

1 Введение

Значение точных прогнозов для проведения макроэкономической политики трудно переоценить. Существование лагов политики приводит к тому, что решения, принятые сегодня, повлияют на экономику только через некоторое время, поэтому фискальным и монетарным властям при принятии решений приходится ориентироваться не на текущие, а 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 её 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,¹ и предложил использовать более простую по построению динамическую модель, основанную на разложении Вольда и не требующую введения никаких ограничений на взаимную динамику переменных — VAR.

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 количество лагов, что, в свою очередь, может привести к высоким ошибкам прогноза. Проблема усугубляется тем, что в реальности при проведении политики центральные банки развитых стран ориентируются на большое количество показателей, и VAR малой размерности не может отразить всей информации, доступной центральным банкам, что означает, что использование моделей высокой размерности потенциально может улучшить качество прогноза. При этом увеличение числа переменных в VAR приводит к тому, что количество оцениваемых параметров, растет нелинейно. Это усугубляет проблему неэффективности оценивания и высоких ошибок прогноза. Одним из решений этой проблемы стало использование априорной информации относительно распределения параметров и ковариационной матрицы ошибок, т. е. переход от обычных² VAR к Байесовским (Bayesian VAR, BVAR).

Исследователи выделяют два преимущества BVAR по отношению к обычным. Во-первых, этот класс моделей предлагает решение проблемы избыточной параметризации и благодаря этому позволяет включать в модель большее количество переменных. При этом априорные веры позволяют снизить неопределенность в распределении параметров модели и улучшить её прогнозные способности. Во-вторых,

¹Под традиционными моделями имеются в виду 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) и Малаховская и Пекарский (2012))

²В англоязычной литературе обычные VAR (без наложения априорных распределений) называются частотными — frequentist.

распространенные в настоящее время априорные распределения отражают современные представления о долгосрочной динамике переменных, не проявляющиеся в коротких выборках, обычно используемых для анализа. Это, в свою очередь, также улучшает точность полученных прогнозов. Кроме того, современные компьютеры осуществляют симуляции настолько быстро, что исследователи более не ограничены необходимостью использования только сопряженных распределений (т.е. распределений, при котором априорное распределение, функция правдоподобия и апостериорное распределение принадлежат одному классу), позволяющих получить явное аналитическое решение, что, безусловно, увеличивает привлекательность байесовского подхода и способствует его быстрому распространению, в частности, в макроэкономическом анализе (Karlsson (2012)).

Наиболее часто цитируемый недостаток байесовского подхода — субъективность, с нашей точки зрения, не является существенным. Действительно, изменение априорного распределения влияет на результаты анализа³. Однако обычная частотная VAR (как и любая другая эконометрическая модель) является в неменьшей степени отражением субъективных представлений исследователя. Выбор весьма ограниченного набора переменных в модели, разделение их на эндогенные и экзогенные, определение числа лагов в модели⁴ и т. д. в любом случае отражают собственные представления исследователя о правильной спецификации модели. При этом, в отличие от частотной VAR, в BVAR эта субъективность задается в явном виде с помощью априорных распределений.

К сожалению, несмотря на широкое распространение BVAR в академических статьях, количество практических обзоров этого метода весьма ограничено. Существующие обзоры Karlsson (2012), Del Negro and 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 and Korobilis (2010) и Blake and Mumtaz (2012), к которым прилагаются коды в среде MATLAB.⁵ Однако Koop and Korobilis (2010) не рассматривают ставший весьма популярным метод задания априорного распределения через добавление дополнительных наблюдений, в т. ч. априорное распределение суммы коэффициентов (sum-of-coefficients prior) и априорное распределение начального наблюдения (initial observation prior). Blake and Mumtaz (2012) используют терминологию, несколько отличающуюся от других работ,

³Т.к. плотность апостериорного распределения представляет собой комбинацию плотности априорного распределения и функции правдоподобия. Этот вопрос подробно освещен в следующем разделе.

⁴На практике для определения количества лагов в частотной VAR, как правило, исследователи ориентируются на информационные критерии. Однако довольно часто разные информационные критерии дают противоречивые результаты. В этом случае предпочтение определенного критерия всем остальным — есть тоже субъективное решение.

⁵Находящиеся в открытом доступе известные нам коды в различных пакетах для работы с BVAR описаны в Приложении 1.

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 же академической литературе на данный момент, насколько нам известно, обзоров BVAR вообще нет.

Данный обзор, содержащий подробную классификацию априорных распределений, наиболее популярных при проведении макроэкономических исследований, может быть полезен для экономистов, обладающих ограниченным опытом в области байесовского анализа. Кроме того, к обзору прилагается код в R, в котором используются те же обозначения, что и в тексте работы, и который может быть использован как в учебных, так и в научных целях.

Мы оставляем за рамками данного обзора 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 переменные⁶ y_{it} , объединенные в вектор $y_t = (y_{1t}, y_{2t}, \dots, y_{mt})'$ размерности m . Векторная авторегрессия в сокращенной форме записывается в виде:

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

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

$$y_t = \Phi' x_t + \varepsilon_t \quad (2)$$

Если же сгруппировать переменные и шоки следующим образом: $Y = [y_1, y_2, \dots, y_T]'$, $X = [x_1, x_2, \dots, x_T]'$, $E = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T]'$, то VAR можно записать как:

$$Y = X\Phi + E \quad (3)$$

Эта же модель может быть записана в векторизованном виде⁷:

$$\text{vec}(Y) = \text{vec}(X\Phi I) + \text{vec}(E) \Leftrightarrow \quad (4)$$

$$y = (I_M \otimes X)\phi + \varepsilon \quad (5)$$

⁶Для удобства читателя обозначения также вынесены в Приложение 2.

