Прикладная эконометрика, 2016, т. 43, с. 118–141.
Applied Econometrics, 2016, v. 43, pp. 118–141.

Б. Б. Демешев, О. А. Малаховская[1]

Картографирование BVAR

Работа представляет собой обзор метода оценивания и прогнозирования с помощью BVAR. Предлагается классификация наиболее часто встречающихся априорных распределений и показывается, как параметры апостериорных распределений могут быть рассчитаны для каждого рассмотренного типа априорных распределений. Отдельный раздел описывает эндогенный выбор гиперпараметров априорных распределений, что в настоящее время является ключевым шагом при построении BVAR большой размерности. Одна из частей работы посвящена прогнозированию с помощью BVAR. Авторы рассматривают как точечное прогнозирование, так и прогнозирование плотности.

Ключевые слова: BVAR; априорные распределения; точечное прогнозирование; прогнозирование плотности.

JEL classification: C11; C32; C53.

1. Введение

Значение точных прогнозов для проведения макроэкономической политики трудно переоценить. Существование лагов политики приводит к тому, что решения, принятые сегодня, могут повлиять на экономику через некоторое время, поэтому при принятии решений приходится ориентироваться не только на текущие, но и на ожидаемые показатели. Точный прогноз макроэкономических показателей, таким образом, является одним из ключевых факторов успешной политики.

В настоящее время основной моделью для прогнозирования макроэкономических временных рядов является модель векторной авторегрессии (VAR) и ее модификации. Использование векторных авторегрессий в макроэкономическом анализе явилось следствием критики активно использовавшихся прежде традиционных эконометрических моделей. В частности, Sims (1980) обратил внимание на необоснованность ограничений, вводимых в рамках традиционных моделей[2], и предложил использовать более простую по построению динамическую модель VAR, основанную на разложении Вольда и не требующую введения никаких ограничений на взаимную динамику переменных.

[1] **Демешев Борис Борисович —** Национальный исследовательский университет «Высшая школа экономики», Москва; boris.demeshev@gmail.com.

Малаховская Оксана Анатольевна — Национальный исследовательский университет «Высшая школа экономики», Москва; omalakhovskaya@hse.ru.

[2] Под традиционными понимаются модели, построенные в рамках подхода комиссии Коулза. Их прогнозная способность резко ухудшилась в начале 1970-х годов, т. е. примерно тогда же, когда «исчезла» базовая кривая Филлипса (подробнее об этом см. (Favero, 2001; Малаховская, Пекарский, 2012)).

Модели этого класса стали широко использоваться как для прогнозирования, так и для структурного анализа благодаря своей логичности и относительной простоте. Однако для того, чтобы правильно отражать динамику фактических временных рядов, VAR часто требуется большое число лагов, что может привести к высоким ошибкам прогноза. Проблема усугубляется тем, что центральные банки развитых стран при проведении политики ориентируются на большое количество показателей, и VAR малой размерности не может отразить всей доступной им информации. Следовательно, использование моделей высокой размерности потенциально может улучшить качество прогноза. При этом увеличение числа переменных в VAR приводит к тому, что число оцениваемых параметров растет нелинейно. Это усугубляет проблему неэффективности оценивания и больши́х ошибок прогноза. Одним из решений этой проблемы стало использование априорной информации относительно распределения параметров и ковариационной матрицы ошибок, т. е. переход от обычных[3] VAR к байесовским (Bayesian VAR, BVAR).

Исследователи выделяют несколько преимуществ байесовского подхода по отношению к частотному. Во-первых, он позволяет преодолеть численные трудности, связанные с максимизацией функции правдоподобия. В моделях с большим числом параметров или в тех, которые приходится оценивать на коротких выборках, функция правдоподобия может быть достаточно плоской или содержать много локальных максимумов. В этом случае стандартные процедуры поиска экстремума могут не дать желаемого результата. Байесовский подход позволяет заменить сложную, иногда нерешаемую задачу численной процедурой генерирования реализаций случайных величин. Во-вторых, введение априорных вероятностей позволяет снизить неопределенность в распределении параметров модели. В частности, распространенные в настоящее время априорные распределения отражают современные представления о долгосрочной динамике переменных, которые не могут быть проверены на коротких выборках, обычно используемых для анализа. Например, исследователь может предполагать наличие единичного корня во временном ряде или коинтегрированность нескольких временных рядов, при этом тесты на коротких выборках могут этого не обнаруживать. Добавление априорной информации, в явном виде учитывающей представления исследователя, может улучшить точность полученных прогнозов. В-третьих, байесовская оценка является более общей по отношению к оценкам МНК или метода максимального правдоподобия. Вопреки распространенному мнению, байесовский подход не более субъективен, чем частотный. Используя частотный подход, исследователь все равно неявно предполагает некоторое априорное распределение и модель для данных. В этой ситуации представляется более разумным задать такое априорное распределение, которое отражает представления исследователя о параметрах модели, или то, которое лучше теоретически обосновано.

К сожалению, несмотря на широкое распространение BVAR в научных статьях, число практических обзоров этого метода весьма ограничено. Существующие обзоры (Karlsson, 2013; Del Negro, Schorfheide, 2011) и изложение в учебнике (Canova, 2007) сильно математизированы и едва ли доступны для экономистов без специальной математической подготовки. При этом ни в одном из них не содержится достаточно подробной классификации априорных распределений, и к большинству из них не прилагается инструкций для реализации предложенных методов в эконометрическом пакете. Исключениями являются обзоры (Koop,

[3] В англоязычной литературе обычные VAR (без априорных распределений) называются частотными — frequentist.

Korobilis, 2010) и (Blake, Mumtaz, 2012), к которым прилагаются коды в среде MATLAB[4]. Однако Koop, Korobilis (2010) не рассматривают ставший весьма популярным метод задания априорного распределения через добавление дополнительных наблюдений, в том числе априорное распределение суммы коэффициентов (sum-of-coefficients prior) и априорное распределение начального наблюдения (initial observation prior). В работе (Blake, Mumtaz, 2012) используется терминология, несколько отличающаяся от других работ, а код фактически содержит пример построения BVAR только с использованием семплирования по Гиббсу. При этом ни в одном из указанных обзоров не рассмотрен достаточно подробно вопрос о прогнозировании с помощью BVAR (а именно это и является обычно целью их построения, по крайней мере, BVAR в сокращенной форме). В российской научной литературе существуют обзоры общих принципов оценки моделей с помощью байесовского подхода (Айвазян, 2008; Слуцкин, 2010), но, насколько известно авторам, нет обзоров BVAR.

Нестандартное название настоящей работы вызвано тем, что приводимый в ней обзор представляет собой своеобразную «карту» априорных распределений для BVAR. Он содержит подробную классификацию априорных распределений, наиболее популярных при проведении макроэкономических исследований, и схему их взаимосвязей. Данная работа может быть полезна для экономистов, обладающих ограниченным опытом в области байесовского анализа. Кроме того, авторами написан пакет *bvarr* для R, в котором используются те же обозначения, что и в данном тексте, и который может быть использован как в учебных, так и в научных целях[5].

За рамками данного обзора остаются построение структурных BVAR (SBVAR), BVAR с меняющимися параметрами (TVP-BVAR), BVAR со стохастической волатильностью, а также проблемы выбора переменных при построении BVAR.

2. Оценивание BVAR

2.1. Байесовская VAR: формулировка модели

Рассмотрим переменные[6] y_{it}, объединенные в вектор $y_t = (y_{1t}, y_{2t}, \ldots, y_{mt})'$ размерности m ($t = 1,...,T$). Векторная авторегрессия в сокращенной форме записывается в виде

$$y_t = \Phi_{const} + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \ldots + \Phi_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma), \tag{1}$$

где $\Phi_{const} = (c_1, \ldots, c_m)'$ — вектор констант размерности m, Φ_l — авторегрессионные матрицы размерности $m \times m$, $l = 1, \ldots, p$. Вектор ε_t — m-мерный нормальный вектор ошибок, некоррелированный с объясняющими переменными. Группируя матрицы параметров в общую матрицу $\Phi = [\Phi_1 \ldots \Phi_p \; \Phi_{const}]'$ и определяя новый вектор $x_t = [y'_{t-1} \ldots y'_{t-p} \; 1]'$, получаем VAR в более компактном виде

$$y_t = \Phi' x_t + \varepsilon_t. \tag{2}$$

---

[4] Находящиеся в открытом доступе известные авторам коды в различных пакетах для работы с BVAR описаны в Приложении 1.

[5] См. Приложение 1, п. 8.

[6] Для удобства читателя все обозначения дополнительно вынесены в Приложение 2.

Если сгруппировать переменные и шоки следующим образом:

$$Y = [y_1, y_2, \ldots, y_T]', \; X = [x_1, x_2, \ldots, x_T]', \; E = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T]',$$

то VAR можно записать как

$$Y = X\Phi + E. \tag{3}$$

Эта модель может быть записана и в векторизованном виде[7]

$$vec(Y) = vec(X\Phi I_m) + vec(E) \;\Leftrightarrow \tag{4}$$

$$\Leftrightarrow \; y = (I_m \otimes X)\phi + \varepsilon, \tag{5}$$

где $\varepsilon \sim \mathcal{N}(0, \Sigma \otimes I_T)$ и вектор $\phi = vec(\Phi)$ имеет размерность $km \times 1$. Матрицы I_m, I_T — единичные размерности $m \times m$ и $T \times T$ соответственно. Константа k обозначает число параметров в отдельном уравнении и равна $k = mp + 1$.

