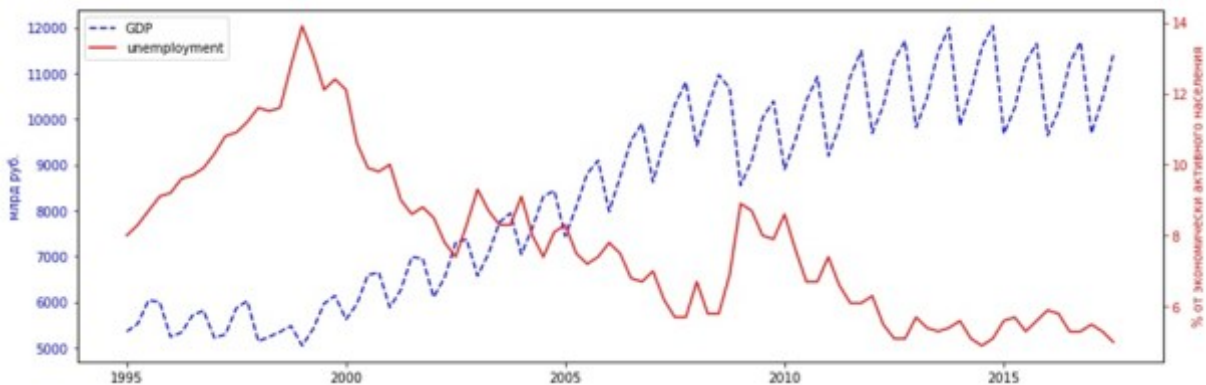


График 1: Временные ряды реального ВВП и безработицы

Поскольку данные демонстрируют ярко выраженную сезонность, сезонная компонента была удалена с помощью эконометрического пакета Eviews10 (процедура Census X-12) на первом шаге работы с данными. Перед этим временной ряд ВВП был прологарифмирован для того, чтобы стабилизировать дисперсию, предполагалась аддитивность сезонной компоненты для логарифма ВВП, и мультипликативная сезонная компонента для уровня безработицы. Далее необходимо удалить тренд. Для выделения тренда был использован фильтр Ходрика-Прескотта [13]. Этот фильтр представляет собой метод сглаживания временного ряда, который используется для выделения длительных тенденций временного ряда – трендов. Сглаженный ряд, с одной стороны, должен быть достаточно близок к исходному ряду, то есть необходимо минимизировать сумму квадратов отклонений. С другой стороны, сглаженный ряд должен быть достаточно гладким, то есть сам ряд должен изменяться как можно менее резко. Элементы сглаженного ряда выбираются таким образом, чтобы минимизировать следующий функционал:

Параметр λ можно интерпретировать как силу сглаживания: при $\lambda = 0$ имеем $s_t = y_t$, с увеличением λ ряд все сильнее сглаживается, а при $\lambda = \infty$ ряд вырождается в линейный тренд. Рекомендованное значение λ для квартальных данных равно 1600. [13]

Таким образом, после удаления тренда мы получили циклические компоненты ВВП и безработицы. Представим эти ряды на одном графике.

$$\sum_{t=1}^T (y_t - s_t)^2 + \lambda \sum_{t=2}^{T-1} ((s_{t+1} - s_t) - (s_t - s_{t-1}))^2 \rightarrow \min$$