⁷Уравнение (5) системы следует из тождества: $\text{vec}(ABC) = (C \otimes A) \text{vec}(B)$

где $\varepsilon \sim \mathcal{N}(0, \Sigma \otimes I_T)$ и вектор $\phi = \text{vec } \Phi$ имеет размерность $km \times 1$.

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

$$p(\Phi, \Sigma|Y) = \frac{p(\Phi, \Sigma)p(Y|\Phi, \Sigma)}{p(Y)} \quad (6)$$

Т.к. $p(Y)$ не зависит от Φ и Σ , то можно записать:

$$p(\Phi, \Sigma|Y) \propto p(\Phi, \Sigma)p(Y|\Phi, \Sigma) \quad (7)$$

Так как $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$, то функция правдоподобия задается как:⁸

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

Другая форма записи функции правдоподобия:

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

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.

2.2 Классификация популярных априорных распределений

Детально каждое из априорных распределений будет описано отдельно, а в этом разделе будет дана общая схема связи априорных распределений. Среди самых популярных априорных распределений можно назвать:

1. Независимое нормальное-обратное Уишарта априорное распределение (Independent Normal-inverse Wishart prior)

$$\begin{cases} \phi \sim \mathcal{N}(\underline{\phi}; \Xi) \\ \Sigma \sim \mathcal{IW}(\underline{S}; \nu) \\ \phi \text{ и } \Sigma \text{ независимы} \end{cases} \quad (10)$$

В общем случае выборку из апостериорного распределения можно получить по схеме Гиббса.

Частными случаями независимого нормального-обратного Уишарта являются:

⁸Здесь и далее $\text{etr}(\cdot) = \exp(\text{tr}(\cdot))$.

- (a) Априорное распределение Миннесоты (Minnesota prior)

$$\begin{cases} \phi \sim \mathcal{N}(\phi; \Xi) \\ \Sigma = \text{const} \end{cases} \quad (11)$$

Получается из независимого нормального-обратного Уишарта при $\underline{S} = (\underline{\nu} - m - 1) \cdot \Sigma$ и $\underline{\nu} \rightarrow \infty$. В этом случае апостериорное распределение выписывается в явной форме. Можно использовать алгоритм Монте-Карло, сразу генерирующий случайную выборку из апостериорного распределения, без необходимости периода «прожига».

Более того, алгоритм симуляции упрощается если матрица Ξ имеет структуру кронекерова произведения $\Xi = \Sigma \otimes \underline{\Omega}$. В этом случае распределение Миннесоты также становится частным случаем сопряжённого нормального-обратного Уишарта (см. ниже).

- (b) Независимое нормальное-Джеффриса априорное распределение (Independent normal Jeffreys prior)

$$\begin{cases} \phi \sim \mathcal{N}(\phi; \Xi) \\ \Sigma \sim |\Sigma|^{-(m+1)/2} \\ \phi \text{ и } \Sigma \text{ независимы} \end{cases} \quad (12)$$

Получается из независимого нормального-обратного Уишарта при $\underline{S} = \underline{\nu}^{1/m} \cdot I$ и $\underline{\nu} \rightarrow 0$. Функция плотности обратного Уишарта распределения имеет вид:

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

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

$$p(\Sigma) \rightarrow \text{const} \cdot |\Sigma|^{-(m+1)/2}$$

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

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

Распределение Миннесоты и независимое нормальное-Джеффриса являются противоположными крайностями независимого-обратного Уишарта.

В распределении Миннесоты матрица Σ предполагается известной, а в нормальном-Джеффриса матрица Σ имеет «размытое» неинформативное распределение.

Частным случаем независимого нормального-Джеффриса распределения является:

- i. Неинформативное-Джеффриса априорное распределение (Diffuse Jeffreys prior)

$$\begin{cases} \phi \sim 1 \\ \Sigma \sim |\Sigma|^{-(m+1)/2} \\ \phi \text{ и } \Sigma \text{ независимы} \end{cases} \quad (13)$$

При задании этого априорного распределения не нужно указывать ни одного гиперпараметра. Оно получается из независимого нормального-Джеффриса при $\underline{\phi} = 0$ и $\underline{\Xi} = a \cdot I$ и $a \rightarrow \infty$ и является также частным случаем сопряженного нормального-Джеффриса априорного распределения.

Для получения выборки из апостериорного распределения можно использовать прямое моделирование по схеме Монте-Карло без алгоритма Гиббса. Распределение получается из общего случая просто подстановкой $\underline{S} = 0$, $\underline{\nu} = 0$, $\underline{\Xi}^{-1} = 0$, $\phi = 0$. При этом формулы существенно упрощаются, в частности исчезает необходимость обращать матрицу размера $km \times km$.

2. Сопряженное нормальное-обратное Уишарта распределение (Conjugate normal-inverse Wishart prior)

$$\begin{cases} \phi|\Sigma \sim \mathcal{N}(\underline{\phi}, \Sigma \otimes \underline{\Omega}) \\ \Sigma \sim \mathcal{IW}(\underline{S}, \underline{\nu}) \end{cases} \quad (14)$$

Для сопряженного нормального-обратного Уишарта распределения нет необходимости использовать алгоритм Гиббса, так как существуют выведенные формулы параметров апостериорного распределения. Это позволяет использовать алгоритм Монте-Карло, сразу генерирующий случайную выборку из апостериорного распределения, что исключает период «прожига», необходимый для сходимости алгоритма Гиббса.

Частными случаями сопряженного нормального-обратного Уишарта распределения являются:

- (a) Распределение Миннесоты при $\underline{\Xi} = \Sigma \otimes \underline{\Omega}$.

