

ПРОГНОЗИРОВАНИЕ РЕГИОНАЛЬНЫХ МАКРОПОКАЗАТЕЛЕЙ С ПОМОЩЬЮ ПРОСТРАНСТВЕННЫХ VAR

Т. Горшкова, н.с., РАНХиГС

С начала 1980-х годов для моделирования региональных показателей все чаще используются модели пространственной корреляции, учитывающие географические, социальные и экономические связи между соседними регионами. Кроме того, с развитием методов машинного обучения стали появляться работы, посвященные исследованию возможности применения методов обучения больших данных к макроэкономическим показателям.

Чаще всего подобные исследования проводятся на примерах данных США, в которых большое количество штатов в совокупности с длинным временным рядом наблюдений позволяет отнести имеющуюся информацию к понятию «большие данные»¹. На российские данные это понятие может быть распространено с большим трудом в связи с отсутствием данных до 2000 г., что, однако, может быть частично нивелировано большим количеством регионов.

Начиная с 1990-х гг. Федеральная служба государственной статистики (Росстат) проводит регулярный сбор и консолидацию информации по большому числу показателей экономического развития всех регионов, которые затем, в виде объединенных социально-экономических показателей для всей страны, используются при оценке и прогнозировании развития страны. Такой подход не учитывает региональные особенности Российской Федерации и снижает точность полученных выводов, что, в свою очередь, может привести к принятию неверных управленческих решений и ухудшению социально-экономического положения отдельных регионов.

Все более распространенным методом исследования региональных показателей становится моделирование с учетом пространственных связей между объектами, входящими в состав одной экономической зоны. Исследования пространственных данных позволили сделать вывод, что экономические показатели страны зависят от показателей других стран, с которыми установлены торговые отношения. Возможности безбарьерной торговли выравнивают цены, способствуют установлению производственной специализации, влияют на торговый баланс. Это верно и для региональных показателей, имеющих существенные экономические различия в пределах одной страны. В Российской Федерации большая территория, расстояния между региональными центрами, различия в климатических условиях и в наделенности ресурсами играют существенную роль в развитии отдельных регионов. Однако практически не существует работ, посвященных исследованию социально-экономических показателей российских регионов.

При прогнозировании макроэкономических показателей достаточно точные результаты дают векторные авторегрессии, позволяющие строить структурные модели сразу для нескольких показателей². Для прогнозирования региональных показателей разработан отдельный класс моделей,

¹ На данный момент не разработано единое определение понятия «больших данных», но большинство исследователей определяют это понятие как «данные огромных объемов и значительного многообразия, эффективно обрабатываемых горизонтально масштабируемыми программными инструментами» (<https://dis-group.ru/company-news/articles/chto-takoe-big-data-azbuka-bolshih-danniyh-ot-a-do-ya-chast-1/>)

² См, например: Badinger H., Muller W. G., Tondl, G. Regional convergence in the European Union 1985-1999: a spatial dynamic panel analysis//Regional Studies. – 2004. – Vol. 38. P. 241–253; Schanne N. A Global Vector Autoregression (GVAR) model for regional labour markets and its forecasting performance with leading indicators in Germany//AB Discussion Paper – 2015. – Vol. 13; Patuelli R., Mayor M. Short-Run Regional Forecasts: Spatial Models Through Varying Cross-Sectional and Temporal Dimensions//from book From Administrative Spatial Units to Local Labour Market Areas. Some Remarks on the Unit of Investigation of Regional Economics with Particular Reference to the Applied Research in Italy. – 2012. – P. 173–192.

в которых учитываются как теоретические и эмпирические взаимосвязи между самими экономическими показателями, так и взаимосвязи между регионами / странами, в которых данные показатели наблюдаются¹.

Данная статья построена следующим образом: в первой части проводится обзор эмпирических работ, посвященных построению пространственных векторных авторегрессий (Spatial Vector Autocorrelation – SpVAR) для различных регионов. Во второй части SpVAR модели применяются к российским региональным данным. В заключении описаны полученные результаты и даны рекомендации по дальнейшему применению SpVAR моделей.

Пространственные векторные авторегрессии

Понятие «пространственная векторная авторегрессия» введено в работе Beenstock, Fisher (2007)², хотя исследование пространственных динамических моделей проводилось и в более ранних работах³.

Основное различие между стандартными и пространственными векторными авторегрессиями заключается в том, что помимо зависимости показателей от друг друга во времени в SpVAR также учитывается зависимость показателей от географического положения регионов. В SpVAR модели значения показателей в момент времени t могут зависеть от значений объясняющих переменных в тот же момент времени t в соседних регионах (по аналогии со стандартными пространственными моделями панельных данных) и от значений объясняющих переменных в момент времени $t-\tau$ ($\tau > 0$) в соседних регионах. Таким образом, VAR является частным случаем SpVAR при отсутствии пространственных лагов, а пространственная модель панельных данных является частным случаем пространственной VAR при отсутствии временных лагов.

По аналогии с обычными пространственными моделями в SpVAR учет пространственной корреляции может происходить двумя способами: включение пространственной зависимости в авторегрессионную структуру (Spatial Autocorrelation model – SAC), при которой ошибки модели пространственно некоррелированы, и в структуру ошибки (Spatial Error Model – SEM).