Задача байесовского оценивания заключается в поиске апостериорных распределений параметров $p(\Phi, \Sigma \mid Y)$ с использованием функции максимального правдоподобия $p(Y \mid \Phi, \Sigma)$ и априорного распределения $p(\Phi, \Sigma)$. Для этого используется правило Байеса

$$p(\Phi, \Sigma \mid Y) = \frac{p(\Phi, \Sigma)\, p(Y \mid \Phi, \Sigma)}{p(Y)}. \tag{6}$$

Так как $p(Y)$ не зависит от Φ и Σ, то апостериорную плотность можно представить в виде

$$p(\Phi, \Sigma \mid Y) \propto p(\Phi, \Sigma)\, p(Y \mid \Phi, \Sigma), \tag{7}$$

где символ $\propto$ означает, что правая часть выражения равна его левой части с точностью до умножения на константу, не зависящую от аргументов функции, в данном случае от Φ и Σ. Поскольку $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$, то функция правдоподобия задается как

$$p(Y \mid \Phi, \Sigma) \propto |\Sigma|^{-T/2} \, etr\left\{-\frac{1}{2}\left[\Sigma^{-1}(Y - X\Phi)'(Y - X\Phi)\right]\right\}, \tag{8}$$

где $etr(\cdot) = \exp(tr(\cdot))$. Другая форма записи функции правдоподобия:

$$p(Y \mid \Phi, \Sigma) \propto |\Sigma|^{-T/2} \, etr\left\{-\frac{1}{2}\left[\Sigma^{-1}\hat{E}'\hat{E}\right]\right\} \times etr\left\{-\frac{1}{2}\left[\Sigma^{-1}(\Phi - \hat{\Phi})' X'X (\Phi - \hat{\Phi})\right]\right\}, \tag{9}$$

где $\hat{E} = Y - X\hat{\Phi}$ и $\hat{\Phi} = (X'X)^{-1}X'Y$.

В двух следующих разделах будут разобраны наиболее известные априорные и построенные на их основе апостериорные распределения.

[7] Уравнение (5) следует из тождества $vec(ABC) = (C \otimes A)vec(B)$.

2.2. Оценка моделей с различными априорными распределениями

2.2.1. Априорное распределение Миннесоты

Решение проблемы избыточной идентификации на основе байесовских методов было предложено в работе (Litterman, 1979), где показано, что введение ограничений в форме априорных распределений параметров увеличивает точность оценок и прогнозов. Априорное распределение, получившее название «априорное распределение Миннесоты» (Minnesota prior), было предложено в работе (Litterman, 1986) и с некоторыми модификациями в (Doan et al., 1984).

Ковариационная матрица Σ вектора ε_t предполагается диагональной и постоянной. Априорное распределение параметров предполагается многомерным нормальным, зависящим от нескольких гиперпараметров:

$$\phi \sim \mathcal{N}(\underline{\phi}, \underline{\Xi}). \tag{10}$$

Параметры ϕ предполагаются независимыми, следовательно, их ковариационная матрица $\underline{\Xi}$ диагональна. Априорная плотность распределения ϕ может быть записана как

$$p(\phi) = \frac{1}{(2\pi)^{km/2}\,|\underline{\Xi}|^{1/2}} \exp\left\{-\frac{1}{2}(\phi - \underline{\phi})'\underline{\Xi}^{-1}(\phi - \underline{\phi})\right\}. \tag{11}$$

Комбинируя ее с функцией правдоподобия (8), получаем, что апостериорное распределение параметров задается в следующем виде:

$$\phi \,|\, Y \sim \mathcal{N}(\overline{\phi}, \overline{\Xi}), \tag{12}$$

где $\overline{\Xi} = [\underline{\Xi}^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1}$, $\overline{\phi} = \overline{\Xi}[\underline{\Xi}^{-1}\underline{\phi} + (\Sigma^{-1} \otimes X')y]$.

Если $\underline{\Xi}$ имеет структуру кронекерова произведения $\underline{\Xi} = \Sigma \otimes \underline{\Omega}$, то формулы можно существенно упростить и обойтись обращением матриц меньшей размерности:

$$\overline{\Xi} = [\underline{\Xi}^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1} = [(\Sigma \otimes \underline{\Omega})^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1} =$$

$$= [\Sigma^{-1} \otimes \underline{\Omega}^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1} = \Sigma \otimes (\underline{\Omega}^{-1} + X'X)^{-1} = \Sigma \otimes \overline{\Omega}, \text{ где } \overline{\Omega} = \left(\underline{\Omega}^{-1} + X'X\right)^{-1}. \tag{13}$$

В результате получаем $\Phi \,|\, Y \sim \mathcal{N}\left(\overline{\Phi}, \Sigma \otimes \overline{\Omega}\right)$.

На практике в качестве матрицы Σ используют ее оценку $\hat{\Sigma}$, диагональные элементы которой равны $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_m^2$, где $\hat{\sigma}_i^2$ — оценка дисперсии случайной составляющей в $\mathrm{AR}(p)$ модели для ряда i. При этом некоторые авторы для подсчета оценки дисперсии используют AR(1) модель, даже если сама VAR имеет большее число лагов.

Математическое ожидание априорного распределения параметров может быть записано с помощью матрицы $\underline{\Phi} = \mathbf{E}(\Phi)$ размерности $k \times m$, где $\underline{\Phi} = [\underline{\Phi}_1 \ldots \underline{\Phi}_p \; \underline{\Phi}_{const}]'$ и $\underline{\phi} = vec(\underline{\Phi})$:

$$\left(\underline{\Phi}_l\right)_{ij} = \begin{cases} \delta_i, & \text{если } i = j,\ l = 1, \\ 0, & \text{в остальных случаях.} \end{cases} \tag{14}$$

Распределение Миннесоты было задумано таким образом, чтобы учесть нестационарность многих макроэкономических временных рядов. В настоящее время широкое распространение получила практика назначать $\delta_i = 1$ для нестационарных рядов и $\delta_i < 1$ для стационарных.

Распределение Миннесоты предполагает, что априорная ковариационная матрица параметров $\underline{\Xi}$ диагональна. Диагональ матрицы $\underline{\Xi}$ разбивается на блоки $\underline{\Xi}_1, \underline{\Xi}_2, \ldots, \underline{\Xi}_m$ размерности $k \times k$. В свою очередь, каждый блок $\underline{\Xi}_i$, $i = 1, \ldots, m$ может быть разбит на диагональные подблоки размерности $m \times m$: $\underline{\Xi}_{i,lag=l}$, $l = 1, \ldots, p$ с константой $\underline{\Xi}_{i,const}$ в конце главной диагонали:

$$\underline{\Xi} = \begin{pmatrix} \underline{\Xi}_1 & 0_{k\times k} & \cdots & 0_{k\times k} & 0_{k\times k} \\ 0_{k\times k} & \underline{\Xi}_2 & \cdots & 0_{k\times k} & 0_{k\times k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{k\times k} & 0_{k\times k} & \cdots & \underline{\Xi}_{m-1} & 0_{k\times k} \\ 0_{k\times k} & 0_{k\times k} & \cdots & 0_{k\times k} & \underline{\Xi}_m \end{pmatrix}, \quad \underline{\Xi}_i = \begin{pmatrix} \underline{\Xi}_{i,lag=1} & 0_{m\times m} & \cdots & 0_{m\times m} & 0_{m\times 1} \\ 0_{m\times m} & \underline{\Xi}_{i,lag=2} & \cdots & 0_{m\times m} & 0_{m\times 1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{m\times m} & 0_{m\times m} & \cdots & \underline{\Xi}_{i,lag=p} & 0_{m\times 1} \\ 0_{1\times m} & 0_{1\times m} & \cdots & 0_{1\times m} & \underline{\Xi}_{i,const} \end{pmatrix}.$$

Диагональные элементы $\underline{\Xi}_{i,lag=l}$ определяются по формулам

$$(\underline{\Xi}_{i,lag=l})_{jj} = \begin{cases} \left(\lambda_{tight} \Big/ l^{\lambda_{lag}}\right)^2, & j = i, \\ \left(\lambda_{tight}\lambda_{kron}\sigma_i\right)^2 \Big/ \left(l^{\lambda_{lag}}\sigma_j\right)^2, & j \neq i, \end{cases} \qquad \underline{\Xi}_{i,const} = \lambda_{tight}^2 \lambda_{const}^2 \sigma_i^2. \tag{15}$$

Как можно видеть из формулы (15), априорная дисперсия параметров зависит от нескольких задаваемых исследователем гиперпараметров. Проинтерпретируем гиперпараметры.

Параметр регуляризации λ_{tight} отражает общую «жесткость» априорного распределения. Если $\lambda_{tight} \to 0$, то априорное распределение совпадает с апостериорным, и данные не играют никакой роли при оценке параметров. В этом случаем считается, что параметры точно известны, т. е.

$$\Phi \sim \mathcal{N}\left(\underline{\Phi}, 0\right), \quad \Phi \mid Y \sim \mathcal{N}\left(\underline{\Phi}, 0\right).$$

Если $\lambda_{tight} \to \infty$, то апостериорное математическое ожидание параметров сходится к МНК-оценке. В этом случае $\underline{\Xi}^{-1} = 0$, поэтому $\overline{\Xi} = (0 + \Sigma^{-1} \otimes (X'X))^{-1} = \Sigma \otimes (X'X)^{-1}$.

Отсюда

$$\begin{aligned} \overline{\phi} &= (\Sigma \otimes (X'X)^{-1}) \cdot (\Sigma^{-1} \otimes X') \cdot y = (I_m \otimes ((X'X)^{-1}X')) \cdot vec(Y) = \\ &= vec((X'X)^{-1}X'Y \cdot I_m) = vec((X'X)^{-1}X'Y). \end{aligned} \tag{16}$$

Параметр кросс-регуляризации λ_{kron} добавляет дополнительную жесткость лагам других переменных по сравнению с лагами зависимой переменной. Условие $\lambda_{kron} < 1$ отражает предположение о том, что собственные лаги зависимой переменной помогают предсказывать значение переменной лучше, чем лаги других переменных. При этом коэффициенты при лагах других переменных оказываются распределены ближе к нулю.