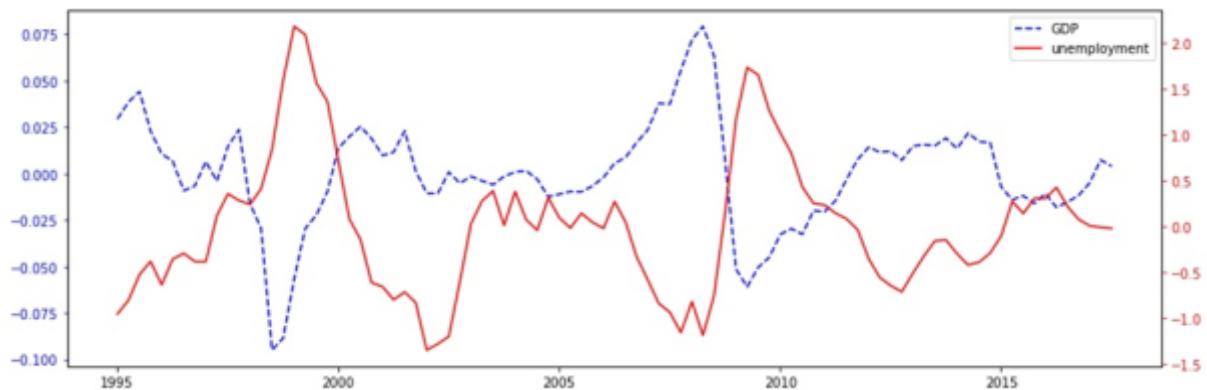


График 2: циклические компоненты ВВП и безработицы

Стоит отметить, что визуально видна отрицательная корреляция. Подъем одного показателя совпадают с падением другого, кризисы 1998, 2008 и 2014 года сопровождаются резким увеличением безработицы и столь же резким падением реального ВВП.

После выделения циклических компонент ВВП и безработицы, необходимо определить пространство признаков (объясняющих переменных). В качестве признаков были выбраны: запаздывания(лаги) безработицы и ВВП, их скользящие средние за последние несколько периодов, а также темпы роста за последние несколько периодов. Так как некоторые модели чувствительны к масштабу, то данные были масштабированы на отрезок $[0,1]$.

Вследствие того, что обучающая выборка мала и многие признаки были сильно коррелированы, то было необходимо провести отбор признаков. Для этих целей была использована LASSO регрессия. Как уже говорилось выше, такого рода модели обнуляют коэффициенты при незначимых или избыточных признаках. А так как все признаки были приведены к одному масштабу, то коэффициент при соответствующем признаке можно интерпретировать как важность этого признака. Таким образом, количество признаков сократилось до 7.

Теперь, когда есть качественное признаковое описание, можно уже более формально показать нелинейную взаимосвязь между признаками и целевой переменной. Для этих целей можно использовать тест, предложенный в [16] на основе нейронной сети. Рассматривается модель:

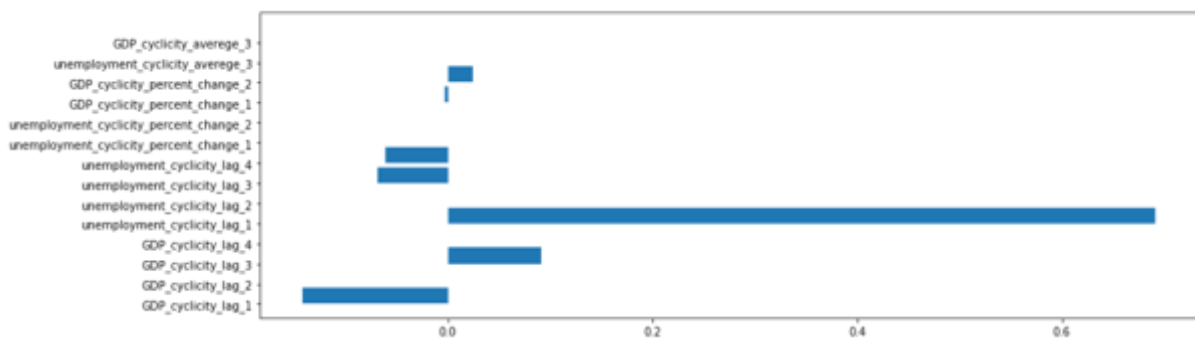


График 3: Веса вспомогательной модели LASSO

$$y_t = x_t' \phi + \sum_{k=1}^q \beta_k G(x_t' \gamma_k) + \varepsilon_t$$

Нулевая гипотеза заключается в том, что коэффициенты при нейронах равны нулю, что означает линейную зависимость между признаками и объясняемой переменной $H_0: \beta_1 = \beta_2 = \dots = \beta_q$. Данный тест состоит из нескольких этапов:

1. Строится регрессия y_t на признаки и вычисляются остатки \hat{u}_t
2. После чего оценивается модель:

$$\hat{u}_t = x_t' \alpha + \sum_{k=1}^q \delta_k G(x_t' \gamma_k) + \eta_t$$

3. Вычисляется коэффициент детерминации R^2 этой модели. И статистика равна nR^2 , где n равно количеству наблюдений

Статистика имеет асимптотическое χ^2 распределение с количеством степеней свободы равным $q - 1$, где q – количество нейронов в скрытом слое

Стоит отметить, что результат теста в некоторой степени случаен. Для того, чтобы отвержение или не отвержение нулевой гипотезы не носило случайный характер, можно использовать следующую стратегию: вычислять значение теста несколько раз, а после этого использовать метод множественной проверки гипотез, например метод Холма. Lee, White и Granger (1993) показали на сгенерированных данных, что тест нейронной сети является эффективной проверкой на нелинейную зависимость. [16]

Для российских данных данный тест дал $p\text{-value} = 6 * 10^{-6}$, т.е. нулевая гипотеза о линейной зависимости отвергается на любом разумном уровне значимости.

Как уже говорилось выше, для многих моделей необходимо подбирать различные гиперпараметры и делать это на отдельной выборке чтобы избежать переобучения. Поэтому вся выборка была разделена на 3 части: обучение, на ней непосредственно определяются коэффициенты модели, валидацию, на ней определяются гиперпараметры и тест, на нем измеряется окончательное качество модели. Причем валидация происходит на 1 шаг вперед, т.е. сначала модель обучается только на обучающей выборке и вычисляется предсказание для первого периода валидации, затем модель оценивается на обучающей выборке и первом периоде валидации и вычисляется предсказание для второго периода валидации и так далее. Таким образом получается предсказание для всего периода валидации.

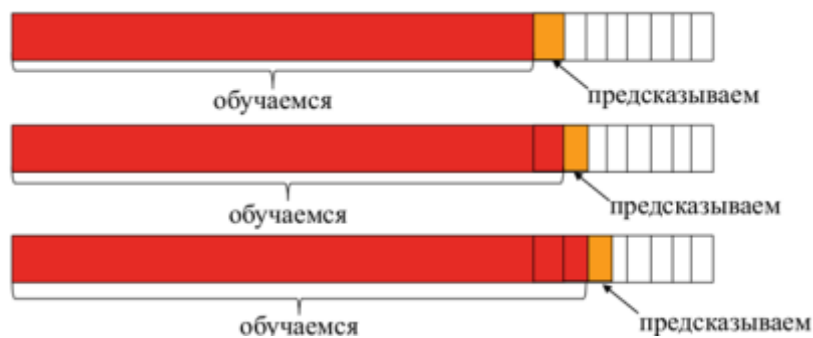


Рис. 1: Описание валидации

Для того чтобы корректно сравнивать модели с различными параметрами необходимо зафиксировать критерий качества предсказания. В данной работе используется средняя абсолютная ошибка (MAE) и корень из средней квадратической ошибки (RMSE).

4. Результаты

Базовые модели

Оценим несколько базовых традиционных моделей, с которыми будем сравнивать все следующие модели. В качестве базовых моделей используются: модель наивного предсказания,

AR(1), а также линейная модель построенная на признаках, отобранных Lasso. Ниже приведены остатки этой модели.