В работе Beenstock, Fisher (2007)⁴ авторы сравнивают между собой пространственную VAR, в которой нет пространственных лагов в самой модели, но есть пространственная корреляция в ошибках, и модели, в которых ошибки пространственно независимы, но в модели учтены пространственные лаги. Авторы отмечают, что первый тип модели (с пространственными ошибками) является вложенной моделью во второй тип, и для разделения этих моделей может быть использован метод главных компонент. Пространственная VAR рассматривается как гомогенная стационарная панель, в которой регионы имеют специфические эффекты.

В работе Beenstock, Fisher (2007)⁵ рассматривается одномерная модель с одним временным лагом, которая записывается с помощью уравнения (1), и ошибками, моделирующимися с помощью формулы (2):

$$Y_{nt} = \mu_n + \Theta Y_{nt}^* + \beta Y_{nt-1} + \lambda Y_{nt-1}^* + u_{nt}, \quad (1)$$

$$u_{nt} = \rho u_{nt-1} + \delta u_{nt} + \gamma u_{nt-1} + \varepsilon_{nt}, \sigma_{ni} = cov(\varepsilon_n \varepsilon_i), \quad (2)$$

где μ_n – регионально-специфические эффекты, Θ – коэффициенты при значениях объясняемой переменной в соседних регионах, β – коэффициент при временных лагах, λ – «запаздывающий пространственный лаг», который представляет собой временной лаг в пространственных данных,

¹ См. например: Girardin, E., Kholodilin, K.A., 2011. How helpful are spatial effects in forecasting the growth of Chinese provinces? Journal of Forecasting 30, 622–643; Longhi, S., Nijkamp, P., 2007. Forecasting regional labor market developments under spatial heterogeneity and spatial correlation. International Regional Science Review 30, 100–119; Marques, H., Pino, G., Horrillo J., 2014. Regional inflation dynamics using space-time models. Empirical Economics, Springer, vol. 47(3), November. P. 1147–1172.

² Beenstock M.,Felsenstein D. Spatial Vector Autoregressions//Spatial Economic Analysis/ - 2007. – Vol. 2(2). P. 167–96.

³ См. например: Badinger H., Muller W. G.,Tondl, G. Regional convergence in the European Union 1985-1999: a spatial dynamic panel analysis//Regional Studies. – 2004. – Vol. 38. P. 241–253; Lee, L.-F. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models//Econometrica. – 2004 – Vol. 72. P. 1899–1925.

⁴ Beenstock M.,Felsenstein D. Spatial Vector Autoregressions//Spatial Economic Analysis/ - 2007. – Vol. 2(2). P. 167–96.

⁵ Ibid.

δ – коэффициент пространственной автокорреляции в ошибках, γ – коэффициент запаздывающей пространственной автокорреляции в ошибках, ρ – коэффициент временной автокорреляции.

Также в данных может наблюдаться пространственная корреляция между остатками ε , если $\sigma_n^i \neq 0$, т.е. если ковариационная матрица остатков, Ω , является недиагональной.

Модель (1) – (2) строится на данных о реальной зарплате, численности населения, цен на жилье и предложении жилья в девяти регионах Израиля за 1987–2004 гг. На основе полученных оценок симулируются импульсные отклики. Согласно симуляции, шок зарплаты в одном регионе влияет как на другие экономические показатели данного региона, так и на показатели других регионов. Таким образом, авторы делают вывод, что учет пространственной корреляции в данных существенно влияет как на величину, так и на знаки импульсных откликов и повышает точность моделей, а пространственная автокорреляционная структура может усиливать или компенсировать импульсы при предположении, что шоки пространственно некоррелированы.

Если же шоки коррелированы, то применение SpVAR к таким данным может привести к смешению оценок. Однако авторы работы Badinger, Muller, Tondl, (2004)¹ считают, что метод оценивания динамических пространственно некоррелированных панелей, предложенный в работе Beenstock, Fisher (2007)², может быть применим к коррелированным данным, если эти данные предварительно отфильтрованы от пространственной зависимости. В работе Badinger, Muller, Tondl, (2004)³ предлагается подход, согласно которому пространственная зависимость в данных рассматривается как шум. То есть пространственную модель можно рассматривать как базовую модель без пространственных связей с добавлением пространственно коррелированного шума. Однако если это предположение неверно, то двухэтапная процедура, описанная в работе Badinger, Muller, Tondl, (2004)⁴, может отфильтровать важные компоненты в базовой модели.

В работе Schanne (2015)⁵ для рынка труда в Германии строится глобальная векторная авторегрессия (GVAR), с помощью которой анализируется пространственно-временная динамика показателей. GVAR была разработана Pesaran, Schuermann, Weiner (2004)⁶, Dees и др. (2007)⁷ для данных с высокой кросс-секционной зависимостью, т.е. для ситуаций, когда существует регион, показатели в котором коррелируют с показателями всех других регионов. Тогда считается, что все регионы зависят от «доминирующего».

