
Эконометрическое моделирование

ОБ ИСПОЛЬЗОВАНИИ МОДЕЛЕЙ ПАНЕЛЬНЫХ ДАННЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ ТЕМПОВ РОСТА ОТРАСЛЕЙ РОССИЙСКОЙ ОБРАБАТЫВАЮЩЕЙ ПРОМЫШЛЕННОСТИ

А. В. ПОЛБИН
А. В. ШУМИЛОВ

В работе исследуется вопрос использования моделей панельных данных для прогнозирования темпов роста производства отраслей российской обрабатывающей промышленности. На месячных данных 2015–2021 гг. показано, что одношаговые прогнозы панельных моделей качественно превосходят наивные прогнозы, основанные на усреднении значений роста предыдущих периодов. По сравнению с индивидуальными отраслевыми моделями панельные модели дают более точные прогнозы на 1–6 месяцев вперед для ряда отраслей, однако совокупная точность прогнозирования при этом улучшается незначительно.

Статья подготовлена в рамках выполнения научно-исследовательской работы государственного задания РАНХиГС.

Ключевые слова: прогнозирование, динамическая модель панельных данных, рост производства, отрасли обрабатывающей промышленности России.

JEL: C22, C23, C53.

Введение

Существует достаточно много эмпирических свидетельств того, что использование моделей анализа панельных данных в качестве альтернативы моделям временных рядов способно существенно улучшить качество прогнозирования различных экономических показателей. Так, в пионерской эмпирической работе Балтаги и Гриффина [2] на данных по 18 странам ОЭСР за период 1960–1990 гг. было оценено динамическое уравнение удельного спроса на бензин как для отдельных стран, так и на основе объединения данных в панель. Сопоставление оценок эластичностей спроса по доходу и цене бензина разных моделей показало, что в моделях временных рядов оценки зачастую принимают неправдоподобно высокие или низкие значения, в то время как в моделях панельных данных оценки эластичностей достаточно стабильны. Из сопоставлений модельных вневыборочных прогнозов спроса на бен-

зин разных моделей на горизонтах от 1 до 10 лет следует, что панельные модели дают меньшую (среднюю по всем странам) квадратичную ошибку прогнозирования, чем индивидуальные модели для стран.

Аналогичный результат о превосходстве моделей анализа панельных данных над моделями временных рядов по качеству вневыборочного прогнозирования был получен в исследованиях душевого потребления табачной продукции в 46 штатах США в 1963–1992 гг. [3], квартальных инвестиций в основные средства 48 отраслей британской промышленности в 1978–1999 гг. [4], обменных курсов к доллару США 18 развитых стран (квартальные данные 1991–2006 гг.) [7], роста валового регионального продукта 18 германских земель (1991–2006 гг.) [6] и 31 китайской провинции (1979–2007 гг.) [5], а также ряда других показателей экономических субъектов (см., например, обзор [1]).

Полбин Андрей Владимирович, заведующий лабораторией математического моделирования экономических процессов РАНХиГС при Президенте Российской Федерации; заместитель заведующего международной лабораторией математического моделирования экономических процессов Института экономической политики имени Е.Т. Гайдара, канд. экон. наук (Москва), e-mail: arolbin@ier.ru; Шумилов Андрей Валерьевич, старший научный сотрудник РАНХиГС при Президенте Российской Федерации, канд. физ.-мат. наук (Москва), e-mail: shumilov-av@ranepa.ru

В настоящем исследовании мы задаемся вопросом: могут ли модели панельных данных улучшить качество прогнозирования роста производства отраслей российской обрабатывающей промышленности? Этот вопрос важен, поскольку задача прогнозирования отраслевого роста предельно актуальна для мониторинга тенденций в отраслях промышленности и для разработки оперативных мер экономической политики. В следующем разделе работы описываются используемые данные, спецификации моделей временных рядов и панельных данных, методики их оценивания и построения прогнозов. Далее обсуждаются результаты прогнозирования с помощью альтернативных моделей отраслевого роста. В заключении представлены выводы исследования.

Описание данных и моделей

Для целей прогнозирования в работе используются месячные временные ряды роста производства 24-х отраслей российской обрабатывающей промышленности за период с января 2015 г. по ноябрь 2021 г., построенные на основе данных Росстата по индексам производства по отдельным видам экономической деятельности ОКВЭД2¹. Темпы роста взяты «в годовом выражении» как процентный рост производства в текущем месяце по отношению к соответствующему месяцу предыдущего года. Такой подход к их расчету позволяет избавиться от сезонности в рядах данных без применения сложных методик, затрудняющих прогнозирование.