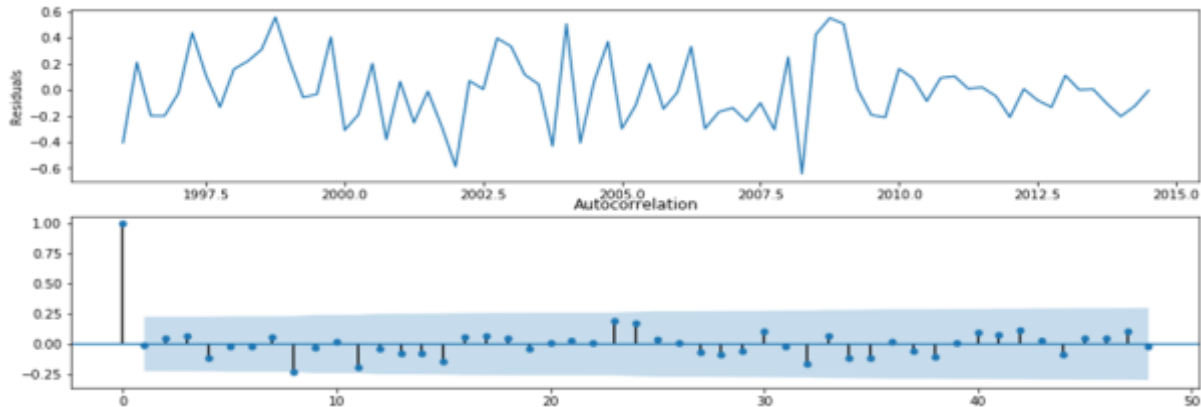


График 4: Остатки линейной модели

На коррелограмме нет значимой автокорреляции. Визуально структуры в остатках не наблюдается, что говорит о качественно подобранных гиперпараметрах и об адекватности модели в целом. Результаты предсказаний этих моделей на тестовой выборке можно увидеть в таблице 1.

Линейная модель с регуляризацией

Для каждой из моделей (lasso, ridge, elastic net) сила регуляризации определялась на валидации с шагом 1. После выбора этого коэффициента, вычислялись предсказания, также с шагом 1. Остатки данных моделей также свидетельствует об адекватности моделей. Каждая из моделей с регуляризацией превзошли обычную линейную регрессию. Отдельно стоит заметить, что модель с обоими видами регуляризации (elastic net) показала лучшее качество среди всех линейных моделей на каждом из рассмотренных горизонтах прогнозирования.

Ансамбли

В качестве реализации случайного леса использовалась модель из библиотеки sklearn, а для бустинга использовалась реализация из xgboost. Для моделей, основанных на решающих деревьях, был произведен несколько другой отбор признаков, чем для линейных моделей. Здесь важностью некоторого признака можно считать количество разбиений в узлах решающих деревьев по этому признаку. Поэтому сначала была обучена базовая модель по которой выделялись наиболее важные признаки. Распределение разбиений по признакам представлен ниже.

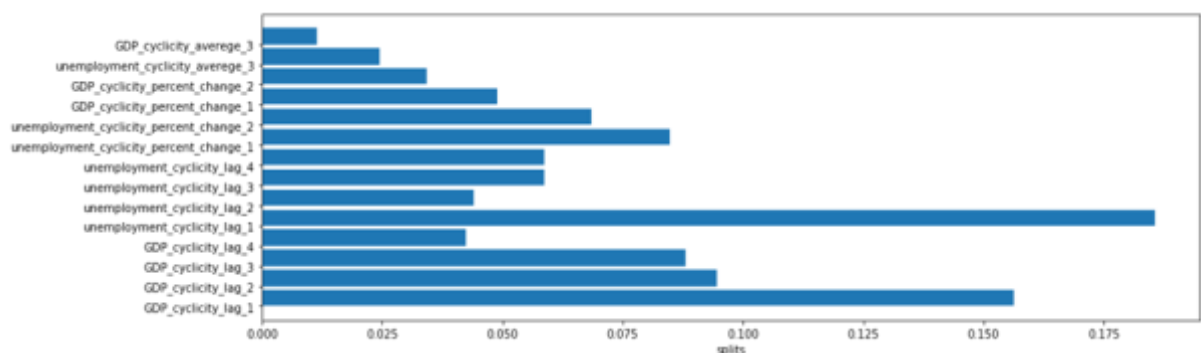


График 5: Доля разбиений по каждому признаку

Аналогично линейным моделям на валидации определялся ряд важных параметров, таких как глубина деревьев, их количество и некоторые другие. Качество прогноза бустинга оказалась несколько выше, чем случайного леса. Однако предсказания обоих алгоритмов оказались хуже, чем линейные модели. Отдельно стоит заметить, что предсказания случайного леса не превзошли наивный прогноз.

Нейронная сеть

Для реализации данной модели использовалась библиотека tensorflow(1.11).