Рассматривается система уравнений, состоящая из двух уравнений: уравнения для уровня занятости и для уровня безработицы. Уравнение для безработицы имеет следующий вид:

$$\Delta U_{it} = \Delta N_{it} - [m(\theta_i)(1 - a_i) + m(\theta_i^*)a_i U_{i,t-1} + \delta_i(1 - a_i)L_{i,t-1} + \delta_i^*a_i L_{i,t-1}^*], \quad (3)$$

где U_{it} – уровень безработицы в регионе i в момент времени t , N_{it} – размер рабочей силы при предположении, что все люди, переехавшие в регион i становятся безработными, ищущими работу, Θ_i – уровень напряженности на рынке труда, рассчитывающийся как отношение свободных вакансий к количеству безработных в регионе i , Θ_i^* – в соседних регионах, $1 - a_i$ – вероятность нахождения работника из региона i работы в своем регионе, a_i – в других регионах, L_{it-1} – уровень занятости, δ_i – уровень увольнения.

Уравнение для занятости имеет следующий вид:

$$\Delta L_{i,t} = -\delta_i L_{i,t-1} + m(\theta_i^*)(1 - a_i))U_{i,t-1} + m(\theta_i^*)a_i U_{i,t-1}^*. \quad (4)$$

¹ Badinger H., Muller W. G., Tondl, G. Regional convergence in the European Union 1985-1999: a spatial dynamic panel analysis// Regional Studies. – 2004. – Vol. 38. P. 241–253.

² Beenstock M., Felsenstein D. Spatial VectorAutoregressions//Spatial Economic Analysis / - 2007. – Vol. 2(2). P. 167–96.

³ Badinger H., Muller W. G., Tondl, G. Regional convergence in the European Union 1985-1999: a spatial dynamic panel analysis// Regional Studies. – 2004. – Vol. 38. P. 241–253.

⁴ Ibid.

⁵ Schanne N. A Global Vector Autoregression (GVAR) model for regional labour markets and its forecasting performance with leading indicators in Germany//AB Discussion Paper – 2015. – Vol. 13.

⁶ Pesaran, M. Hashem, Schuermann, T., Weiner, Scott M. Modeling Regional Interdependencies using a Global Error-Correcting Macroeconometric Model//Journal of Business and Economics Statistics – 2004 – Vol. 22. P. 129–162.

⁷ Dees, Stephane, Di Mauro, Filippo, Pesaran, M. Hashem, Smith, L. Vanessa. Exploring the International Linkages of the Euro Area: A Global VAR Analysis//Journal of Applied Econometrics – 2007 – Vol. 22. P. 1–38.

При построении GVAR вводятся следующие определения: y_{it} – вектор размерности $m \times 1$, состоящий из значений целевых переменных в регионе $i \in \{0, \dots, n\}$ ($y_{it} = (\ln U_{it}, \ln L_{it})$). Вектор $\xi_{it} = B_i x_{it}$ содержит переменные x_{it} , объясняющие безработицу и занятость. $Y_t = (y_{0t}, \dots, y_{nt})$, $\Xi_t = (\xi_{0t}, \dots, \xi_{nt})$, Γ_t – вектор случайных ошибок, φ_i – матрица коэффициентов при запаздывающих значениях переменных ($i=1, 2$).

Тогда GVAR для всех регионов записывается как:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 + \Xi_t. \quad (5)$$

Авторы отмечают, что данная система уравнений не решается без дополнительных ограничений в связи с небольшим количеством регионов и квадратичным ростом ковариационной матрицы остатков. Для введения ограничений для GVAR в работе Schanne (2015)¹ предлагается основываться на географическом положении регионов. Авторы предполагают, что информация о географической близости регионов позволит учесть одновременные колебания системы уравнений за счет коррелированных остатков.

Также в работе предполагается, что несмотря на то, что влияние всех регионов на исследуемый достаточно сильное, вклад одного соседнего региона в изменение рынка труда незначим. В связи с этим предлагается моделировать вклад других регионов следующим образом: влияние одного доминирующего региона моделируется в явном виде, для учета влияния остальных регионов используется их средневзвешенное влияние с весами w_{ijk} , где i – исследуемый регион, j – соседний, k – объясняющие переменные.

На основе оцененной системы уравнений авторы строят 3- и 12-месячные прогнозы с октября 2005 г. по декабрь 2011 г. В работе оценивались три модели: GVAR в разностях, VAR в уровнях и VAR в разностях. Полученные оценки позволяют сделать выводы, что ошибки прогнозов на 12 месяцев в 2,5–4,5 раз выше ошибок 3-месячных прогнозов. Авторы отмечают более высокую точность GVAR по сравнению с обеими VAR моделями, хотя разница между ошибками прогнозов незначительна. При прогнозировании уровня безработицы ошибка прогноза MAPE равна 0,979 при использовании GVAR против 0,987 при использовании VAR в разностях на 3-месячном горизонте и 0,935 против 0,978 соответственно на годовом горизонте.

В работе Patuelli, Mayor (2012)² исследуется точность пространственных моделей, в том числе пространственных векторных авторегрессий, при построении краткосрочных прогнозов региональной безработицы.