При $\lambda_{kron}=1$ матрица $\underline{\Xi}$ имеет структуру кронекерова произведения и представима в виде $\underline{\Xi}=\Sigma\otimes\underline{\Omega}$, где $\underline{\Omega}$ — матрица размерности $k\times k$, соответствующая отдельному уравнению. Кронекерово умножение слева на матрицу Σ для i-го уравнения означает умножение дисперсий, указанных в матрице $\underline{\Omega}$, на коэффициент σ_i^2. Сама матрица $\underline{\Omega}$ представима в виде

$$\underline{\Omega}=\begin{pmatrix} \underline{\Omega}_{lag=1} & 0_{m\times m} & \cdots & 0_{m\times m} & 0_{m\times 1} \\ 0_{m\times m} & \underline{\Omega}_{lag=2} & \cdots & 0_{m\times m} & 0_{m\times 1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{m\times m} & 0_{m\times m} & \cdots & \underline{\Omega}_{lag=p} & 0_{m\times 1} \\ 0_{1\times m} & 0_{1\times m} & \cdots & 0_{1\times m} & \underline{\Omega}_{const} \end{pmatrix}. \qquad (17)$$

При этом матрица $\underline{\Omega}_{lag=l}$ имеет размерность $m\times m$, и ее диагональные элементы определяются по формулам

$$(\underline{\Omega}_{lag=l})_{jj}=\lambda_{tight}^2\Big/\left(l^{\lambda_{lag}}\sigma_j\right)^2, \qquad \underline{\Omega}_{const}=\lambda_{tight}^2\lambda_{const}^2. \qquad (18)$$

Параметр λ_{const} отражает относительную жесткость распределения константы, а параметр λ_{lag} отвечает за то, насколько быстро убывает априорная дисперсия с увеличением номера лага.

Алгоритм генерации случайной выборки непосредственно из апостериорного распределения использует метод Монте-Карло:

1) на s-м шаге сгенерировать очередную итерацию согласно $\phi^{[s]}\sim\mathcal{N}\left(\overline{\phi},\overline{\Xi}\right)$;

2) увеличить s на единицу и перейти к п. 1.

Если $\underline{\Xi}$ имеет структуру кронекерова произведения $\underline{\Xi}=\Sigma\otimes\underline{\Omega}$, то вместо вектора $\phi^{[s]}$ можно генерировать матрицу $\Phi^{[s]}$ численно более простым алгоритмом:

1) сгенерировать матрицу $V^{[s]}$ размера $k\times m$ из независимых стандартных нормальных величин;

2) вычислить матрицу $\Phi^{[s]}$ по формуле $\Phi^{[s]}=\overline{\Phi}+chol(\overline{\Omega})\cdot V^{[s]}\cdot chol(\Sigma)'$, где $chol(\overline{\Omega})$ и $chol(\Sigma)$ — верхне-треугольные матрицы, полученные путем разложения Холецкого матриц $\overline{\Omega}$ и $\Sigma^{[s]}$ соответственно. Можно использовать теоретически эквивалентный, но численно менее устойчивый способ, а именно сгенерировать вектор $vec\left(\Phi^{[s]}\right)\sim\mathcal{N}\left(vec(\overline{\Phi}),\Sigma\otimes\overline{\Omega}\right)$ по формуле $vec\left(\Phi^{[s]}\right)=vec\left(\overline{\Phi}\right)+chol(\Sigma\otimes\overline{\Omega})\times v$, где v — вектор независимых стандартных нормальных случайных величин.

Можно выделить несколько преимуществ априорного распределения Миннесоты. Прежде всего, оно просто задается и успешно применяется для решения различных задач. И наконец, получившееся апостериорное распределение является нормальным, а значит, можно легко получить значение любой функции параметров с помощью метода Монте-Карло. Однако существенным недостатком этого распределения является то, что оно не предполагает использования байесовской процедуры для оценки ковариационной матрицы Σ.

2.2.2. Независимое нормальное — обратное Уишарта распределение

Обобщением распределения Миннесоты является независимое нормальное — обратное Уишарта априорное распределение (independent normal inverse Wishart prior). Оно предполагает, что ковариационная матрица параметров может быть произвольной формы:

$$\phi \sim \mathcal{N}\left(\underline{\phi}, \underline{\Xi}\right), \quad \Sigma \sim \mathcal{IW}\left(\underline{S}, \underline{\nu}\right), \ \phi, \Sigma \text{ — независимы.} \tag{19}$$

Распределение Миннесоты получается из независимого нормального — обратного Уишарта распределения при $\underline{S} = (\underline{\nu} - m - 1) \cdot \Sigma$ и $\underline{\nu} \to \infty$.

Можно показать (Karlsson, 2013), что условные апостериорные распределения имеют вид

$$\phi \,|\, \Sigma, Y \sim \mathcal{N}\left(\overline{\phi}, \overline{\Xi}\right), \quad \Sigma \,|\, \phi, Y \sim \mathcal{IW}\left(\overline{S}, \overline{\nu}\right), \tag{20}$$

где гиперпараметры апостериорного распределения задаются как

$$\overline{\nu} = \underline{\nu} + T,$$
$$\overline{S} = \underline{S} + E'E, \quad \text{где } E = Y - X\Phi,$$
$$\overline{\Xi} = (\underline{\Xi}^{-1} + \Sigma^{-1} \otimes X'X)^{-1},$$
$$\overline{\phi} = \overline{\Xi} \cdot (\underline{\Xi}^{-1}\underline{\phi} + vec(X'Y\Sigma^{-1})).$$

Гиперпараметры априорного распределения могут быть выбраны точно так же, как и в априорном распределении Миннесоты, см. формулы (17) и (18). При необходимости неинформативное априорное распределение для коэффициентов при константе можно задать, обнулив соответствующее значение в матрице $\underline{\Xi}^{-1}$. Использование произвольной ковариационной матрицы приводит к тому, что исследователю оказываются известны только условные апостериорные распределения для ϕ и Σ. Это обуславливает необходимость использования алгоритма Гиббса для получения реализаций из совместного апостериорного распределения:

1) сгенерировать произвольно начальную матрицу $\Sigma^{[0]}$ (например единичную) и вектор $\phi^{[0]}$ (например нулевой);

2) на s-м шаге сгенерировать очередную итерацию согласно формулам:

$$\phi^{[s]} \sim \mathcal{N}\left(\overline{\phi}^{[s-1]}, \overline{\Xi}^{[s-1]}\right), \text{ где } \overline{\phi}^{[s-1]}, \overline{\Xi}^{[s-1]} \text{ рассчитываются через } \Sigma^{[s-1]}, \tag{21}$$

$$\Sigma^{[s]} \sim \mathcal{IW}\left(\overline{S}^{[s]}, \overline{\nu}\right), \text{ где } \overline{S}^{[s]} \text{ рассчитываются с помощью } \phi^{[s]}; \tag{22}$$

3) увеличить s на единицу и перейти к п. 2.

2.2.3. Сопряженное нормальное — обратное Уишарта априорное распределение

Недостатка априорного распределения Миннесоты, состоящего в отсутствии байесовской процедуры для оценки матрицы Σ, также можно избежать, если рассматривать сопряженное априорное распределение.

Сопряженное нормальное — обратное Уишарта априорное распределение (conjugate normal inverse Wishart prior) может быть записано как

$$\Sigma \sim \mathcal{IW}\left(\underline{S}, \underline{\nu}\right), \quad \phi \,|\, \Sigma \sim \mathcal{N}\left(\underline{\phi}, \Sigma \otimes \underline{\Omega}\right). \tag{23}$$

Распределение Миннесоты будет являться частным случаем (23), если ковариационная матрица параметров в уравнении (10) имеет кронекерову структуру $\underline{\Xi} = \Sigma \otimes \underline{\Omega}$, т. е. при λ_{kron}, равной единице.

Гиперпараметры вектора математического ожидания $\underline{\phi}$ и ковариационной матрицы $\underline{\Omega}$ условного априорного распределения могут быть заданы точно так же, как и для распределения Миннесоты в случае $\lambda_{kron} = 1$ — см. формулы (14), (17) и (18). Матрица $\underline{S}$ выбирается так, чтобы среднее Σ совпадало с фиксированной ковариационной матрицей Σ в априорном распределении Миннесоты. Так как математическое ожидание и дисперсия параметров имеют вид[8]

$$\mathbf{E}(\phi) = \underline{\phi}, \quad \mathrm{Var}(\phi) = (\underline{\nu} - m - 1)^{-1}(\underline{S} \otimes \underline{\Omega}), \tag{24}$$

то диагональные элементы $\underline{S}$ выбираются следующим образом:

$$(\underline{S})_{ii} = (\underline{\nu} - m - 1)\hat{\sigma}_i^2 . \tag{25}$$

Число степеней свободы обратного Уишарта распределения $\underline{\nu}$ выбирается так, чтобы

$$\underline{\nu} \geq \max\{m + 2, m + 2h - T\}, \tag{26}$$

что обеспечивает существование как априорной дисперсии параметров, так и апостериорной дисперсии прогнозов на горизонте h, см. (Kadiyala, Karlsson, 1997).