В качестве эталона для сопоставлений прогнозов отраслевого роста берутся авторегрессионные модели первого порядка (**Ind.**), оцениваемые для каждой отрасли i в отдельности:

$$y_{it} = \alpha_i + \beta_i y_{it-1} + \varepsilon_{it}$$

где y_{it} – рост производства i в период t ;

$\varepsilon_{it} \sim N(0, \sigma_i^2)$ – случайная ошибка с гетерогенной дисперсией по секторам.

Для данных по отраслевому росту, объединенных в панель, рассматриваются две следующие динамические модели.

Объединенная модель регрессии (**Pooled**), имеющая вид

$$y_{it} = \alpha + \beta y_{it-1} + \lambda_t + \varepsilon_{it}$$

где λ_t – временные эффекты²; $\varepsilon_{it} \sim N(0, \sigma^2)$ – случайная ошибка с гомогенной дисперсией по секторам. Здесь свободный член α и угловой коэффициент β принимают одни и те же значения для всех отраслей.

Вторая модель анализа панельных данных – модель с фиксированными эффектами (**FE**):

$$y_{it} = \alpha_i + \beta y_{it-1} + \lambda_t + \varepsilon_{it}$$

где величины α_i (индивидуальные эффекты) могут различаться между отраслями, а у случайной ошибки $\varepsilon_{it} \sim N(0, \sigma^2)$, так же как и в предыдущем случае, предполагается гомогенная дисперсия по секторам.

Модели авторегрессии для отдельных отраслей и модель объединенной регрессии оцениваются методом наименьших квадратов, модель с фиксированными эффектами – с помощью МНК с фиктивными переменными (LSDV). Поскольку число временных периодов в нашем случае достаточно велико, LSDV-оценку параметра β можно считать состоятельной.

Для оценки качества прогнозов указанных выше моделей на горизонтах от 1 до 6 месяцев используется тестовая выборка с января 2019 г. по ноябрь 2021 г. Прогнозирование осуществляется согласно схеме расширяющегося окна оценивания, описываемой следующим образом. Сначала модели оцениваются на исходной обучающей выборке 2015–2018 гг. На основе оцененных параметров моделей

¹ URL: https://rosstat.gov.ru/enterprise_industrial; https://rosstat.gov.ru/storage/mediabank/ind_baza%202018.xlsx

² Были также рассмотрены варианты моделей панельных данных без временных эффектов. Расчеты показали, что они обладают чуть более слабой прогнозной силой, чем спецификации с временными эффектами, поэтому соответствующие результаты в работе не приводятся.

последовательно рассчитываются вневыборочные прогнозы на 1–6 шагов вперед³ (январь-июнь 2019 г.). Затем модели переоцениваются на обучающей выборке с добавлением января 2019 г. и строятся вневыборочные прогнозы разных горизонтов на февраль-июль 2019 г. Итерации расчета прогнозов продолжают до тех пор, пока правая граница периода оценки моделей не достигнет ноября 2021 г. Для каждой отрасли итоговое количество построенных прогнозов той или иной модели на 1 шаг вперед равно 35, на 6 шагов вперед – 30.

По своим прогнозным свойствам модели панельных данных дополнительно сравниваются с простыми наивными моделями, основанными на усреднении данных предыдущих периодов. В первой наивной модели (**Naive1**) в качестве прогноза на h шагов вперед берется отраслевой темп роста за последний месяц обучающей выборки. Во второй наивной модели (**Naive2**) прогноз на h шагов вперед рассчитывается как среднее арифметическое отраслевого темпа роста за последний год обучающей выборки.

Результаты прогнозирования

В качестве меры качества прогнозирования мы берем квадратный корень среднеквадратичной ошибки прогноза на тестовой выборке (RMSE). Результаты расчета RMSE (как для каждой отдельной отрасли, так и для всех отраслей в совокупности) для прогнозов на 1, 3 и 6 шагов вперед рассмотренных моделей отраслевого роста представлены в таблице. Как видно из нее, наихудшей по качеству является наивная модель 1, где прогнозом считается рост производства за последний месяц обучающей выборки. На горизонтах прогнозирования 1, 3 и 6 шагов все регрессионные модели превосходят ее по совокупной точности предсказания, как минимум, на 14, 22 и 25% соответ-

ственно. Что касается второй наивной модели (прогноз – средний рост производства на обучающей выборке), то при $h = 1$ регрессионные модели также значительно превосходят ее (как минимум, на 8,5%), однако при $h = 3$ и $h = 6$ наивная модель 2 по качеству прогнозирования становится сопоставимой с регрессионными.