Для нейронной сети необходимо было определить ряд гиперпараметров:

1. Количество нейронов в скрытом слое.

2. Силу регуляризации $r_\phi, r_\beta, r_\gamma$:

$$Q_n(\theta) = \sum_{t=1}^t [y_t - \hat{y}_t(\theta)]^2 + r_\phi \sum_{i=0}^k \phi_i^2 + r_\beta \sum_{j=0}^q \beta_j^2 + r_\gamma \sum_{j=0}^q \sum_{i=0}^k \gamma_{i,j}^2$$

3. Функцию активации.

Все эти параметры определяли по валидации. Причем при обучении после каждой эпохи сохранялась лучшая по качеству на валидации модель и в дальнейшем рассматривалась именно она. Это связано с тем, что ошибка на обучающей выборке уменьшается с каждой эпохой, в то время как ошибка на валидации сначала уменьшается, а потом увеличивается, то есть модель начинает подстраиваться под шумы и терять обобщающую способность.

Модель на основе нейронной сети показала наилучшее качество среди всех моделей при краткосрочном горизонте предсказаний (1 квартал). Если горизонт предсказания достигает 3 кварталов, то предсказания данной модели уступают моделям с регуляризацией, в частности elastic net. Предсказания различных моделей на 1 квартал вперед представлены на графике ниже.

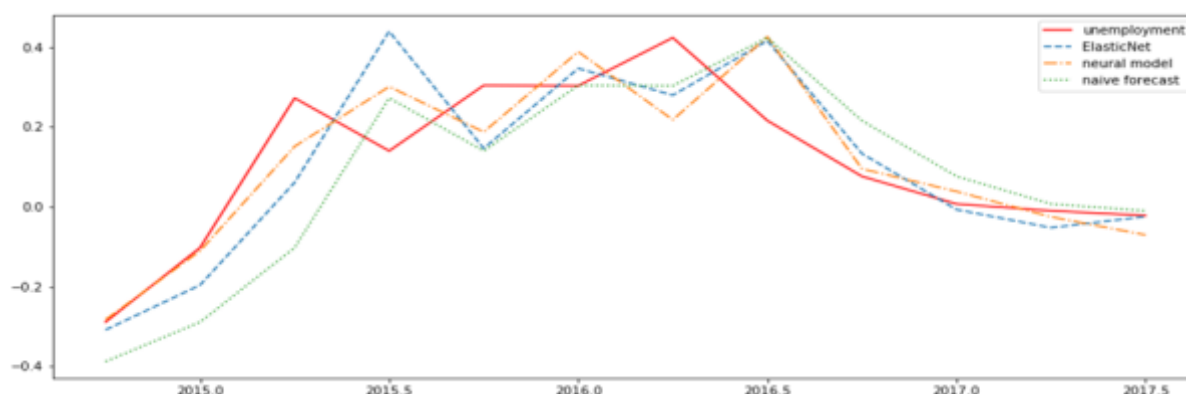


График 6: Предсказания моделей на тестовых данных

К сожалению, невозможно проверить гипотезу о статистической значимой разнице между предсказаниями разных моделей, вследствие малого объема выборки.

Таблица 1

Качество моделей на тестовой выборке

	1 квартал		2 квартала		3 квартала	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
наивный прогноз	0.162	0.127	0.249	0.197	0.341	0.287
AR(1)	0.150	0.119	0.206	0.168	0.242	0.209
Линейная модель	0.159	0.117	0.189	0.165	0.221	0.187
Ridge	0.151	0.111	0.180	0.154	0.209	0.172
Lasso	0.141	0.109	0.182	0.146	0.232	0.190
ElasticNet	0.141	0.107	0.175	0.139	0.206	0.163
RandomForest	0.186	0.149	0.331	0.259	0.323	0.261
xgboost	0.162	0.125	0.229	0.178	0.283	0.244
neural model	0.113	0.086	0.176	0.139	0.219	0.181

5. Заключение

Любая модель, какой бы сложной она ни была, не может охватить все зависимости в данных. Экономическое прогнозирование – это процесс, в котором ищутся лучшие подмножества аппроксимирующих моделей для истинного базового процесса.