Можно показать, что с учетом функции правдоподобия (8) апостериорное распределение принадлежит тому же классу, см., например, (Zellner, 1996):

$$\Sigma \mid Y \sim \mathcal{IW}\left(\overline{S}, \overline{\nu}\right), \quad \Phi \mid \Sigma, Y \sim \mathcal{N}\left(\overline{\Phi}, \Sigma \otimes \overline{\Omega}\right), \tag{27}$$

где гиперпараметры апостериорного распределения задаются как

$$\overline{\nu} = \underline{\nu} + T, \ \overline{\Omega} = (\underline{\Omega}^{-1} + X'X)^{-1}, \ \overline{\Phi} = \overline{\Omega} \cdot (\underline{\Omega}^{-1}\underline{\Phi} + X'Y),$$

$$\overline{S} = \underline{S} + \hat{E}'\hat{E} + \hat{\Phi}'X'X\hat{\Phi} + \underline{\Phi}'\underline{\Omega}^{-1}\underline{\Phi} - \overline{\Phi}'\overline{\Omega}^{-1}\overline{\Phi} = \underline{S} + \hat{E}'\hat{E} + (\underline{\Phi} - \hat{\Phi})'(\underline{\Omega} + (X'X)^{-1})^{-1}(\underline{\Phi} - \hat{\Phi}),$$

$$\hat{\Phi} = (X'X)^{-1}X'Y, \ \hat{E} = Y - X\hat{\Phi}.$$

Как и в случае распределения Миннесоты, нет необходимости использовать алгоритм Гиббса, а можно генерировать случайную выборку непосредственно из апостериорного распределения. Например, можно применять такой алгоритм:

1) на s-м шаге сгенерировать очередную итерацию согласно соотношениям

$$\Sigma^{[s]} \sim \mathcal{IW}\left(\overline{S}, \overline{\nu}\right), \ \phi^{[s]} \sim \mathcal{N}\left(\overline{\phi}, \Sigma^{[s]} \otimes \overline{\Omega}\right);$$

2) увеличить s на единицу и перейти к п. 1.

На практике вместо вектора $\phi^{[s]}$ генерируют сразу матрицу $\Phi^{[s]}$ в два шага:

1) сгенерировать матрицу $V^{[s]}$ размера $k \times m$ из независимых стандартных нормальных величин;

2) вычислить матрицу $\Phi^{[s]}$ по формуле $\Phi^{[s]} = \overline{\Phi} + chol(\overline{\Omega}) \cdot V^{[s]} \cdot chol(\Sigma^{[s]})'$.

Желание задать сопряженное нормальное — обратное Уишарта априорное распределение с помощью нескольких гиперпараметров приводит к тому, что моменты априорных распределений параметров для разных уравнений оказываются зависимыми друг от друга. В частности, все коэффициенты при первом лаге зависимой переменной априорно имеют

[8] Правая часть (24) следует из того, что $\mathrm{Var}(\phi) = \mathrm{Var}(\mathbf{E}(\phi \mid \Sigma)) + \mathbf{E}(\mathrm{Var}(\phi \mid \Sigma)) =$
$= \mathrm{Var}(\underline{\phi}) + \mathbf{E}(\Sigma \otimes \underline{\Omega}) == \mathbf{E}(\Sigma \otimes \underline{\Omega}) = \mathbf{E}(\Sigma) \otimes \underline{\Omega} = (\nu - m - 1)^{-1}\underline{S} \otimes \underline{\Omega}.$

одну и ту же дисперсию λ_{tight}^2. Хотя обычно эта предпосылка не является слишком ограничительной, в реальности легко встретиться с задачами, в которых ковариационная матрица априорного распределения не должна быть симметрично сформирована для разных уравнений. Например, довольно известным в литературе является следующий пример (Kadiyala, Karlsson, 1997). Предположим, что исследователь хочет учесть в VAR наличие нейтральности денег. При построении модели эта предпосылка может быть учтена таким априорным распределением, в котором все коэффициенты при лагах денег в уравнении для выпуска имеют нулевое математическое ожидание и маленькую дисперсию. Однако это означает, что и в других уравнениях дисперсия коэффициентов в априорном распределении будет относительно низкой. Это характеристика может быть нежелательной, и, чтобы этого избежать, априорное распределение можно задать как независимое нормальное — обратное Уишарта.

2.2.4. Добавление наблюдений и модификации сопряженного нормального — обратного Уишарта априорного распределения

Существует достаточно популярный альтернативный подход для подсчета апостериорных гиперпараметров для сопряженного нормального — обратного Уишарта априорного распределения.

Обнуляются матрицы $\underline{S}$ и $\underline{\Omega}^{-1}$, при этом матрица $(\underline{\Omega}+(X'X)^{-1})^{-1}$ оказывается равной 0, а матрица $\underline{\Phi}$ исчезает из формул. Чтобы компенсировать разницу, добавляем фиктивные наблюдения (dummy observations) в матрицу X и в матрицу Y:

$$X^* = \begin{bmatrix} X^+ \\ X \end{bmatrix}, \quad Y^* = \begin{bmatrix} Y^+ \\ Y \end{bmatrix}.$$

При добавлении наблюдений матрицы скалярных произведений $X^{*\prime}X^*$ и $X^{*\prime}Y^*$ разлагаются в сумму $X^{*\prime}X^* = X^{+\prime}X^+ + X'X$, $X^{*\prime}Y^* = X^{+\prime}Y^+ + X'Y$. В частности, добавление нулевых фиктивных наблюдений никак не изменяет матрицы $X'X$, $X'Y$ и $Y'Y$. Матрицы X и Y входят в гиперпараметры апостериорного распределения только в составе матриц $X'X$, $X'Y$ и $Y'Y$, поэтому абсолютно не важно, в каком порядке и каким образом по отношению к матрицам X и Y добавлять фиктивные наблюдения. Их можно добавить в конец матриц X и Y, в начало или в середину.

Получим новые формулы для апостериорных гиперпараметров:

$$\overline{\nu} = \underline{\nu} + T, \ \overline{\Omega} = (X^{*\prime}X^*)^{-1} = (X^{+\prime}X^+ + X'X)^{-1},$$
$$\overline{\Phi} = \overline{\Omega}\cdot(X^{*\prime}Y^*) = \overline{\Omega}\cdot(X^{+\prime}Y^+ + X'Y) = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^*,$$
$$\overline{S} = \hat{E}^{*\prime}\hat{E}^*, \ \hat{E}^* = Y^* - X^*\overline{\Phi}.$$

Наблюдения добавляются так, чтобы гиперпараметры апостериорных наблюдений не изменились. Для этого необходимо, чтобы

$$X^{+\prime}X^+ = \underline{\Omega}^{-1}, \quad X^{+\prime}Y^+ = \underline{\Omega}^{-1}\underline{\Phi}, \quad (Y^+ - X^+\underline{\Phi})'(Y^+ - X^+\underline{\Phi}) = \underline{S}. \tag{28}$$

Взаимосвязь «новых» формул с матрицами, определяющими априорное и апостериорное распределения, может быть представлена в виде табл. 1.

Таблица 1. Интерпретация параметров с помощью дополнительных наблюдений

Обозначение	Интерпретация	Формула
$\underline{\Phi}$	Оценки коэффициентов регрессий Y^+ на X^+	$(X^{+\prime}X^+)^{-1}\cdot(X^{+\prime}Y^+)=\underline{\Phi}$
$\underline{S}$	Скалярные произведения остатков этих регрессий	$\hat{E}^{+\prime}\hat{E}^+$, где $\hat{E}^+=Y^+-X^+\underline{\Phi}$
$\underline{\Omega}^{-1}$	Скалярные произведения регрессоров из X^+	$X^{+\prime}X^+$
$\overline{\Phi}$	Оценки коэффициентов регрессий Y^* на X^*	$(X^{*\prime}X^*)^{-1}\cdot(X^{*\prime}Y^*)=\overline{\Phi}$
$\overline{S}$	Скалярные произведения остатков этих регрессий	$\hat{E}^{*\prime}\hat{E}^*$, где $\hat{E}^*=Y^*-X^*\overline{\Phi}$
$\overline{\Omega}^{-1}$	Скалярные произведения регрессоров из X^*	$X^{*\prime}X^*$

Чтобы реализовать структуру матрицы $\underline{\Omega}$, задаваемую уравнениями (17) и (18), можно добавить фиктивные наблюдения по схеме[9]

$$Y^{NIW}=\begin{bmatrix}\dfrac{diag(\delta_1\sigma_1,\ldots,\delta_m\sigma_m)}{\lambda_{tight}}\\ 0_{m(p-1)\times m}\\ diag(\sigma_1,\ldots,\sigma_m)\\ 0_{1\times m}\end{bmatrix},$$

$$X^{NIW}=\begin{bmatrix}\dfrac{diag(1,2^{\lambda_{lag}},\ldots,p^{\lambda_{lag}})\otimes diag(\sigma_1,\ldots,\sigma_m)}{\lambda_{tight}} & 0_{mp\times 1}\\ 0_{m\times mp} & 0_{m\times 1}\\ 0_{1\times mp} & \left(\lambda_{tight}\lambda_{const}\right)^{-1}\end{bmatrix}. \tag{29}$$

Однако ничто не мешает исследователю добавить фиктивные наблюдения по другой схеме. В работах (Doan et al., 1984) и (Sims, 1993) было предложено добавить к априорному распределению дополнительную характеристику, обусловленную возможным наличием во временных рядах единичных корней и коинтеграционных соотношений. Это позволяет исключить появление неправдоподобно большой доли внутривыборочной дисперсии, объясняемой экзогенными переменными (Carriero et al., 2015). Априорное распределение суммы коэффициентов (sum-of-coefficients prior) было предложено в работе (Doan et al., 1984). Это распределение отражает следующую идею: если переменные в VAR имеют единичный корень, то можно учесть эту информацию, задав априорное распределение, в котором сумма коэффициентов при всех лагах зависимой переменной равна единице, см. (Robertson, Tallman, 1999; Blake, Mumtaz, 2012). Другими словами, среднее значение лагированных значений какой-либо переменной является хорошим прогнозом для будущих значений этой переменной.

[9] Аналогичные формулы, приведенные в работах (Banbura et al., 2010; Berg, Henzel, 2013) для задания сопряженного нормального — обратного Уишарта априорного распределения, являются частным случаем (29) для $\lambda_{lag}=1$ и $\lambda_{const}\to\infty$.