Сравнения совокупной предсказательной силы регрессионных моделей показывают, что, хотя на всех горизонтах прогнозирования они ранжируются от худшей к лучшей как: 3) индивидуальные модели роста, 2) модель объемной регрессии, 1) модель с фиксированными эффектами, различия между моделями 3 и 1 малы (улучшение RMSE для модели с фиксированными эффектами не превышает 2%). Не являясь лидером по точности прогнозирования по всем отраслям в среднем, индивидуальные модели отраслевого роста тем не менее позволяют получить для некоторых отдельных секторов промышленности более точные прогнозы по сравнению с обеими моделями панельных данных: при $h = 1$, $h = 3$ и $h = 6$ количество таких отраслей равно 6, 7 и 4 соответственно⁴. Таким образом, можно сделать общий вывод о том, что выгоды от использования моделей панельных данных вместо индивидуальных моделей для прогнозирования роста отраслей российской обрабатывающей промышленности нельзя назвать существенными. Отметим, что качественные результаты сопоставлений прогнозных сил различных моделей отраслевого роста остаются неизменными при сокращении тестовой выборки до периода 2020–2021 гг.

Заключение

В настоящей работе мы сопоставили по способности прогнозировать рост отраслей российской обрабатывающей промышленности на горизонтах от 1 до 6 месяцев динамические

³ В моделях панельных данных оценки временных эффектов для периодов вне обучающей выборки, необходимые для расчета прогнозов, рассчитываются как среднее значение оценок временных эффектов на обучающей выборке.

⁴ Например, при $h = 1$ в список отраслей, для которых индивидуальные модели превосходят по качеству прогнозов панельные модели, входят производства текстильных изделий, кокса и нефтепродуктов, химических веществ и продуктов, готовых металлоизделий, компьютеров и электронных и оптических изделий, мебели.

Квадратные корни среднеквадратичных ошибок прогнозов моделей отраслевого роста (RMSE, 2019–2021 гг.)