Авторы рассматривают страну, состоящую из N регионов, и для каждого региона $i=1, \dots, N$ моделируют уровень безработицы, y_{it} , следующего вида:

$$y_{it} = c_i + \sum_{p=1}^p \beta_{ip} y_{i,t-p} + \sum_{s=1}^s \delta_{is} W_s y_{i,t} + \sum_{s=1}^s \sum_{p=1}^s \gamma_{isp} W_s y_{i,t-p} + \varepsilon_{it}, \quad (6)$$

где W_s – пространственная весовая матрица, p – количество временных лагов, s – количество пространственных лагов.

Авторы отмечают, что в модель включены пространственные кросс-временные лаги, которые получаются перемножением временных лагов на матрицу пространственных весов. Так как параметры $W_s y_{it}$ и ε_{it} зависят от значений в соседних регионах, система уравнений оценивается с помощью внешне несвязанных уравнений (Seemingly Unrelated Regressions (SUR)), при этом коэффициенты пространственных лагов для одного периода времени оцениваются 2-шаговым методом с использованием предсказанных значений объясняемой переменной в качестве лага.

Авторы сравнивают результаты, полученные по VAR, с результатами динамической панели с гетерогенными коэффициентами и метода пространственного фильтра. Метод пространственного фильтра позволяет учитывать пространственную неоднородность и / или автокорреляцию. Данный подход основан на значении / Морана, который рассчитывается по формуле (7):

¹ Schanne N. A Global Vector Autoregression (GVAR) model for regional labour markets and its forecasting performance with leading indicators in Germany//AB Discussion Paper – 2015. – Vol. 13.

² Patuelli R., Mayor M. Short-Run Regional Forecasts: Spatial Models Through Varying Cross-Sectional and Temporal Dimensions//From Administrative Spatial Units to Local Labour Market Areas. Some Remarks on the Unit of Investigation of Regional Economics with Particular Reference to the Applied Research in Italy. – 2012. – P. 173–192.

$$I = \frac{N \sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i \sum_j w_{ij} \sum_i (x_i - \bar{x})^2}, \quad (7)$$

где x_i – значение показателя в регионе i , \bar{x} – среднее значение показателя во всех регионах, w_{it} – элемент весовой матрицы.

Формулой (8) авторы описывают матрицу C , которая может быть использована в качестве числителя для / Морана, и собственные значения которой представляют собой экстремальные значения / Морана:

$$C = (I_n - 11^T/n)W(I_n - 11^T)', \quad (8)$$

где 1 – вектор единиц размерностью $n \times 1$.

Авторы отмечают, что собственные вектора матрицы C представляют все ортогональные и независимые пространственные структуры, предполагаемые выбранной весовой матрицей. Для выбора наиболее значимых собственных векторов используется порог 0.25 для отношения $I(E_k)/\max_k I_k$. Линейная комбинация k собственных векторов называется пространственным фильтром.

Пространственный фильтр также может быть использован в качестве прокси переменной для гетерогенных коэффициентов, если строить фильтр на основании вектора авторегрессионных коэффициентов. В этом случае фильтр можно получить, добавляя каждому потенциальному собственному вектору дополнительный лаг зависимой переменной T раз. Таким образом будет создан набор новых переменных, представляющих пространственное разложение собственного вектора на ортогональные компоненты. Коэффициенты регрессии, полученные на основе этих новых переменных, отражают актуальность пространственных закономерностей в процессе корректировки.

Если метод пространственных фильтров применяется одновременно к коэффициенту серийной корреляции и константе, полученная модель записывается уравнением (9):

$$y_{it} = c + \beta y_{i,t-1} + \sum_{m=1}^k \beta_m E_{im} y_{i,t-1} + \sum_{m'=1}^{k'} \beta_{m'} E_{im'} + \varepsilon_{it}, \quad (9)$$

где k и k' – количество собственных векторов, выбранных для лагов объясняемой переменной и для константы, соответственно.

Авторы строят обе модели на региональных данных Испании и Швейцарии за 1975–2008 гг. Для сравнения точности прогнозов по SVAR и модели с пространственными фильтрами используются MAPE ошибки вневыборочных прогнозов за 2007–2008 гг. По результатам проведенного анализа авторы приходят к выводу, что модели SVAR демонстрируют более высокую точность в случаях, когда количество временных периодов значительно выше числа пространственных объектов.

В работе Di Giacinto (2002)¹ эффекты денежно-кредитной политики на штаты США оцениваются с помощью структурной VAR. В SVAR авторы включают три набора переменных:

- вектор x_t состоит из K макроэкономических переменных, таких как инфляция, цены на нефть и других переменных, учитывающих влияние шоков монетарной политики на спрос и предложение;
- вектор y_t состоит из значений объясняемой переменной во всех регионах;
- третий набор состоит из одной переменной r_t , которая представляет собой инструмент монетарной политики.

Таким образом, SVAR записывается с помощью уравнения (10):

$$C_0 z_t = C_1 z_{t-1} + \dots + C_p z_{t-p} + u_t, \quad (10)$$

где $z_t = [x'_t, y'_t, r_t]$, $u_t = [u_{1t}^x, \dots, u_{Kt}^x, u_{1t}^y, \dots, u_{Nt}^y, u_t^r]$ – многомерный ортогональный белый шум, т.е. $E(u_t) = 0$, и $E(u_t u'_{t-h}) = \Omega = diag([\sigma_{x1}^2, \dots, \sigma_{xK}^2, \sigma_{y1}^2, \dots, \sigma_{yN}^2, \sigma_r^2]')$, если $h = 0$ и $E(u_t u'_{t-h}) = 0$ иначе.