Внедрение этого априорного распределения производится путем добавления фиктивных наблюдений по следующей схеме:

$$Y^{SC} = \frac{1}{\lambda_{sc}} diag(\delta_1\mu_1,\ldots,\delta_m\mu_m), \tag{30}$$

$$X^{SC} = \frac{1}{\lambda_{sc}}\left[1_{1\times p} \otimes diag(\delta_1\mu_1,\ldots,\delta_m\mu_m) \quad 0_{m\times 1}\right], \tag{31}$$

где $1_{1\times p}$ — вектор-строка из единиц размера p, μ_i есть i-я компонента вектора μ, который состоит из средних начальных значений всех переменных[10]: $\mu = \frac{1}{p}\sum_{t=1}^{p} y_t$.

Априорное распределение фиктивного начального наблюдения (dummy initial observation prior), предложенное в работе (Sims, 1993), отражает априорную веру в то, что переменные имеют общий стохастический тренд. Для этого вводится единственное фиктивное наблюдение, так что значения всех переменных равны соответствующему среднему начальных значений μ_i с точностью до коэффициента масштаба λ_{io}:

$$Y^{IO} = \frac{1}{\lambda_{io}}\left[\delta_1\mu_1,\ldots,\delta_m\mu_m\right], \tag{32}$$

$$X^{IO} = \frac{1}{\lambda_{io}}\left[1_{1\times p} \otimes (\delta_1\mu_1,\ldots,\delta_m\mu_m) \quad 1\right]. \tag{33}$$

Это априорное распределение предполагает, что среднее по каждой переменной есть линейная комбинация всех остальных средних.

Гиперпараметр λ_{io} отражает жесткость указанного априорного распределения. Когда λ_{io} стремится к нулю, модель принимает вид, в котором либо все переменные стационарны со средним, равным выборочному среднему начальных условий, либо нестационарны без дрейфа и коинтегрированы.

Таким образом, к исходным наблюдениям добавляются три блока фиктивных наблюдений: Y^{NIW} и X^{NIW}, Y^{SC} и X^{SC}, Y^{IO} и X^{IO}. Поскольку, помимо Y^{NIW} и X^{NIW}, добавлены еще два блока наблюдений, структура получающейся матрицы $\underline{\Omega}$ оказывается модифицированной по сравнению с формулами (17) и (18).

2.2.5. Априорные распределения Джеффриса

В этих распределениях предполагают, что априорное распределение ковариационной матрицы ошибок не содержит гиперпараметров и имеет вид

$$\Sigma \sim |\Sigma|^{-(m+1)/2}. \tag{34}$$

Различают независимое нормальное — Джеффриса априорное распределение (independent normal Jeffreys prior) и сопряженное нормальное — Джеффриса априорное распределение (conjugate normal Jeffreys prior).

[10] Некоторые авторы для расчета μ усредняют все наблюдения в выборке: $\mu = \frac{1}{T}\sum_{t=1}^{T} y_t$ (Banbura et al., 2010; Carriero et al., 2015). Однако, в соответствии с работой (Sims, Zha, 1998), для расчета среднего следует использовать только первые p наблюдений.

1. *Независимое нормальное — Джеффриса априорное распределение*:

$$\phi \sim \mathcal{N}\left(\underline{\phi}, \underline{\Xi}\right), \quad \Sigma \sim |\Sigma|^{-(m+1)/2}, \quad \phi, \Sigma \text{ — независимы.} \tag{35}$$

Это распределение получается из независимого нормального — обратного Уишарта при $\underline{S} = \underline{\nu}^{1/m} \cdot I$ и $\underline{\nu} \to 0$. Функция плотности обратного Уишарта распределения имеет вид:

$$p(\Sigma) = \frac{1}{\Gamma_m(\underline{\nu}/2)} |\underline{S}|^{\nu/2} |\Sigma|^{-(\underline{\nu}+m+1)/2} \, 2^{-\underline{\nu}m/2} etr\left(-\frac{1}{2}\underline{S}\Sigma^{-1}\right),$$

где $\Gamma_m(\cdot)$ обозначает m-мерную гамма-функцию.

Если $\underline{S} = \nu^{1/m} \cdot I$ и $\underline{\nu} \to 0$, то одновременно $|\underline{S}|^{\nu} \to 1$ и $etr\left(-\frac{1}{2}\underline{S}\Sigma^{-1}\right) \to 1$.

Поэтому $p(\Sigma) \to const \cdot |\Sigma|^{-(m+1)/2}$.

В данном случае распределение Джеффриса является несобственным, т. е. интеграл под всей функцией плотности невозможно отнормировать так, чтобы он равнялся единице. Тем не менее, апостериорное распределение будет собственным при достаточном ($T > m-1$) числе наблюдений (Alvarez et al., 2014).

Для получения выборки из апостериорного распределения можно использовать схему Гиббса. Необходимые формулы для гиперпараметров апостериорного распределения получаются из общего случая независимого нормального — обратного Уишарта подстановкой $\underline{S} = 0$, $\underline{\nu} = 0$.

Распределение Миннесоты и независимое нормальное — Джеффриса распределение являются противоположными крайностями независимого нормального — обратного Уишарта распределения. В распределении Миннесоты матрица Σ предполагается известной, а в нормальном — Джеффриса матрица Σ имеет «размытое» неинформативное распределение.

2. *Сопряженное нормальное — Джеффриса априорное распределение*:

$$\phi | \Sigma \sim \mathcal{N}\left(\underline{\phi}, \Sigma \otimes \underline{\Omega}\right), \quad \Sigma \sim |\Sigma|^{-(m+1)/2}. \tag{36}$$

Это распределение является частным случаем сопряженного нормального — обратного Уишарта распределения при $\underline{S} = \underline{\nu}^{1/m} \cdot I$ и $\underline{\nu} \to 0$. Формулы для гиперпараметров апостериорного распределения получаются подстановкой $\underline{\nu} = 0$, $\underline{S} = 0$ в общие формулы для сопряженного нормального — обратного Уишарта распределения.

Частным случаем двух указанных распределений Джеффриса является неинформативное — Джеффриса распределение (diffuse Jeffreys prior):

$$\phi \sim 1, \quad \Sigma \sim |\Sigma|^{-(m+1)/2}, \quad \phi, \Sigma \text{ — независимы.} \tag{37}$$

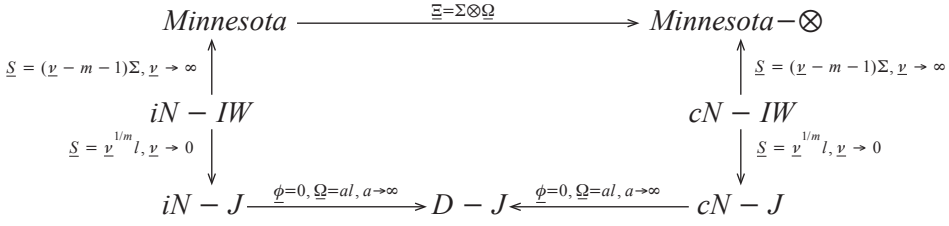
При задании этого априорного распределения не нужно указывать ни одного гиперпараметра. Оно получается из независимого нормального — Джеффриса при $\underline{\phi} = 0$, $\underline{\Xi} = a \cdot I$ и $a \to \infty$. Для получения выборки из апостериорного распределения можно использовать прямое симулирование по схеме Монте-Карло без алгоритма Гиббса. Распределение получается из общего случая независимого нормального — обратного Уишарта распределения подстановкой $\underline{S} = 0$, $\underline{\nu} = 0$, $\underline{\Xi}^{-1} = 0$, $\underline{\phi} = 0$. При этом формулы существенно упрощаются, в частности, исчезает необходимость обращать матрицу размера $km \times km$.

Также неинформативное — Джеффриса распределение получается из сопряженного нормального — Джеффриса при $\underline{\phi} = 0$ и $\underline{\Omega} = a \cdot I$ и $a \to \infty$. Формулы для гиперпараметров апостериорного распределения получаются подстановкой $\underline{\phi} = 0$, $\underline{\Omega}^{-1} = 0$, $\underline{\nu} = 0$, $\underline{S} = 0$.

Достоинство и недостаток априорных распределений Джеффриса состоит в том, что они зависят от малого числа гиперпараметров. С одной стороны, исследователю надо меньше думать о выборе гиперпараметров, с другой стороны, малое количество гиперпараметров означает негибкость данных априорных распределений.

2.3. Взаимосвязь априорных распределений

Взаимосвязи всех упомянутых априорных распределений представлены на схеме:



Minnesota — распределение Миннесоты,
Minnesota $-\otimes$ — распределение Миннесоты при условии $\underline{\Xi} = \Sigma \otimes \underline{\Omega}$,
$iN - IW$ — независимое нормальное — обратное Уишарта распределение,
$cN - IW$ — сопряженное нормальное — обратное Уишарта распределение,
$iN - J$ — независимое нормальное — Джеффриса распределение,
$cN - J$ — сопряженное нормальное — Джеффриса распределение,
$D - J$ — неинформативное — Джеффриса распределение.

Стрелки идут от более общих распределений к частным случаям, над стрелками указаны соответствующие ограничения.

2.4. Эндогеннное определение гиперпараметров: оценка BVAR большой размерности

2.4.1. Регуляризация с помощью неограниченной VAR

В некоторых случаях гиперпараметры априорного распределения определяются эндогенно. В частности, это происходит при оценке модели большой размерности. Для решения ряда прикладных макроэкономических задач, связанных с прогнозированием и структурным анализом, исследователям требуется работать с выборками большой размерности. И хотя BVAR широко использовались в моделях малой размерности практически с момента их появления в начале 1980-х годов, их применение в моделях большой размерности до недавнего времени было весьма ограничено. Причина заключалась в существовании консенсуса о том, что байесовская регуляризация сама по себе недостаточна для решения проблемы избыточной параметризации в моделях большой размерности и требует введения дополнительных небайесовских ограничений.