| Отрасль | Модель | | | | | | | | | | | | | | |
|--|-------------------------|--------------|--------------|--------------|--------------|--------------------------|--------------|--------------|--------------|--------------|---------------------------|--------------|--------------|--------------|--------------|
| | Прогноз на 1 шаг вперед | | | | | Прогноз на 3 шага вперед | | | | | Прогноз на 6 шагов вперед | | | | |
| | Naive1 | Naive2 | Ind. | Pooled | FE | Naive1 | Naive2 | Ind. | Pooled | FE | Naive1 | Naive2 | Ind. | Pooled | FE |
| Пищевые продукты | 2,90 | 3,07 | 2,67 | 2,54 | 2,63 | 3,33 | 3,13 | 3,08 | 2,87 | 3,07 | 4,43 | 3,24 | 3,24 | 3,12 | 3,27 |
| Напитки | 5,74 | 6,87 | 5,66 | 5,24 | 5,59 | 8,78 | 6,78 | 6,77 | 6,25 | 6,69 | 7,72 | 7,00 | 6,97 | 6,42 | 6,89 |
| Табачные изделия | 13,91 | 12,39 | 11,92 | 11,59 | 11,43 | 19,59 | 12,68 | 13,11 | 13,28 | 13,29 | 21,25 | 12,66 | 13,40 | 12,43 | 13,67 |
| Текстильные изделия | 6,59 | 7,24 | 5,99 | 6,15 | 6,07 | 7,56 | 7,39 | 6,75 | 7,63 | 7,16 | 9,68 | 7,04 | 8,09 | 7,07 | |
| Одежда | 12,82 | 10,08 | 11,29 | 10,33 | 10,24 | 16,89 | 10,23 | 12,35 | 10,63 | 10,61 | 15,39 | 10,48 | 10,92 | 10,54 | 10,58 |
| Кожа и изделия из кожи | 16,77 | 17,32 | 15,00 | 14,81 | 14,97 | 22,17 | 17,96 | 17,76 | 17,60 | 17,81 | 24,53 | 18,96 | 18,93 | 18,71 | 19,01 |
| Обработка древесины | 6,64 | 8,12 | 6,18 | 6,07 | 6,46 | 9,71 | 8,46 | 8,13 | 7,77 | 8,37 | 10,77 | 8,99 | 8,74 | 8,25 | 9,06 |
| Бумага и бумажные изделия | 3,78 | 4,05 | 3,47 | 3,59 | 3,44 | 5,35 | 4,19 | 4,14 | 4,72 | 4,16 | 5,40 | 4,36 | 4,33 | 5,13 | 4,34 |
| Полиграфия | 10,67 | 11,35 | 9,71 | 9,47 | 9,72 | 10,84 | 11,73 | 11,17 | 10,66 | 11,25 | 15,60 | 12,30 | 12,27 | 11,64 | 12,22 |
| Кокс и нефтепродукты | 4,02 | 6,66 | 4,03 | 4,79 | 4,79 | 7,02 | 6,96 | 6,38 | 7,20 | 6,79 | 9,11 | 7,32 | 7,38 | 7,79 | 7,33 |
| Химические продукты | 2,77 | 2,94 | 2,49 | 2,66 | 2,50 | 4,31 | 2,84 | 2,90 | 3,53 | 2,86 | 3,63 | 2,85 | 2,82 | 3,72 | 2,80 |
| Лекарственные средства | 23,78 | 22,11 | 20,62 | 22,05 | 20,07 | 30,72 | 22,68 | 22,62 | 27,12 | 22,39 | 29,59 | 22,09 | 21,90 | 27,52 | 21,75 |
| Резиновые и пластмассовые изделия | 7,72 | 8,53 | 7,12 | 6,86 | 7,05 | 10,26 | 8,55 | 8,37 | 7,83 | 8,25 | 10,65 | 8,90 | 8,46 | 7,98 | 8,67 |
| Прочая неметаллическая минеральная продукция | 5,93 | 8,49 | 5,65 | 5,58 | 6,32 | 9,47 | 8,78 | 8,23 | 7,23 | 8,47 | 9,55 | 8,61 | 8,11 | 7,31 | 8,50 |
| Производство металлургическое | 4,51 | 4,22 | 3,93 | 4,02 | 3,84 | 4,83 | 4,26 | 4,04 | 4,88 | 4,17 | 5,76 | 4,46 | 4,44 | 5,32 | 4,46 |
| Готовые металлоизделия | 25,73 | 18,41 | 19,28 | 19,87 | 19,80 | 21,98 | 17,03 | 16,85 | 16,88 | 16,85 | 25,84 | 16,67 | 16,70 | 17,03 | 16,75 |
| Компьютеры, электронные и оптические изделия | 27,19 | 20,65 | 21,03 | 21,66 | 21,38 | 30,34 | 21,12 | 21,18 | 21,33 | 21,25 | 30,40 | 22,02 | 22,05 | 21,83 | 22,00 |
| Электрооборудование | 11,76 | 9,71 | 9,81 | 9,53 | 9,41 | 14,15 | 9,85 | 9,94 | 9,83 | 9,85 | 15,28 | 10,24 | 10,23 | 10,16 | 10,21 |
| Прочие машины и оборудование | 13,61 | 12,47 | 11,44 | 11,46 | 11,28 | 13,19 | 12,72 | 11,88 | 12,36 | 12,08 | 15,12 | 13,07 | 12,83 | 13,08 | 12,74 |
| Автотранспорт | 30,24 | 33,50 | 28,32 | 27,31 | 27,88 | 43,85 | 34,91 | 36,62 | 33,91 | 34,58 | 51,28 | 36,79 | 38,82 | 35,98 | 36,82 |
| Прочие транспортные средства | 20,88 | 17,71 | 17,62 | 16,89 | 17,17 | 23,71 | 16,33 | 16,40 | 16,22 | 16,49 | 24,82 | 16,71 | 16,75 | 16,75 | 16,82 |
| Мебель | 15,13 | 19,60 | 14,38 | 14,55 | 15,11 | 23,48 | 20,48 | 20,62 | 19,39 | 19,99 | 27,69 | 21,62 | 22,54 | 20,74 | 21,50 |
| Прочие готовые изделия | 18,35 | 15,67 | 15,87 | 14,83 | 14,88 | 22,50 | 16,20 | 16,49 | 15,69 | 16,06 | 22,80 | 16,81 | 16,59 | 15,86 | 16,59 |
| Ремонт машин и оборудования | 14,60 | 12,37 | 12,54 | 11,99 | 11,91 | 19,07 | 12,76 | 12,76 | 12,87 | 12,91 | 18,00 | 13,30 | 13,30 | 13,11 | 13,27 |
| Total RMSE | 15,08 | 14,16 | 12,96 | 12,87 | 12,83 | 18,78 | 14,42 | 14,59 | 14,47 | 14,30 | 20,38 | 14,88 | 15,13 | 15,01 | 14,87 |
| Ранг по Total RMSE | 5 | 4 | 3 | 2 | 1 | 5 | 2 | 4 | 3 | 1 | 5 | 2 | 4 | 3 | 1 |