Авторы отмечают, что помимо ортогональной структуры для идентификации параметров также требуется выполнение ограничения на матрицу C_0 размерности $(R \times R): R = K + N + 1$. Для выполнения

¹ Di Giacinto V., Differential regional effects of monetary policy: a geographical SVAR approach//ERSA conference papers ersa02p257, European Regional Science Association, 2002.

ограничения в соответствии с исследованием [58] идентификация параметров возможна при использовании следующих предположений в модели:

- шоки, специфичные для региона, влияют на данный регион только в тот период, когда шок произошел, в то время как влияние на соседние регионы может быть и в следующих периодах;
- шоки, связанные с кредитно-денежной политикой и с макропеременными, влияют на региональные доходы не раньше, чем в следующем периоде после того, как эти шоки произошли;
- шоки макропеременных (вектор x_t) не подвержены влиянию шоков других переменных в модели и не влияют друг на друга.

С учетом данных предположений матрица C_0 принимает следующий вид (11):

$$C_0 = \begin{bmatrix} I_K & 0 & 0 \\ 0 & C_0^{yy} & 0 \\ -C_0^{rx} & -C_0^{ry} & 1 \end{bmatrix}, \quad (11)$$

где I_k – единичная матрица размерности K , C_0^{rx} – вектор коэффициентов, относящихся к инструментам политики для значений макропеременных размерности $(1 \times K)$, $C_0^{ry} = a_0^{ry} w'$, a_0^{ry} – скалярная переменная, которую необходимо оценить, w – вектор N долей каждого региона в экономике страны, $C_0^{yy} = I_N - \sum_{k=1}^{\lambda_0} \Phi_{0k} W_k$, где $\Phi_{0k} = \text{diag}([\varphi_{0k}^1, \varphi_{0k}^2, \dots, \varphi_{0k}^N])$, W_k – матрица пространственных весов с порядком лагов k размерностью $N \times N$, φ_{0k}^i – значение объясняемой переменной в регионе i .

Данная модель оценивается с помощью метода максимального правдоподобия. В качестве объясняемой переменной в модели для штатов США применяются ряды реальных доходов населения в каждом штате за период со II кв. 1958 г. по IV кв. 2000 г. Авторы использовали различные спецификации моделей, различающиеся объясняющими переменными. В первой спецификации набор макроэкономических переменных включает только цены на нефть, которые рассчитываются как отношение индекса цен производителей лишь по нефтяным товарам и индекса цен производителей по всем товарам.

По результатам оценивания авторы делают вывод, что изменения в одном регионе значимо влияют на соседние регионы, где матрица соседей состоит из штатов, имеющих общую границу. Все коэффициенты φ_{0k}^i значимы и в среднем равны 0,57 при среднем стандартном отклонении 0,11. Импульсные отклики также оказались значительными и в среднем достигают долгосрочного уровня через 12 кварталов после шока.

Во второй спецификации помимо цен на нефть в качестве объясняющих факторов использовались ИПЦ и отрасль. Корреляция остатков, как и в первой модели, незначительна, среднее значение φ_{0k}^i составило 0,56 при стандартном отклонении 0,11. После шока в одном штате долгосрочное равновесие в других штатах достигается примерно через 10 кварталов. Таким образом, авторы приходят к выводу, что в целом структурная VAR дает устойчивые оценки региональных показателей.

По результатам проведенного обзора эмпирических работ можно сделать вывод, что VAR, учитывающие пространственную корреляцию между данными дают более точные результаты по сравнению с другими методами.

Построение SpVAR на российских региональных данных

Поскольку пространственные модели дают точные прогнозы макроэкономических показателей для стран с большим количеством регионов, во второй части данной статьи указанные модели применялись к российским региональным данным. По аналогии с работами Beenstock, Fisher (2007)¹ и Patuelli, Mayor (2012)² для региональных показателей помимо стандартных VAR также строятся модели, учитывающие пространственную зависимость между данными в соседних регионах. Для анализа применяются годовые показатели за 2000–2017 гг. В качестве эндогенных переменных используются

¹ Beenstock M., Felsenstein D. Spatial VectorAutoregressions//Spatial Economic Analysis/ - 2007. – Vol. 2(2). P. 167–96.

² Patuelli R., Mayor M. Short-Run Regional Forecasts: Spatial Models Through Varying Cross-Sectional and Temporal Dimensions//from book From Administrative Spatial Units to Local Labour Market Areas. Some Remarks on the Unit of Investigation of Regional Economics with Particular Reference to the Applied Research in Italy. – 2012. – P. 173–192.

Прогонозирование региональных макропоказателей

значения ВРП и регионального ИПЦ, зависящие от лагов этих же переменных, а также от экзогенной переменной – уровня безработицы.