При $\underline{\Xi} = \Sigma \otimes \underline{\Omega}$ распределение Миннесоты является одновременно и частным случаем независимого нормального-обратного Уишарта, и сопряженного нормального-обратного Уишарта. Для получения выборки из апостериорного распределения можно использовать и схему Гиббса для

независимого нормального-обратного Уишарта, и алгоритм Монте-Карло без «прожига».

- (b) Сопряжённое нормальное-Джеффриса априорное распределение (Conjugate normal Jeffreys prior)

$$\begin{cases} \phi | \Sigma \sim \mathcal{N}(\underline{\phi}; \Sigma \otimes \underline{\Omega}) \\ \Sigma \sim |\Sigma|^{-(m+1)/2} \end{cases} \quad (15)$$

Это распределение является частным случаем сопряжённого нормального-обратного Уишарта при $\underline{S} = \underline{\nu}^{1/m} \cdot I$ и $\underline{\nu} \rightarrow 0$. Формулы для гиперпараметров апостериорного распределения получаются подстановкой $\underline{\nu} = 0$, $\underline{S} = 0$.

Частным случаем сопряжённого нормального-Джеффриса является

- i. Неинформативное-Джеффриса (Diffuse Jeffreys prior)

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

В нашей работе мы всегда явно указываем, идёт ли речь о независимом или сопряжённом априорном распределении. Однако при чтении многих других работ надо быть внимательным, зачастую авторы говорят о «нормальном-обратном Уишарта» распределении, не уточняя, какое именно имеется в виду.

2.3 Оценка моделей с различными априорными распределениями

2.3.1 Априорное распределение Миннесоты

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

Априорное распределение параметров предполагается многомерным нормальным, зависящим от нескольких гиперпараметров. Параметры предполагаются независимыми, следовательно, их ковариационная матрица Ξ диагональна. Ковариационная матрица вектора $\varepsilon_t(\Sigma)$ также предполагается диагональной и постоянной. Тогда вектор ϕ не зависит от Σ :

$$\phi \sim \mathcal{N}(\underline{\phi}, \Xi) \quad (16)$$

Априорная плотность распределения ϕ может быть записана как:

$$p(\phi) = \frac{1}{(2\pi)^{km/2} |\Xi|^{1/2}} \exp \left\{ -\frac{1}{2} (\phi - \underline{\phi})' \Xi^{-1} (\phi - \underline{\phi}) \right\}. \quad (17)$$

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

$$\phi|Y \sim \mathcal{N}(\bar{\phi}, \bar{\Xi}) \quad (18)$$

где

$$\begin{aligned} \bar{\Xi} &= [\Xi^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1} \\ \bar{\phi} &= \bar{\Xi}[\Xi^{-1}\underline{\phi} + (\Sigma^{-1} \otimes X')y]. \end{aligned}$$

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

$$\begin{aligned} \bar{\Xi} &= [\Xi^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1} = [(\Sigma \otimes \underline{\Omega})^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1} = \\ &= [\Sigma^{-1} \otimes \underline{\Omega}^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1} = \Sigma \otimes (\underline{\Omega}^{-1} + X'X)^{-1} = \Sigma \otimes \bar{\Omega} \end{aligned} \quad (19)$$

В результате получаем:

$$\Phi|Y \sim \mathcal{N}(\bar{\Phi}, \Sigma \otimes \bar{\Omega})$$

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

Математическое ожидание априорного распределения параметров может быть записано с помощью матрицы $\underline{\Phi} = \mathbb{E}(\Phi)$ размерности $k \times m$, где $\underline{\Phi} = [\Phi_1 \dots \Phi_p \ \Phi_{const}]'$ и $\underline{\phi} = \text{vec } \underline{\Phi}$.

$$(\Phi_l)_{ij} = \begin{cases} \delta_i & \text{если } i = j, l = 1; \\ 0, & \text{в остальных случаях} \end{cases} \quad (20)$$

Распределение Миннесоты было задумано таким образом, чтобы учесть нестационарность многих макроэкономических временных рядов. В этом случае δ_i принимают значение единица.

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 Ξ диагональна. Диагональ матрицы Ξ разбивается на блоки $\Xi_1, \Xi_2, \dots, \Xi_m$ размерности $k \times k$. В свою очередь, каждый блок Ξ_i , $i = 1, \dots, m$ может быть разбит на диагональные подблоки размерности $m \times m$: $\Xi_{i, \text{lag}=l}$, $l = 1, \dots, p$ с константой $\Xi_{i, \text{const}}$ в конце главной диагонали:

$$\Xi = \begin{pmatrix} \Xi_1 & 0_{k \times k} & \cdots & 0_{k \times k} & 0_{k \times k} \\ 0_{k \times k} & \Xi_2 & \cdots & 0_{k \times k} & 0_{k \times k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{k \times k} & 0_{k \times k} & \cdots & \Xi_{m-1} & 0_{k \times k} \\ 0_{k \times k} & 0_{k \times k} & \cdots & 0_{k \times k} & \Xi_m \end{pmatrix} \quad \Xi_i = \begin{pmatrix} \Xi_{i,lag=1} & 0_{m \times m} & \cdots & 0_{m \times m} & 0_{m \times 1} \\ 0_{m \times m} & \Xi_{i,lag=2} & \cdots & 0_{m \times m} & 0_{m \times 1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{m \times m} & 0_{m \times m} & \cdots & \Xi_{i,lag=p} & 0_{m \times 1} \\ 0_{1 \times m} & 0_{1 \times m} & \cdots & 0_{1 \times m} & \Xi_{i,const} \end{pmatrix}$$

Тогда диагональные элементы $\Xi_{i,lag=l}$ определяются по формулам:

$$(\Xi_{i,lag=l})_{jj} = \begin{cases} \left(\frac{\lambda_{tight}}{l^{\lambda_{lag}}} \right)^2, & j = i \\ \left(\frac{\lambda_{tight} \cdot \lambda_{kron} \sigma_i}{l^{\lambda_{lag} \sigma_j}} \right)^2, & j \neq i \end{cases} \quad \Xi_{i,const} = \lambda_{tight}^2 \lambda_{const}^2 \sigma_i^2, \quad (21)$$

Как можно видеть из приведенной выше формулы (21) априорная дисперсия параметров зависит от нескольких гиперпараметров, задаваемых исследователем. Гиперпараметры имеют следующую интерпретацию:⁹ λ_{tight} (параметр регуляризации) отражает общую «жесткость» априорного распределения.