Примечание. Жирным шрифтом выделена наилучшая модель для каждого горизонта прогнозирования.

Источник: расчеты авторов.

модели панельных данных (с фиксированными эффектами и объединенной регрессии) с моделями временных рядов AR(1) и наивными моделями среднего роста за прошедшие периоды. Результаты оценки качества прогнозов показали, что модели панельных данных значительно превосходят лучшую из наивных моделей только на самом коротком горизонте прогнозирования в 1 месяц. По сравнению с индивидуальными отраслевыми моделями панельные модели дают более точные прогнозы на 1–6 месяцев вперед для ряда отраслей (обе на всех горизонтах прогнозирования одновременно – для производства напитков, одежды, электрооборудования и прочих готовых изделий), однако совокупная точность

прогнозирования при этом улучшается незначительно. Таким образом, применительно к прогнозированию роста отраслей российской промышленности использование рассмотренных панельных моделей не дает существенно-го выигрыша относительно использования моделей временных рядов.

Одним из возможных направлений дальнейших исследований в данной области является проверка предсказательной способности более сложных моделей панельных данных с пространственными взаимосвязями между экономическими единицами (см., например, [5]) на основе построения какой-либо меры «близости» отраслей российской промышленности друг к другу. ■

Литература / References

1. Baltagi B.H. Panel data forecasting // Handbook of economic forecasting. 2013. Vol. 2. Pp. 995–1024.
2. Baltagi B.H., Griffin J.M. Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline // Journal of Econometrics. 1997. Vol. 77. No. 2. Pp. 303–327.
3. Baltagi B.H., Griffin J.M., Xiong W. To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand // Review of Economics and Statistics. 2000. Vol. 82. No. 1. Pp. 117–126.
4. Driver C., Imai K., Temple P., Urga G. The effect of uncertainty on UK investment authorisation: Homogenous vs. heterogeneous estimators // Empirical Economics. 2004. Vol. 29. No. 1. Pp. 115–128.
5. Girardin E., Kholodilin K.A. How helpful are spatial effects in forecasting the growth of Chinese provinces? // Journal of Forecasting. 2011. Vol. 30. No. 7. Pp. 622–643.
6. Kholodilin K., Siliverstovs B., Kooths S. A dynamic panel data approach to the forecasting of the GDP of German Lander // Spatial Economic Analysis. 2008. Vol. 3. No. 2. Pp. 195–207.
7. Rapach D.E., Wohar M.E. Testing the monetary model of exchange rate determination: a closer look at panels // Journal of international Money and Finance. 2004. Vol. 23. No. 6. Pp. 867–895.

Forecasting Output Growth of Russian Manufacturing Industries Using Panel Data Models

Andrey V. Polbin – Head of Department of Mathematical Modeling of Economic Processes, Russian Presidential Academy of National Economy and Public Administration; Deputy Head of International Department of Mathematical Modeling of Economic Processes, Gaidar Institute for Economic Policy, Candidate of Economic Sciences (Moscow, Russia). E-mail: apolbin@iep.ru

Andrey V. Shumilov – Senior Researcher of the Russian Presidential Academy of National Economy and Public Administration, Candidate of Physical and Mathematical Sciences, Associate Professor (Moscow, Russia). E-mail: shumilov-av@ranepa.ru

In this paper, we utilize panel data models for forecasting output growth rates of Russian manufacturing industries. Using monthly data for 2015–2021, we find that one-month-ahead forecasts of panel models are superior to corresponding naive forecasts based on averaging past growth rates. Compared to individual industry models, panel data models yield better forecasts at 1–6 months horizons for a number of industries, but the overall forecasting accuracy improves only slightly.

The article was written on the basis of the RANEPА state assignment research programme.

Key words: forecasting, dynamic panel data model, output growth, Russian manufacturing industries.

JEL-codes: C22, C23, C53.