Стандартная VAR, таким образом, записывается с помощью формулы (12):

$$\begin{cases} \pi_{it} = \alpha_{1i} + \alpha_{2i} Y_{it-1} + \alpha_{3i} \pi_{it-1} + \alpha_{4i} \text{unem}_{it-1} + \varepsilon_{it} \\ Y_{it} = \beta_{1i} + \beta_{2i} Y_{it-1} + \beta_{3i} \pi_{it-1} + \beta_{4i} \text{unem}_{it-1} + \mu_{it} \end{cases} \quad (12)$$

где π_{it} – значение инфляции в регионе i в момент времени t , Y_{it} – значение ВРП в регионе i в момент времени t , unem_{it-1} – значение безработицы в регионе i в предыдущий момент времени $t - 1$.

Модель, соответствующая структуре пространственной VAR модели, записывается как:

$$\begin{cases} \pi_{it} = \alpha_{1i} + \alpha_{2i} Y_{it-1} + \alpha_{3i} \pi_{it-1} + \alpha_{4i} \text{unem}_{it-1} + \alpha_5 \sum_{j=1}^N w_{ij} \pi_{jt} + \varepsilon_{it} \\ Y_{it} = \beta_{1i} + \beta_{2i} Y_{it-1} + \beta_{3i} \pi_{it-1} + \beta_{4i} \text{unem}_{it-1} + \beta_5 \sum_{j=1}^N w_{ij} Y_{jt} + \mu_{it} \end{cases}, \quad (13)$$

где w_{ij} – элемент весовой матрицы, характеризующий взаимосвязь между регионами i и j .

Для учета значений показателей в соседних регионах используется весовая матрица смежности, в которой регионам, имеющим общую морскую или сухопутную границу, присвоено значение 1, остальным – 0.

При построении моделей предполагается, что большее влияние на значение определенного показателя оказывает значение того же показателя, как в исследуемом регионе в предыдущем периоде, так и в соседних регионах. При этом значение второго эндогенного и экзогенного показателей в соседних регионах незначимо влияет на исследуемый показатель. В связи с этим в уравнение для инфляции включены только значения инфляции в соседних регионах и исключены значения ВРП и безработицы.

Помимо модели, учитывающей значения показателей в соседних регионах в тот же момент времени, также построена модель, учитывающая лаги исследуемых переменных. Данная модель имеет вид (14):

$$\begin{cases} \pi_{it} = \alpha_{1i} + \alpha_{2i} Y_{it-1} + \alpha_{3i} \pi_{it-1} + \alpha_{4i} \text{unem}_{it-1} + \alpha_5 \sum_{j=1}^N w_{ij} \pi_{jt-1} + \varepsilon_{it} \\ Y_{it} = \beta_{1i} + \beta_{2i} Y_{it-1} + \beta_{3i} \pi_{it-1} + \beta_{4i} \text{unem}_{it-1} + \beta_5 \sum_{j=1}^N w_{ij} Y_{jt-1} + \mu_{it} \end{cases}. \quad (14)$$

Все модели были оценены с помощью пакета panelvar в языке программирования R методом обобщенных моментов. Результаты оценивания моделей представлены в табл. 1.

Таблица 1

Результат оценивания моделей

		Лаг инфляции	Лаг ВВП	Лаг безработицы	Инфляция соседей	ВВП соседей
VAR (12)	Инфляция	0,6534***	-0,1381*	0,5375*		
	ВРП	-0,018	0,8132***	-0,3141		
SpVAR (13)	Инфляция	0,6406***	-0,2055*	0,2344	39,6263**	0,6871**
	ВРП	0,004	0,704***	0,1035	-1,307	0,3902

*коэффициент значим на уровне значимости 5%, **коэффициент значим на уровне значимости 1%, ***коэффициент значим на уровне значимости 0,1%.

Источник: расчеты авторов.

Согласно полученным результатам, в стандартной VAR модели (12) и на инфляцию, и на ВРП значимо влияют их предыдущие значения, а на инфляцию также положительно влияет рост уровня безработицы в предыдущем периоде и отрицательно влияет рост ВРП в предыдущем периоде, что соответствует экономической теории. На ВРП не влияют ни лаг инфляции, ни лаг безработицы. Согласно тесту Хансена, инструменты, примененные при использовании ОММ, не сверхидентифицированы.

При добавлении инфляции и ВРП соседей (модель 13) значение безработицы в предыдущем периоде становится незначимым для инфляции, в то время как лаги инфляции и ВРП в рассматриваемом регионе и текущие значения ВРП и инфляции в соседних регионах значимы на 5% уровне значимости. Для ВРП единственным значимым показателем по-прежнему остается лаг ВРП в рассматриваемом

регионе. Согласно тесту Хансена, инструменты, примененные при использовании ОММ, не сверхидентифицированы. В табл. 2 приведены оценки коэффициентов.

Таблица 2

Результат оценивания модели с пространственными лагами

		Лаг инфляции	Лаг ВВП	Лаг безработицы	Лаг инфляции соседей	Лаг ВВП соседей
Модель 14	Инфляция	0,6884***	0,1286*	0,1622	-20,1772**	-0,8164***
	ВРП	-0,0048	0,647***	0,1947	0,1716	0,4526*

*коэффициент значим на уровне значимости 5%, **коэффициент значим на уровне значимости 1%, ***коэффициент значим на уровне значимости 0,1%.

Источник: расчеты авторов.