Ключевую роль в развитии подхода сыграли работы (De Mol et al., 2008) и (Banbura et al., 2010), в которых было показано, что BVAR вполне могут быть применены к выборкам большой размерности без введения дополнительных ограничений. Однако использование большого числа временных рядов требует уменьшения параметра λ_{tight} с увеличением размерности выборки, что означает введение более жесткого априорного распределения. На данный момент в литературе используется два подхода к определению оптимальной величины λ_{tight}.

Первый алгоритм, предложенный в работе (Banbura et al., 2010), основан на идее о том, что регуляризация должна быть настолько жесткой, чтобы исключить возможность избыточной параметризации модели. При этом предполагается, что трехмерная VAR — достаточно простая модель, не содержащая слишком большого количества параметров, и поэтому она не требует дополнительной регуляризации. Это означает, что гиперпараметр λ_{tight} должен быть выбран таким образом, чтобы модель демонстрировала такую же внутривыборочную подгонку, как и VAR с тремя переменными. Другими словами, каждая модель регуляризуется до размера простой неограниченной VAR.

Приведем детальное описание процедуры. Обозначим фактическое значение переменной var в момент $T+h$ как $y_{var,T+h}$, а прогноз переменной var, осуществленный в момент T на горизонт h в модели с m переменными и параметром жесткости λ_{tight}, как $y^{\lambda,m}_{var,T+h|T}$ (для краткости в формулах используется λ вместо λ_{tight}). Схема выбора λ_{tight} состоит из следующих этапов.

1. С помощью BVAR строятся внутривыборочные одноперiодные прогнозы на обучающей выборке и рассчитывается среднеквадратичная ошибка прогноза для M переменных, объединенных в набор $\tilde{M}$ и представляющих особый интерес[11]:

$$MSFE^{\lambda,m}_{var,1} = \frac{1}{T-p}\sum_{t=p}^{T-1}\left(y^{\lambda,m}_{var,t+1|t} - y_{var,t+1}\right)^2. \tag{38}$$

2. Аналогично рассчитываются одноперiодные прогнозы в соответствии с моделью случайного блуждания с дрейфом[12] для тех же самых переменных $\left(MSFE^{0}_{var,1}\right)$, и, кроме того, рассчитывается новый индикатор $FIT^{\lambda,m}$:

$$FIT^{\lambda,m} = \frac{1}{M}\sum_{var\in\tilde{M}}\frac{MSFE^{\lambda,m}_{var,1}}{MSFE^{0}_{var,1}}. \tag{39}$$

3. Оценивается трехмерная VAR для тех же M переменных, представляющих особый интерес при прогнозировании[13] и рассчитываются MSFE и индикатор $FIT^{\infty,M}$:

---

[11] В базовый набор переменных $\tilde{M}$, представляющих особый интерес, в статье (Banbura et al., 2010) включается индекс промышленного производства, индекс потребительских цен и межбанковская процентная ставка, т. е. $M=3$.

[12] MSFE для моделей BVAR и VAR нормализуются на MSFE, полученные с помощью модели RW, чтобы учесть тот факт, что различные временные ряды имеют разные единицы измерения. Для указания на модель RW используется надстрочный индекс 0, потому что RW может рассматриваться как частный случай BVAR для $\lambda_{tight}=0$ и $\delta_i=1,\ i=1,\ldots,k$.

[13] Для указания на неограниченную VAR используется надстрочный индекс ∞, т. к. неограниченная VAR является частным случаем BVAR при $\lambda_{tight}\to\infty$. В этом случае апостериорное распределение совпадает с функцией правдоподобия.

$$FIT^{\infty,M} = \frac{1}{M}\sum_{var\in\tilde{M}}\frac{MSFE^{\infty,M}_{var,1}}{MSFE^{0}_{var,1}}. \tag{40}$$

4. Оптимальным λ_{tight} считается значение, минимизирующее абсолютную разность между $FIT^{\lambda,m}$ и $FIT^{\infty,M}$:

$$\lambda^{*}_{m} = \underset{\lambda}{\operatorname{argmin}}\left|FIT^{\lambda,m} - FIT^{\infty,M}\right|. \tag{41}$$

После того как выбрано оптимальное λ_{tight} для каждой модели, происходит построение вневыборочных прогнозов на оценивающей выборке.

2.4.2. Регуляризация с помощью одноперiодного прогноза

Второй алгоритм предложен в работе (Doan et al., 1984) и представляет собой выбор такого параметра λ_{tight}, который максимизирует точность одноперiодного вневыборочного прогноза на обучающей выборке. Этот выбор сводится к максимизации функции плотности (маржинальной функции правдоподобия, marginal likelihood):

$$\lambda^{*} = \underset{\lambda}{\operatorname{argmax}}\ \ln p(Y). \tag{42}$$

При этом функция плотности может быть получена путем интегрирования коэффициентов модели:

$$p(Y) = \int p(Y\,|\,\phi)\,p(\phi)\,d\phi. \tag{43}$$

Если априорное распределение является сопряженным нормальным — обратным Уишарта, то плотность $p(Y)$ может быть получена аналитически (Zellner, 1996; Bauwens et al., 2000; Carriero et al., 2015):

$$\begin{aligned} p(Y) = \pi^{-\frac{Tm}{2}}\left|(I + X\underline{\Omega}X')^{-1}\right|^{\frac{N}{2}}\,|\,\underline{S}\,|^{\frac{\underline{\nu}}{2}}\cdot\frac{\Gamma_N\left((\underline{\nu}+T)/2\right)}{\Gamma_N\left(\underline{\nu}/2\right)}\cdot \\ \cdot\left|\underline{S} + (Y - X\underline{\Phi})'(I + X\underline{\Omega}X')^{-1}(Y - X\underline{\Phi})\right|^{-\frac{\underline{\nu}+T}{2}}. \end{aligned} \tag{44}$$

Выбор числа лагов происходит аналогично путем максимизации по p функции плотности (44):

$$p^{*} = \underset{p}{\operatorname{argmax}}\ \ln p(Y). \tag{45}$$

2.4.3. Особенности кодирования

При практической реализации алгоритма Гиббса или Монте-Карло часто приходится обращать положительно определенные симметричные матрицы. Некоторые из этих матриц могут иметь определитель, близкий к нулю, что мешает практическому обращению этих матриц на компьютере. Если необходимо обратить матрицу A вида $A = X'X$, то можно сразу получить обратную к A без вычисления самой A:

1) получить сингулярное разложение матрицы X, $X = U\Sigma V'$;

2) получить обратную к $A = X'X$ по формуле $A^{-1} = V\Sigma^{-2}V'$.

Если матрица X неизвестна, то обратить матрицу A можно так:

1) получить разложение Холецкого матрицы A, $A = U'U$;

2) обратить верхне-треугольную матрицу U с помощью специального более устойчивого алгоритма;

3) получить A^{-1} по формуле $A^{-1} = U^{-1}U^{-1'}$.

Однако данный способ сопряжен с численными трудностями, если обращаемая матрица плохо обусловлена. В таком случае оправдано использовать псевдообратную матрицу Мура–Пенроуза.

3. Прогнозирование с помощью BVAR

Основная цель оценки неструктурных байесовских VAR — это построение прогнозов. BVAR позволяют строить точечные прогнозы, интервальные прогнозы и прогнозы плотности. Если задача состоит не только в построении прогноза по определенной модели, но и в оценке качества этого прогноза, то модель оценивается на некоторой исторической выборке, и прогноз строится на такие моменты времени, по которым фактические данные уже есть. Оценка модели может происходить как по сдвигающейся выборке (rolling window scheme), так и по растущей выборке (recursive regression). В первом случае оценка происходит по фиксированному числу наблюдений, но на каждом шаге начало выборки и ее окончание сдвигаются на один шаг вперед. Прогнозы строятся на каждом шаге на выбранный горизонт. Так происходит до тех пор, пока не будут исчерпаны все наблюдения, по которым можно сопоставить прогнозы и фактические данные. В случае растущей выборки фиксируется ее начало, и на каждом шаге ее длина увеличивается на одно наблюдение.

Ключевой концепцией при построении прогноза с помощью байесовской модели является функция апостерионой прогнозной плотности (posterior predictive density), которую вслед за работой (Karlsson, 2013) будем обозначать $p(y_{T+1:T+H} \mid Y_T)$. Эта запись означает, что прогноз строится на все моменты времени, начиная с момента $T+1$ и заканчивая моментом $T+H$ при известных наблюдениях до момента T. Здесь матрица $y_{T+1:T+H} = (y_{T+1} \ldots, y_{T+H})'$ — это все будущие наблюдения, а матрица $Y_T = (y_1 \ldots, y_T)'$ — это все наблюдения, по которым оценивалась модель. Функция апостериорной прогнозной плотности может быть представлена как

$$p(y_{T+1:T+H} \mid Y_T) = \int p(y_{T+1:T+H} \mid Y_T, \phi) p(\phi \mid Y_T) d\phi, \tag{46}$$

где $p(y_{T+1:T+H} \mid Y_T, \phi)$ — функция плотности будущих наблюдений при условии фиксированных параметров ϕ и данных вплоть до периода T, а $p(\phi \mid Y_T)$ — апостериорная плотность параметров.

Как правило, функция апостериорной прогнозной плотности не выражается аналитически для прогнозирования дальше, чем на один период вперед. Однако ее расчет возможен по формуле (46) с помощью численных методов. Для этого необходимо для каждой реализации вектора параметров из апостериорного распределения $p(\phi \mid Y_T)$ вычислить прогнозные

значения переменных $\tilde{y}_{T+h}$ на горизонте $h=1,\ldots,H$ с помощью функции условной плотности $p(y_{T+1:T+H} \mid Y_T, \phi)$. При этом при построении прогноза на горизонт h прогнозы, построенные на горизонт $\tilde{h} < h$, становятся как будто известными значениями. Повторение описанной процедуры большое число раз позволяет получить выборку из апостериорного прогнозного распределения для каждого h.