Если $\lambda_{tight} \rightarrow 0$, то априорное распределение совпадает с апостериорным распределением, и данные не играют никакой роли при оценке параметров. В этом случае мы считаем, что доподлинно знаем параметры, то есть:

$$\Phi \sim \mathcal{N}(\underline{\Phi}, 0), \quad \Phi|Y \sim \mathcal{N}(\underline{\Phi}, 0)$$

Если $\lambda_{tight} \rightarrow \infty$, то апостериорное математическое ожидание параметров сходится к МНК-оценке. В этом случае $\Xi^{-1} = 0$, поэтому

$$\bar{\Xi} = (\Sigma^{-1} \otimes (X'X))^{-1} = \Sigma \otimes (X'X)^{-1}$$

Отсюда

$$\begin{aligned} \bar{\phi} &= 0 + (\Sigma \otimes (X'X)^{-1}) \cdot (\Sigma^{-1} \otimes X') \cdot y = (I \otimes (X'X)^{-1} X') \cdot \text{vec}(Y) = \\ &= \text{vec}((X'X)^{-1} X'Y \cdot I') = \text{vec}((X'X)^{-1} X'Y) \end{aligned} \quad (22)$$

Параметр кросс-регуляризации λ_{kron} добавляет дополнительную жесткость лагам других переменных по сравнению с лагами зависимой переменной. Если $\lambda_{kron} < 1$, то собственные лаги зависимой переменной помогают предсказывать значение переменной лучше, чем лаги других переменных, поэтому коэффициенты при лагах других переменных оказываются жестче регуляризованы к нулю.

При $\lambda_{kron} = 1$ матрица Ξ имеет структуру кронекерова произведения и представляема в виде:

$$\Xi = \Sigma \otimes \underline{\Omega},$$

⁹Соотношение гиперпараметров, определяющих дисперсию в различных работах, можно найти в Приложении 3.

где $\underline{\Omega}$ — матрица размера $k \times k$, соответствующая отдельному уравнению. Кронекерово домножение слева на матрицу Σ для i -го уравнения означает домножение дисперсий, указанных в матрице $\underline{\Omega}$, на коэффициент σ_i^2 . Сама матрица $\underline{\Omega}$ представима в виде:

$$\underline{\Omega} = \begin{pmatrix} \underline{\Omega}_{lag=1} & 0_{m \times m} & \cdots & 0_{m \times m} & 0_{m \times 1} \\ 0_{m \times m} & \underline{\Omega}_{lag=2} & \cdots & 0_{m \times m} & 0_{m \times 1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{m \times m} & 0_{m \times m} & \cdots & \underline{\Omega}_{lag=p} & 0_{m \times 1} \\ 0_{1 \times m} & 0_{1 \times m} & \cdots & 0_{1 \times m} & \underline{\Omega}_{const} \end{pmatrix} \quad (23)$$

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

$$(\underline{\Omega}_{lag=l})_{jj} = \left(\frac{\lambda_{tight}}{l^{\lambda_{lag}} \sigma_j} \right)^2 \quad \underline{\Omega}_{const} = \lambda_{tight}^2 \lambda_{const}^2 \quad (24)$$

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

Как уже говорилось выше, при использовании априорного распределения Миннесоты, нет необходимости применять алгоритм Гиббса для получения апостериорного распределения. Алгоритм генерации случайной выборки непосредственно из апостериорного распределения происходит методом Монте-Карло:

1. На s -ом шаге сгенерировать очередную итерацию согласно:

$$\phi^{[s]} \sim \mathcal{N}(\bar{\phi}; \bar{\Xi}) \quad (25)$$

2. Увеличить s на единицу и перейти к пункту 1

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

1. Сгенерировать матрицу V размера $k \times m$ из независимых стандартных нормальных величин
2. Посчитать матрицу $\Phi^{[s]}$ по формуле:

$$\Phi^{[s]} = \bar{\Phi} + \text{chol}(\bar{\Omega}) \cdot V \cdot \text{chol}(\Sigma^{[s]})', \quad (26)$$

где $\text{chol}(\bar{\Omega})$ и $\text{chol}(\Sigma^{[s]})$ — верхнетреугольные матрицы, полученные путем разложения Холецкого матриц $\bar{\Omega}$ и $\Sigma^{[s]}$ соответственно.

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

2.3.2 Сопряженное нормальное-обратное Уишарта априорное распределение

Недостатка априорного распределения Миннесоты, состоящего в отсутствии байесовской процедуры для оценки матрицы Σ , можно избежать, если рассматривать сопряженное априорное распределение. При известной ковариационной матрицы остатков функция правдоподобия может быть разбита на две части: часть, соответствующую нормальному распределению, и часть, соответствующую обратному распределению Уишарта. Поэтому сопряженным априорным распределением для рассматриваемой модели будет также нормальное-обратное Уишарта распределение.