При учете в модели не текущих, а запаздывающих значений ВРП и инфляции в соседних регионах результаты моделирования инфляции практически не изменяются: все показатели, кроме лага безработицы, оказываются значимыми на 5%-ном уровне значимости. Для ВРП, в свою очередь, кроме лага ВРП в рассматриваемом периоде значимыми на 5%-ном уровне значимости становятся лаги ВРП в соседних регионах. Тест Хансена по-прежнему не выявляет сверхидентифицируемость инструментов.

Помимо модели с одной экзогенной переменной – безработицей – также была построена модель, в которой в качестве объясняющих переменных используются индекс промышленного производства и среднедушевые доходы населения. Оба показателя также нормализованы.

Однако при построении модели, учитывающей в качестве экзогенного фактора индекс промышленного производства, возникает ошибка, связанная с вырожденностью матрицы показателей. Одной из причин такой ошибки является возможная высокая корреляция между ИПП и ИПЦ, которая приводит к невозможности оценивания параметров модели.

В связи с этим оценивалась модель, в которой в качестве экзогенных переменных использовались запаздывающее значение безработицы и значение среднедушевых доходов. Коэффициенты, полученные при моделировании, указаны в табл. 3.

Таблица 3

Результат оценивания модели с пространственными лагами

		Лаг инфляции	Лаг ВВП	Лаг безработицы	Лаг инфляции соседей	Лаг ВВП соседей	Доходы населения
Модель с доходами	Инфляция	0,6879***	0,1512*	0,2166	-19,7502**	-0,9841***	0,1595
	ВРП	-0,005	0,6566***	0,219	0,3635	0,3816	0,0685

*коэффициент значим на уровне значимости 5%, **коэффициент значим на уровне значимости 1%, ***коэффициент значим на уровне значимости 0,1%.

Источник: расчеты авторов.

Добавление в модель среднедушевых доходов не улучшает результаты моделирования: для инфляции результаты совпадают с результатами моделей, описанных выше. Для ВРП значимым становится только лаг ВРП в рассматриваемом регионе.

По результатам полученных оценок и в соответствии с исследованием Baltagi (2004)¹, согласно которому даже при отсутствии значимой пространственной связи между регионами для получения эффективных оценок лучше использовать пространственные модели, для моделирования региональных показателей предлагается использовать модели (12)–(14), в которых учитывается пространственная связь между регионами. Так как из указанных моделей только экономически ожидаемые знаки² при коэффициентах ВРП и инфляции дает только модель (14), предлагается именно ее ис-

¹ Baltagi,B.H., Li,D. Prediction in the panel data model with spatial correlation. In: Anselin, L., Florax, R.J.G.M., Rey, S.J. (Eds.), Advances in Spatial Econometrics: Methodology, Tools and Applications//Springer, Berlin – 2004 – P. 283–295 (Chapter 13).

² <https://cyberleninka.ru/article/n/vliyanie-inflyatsii-na-tempy-ekonomicheskogo-rosta>

Прогонозирование региональных макропоказателей

пользовать для анализа российских региональных данных. Однако для получения окончательного вывода необходимо также рассмотреть ошибки прогнозов, полученные по всем моделям.

Качество полученных внутривыборочных прогнозов анализировалось с помощью MAPE (средней абсолютной ошибки прогноза, выраженной в процентах), которая рассчитывается по формуле (15):

$$MAPE = \frac{1}{N} \sum_{t=1}^N \frac{|y_t - \hat{y}_t|}{y_t} * 100\%, \quad (15)$$

где y_t – фактическое значение прогнозируемого показателя, \hat{y}_t – спрогнозированное значение.

Несмотря на то, что на основе значимости полученных оценок было выбрано две из четырех моделей для прогнозирования региональных ВРП и ИПЦ, для сопоставимости ошибки прогноза были рассчитаны для всех проанализированных VAR.

Для сравнения качества моделей для обоих показателей были построены AR(1) модели по формуле (16) и были посчитаны усредненные по регионам ошибки внутривыборочных прогнозов:

$$y_{it} = \alpha_i + \beta_i y_{it-1} + \varepsilon_{it}, \varepsilon_{it} \sim NID(0, \sigma_i^2), \quad (16)$$

где y_{it} – значение ВРП / инфляции для субъекта i в период времени t .

В табл. 4 приведены значения всех посчитанных ошибок прогноза.

Таблица 4
Ошибки прогнозов VAR моделей

Модель	MAPE	
	инфляции	ВРП
AR(1)	2%	13%
Непространственная (12)	4,56%	2,06%
Пространственная с текущими показателями соседей (13)	9,17%	10%
Пространственная с запаздывающими показателями соседей (14)	8,67%	2,88%
Пространственная с учетом доходов	7,87%	4,08%

Источник: расчеты авторов.

Из табл. 4 видно, что модели, учитывающие в качестве объясняющих переменных безработицу и пространственную зависимость между региональными данными, дают наибольшие ошибки прогнозов. AR(1) модель для региональной инфляции дает результаты, сопоставимые по точности с VAR моделями, в то время как точность AR(1) модели для ВРП ниже всех рассмотренных VAR. Для региональной инфляции MAPE ошибки по четырем VAR моделям сопоставимы друг с другом, в то время как для ВРП модель с текущими значениями пространственных показателей в 2-3 раза выше ошибок других моделей. Интересно отметить, что наименьшие ошибки прогнозов дает непространственная модель, что, однако, не соответствует тому, что оценки коэффициентов при пространственных показателях оказываются значимыми.