Другими словами, для BVAR построение прогнозной плотности происходит по следующей схеме (Karlsson, 2013, pp. 800, 811):

1) сгенерировать параметры апостериорного распределения;

2) сгенерировать $\varepsilon_{T+1}^{[s]},\ldots,\varepsilon_{T+H}^{[s]}$ из распределения $\varepsilon_t \sim \mathcal{N}\left(0, \Sigma^{[s]}\right)$ (в случае априорного распределения Миннесоты $\Sigma^{[s]} = \Sigma$) и вычислить рекурсивно:

$$\tilde{y}_{T+h}^{[s]} = \Phi_{const}^{[s]} + \sum_{i=1}^{h-1} \Phi_i^{[s]} \tilde{y}_{T+h-i}^{[s]} + \sum_{i=h}^{p} \Phi_i^{[s]} y_{T+h-i}^{[s]} + \varepsilon_{T+h}^{[s]}; \tag{47}$$

3) увеличить s на единицу и перейти к п. 1.

Далее следует «забыть» о том, что разные прогнозы были сделаны для разных значений параметров. Для априорного распределения Миннесоты и сопряженного нормального — обратного Уишарта распределения следует рассматривать $\{\tilde{y}_{T+1}^{[s]},\ldots,\tilde{y}_{T+H}^{[s]}\}_{s=1}^{S}$ как выборку независимых реализаций из совместного прогнозного распределения. В случае независимого нормального — обратного Уишарта распределения генерация $\varepsilon^{[s]}$ происходит только при s больше некоторого (заранее заданного) B, т. к. первые B реализаций апостериорных параметров используются для сходимости цепи (т. н. период прожига, burn-in). Соответственно, прогнозные реализации также рассматриваются только для $s > B$, т. е. $\{\tilde{y}_{T+1}^{[s]},\ldots,\tilde{y}_{T+H}^{[s]}\}_{s=B+1}^{S}$.

Прогноз на один шаг вперед представляет собой линейную функцию параметров, поэтому апостериорная прогнозная плотность для $h=1$ может быть задана аналитически. Она принимает вид матричного t-распределения (MT), параметры которого зависят от конкретного используемого априорного распределения. Например, для случая сопряженного нормального распределения (Carriero et al., 2015, p. 54) прогноз имеет следующее многомерное t-распределение:

$$y_{T+1}' \mid x_{T+1}' \sim MT\left(x_{T+1}'\bar{\Phi}, \left(x_{T+1}'\bar{\Omega}x_{T+1}\right)^{-1}, \bar{S}, \bar{\nu}\right). \tag{48}$$

Точечные прогнозы рассчитываются после построения случайной выборки из апостериорного прогнозного распределения. Выбор конкретного типа точечного прогноза, например, мода или медиана прогнозной плотности, может быть сделан на основании функции потерь. С формальной точки зрения, существует функция потерь $\mathbb{L}(b, y_{T+1:T+H})$, определяющая, какая матрица значений b будет выбрана в качестве точечного прогноза. Матрица значений b выбирается таким образом, чтобы минимизировать ожидаемые потери при условии доступных данных Y_T (Karlsson, 2013, p. 795):

$$\mathbf{E}[\mathbb{L}(b, y_{T+1:T+H}) \mid Y_T] = \int \mathbb{L}(b, y_{T+1:T+H}) p(y_{T+1:T+H} \mid Y_T) dy_{T+1:T+H}. \tag{49}$$

При заданных функции потерь и прогнозной плотности решение задачи минимизации есть функция только доступных данных, $b(Y_T)$. Для частных случаев функции потерь решение принимает простую форму. Например, для квадратичной функции потерь

$\mathbb{L}(b, y_{T+1:T+H}) = (b - y_{T+1:T+H})'(b - y_{T+1:T+H})$ решение принимает вид условного математического ожидания $b(Y_T) = \mathbf{E}(y_{T+1:T+H} \mid Y_T)$, а для функции потерь в виде абсолютного значения решение представляет собой медиану прогнозного распределения.

4. Заключение

Представленный обзор посвящен механизму оценки байесовских неструктурных векторных авторегрессий и их применению для построения прогнозов. Авторы подробно останавливаются на наиболее часто используемых в макроэкономических прикладных работах априорных распределениях и строят для них «карту», содержащую подробное представление их взаимосвязей.

Отдельный раздел работы посвящен заданию сопряженного нормального — обратного Уишарта распределения с помощью искусственных наблюдений — способу, часто применяемому на практике, но не затронутому во многих из существующих обзоров. В разделе, посвященном прогнозированию с помощью BVAR, рассмотрены как точечные прогнозы, так и прогнозы плотности.

Благодарности. Исследование осуществлено в рамках программы фундаментальных исследований НИУ ВШЭ в 2016 году.

Список литературы

Айвазян С. А. (2008). Байесовский подход в эконометрическом анализе. *Прикладная эконометрика*, 9 (1), 93–130.

Малаховская О. А., Пекарский С. Э. (2012). Исследования причинно-следственных взаимосвязей в макроэкономике: Нобелевская премия по экономике 2011 года. *Экономический журнал Высшей школы экономики*, 16 (1), 3–30.

Слуцкин Л. Н. (2010). Байесовский анализ, когда оцениваемый параметр является случайным нормальным процессом. *Прикладная эконометрика*, 20 (4), 119–131.

Alvarez I., Niemi J., Simpson M. (2014). Bayesian inference for a covariance matrix. *ArXiv preprint*. https://arxiv.org/abs/1408.4050.

Banbura M., Giannone D., Reichlin L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25 (1), 71–92.

Bauwens L., Lubrano M., Richard J.-F. (2000). *Bayesian inference in dynamic econometric models*. Oxford University Press.

Berg T., Henzel S. (2013). Point and density forecasts for the euro area using many predictors: Are large BVARs really superior? *Ifo Working Paper* 155. https://www.cesifo-group.de/portal/page/portal/lang-en/DocBase_Content/WP/WP-Ifo_Working_Papers/wp-ifo-2013/IfoWorkingPaper-155.pdf.

Blake A., Mumtaz H. (2012). *Applied Bayesian econometrics for central bankers*. Centre for Central Banking Studies, Bank of England.

Canova F. (2007). *Methods for applied macroeconomic research*. Princeton University Press.

Carriero A., Clark T., Marcellino M. (2015). Bayesian VARs: Specification choices and forecast accuracy. *Journal of Applied Econometrics*, 30 (1), 46–73.

De Mol C., Giannone D., Reichlin L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146 (2), 318–328.

Del Negro M., Schorfheide F. (2011). Bayesian macroeconometrics. *The Oxford Handbook of Bayesian Econometrics*, 293–389.

Doan T., Litterman R., Sims C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric reviews*, 3 (1), 1–100.

Favero C. (2001). *Applied macroeconometrics*. Oxford University Press.

Kadiyala K., Karlsson S. (1997). Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics*, 12 (2), 99–132.

Karlsson S. (2013). Forecasting with Bayesian vector autoregressions. In: *Handbook of Economic Forecasting*, vol. 2, Part B, 791–897.

Koop G., Korobilis D. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends (R) in Econometrics*, 3 (4), 267–358.

Litterman R. (1979). Techniques of forecasting using vector autoregressions. *Working Paper* 115, Federal Reserve Bank of Minneapolis.

Litterman R. (1986). Forecasting with Bayesian vector autoregressions — five years of experience. *Journal of Business and Economic Statistics*, 4 (1), 25–38.

Robertson J., Tallman E. (1999). Vector autoregressions: Forecasting and reality. *Economic Review*, *Federal Reserve Bank of Atlanta*, 84 (1), 4.

Sims C. (1980). Macroeconomics and reality. *Econometrica*, 48 (1), 1–48.

Sims C. (1993). A nine-variable probabilistic macroeconomic forecasting model. *Business Cycles, Indicators and Forecasting*. University of Chicago Press, 179–212.

Sims C, Zha T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 39 (4), 949–968.

Zellner A. (1996). *An introduction to Bayesian inference in econometrics.* Wiley Classics Library, Wiley.

Поступила в редакцию 02.06.2016;
принята в печать 15.08.2016.

Приложение 1. Доступные реализации кода

1. Страница CReMFi — Центра исследований в макроэкономике и финансах Университета Лондона. Код в Матлаб. http://cremfi.econ.qmul.ac.uk/efp/info. php.

Фиктивные наблюдения вводятся как и для сопряженного нормального — обратного Уишарта распределения, но выполняется семплирование Гиббса; при этом $\bar{\Phi}$ одно и то же, а Σ пересчитывается на каждом шаге в зависимости от предыдущего Φ. При плохо обусловленной матрице $X'X$ используется псевдообратная.

2. Страница Банка Англии с описанием работы (Blake, Mumtaz, 2012). Код в Матлаб.

http://www.bankofengland.co.uk/education/Pages/ccbs/technical_handbooks/techbook4.aspx.

Независимое нормальное — обратное Уишарта распределение названо Миннесотой. Код для сопряженного нормального — обратного Уишарта распределения построен по тому же принципу, что и представленный на странице CReMFi, но добавлено две одинаковых строки для фиктивных наблюдений при определении коэффициента при константе.

3. Страница с кодами к работам D. Korobilis. Код в Матлаб. https://sites.google.com/site/dimitriskorobilis/matlab.

Базовый код негибкий, без правки кода нет возможности прогнозировать больше чем на один шаг, базовые сопряженное и независимое нормальные — обратные Уишарта априорные распределения не содержат гиперпараметров и задаются через фиксированные матрицы.

4. Страница с кодами к работам T. Zha. Код в Матлаб. http://www.tzha.net/code.

На указанной странице можно найти коды для оценивания как BVAR в сокращенной форме, так и структурной BVAR.

5. Страница C. Sims с кодами для оценки VAR. Коды в R и Matlab. http://sims.princeton.edu/yftp/VARtools.

Недостаточно подробное описание. Для модификации необходимо прочитать весь код.