Сопряженное нормальное-обратное Уишарта априорное распределение может быть записано как:

$$\begin{cases} \Sigma \sim \mathcal{IW}(\underline{S}, \underline{\nu}) \\ \phi | \Sigma \sim \mathcal{N}(\underline{\phi}, \Sigma \otimes \underline{\Omega}) \end{cases} \quad (27)$$

В отличие от распределения Миннесоты, система (27) записана для случая, когда ковариационная матрица параметров имеет кронекерову структуру, то есть λ_{kron} полагается равной единице.

Гиперпараметры вектора математического ожидания ($\underline{\phi}$) и ковариационной матрицы $\underline{\Omega}$ условного априорного распределения могут быть заданы точно так же, как и в случае распределения Миннесоты для случая $\lambda_{kron} = 1$ (см. (20), (23) и (24)). \underline{S} выбирается так, чтобы среднее Σ совпадало с фиксированной ковариационной матрицей Σ в априорном распределении Миннесоты. Т.к. математическое ожидание и дисперсия параметров имеют вид¹⁰:

$$\mathbb{E}(\phi) = \underline{\phi} \quad \text{Var}(\phi) = (\underline{\nu} - m - 1)^{-1}(\underline{S} \otimes \underline{\Omega}), \quad (28)$$

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

$$(\underline{S})_{ii} = (\underline{\nu} - m - 1)\hat{\sigma}_i^2 \quad (29)$$

Выбор степеней свободы обратного Уишарта распределения $\underline{\nu}$ в соответствии с:

$$\underline{\nu} \geq \max\{m + 2, m + 2h - T\} \quad (30)$$

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

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

$$\begin{cases} \Sigma | Y \sim \mathcal{IW}(\bar{S}, \bar{\nu}) \\ \Phi | \Sigma, Y \sim \mathcal{N}(\bar{\Phi}, \Sigma \otimes \bar{\Omega}) \end{cases} \quad (31)$$

где гиперпараметры апостериорного распределения задаются как:

¹⁰Правая часть 28 следует из того, что: $\text{Var}(\phi) = \text{Var}(\mathbb{E}(\phi | \Sigma)) + \mathbb{E}(\text{Var}(\phi | \Sigma)) = \text{Var}(\underline{\phi}) + \mathbb{E}(\Sigma \otimes \underline{\Omega}) = \mathbb{E}(\Sigma \otimes \underline{\Omega}) = \mathbb{E}(\Sigma) \otimes \underline{\Omega} = (\underline{\nu} - m - 1)^{-1} \underline{S} \otimes \underline{\Omega}$

$$\begin{aligned}
\bar{\nu} &= \underline{\nu} + T \\
\bar{\Omega} &= (\underline{\Omega}^{-1} + X'X)^{-1} \\
\bar{\Phi} &= \bar{\Omega} \cdot (\underline{\Omega}^{-1}\underline{\Phi} + X'Y) \\
\bar{S} &= \underline{S} + \hat{E}'\hat{E} + \hat{\Phi}'X'X\hat{\Phi} + \underline{\Phi}'\underline{\Omega}^{-1}\underline{\Phi} - \bar{\Phi}'\bar{\Omega}^{-1}\bar{\Phi} \\
&= \underline{S} + \hat{E}'\hat{E} + (\underline{\Phi} - \hat{\Phi})'(\underline{\Omega} + (X'X)^{-1})^{-1}(\underline{\Phi} - \hat{\Phi}), \text{ где:} \\
\hat{\Phi} &= (X'X)^{-1}X'Y \text{ и } \hat{E} = Y - X\hat{\Phi}
\end{aligned}$$

Существует достаточно популярный альтернативный подход для подсчёта гиперпараметров апостериорного распределения.

Мы обнуляем 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 дополнительные наблюдения в матрицу X и в матрицу Y :

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

При добавлении наблюдений матрицы скалярных произведений $X^{*'}X^*$ и $X^{*'}Y^*$ разлагаются в сумму: $X^{*'}X^* = X^{+'}X^+ + X'X$, $X^{*'}Y^* = X^{+'}Y^+ + X'Y$. В частности, добавление нулевых искусственных наблюдений никак не изменяет матрицы $X'X$, $X'Y$ и $Y'Y$. Отметим, что 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 добавлять искусственные наблюдения и каким образом по отношению к матрицам X и Y . Можно их добавить в конец матриц X и Y , можно в начало, можно посередине.

Получим новые формулы для апостериорных гиперпараметров:

$$\begin{aligned}
\bar{\nu} &= \underline{\nu} + T \\
\bar{\Omega} &= (X^{*'}X^*)^{-1} = (X^{+'}X^+ + X'X)^{-1} \\
\bar{\Phi} &= \bar{\Omega} \cdot (X^{*'}Y^*) = \bar{\Omega} \cdot (X^{+'}Y^+ + X'Y) = (X^{*'}X^*)^{-1}X^{*'}Y^* \\
\bar{S} &= \hat{E}^{*'}\hat{E}^* \\
\hat{E}^* &= Y^* - X^*\bar{\Phi}
\end{aligned}$$

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

$$\begin{cases} X^{+'}X^+ = \underline{\Omega}^{-1} \\ X^{+'}Y^+ = \underline{\Omega}^{-1}\underline{\Phi} \\ (Y^+ - X^+\underline{\Phi})'(Y^+ - X^+\underline{\Phi}) = \underline{S} \end{cases} \quad (32)$$

Заметим, что взаимосвязь «новых» формул с матрицами, определяющими априорное и апостериорное распределения, может быть представлена в виде таблицы:

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

Указанные условия будут выполнены если добавить наблюдения по схеме¹¹:

$$Y^{NIW} = \begin{bmatrix} \frac{\text{diag}(\delta_1\sigma_1, \dots, \delta_m\sigma_m)}{\lambda_{tight}} \\ 0_{m(p-1) \times m} \\ \text{diag}(\sigma_1, \dots, \sigma_m) \\ 0_{1 \times m} \end{bmatrix} \quad X^{NIW} = \begin{bmatrix} \frac{\text{diag}(1, 2^{\lambda_{lag}}, \dots, p^{\lambda_{lag}}) \otimes \text{diag}(\sigma_1, \dots, \sigma_m)}{\lambda_{tight}} & 0_{mp \times 1} \\ 0_{m \times mp} & 0_{m \times 1} \\ 0_{1 \times mp} & \frac{1}{\lambda_{tight}\lambda_{const}} \end{bmatrix} \quad (33)$$

Модификации априорного распределения

В работах Doan, Litterman, and Sims (1984) и Sims (1993) было предложено добавить к априорному распределению дополнительную характеристику, введение которой обуславливается возможным наличием во временных рядах единичных корней и коинтеграционных соотношений. Это позволяет исключить появление неправдоподобно большой доли внутривыборочной дисперсии, объясняемой экзогенными переменными (Carriero, Clark, and Marcellino (2015)).

Априорное распределение суммы коэффициентов (sum-of-coefficients prior) было предложено в работе Doan, Litterman, and Sims (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 and Tallman (1999) и Blake and Mumtaz (2012)). Другими словами, среднее значение лагированных значений какой либо переменной является хорошим прогнозом для будущих значений этой переменной.

Внедрение этого априорного распределения производится путем добавления искусственных дамми-наблюдений по следующей схеме:

¹¹ Аналогичные формулы, приведенные в работах Bańbura, Giannone, and Reichlin (2010), Berg and Henzel (2013) для задания сопряженного нормального-обратного Уишарта априорного распределения являются частным случаем (33) для $\lambda_{lag} = 1$ и $\lambda_{const} \rightarrow \infty$.

$$Y^{SC} = \frac{1}{\lambda_{sc}} [\text{diag}(\delta_1 \mu_1, \dots, \delta_m \mu_m)] \quad (34)$$

$$X^{SC} = \frac{1}{\lambda_{sc}} [(1_{1 \times p}) \otimes \text{diag}(\delta_1 \mu_1, \dots, \delta_m \mu_m) \quad 0_{m \times 1}], \quad (35)$$

где $(1_{1 \times p})$ — вектор-строка из единиц длиной p , μ_i есть i -ая компонента вектора μ , который состоит из средних начальных значений всех переменных¹²: $\mu = \frac{1}{p} \sum_{t=1}^p y_t$

Априорное распределение 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 дамми-наблюдение, т. ч. все значения всех переменных равны соответствующему среднему начальных значений μ_i с точностью до коэффициента масштаба λ_{io} :

$$Y^{IO} = \frac{1}{\lambda_{io}} [\delta_1 \mu_1, \dots, \delta_m \mu_m] \quad (36)$$

$$X^{IO} = \frac{1}{\lambda_{io}} [(1_{1 \times p}) \otimes (\delta_1 \mu_1, \dots, \delta_m \mu_m) \quad 1], \quad (37)$$

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

Гиперпараметр λ_{io} отражает жесткость указанного априорного распределения. Когда $\lambda_{io} \rightarrow 0$, модель принимает вид, в котором либо все переменные стационарны со средним, равным выборочному среднему начальных условий, либо нестационарны без дрейфа и коинтегрированы.

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

1. На s -ом шаге сгенерировать очередную итерацию согласно:

$$\begin{aligned} \Sigma^{[s]} &\sim \mathcal{IW}(\bar{S}, \bar{\nu}) \\ \phi^{[s]} &\sim \mathcal{N}(\bar{\phi}; \Sigma^{[s]} \otimes \bar{\Omega}) \end{aligned}$$

2. Увеличить j на единицу и перейти к пункту 1.

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

1. Сгенерировать матрицу V размера $k \times m$ из независимых стандартных нормальных величин

¹²Некоторые авторы для расчета μ усредняют все наблюдения в выборке: $\mu = \frac{1}{T} \sum_{t=1}^T y_t$ (см. Baibura, Giannone, and Reichlin (2010) и Carriero, Clark, and Marcellino (2015)). Однако, в соответствии с работой C. A. Sims and Zha (1998) для расчета среднего следует использовать только первые p наблюдений.

2. Посчитать матрицу $\Phi^{[s]}$ по формуле:¹³

$$\Phi^{[s]} = \bar{\Phi} + \text{chol}(\bar{\Omega}) \cdot V \cdot \text{chol}(\Sigma^{[s]})'$$

Желание задать сопряженное нормальное 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 одну и ту же дисперсию λ_{tight}^2 . Хотя обычно эта предпосылка не является слишком ограничивающей, 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 (см, Kadiyala and Karlsson (1997)). Положим, исследователь хочет учесть в VAR наличие нейтральности денег. При построении модели эта предпосылка может быть учтена наложением такого априорного распределения, в котором все коэффициенты при лагах денег в уравнении для выпуска имеют нулевое математическое ожидание и низкую 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-обратное Уишарта.