* * *

В первой части статьи проведен обзор работ, посвященных построению различных спецификаций векторных авторегрессий. В большинстве работ был получен вывод, что VAR, учитывающие пространственную корреляцию между данными (spatial VAR), дают более точные результаты прогнозирования макроэкономических показателей по сравнению с другими методами. В отличие от моделей с непрерывным временем, для построения VAR можно использовать годовые данные, что позволяет строить пространственные векторные авторегрессии на российских данных.

Во второй части статьи для региональных значений инфляции и ВРП было построено четыре векторных авторегрессии, в том числе три из них – учитывающие пространственную взаимосвязь между регионами. Среди векторных авторегрессий наименьшая ошибка прогноза получается по модели, учитывающей в качестве объясняющих переменных только первые лаги объясняемых переменных и первый лаг безработицы. Однако в моделях, учитывающих пространственную корреляцию между данными, оценки коэффициентов перед пространственными переменными оказываются значимыми.

В связи с этим и в соответствии с исследованием Baltagi [13], согласно которому даже при отсутствии значимой пространственной связи между регионами, для получения эффективных оценок лучше использовать пространственные модели, для моделирования региональных показателей предлагается использовать модели, в которых учитывается пространственная связь между регионами. Из трех рассмотренных пространственных моделей наименьшая ошибка прогноза для инфляции получается при добавлении в список объясняющих показателей показателя среднедушевых доходов населения. Наименьшая ошибка прогноза для ВРП получается в модели, где пространственные показатели используются с первым лагом. Поскольку в модели с учетом дохода оценки коэффициентов при среднедушевых доходах оказываются незначимыми как для инфляции, так и для ВРП, а также учитывая тот факт, что из рассмотренных моделей VAR со следующим набором объясняющих переменных: первые лаги объясняемых переменных в исследуемом и соседних регионах и первый лаг безработицы – дает экономически ожидаемые знаки при коэффициентах ВРП и инфляции, предлагается именно ее использовать для анализа российских региональных данных.

Таким образом, по результатам проведенного исследования можно сделать вывод, что векторные авторегрессии позволяют достаточно точно смоделировать как фактические значения макропоказателей, так и общую ситуацию в регионах России.

Источники

1. Badinger H., Muller W. G., Tondl, G. Regional convergence in the European Union 1985–1999: a spatial dynamic panel analysis//Regional Studies. – 2004. – Vol. 38. P. 241–253.
2. Schanne N. A Global Vector Autoregression (GVAR) model for regional labour markets and its forecasting performance with leading indicators in Germany//AB Discussion Paper – 2015. – Vol. 13.
3. Patuelli R., Mayor M. Short-Run Regional Forecasts: Spatial Models Through Varying Cross-Sectional and Temporal Dimensions//from book From Administrative Spatial Units to Local Labour Market Areas. Some Remarks on the Unit of Investigation of Regional Economics with Particular Reference to the Applied Research in Italy. – 2012. – P. 173–192.
4. Girardin, E., Kholodilin, K.A., 2011. How helpful are spatial effects in forecasting the growth of Chinese provinces? Journal of Forecasting 30. P. 622–643.
5. Longhi, S., Nijkamp, P., 2007. Forecasting regional labor market developments under spatial heterogeneity and spatial correlation. International Regional Science Review 30. P. 100–119.
6. Marques, H., Pino, G., Horrillo J., 2014. Regional inflation dynamics using space–time models. Empirical Economics, Springer, vol. 47(3), November. P. 1147–1172.
7. Beenstock M., Felsenstein D. Spatial VectorAutoregressions//Spatial Economic Analysis/ – 2007. – Vol. 2(2). P. 167–196.
8. Badinger H., Muller W. G., Tondl, G. Regional convergence in the European Union 1985-1999: a spatial dynamic panel analysis//Regional Studies. – 2004. – Vol. 38. P. 241–253.
9. Lee, L.-F. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models//Econometrica – 2004 – Vol. 72. P. 1899–1925.
10. Pesaran, M. Hashem, Schuermann, T., Weiner, Scott M. Modeling Regional Interdependencies using a Global Error-Correcting Macroeconometric Model.//Journal of Business and Economics Statistics – 2004 – Vol. 22. P. 129–162.
11. Dees, Stephane, Di Mauro, Filippo, Pesaran, M. Hashem, Smith, L. Vanessa. Exploring the International Linkages of the Euro Area: A Global VAR Analysis//Journal of Applied Econometrics – 2007 – Vol. 22. P. 1–38.
12. Di Giacinto V., Differential regional effects of monetary policy: a geographical SVAR approach//ERSA conference papers ersa02p257, European Regional Science Association, – 2002.
13. Baltagi, B.H., Li,D. Prediction in the panel data model with spatial correlation. In: Anselin, L., Florax, R.J.G.M., Rey, S.J. (Eds.), Advances in Spatial Econometrics: Methodology, Tools and Applications//Springer, Berlin – 2004 – P. 283–295 (Chapter 13).