6. Страница пакета BMR. Код в R. http://bayes.squarespace.com/bmr/.

Симуляции реализованы в C++, также оценивает DSGE. Реализует TVP-BVAR. Отличная документация.

7. Страница пакета MSBVAR. Код в R. https://cran.r-project.org/web/packages/MSBVAR/.

Симуляции реализованы на Фортране и C++. Оценивает также марковские BVAR с переключением режимов.

8. Страница пакета bvarr. Код в R. https://github.com/bdemeshev/bvarr.

Основной целью написанного авторами пакета bvarr является максимально гибкая реализация сопряженного нормального — обратного Уишарта распределения. Данное распределение не реализовано в пакете BMR, а в пакете MSBVAR реализовано недостаточно гибко. Во-первых, используется сингулярное разложение матрицы X вместо прямого обращения матрицы $X'X$, как в MSBVAR. Это позволяет оценивать BVAR большей размерности. Во-вторых, пользователь bvarr может использовать любые блоки дополнительных наблюдений Y^{NIW} и X^{NIW}, Y^{SC} и X^{SC}, Y^{IO} и X^{IO}, в то время как MSBVAR обязывает пользователя использовать все три блока. В-третьих, пакет bvarr содержит функции для подбора гиперпараметров с помощью максимизации маржинального правдоподобия. Кроме того, пакет bvarr содержит перевод матлабовского кода для Миннесоты и независимого нормального — обратного Уишарта из источника 3.

9. Страница пакета bvarsv. Код в R. https://cran.r-project.org/web/packages/bvarsv/.

Симуляции реализованы на C++. Оценивает TVP-BVAR со стохастической волатильностью.

10. Eviews: встроенная функция.

Код игнорирует тот факт, что коэффициенты были оценены байесовскими методами, прогнозы делаются точно так же, как в обычной модели. Коэффициент λ_{kron} равен по умолчанию 0.99 и не может быть изменен. Большую свободу представляет прямое задание ковариационной матрицы. Математические ожидания коэффициентов при первых лагах могут быть заданы только одинаковыми для всех переменных.

11. Dynare: встроенная функция.

Функция позиционируется как BVAR à la Sims. Оценивается частный случай сопряженного нормального — обратного Уишарта распределения, близкий либо к неинформативному — Джеффриса распределению, либо к несобственному абсолютно плоскому распределению. Без модификации кода можно получить только прогнозы, но не выборку из апостериорного распределения параметров.

Приложение 2. Таблица обозначений

Обозначение	Размерность	Описание	Формула
p	Скаляр	Число лагов	
m	Скаляр	Число эндогенных переменных	
k	Скаляр	Число параметров в одном уравнении	$k = mp + 1$
T	Скаляр	Число наблюдений	
y_t	$m \times 1$	Вектор эндогенных переменных	$y_t = \Phi' x_t + \varepsilon_t$
x_t	$k \times 1$	Вектор всех регрессоров	$x_t = [y'_{t-1} \dots y'_{t-p}\ 1]'$
ε_t	$m \times 1$	Вектор случайных ошибок	$y_t = \Phi' x_t + \varepsilon_t$
Y	$T \times m$	Матрица эндогенных переменных	$Y = [y_1, y_2, \dots, y_T]'$
X	$T \times k$	Матрица регрессоров	$X = [x_1, x_2, \dots, x_T]'$
E	$T \times m$	Матрица ошибок	$E = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T]'$
y , ε	$mT \times 1$	Векторизация Y и E	$y = vec(Y)$, $\varepsilon = vec(E)$
Φ_l	$m \times m$	Коэффициенты VAR	$y_t = \Phi_1 y_{t-1} + \dots + \Phi_{const} + \varepsilon_t$
Φ_{const}	$m \times d$	Вектор свободных членов	$y_t = \Phi_1 y_{t-1} + \dots + \Phi_{const} + \varepsilon_t$
Φ	$k \times m$	«Упаковка» матриц $\Phi_1, \dots$	$\Phi = [\Phi_1 \dots \Phi_p\ \Phi_{const}]'$
$\hat{\Phi}$	$k \times m$	МНК-оценка матрицы Φ	$\hat{\Phi} = (X'X)^{-1} X'Y$
$\hat{E}$	$T \times m$	Остатки МНК-регрессии	$\hat{E} = Y - X\hat{\Phi}$
$\underline{\Phi}$, $\overline{\Phi}$	$k \times m$	Априорное и апостериорное математические ожидания Φ	
ϕ , $\underline{\phi}$, $\overline{\phi}$	$km \times 1$	Векторизация матриц Φ , $\underline{\Phi}$ и $\overline{\Phi}$	$\phi = vec(\Phi)$, $\underline{\phi} = vec(\underline{\Phi})$, $\overline{\phi} = vec(\overline{\Phi})$
$\underline{\Xi}$, $\overline{\Xi}$	$km \times km$	Априорная и апостериорная ковариационные матрицы Φ	$\overline{\Xi} = (\underline{\Xi}^{-1} + \Sigma^{-1} \otimes X'X)^{-1}$
$\underline{\nu}$, $\overline{\nu}$	Скаляр	Априорное и апостериорное число степеней свободы	$\overline{\nu} = T + \underline{\nu}$
$\underline{\Omega}$, $\overline{\Omega}$	$k \times k$	Априорные и апостериорные масштабирующие коэффициенты ковариационной матрицы Φ	$\underline{\Xi} = \Sigma \otimes \underline{\Omega}$, $\overline{\Omega} = (\underline{\Omega}^{-1} + X'X)^{-1}$, $\overline{\Xi} = \Sigma \otimes \overline{\Omega}$
Σ	$m \times m$	Ковариационная матрица ошибок	

Demeshev B., Malakhovskaya O. BVAR mapping. *Applied Econometrics*, 2016, v. 43, pp. 118–141.

Boris Demeshev
National Research University Higher School of Economics, Moscow, Russian Federation;
boris.demeshev@gmail.com

Oxana Malakhovskaya
National Research University Higher School of Economics, Moscow, Russian Federation;
omalakhovskaya@hse.ru

BVAR mapping

This paper reviews estimation and forecasting with Bayesian vector autoregressions (BVARs). In the first part of the paper, we propose a clear classification of the most frequently used prior distributions and we show how the parameters of posterior distributions can be computed for the priors we consider in the paper. A separate section describes the endogenous choice of prior hyperparameters that is currently a key step to estimate a BVAR in a data-rich environment. The second part of this paper is devoted to forecasting with BVARs. We review both point and density forecasting.

Keywords: BVAR; prior distributions; point forecasting; density forecasting.

JEL classification: C11; C32; C53.

References

Aivazian S. (2008). Bayesian methods in econometrics. *Applied Econometrics*, 9 (1), 93–130 (in Russian).

Malakhovskaya O., Pekarsky S. (2012). Causal relationship in macroeconomics: Nobel prize in economics of 2011. *HSE Economic Journal*, 16 (1), 3–30 (in Russian).

Slutskin L. (2010). Bayesian analysis in the case of an estimated parameter following a stochastic process. *Applied Econometrics*, 20 (4), 119–131 (in Russian).

Alvarez I., Niemi J., Simpson M. (2014). Bayesian inference for a covariance matrix. *ArXiv preprint*. https://arxiv.org/abs/1408.4050.

Banbura M., Giannone D., Reichlin L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25 (1), 71–92.

Bauwens L., Lubrano M., Richard J.-F. (2000). *Bayesian inference in dynamic econometric models*. Oxford University Press.

Berg T., Henzel S. (2013). Point and density forecasts for the euro area using many predictors: Are large BVARs really superior? *Ifo Working Paper* 155. https://www.cesifo-group.de/portal/page/portal/lang-en/DocBase_Content/WP/WP-Ifo_Working_Papers/wp-ifo-2013/IfoWorkingPaper-155.pdf.

Blake A., Mumtaz H. (2012). *Applied Bayesian econometrics for central bankers*. Centre for Central Banking Studies, Bank of England.

Canova F. (2007). *Methods for applied macroeconomic research*, Princeton University Press.

Carriero A., Clark T., Marcellino M. (2015). Bayesian VARs: Specification choices and forecast accuracy. *Journal of Applied Econometrics*, 30 (1), 46–73.

De Mol C., Giannone D., Reichlin L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146 (2), 318–328.

Del Negro M., Schorfheide F. (2011). Bayesian macroeconometrics. *The Oxford Handbook of Bayesian Econometrics*, 293–389.

Doan T., Litterman R., Sims C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric reviews*, 3 (1), 1–100.

Favero C. (2001). *Applied macroeconometrics*. Oxford University Press.

Kadiyala K., Karlsson S. (1997). Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics*, 12 (2), 99–132.

Karlsson S. (2013). Forecasting with Bayesian vector autoregressions. In: *Handbook of Economic Forecasting*, vol. 2, Part B, 791–897.

Koop G., Korobilis D. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends (R) in Econometrics*, 3 (4), 267–358.

Litterman R. (1979). Techniques of forecasting using vector autoregressions. *Working Paper* 115, Federal Reserve Bank of Minneapolis.

Litterman R. (1986). Forecasting with Bayesian vector autoregressions — five years of experience. *Journal of Business and Economic Statistics*, 4 (1), 25–38.

Robertson J., Tallman E. (1999). Vector autoregressions: Forecasting and reality. *Economic Review*, *Federal Reserve Bank of Atlanta*, 84 (1), 4.

Sims C. (1980). Macroeconomics and reality. *Econometrica*, 48 (1), 1–48.

Sims C. (1993). A nine-variable probabilistic macroeconomic forecasting model. *Business Cycles, Indicators and Forecasting*. University of Chicago Press, 179–212.

Sims C, Zha T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 39 (4), 949–968.

Zellner A. (1996). *An introduction to Bayesian inference in econometrics.* Wiley Classics Librar, Wiley.

Received 02.06.2016; accepted 15.08.2016.