2.3.3 Независимое нормальное-обратное Уишарта распределение

Этот тип распределений предполагает, что ковариационная матрица параметров может быть произвольной формы:

$$\begin{cases} \phi \sim \mathcal{N}(\phi; \Xi) \\ \Sigma \sim \mathcal{IW}(\underline{S}; \underline{\nu}) \\ \phi \text{ и } \Sigma \text{ независимы} \end{cases} \quad (38)$$

В этом случае можно показать (Karlsson (2012)), что условные апостериорные распределения имеют вид:

$$\begin{cases} \phi | \Sigma, Y \sim \mathcal{N}(\bar{\phi}; \bar{\Xi}) \\ \Sigma | \phi, Y \sim \mathcal{IW}(\bar{S}; \bar{\nu}) \end{cases} \quad (39)$$

¹³Указанная формула следует из того, что генерация вектора $\Phi \sim N(\bar{\Phi}, \Sigma \otimes \bar{\Omega})$: происходит с помощью разложения Холецкого указанной ковариационной матрицы: $\text{vec } \Phi = \text{vec } \bar{\Phi} + \text{chol}(\Sigma \otimes \bar{\Omega}) \times \nu$

где гиперпараметры апостериорного распределения задаются как:

$$\begin{aligned}\bar{\nu} &= \underline{\nu} + T \\ \bar{S} &= \underline{S} + E'E, \text{ где } E = Y - X\Phi \\ \bar{\Xi} &= (\underline{\Xi}^{-1} + \Sigma^{-1} \otimes X'X)^{-1} \\ \bar{\phi} &= \bar{\Xi} \cdot (\underline{\Xi}^{-1}\underline{\phi} + \text{vec}(X'Y\Sigma^{-1})) = \\ &= \bar{\Xi} \cdot (\underline{\Xi}^{-1}\underline{\phi} + (\Sigma^{-1} \otimes (X'X)) \text{vec } \hat{\Phi})\end{aligned}$$

Гиперпараметры априорного распределения могут быть выбраны точно так же, как и в априорном распределении Миннесоты (см. (20) и (21)). При необходимости неинформативное априорное распределение для коэффициентов при детерминированных переменных 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.

Получить Марковскую цепь, сходящуюся к апостериорному распределению можно, например, так :

1. Сгенерировать произвольно стартовую матрицу $\Sigma^{[0]}$, например, единичную
2. На s -ом шаге сгенерировать очередную итерацию согласно:

$$\phi^{[s]} \sim \mathcal{N}(\bar{\phi}^{[s-1]}; \bar{\Xi}^{[s-1]}), \text{ где } \bar{\phi}^{[s-1]} \text{ и } \bar{\Xi}^{[s-1]} \text{ рассчитываются через } \Sigma^{[s-1]} \quad (40)$$

$$\Sigma^{[s]} \sim \mathcal{IW}(\bar{S}^{[s]}; \bar{\nu}), \text{ где } \bar{S}^{[s]} \text{ рассчитываются с помощью } \phi^{[s]} \quad (41)$$

3. Увеличить s на единицу и перейти к пункту 2.

2.4 Эндогенное определение гиперпараметров: оценка BVAR большой размерности

В некоторых случаях гиперпараметры априорного распределения определяются эндогенно. В частности, это происходит при оценке модели большой размерности. Для решения ряда прикладных макроэкономических задач, связанных как с прогнозированием, так и со структурным анализом, исследователям требуется работать с выборками большой размерности. И хотя BVAR широко использовались в моделях низкой размерности практически с момента их появления в начале 1980-х годов, их применение в моделях высокой 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, Giannone, and Reichlin (2008) и Bańbura, Giannone, and Reichlin (2010), в которых было показано, что BVAR вполне могут быть применены к выборкам большой размерности без наложения дополнительных ограничений. Однако использование большого количества временных рядов требует уменьшения параметра λ_{tight} с увеличением размерности выборки, что означает наложение более жесткого априорного распределения. На данный момент в литературе используется два подхода к определению оптимальной величины λ_{tight} .

Первый алгоритм был предложен в работе Bańbura, Giannone, and Reichlin (2010) и он 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 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.

Детальное описание процедуры приведено ниже. Обозначим фактическое значение переменной var в момент $T + h$ как $y_{var,T+h}$, а прогноз переменной var , осуществленный в момент на горизонт h в модели с m переменными и параметром жесткости λ как: $y_{var,T+h|T}^{\lambda,m}$. Схема выбора λ состоит из следующих этапов:

1. На первом этапе с помощью BVAR строятся внутривыборочные однопериодные прогнозы на обучающей выборке и рассчитывается среднеквадратичная ошибка прогноза для M переменных, объединенных в набор (\mathcal{M}) и представляющих особый интерес.¹⁴

$$MSFE_{var,1}^{\lambda,m} = \frac{1}{T_0 - p} \sum_{t=p}^{T_0-1} \left(y_{var,t+1|t}^{\lambda,m} - y_{var,t+1} \right)^2, \quad (42)$$

где коэффициенты BVAR получены в результате оценки на обучающей выборке: $t = p + 1, \dots, T_0$ и T_0 — последнее наблюдение обучающей выборки: $T_0 = p + 120$.

2. Аналогично рассчитываются однопериодные прогнозы в соответствии с моделью случайного блуждания с дрейфом¹⁵ для тех же самых переменных

¹⁴В базовый набор переменных, представляющих особый интерес (\mathcal{M}) включается индекс промышленного производства, индекс потребительских цен и межбанковская процентная ставка, т.е. $M = 3$. Для проверки робастности было проведено исключение процентной ставки из этого набора, что практически не изменило вектор оптимальных λ .

¹⁵MSFE 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 $\lambda = 0$ и $\delta_i = 1, i = 1, \dots, k$.

$(MSFE_{var,1}^0)$, и кроме того рассчитывается новый индикатор $FIT^{\lambda,m}$:

$$FIT^{\lambda,m} = \frac{1}{M} \sum_{var \in \mathcal{M}} \frac{MSFE_{var,1}^{\lambda,m}}{MSFE_{var,1}^0} \quad (43)$$

3. Оценивается трехмерная VAR для тех же самых M переменных, представляющих особый интерес при прогнозировании,¹⁶ и считаются MSFE и индикатор $FIT^{\infty,M}$:

$$FIT^{\infty,M} = \frac{1}{M} \sum_{var \in \mathcal{M}} \frac{MSFE_{var,1}^{\infty,M}}{MSFE_{var,1}^0} \quad (44)$$

4. Оптимальным λ считается значение, минимизирующее разницу между $FIT^{\lambda,m}$ и $FIT^{\infty,M}$:

$$\lambda_m^* = \arg \min_{\lambda} |FIT^{\lambda,m} - FIT^{\infty,M}| \quad (45)$$

После того как выбрано оптимальное λ для каждой модели, происходит построение вневыборочных прогнозов на оценивающей выборке.