Не все методы показали свою состоятельность при решении данной проблемы. Оба ансамблевых метода (случайный лес и бустинг) при прогнозировании безработицы показали результат хуже, чем линейные модели. Более того, предсказательная способность случайного леса оказалась ниже, чем модели наивного прогноза.

Однако модель на основе нейронной сети показала лучшие результаты при краткосрочном прогнозировании. Это показывает конкурентоспособность ML-методов при прогнозировании безработицы с традиционными альтернативами.

Литература

1. Байбуза И. Прогнозирование инфляции с помощью методов машинного обучения // *Деньги и кредит*. – 2018.
2. Богданова Т. К., Шевгунов Т. Я., Уварова О. М. Применение нейронных сетей для прогнозирования платежеспособности российских предприятий обрабатывающих отраслей // *Бизнес-информатика*. – 2013. – №. 2 (24).
3. Вакуленко Е. С., Гурвич Е. Т. Взаимосвязь ВВП, безработицы и занятости: углубленный анализ закона Оукена для России // *Вопросы экономики*. – 2015. – Казакова М. Закон Оукена: теоретические основы и оценки для России // *Экономическое развитие России*. – 2017. – Т. 24. – №. 11. – С. 27-36.
4. Золотова И. Ю., Дворкин В. В. Краткосрочное прогнозирование цен на российском оптовом рынке электроэнергии на основе нейронных сетей // *Проблемы прогнозирования*. – 2017. – №. 6.
5. Колмогоров А. Н. О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения // *Доклады Академии наук*. – Российская академия наук, 1957. – Т. 114. – №. 5. – С. 953-956.
6. Турунцева М. Ю., Киблицкая Т. Р. 2010. Качественные свойства различных подходов к прогнозированию социально-экономических показателей РФ, Москва: ИЭПП, Научные труды № 135Р. – 2010.
7. Фокин Н. Д. VAR-LASSO-модель на большом массиве российских экономических данных // *Экономическое развитие России*. – 2018. – №1(20-30)
8. Adriansson N., Mattsson I. Forecasting GDP Growth, or How Can Random Forests Improve Predictions in Economics?. – 2015.
9. Ball L. M., Leigh D., Loungani P. Okun's law: fit at fifty?. – *National Bureau of Economic Research*, 2013. – №. w18668.
10. Dayhoff J. E., DeLeo J. M. Artificial neural networks: opening the black box // *Cancer: Interdisciplinary International Journal of the American Cancer Society*. – 2001. – Т. 91. – №. S8. – С. 1615-1635.
11. Franses P. H., Draisma G. Recognizing changing seasonal patterns using artificial neural networks // *Journal of Econometrics*. – 1997. – Т. 81. – №. 1. – С. 273-280.
12. Franses P. H., van D. Dick. *Nonlinear Time Series Models in Empirical Finance*. – 2003.
13. Hodrick R. J., Prescott E. C. Postwar US business cycles: an empirical investigation // *Journal of Money, credit, and Banking*. – 1997. – С. 1-16.
14. Kingma D. P., Ba J. Adam: A method for stochastic optimization // *arXiv preprint arXiv:1412.6980*. – 2014.
15. *Learning D. Ian Goodfellow, Yoshua Bengio and Aaron Courville*. – 2016.
16. Lee T. H., White H., Granger C. W. J. Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests // *Journal of Econometrics*. – 1993. – Т. 56. – №. 3. – С. 269-290.
17. McNelis P. D., McAdam P. Forecasting inflation with thick models and neural networks. – 2004.
18. Meyer B., Tasci M. Lessons for forecasting unemployment in the United States: Use flow rates, mind the trend. – 2015.
19. Müller A. C. et al. *Introduction to machine learning with Python: a guide for data scientists*. – O'Reilly Media, Inc., 2016.
20. Okun A. M. (1962). Potential GNP: its measurement and significance. *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, 1962; reprinted in Cowles Foundation, Yale University
21. Ucenic C. I., Atsalakis G. Forecasting CPI using a neural network with fuzzy inference system // *ASMDA. Proceedings of the International Conference Applied Stochastic Models and Data Analysis*. – Vilnius Gediminas Technical University, Department of Construction Economics & Property, 2009. – Т. 13. – С. 309.

References:

1. Baybuza I. Inflation forecasting using machine learning methods // *Money and credit*. - 2018.
2. Bogdanova TK, Shevgunov T. Ya., Uvarova OM. The use of neural networks to predict the solvency of Russian enterprises in the processing industries // *Business Informatics*. - 2013. - №. 2 (24).
3. Vakulenko E., S., Gurvich, E. T. The Relationship of GDP, Unemployment, and Employment: In-depth Analysis of Oken's Law for Russia // *Voprosy Ekonomiki*. - 2015. - Kazakova M. Okun's Law: Theoretical Foundations and Estimates for Russia // *Economic Development of Russia*. - 2017. - V. 24. - №. 11. - p. 27-36.

4. Zolotova I. Yu., Dvorkin V. V. *Short-term prediction of prices on the Russian wholesale market electricity based on neural networks // Forecasting Problems*. - 2017. - no. 6.
5. Kolmogorov A. N. *On the representation of continuous functions of several variables as superpositions of continuous functions of one variable and addition, Report. AS SSSR, 114: 5 (1957), 953–956*
6. Turuntseva M. Yu., Kiblitckaya TR, 2010, *Qualitative properties of various approaches to forecasting the socio-economic indicators of the Russian Federation, Moscow: IET, Scientific works No. 135P*. - 2010.
7. Fokin N. D. *VAR-LASSO-model on a large array of Russian economic data // Economic Development of Russia*. - 2018. - №1 (20-30)
8. Adriansson N., Mattsson I. *Forecasting GDP Growth, or How Can Random Forests Improve Predictions in Economics?*. – 2015.
9. Ball L. M., Leigh D., Loungani P. *Okun's law: fit at fifty?*. – National Bureau of Economic Research, 2013. – №. w18668.
10. Dayhoff J. E., DeLeo J. M. *Artificial neural networks: opening the black box //Cancer: Interdisciplinary International Journal of the American Cancer Society*. – 2001. – Т. 91. – №. S8. – С. 1615-1635.
11. Franses P. H., Draisma G. *Recognizing changing seasonal patterns using artificial neural networks // Journal of Econometrics*. – 1997. – Т. 81. – №. 1. – С. 273-280.
12. Franses P. H., van D. Dick. *Nonlinear Time Series Models in Empirical Finance*. – 2003.
13. Hodrick R. J., Prescott E. C. *Postwar US business cycles: an empirical investigation //Journal of Money, credit, and Banking*. – 1997. – С. 1-16.
14. Kingma D. P., Ba J. *Adam: A method for stochastic optimization //arXiv preprint arXiv:1412.6980*. – 2014.
15. *Learning D. Ian Goodfellow, Yoshua Bengio and Aaron Courville*. – 2016.
16. Lee T. H., White H., Granger C. W. J. *Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests //Journal of Econometrics*. – 1993. – Т. 56. – №. 3. – С. 269-290.
17. McNelis P. D., McAdam P. *Forecasting inflation with thick models and neural networks*. – 2004.
18. Meyer B., Tasci M. *Lessons for forecasting unemployment in the United States: Use flow rates, mind the trend*. – 2015.
19. Müller A. C. et al. *Introduction to machine learning with Python: a guide for data scientists*. – O'Reilly Media, Inc., 2016.
20. Okun A. M. (1962). *Potential GNP: its measurement and significance. Proceedings of the Business and Economic Statistics Section of the American Statistical Association, 1962; reprinted in Cowles Foundation, Yale University*
21. Ucenic C. I., Atsalakis G. *Forecasting CPI using a neural network with fuzzy inference system //ASMDA. Proceedings of the International Conference Applied Stochastic Models and Data Analysis*. – Vilnius Gediminas Technical University, Department of Construction Economics & Property, 2009. – Т. 13. – С. 309.