2.4.1 Регуляризация с помощью однопериодного прогноза

Второй алгоритм предложен в работе Doan, Litterman, and Sims (1984) и представляет собой выбор такого параметра λ_{tight} , который бы максимизировал точность однопериодного вневыборочного прогноза на обучающей выборке. Этот выбор сводится к максимизации функции предельной плотности:

$$\lambda^* = \arg \max_{\lambda} \ln p(Y) \quad (46)$$

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

$$p(Y) = \int p(Y|\phi)p(\phi)d\phi \quad (47)$$

Если априорное распределение является нормальным-обратным Уишарта, то маргинальная плотность $p(Y)$ может быть посчитана аналитически (Zellner (1996); Bauwens, Lubrano, and Richard (2000); Carriero, Kapetanios, and Marcellino (2012)):

¹⁶Для указания на неограниченную VAR используется надстрочный индекс ∞ , так как неограниченная VAR является частным случаем BVAR при $\lambda \rightarrow \infty$. В этом случае апостериорное распределение совпадает с функцией правдоподобия.

$$p(Y) = \pi^{-\frac{Tm}{2}} \times + |(I + X\Omega X')^{-1}|^{\frac{N}{2}} \times |\underline{S}|^{\frac{\nu}{2}} \times \frac{\Gamma_N(\frac{\nu+T}{2})}{\Gamma_N(\frac{\nu}{2})} \times \\ \times |\underline{S} + (Y - X\Phi)'(I + X\Omega X')^{-1}(Y - X\Phi)|^{-\frac{\nu+T}{2}}, \quad (48)$$

где $\Gamma_N(\cdot)$ обозначает N -мерную гамма функцию. Выбор числа лагов происходит аналогично путем максимизации по p функции предельной плотности (48):

$$p^* = \arg \max_p \ln p(Y) \quad (49)$$

2.4.2 Особенности кодирования

При практической реализации алгоритма Гиббса или Монте-Карло часто приходится обращаться положительно определённые симметричные матрицы. Некоторые из этих матриц могут иметь определитель близкий к нулю, что мешает практическому обращению этих матриц на компьютере. В таком случае можно использовать следующий способ:

1. Получить разложение Холецкого для заданной матрицы A

$$A = U'U,$$

где U — верхнетреугольная матрица

2. Обратить матрицу U . Существуют специальные алгоритмы обращения верхнетреугольных матриц.
3. Получить A^{-1} по формуле

$$A^{-1} = U^{-1}U^{-1'}$$

Однако даже данный способ сопряжён с численными трудностями, если обрабатываемая матрица плохо обусловлена. В таком случае оправданно использовано псевдообратную матрицу Мура-Пенроуза.

Если дополнительно известно, что обрабатываемая матрица A имеет вид $A = X'X$, то изложенный алгоритм лучше не применять, а сразу получить обратную к A , без вычисления A :

1. Получить сингулярное разложение матрицы X , $X = U\Sigma V'$
2. Получить обратную к $A = X'X$ по формуле $A^{-1} = V\Sigma^{-2}V'$

3 Прогнозирование с помощью BVAR

3.1 Апостериорная прогнозная плотность

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 определённой модели, но и в оценке качества этого прогноза, то модель оценивается на некоторой исторической выборке и прогноз строится на те 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 количеству наблюдений, но на каждом шаге начало выборки и её окончание сдвигаются на один шаг вперёд. 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 начало выборки, и на каждом шаге её длина увеличивается на одно наблюдение.

Ключевой концепцией при построении прогноза с помощью байесовской модели является функция апостериорной прогнозной плотности (posterior predictive density). Вслед за Karlsson (2012) мы обозначим прогнозную плотность с помощью $p(y_{T+1:T+H}|Y_T)$. Эта запись означает, что прогноз строится на все моменты времени, начиная с $T + 1$ -го и заканчивая моментом $T + H$ -ым при известных наблюдениях с момента 1 по момент T . Здесь матрица $y_{T+1:T+H} = (y'_{T+1}, \dots, y'_{T+H})'$ — это все будущие наблюдения, а матрица $Y_T = (y'_1, \dots, y'_T)'$ — это все наблюдения, по которым оценивалась модель. Функция апостериорной прогнозной плотности может быть представлена как:

$$p(y_{T+1:T+H}|Y_T) = \int p(y_{T+1:T+H}|Y_T, \phi) p(\phi|Y_T) d\phi, \quad (50)$$

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

Как правило, функция апостериорной прогнозной плотности не выражается аналитически для прогнозирования дальше, чем на один период вперёд. Однако её расчет возможен в соответствии с формулой (50) с помощью численных методов. Для этого необходимо для каждой реализации вектора параметров из апостериорного распределения $p(\theta|Y_T)$ посчитать прогнозные значения переменных \tilde{y}_{T+h} на горизонте для $h = 1, \dots, H$ с помощью функции условной плотности $p(y_{T+1:T+H}|Y_T, \theta)$. При этом при построении прогноза на горизонт \tilde{h} прогнозы, построенные на $h < \tilde{h}$ становятся как будто известными значениями. Повторение описанной процедуры большое число раз позволяет получить выборку из апостериорного прогнозного распределения для каждого h .

тильда??? — обозначение для прогнозов

Другими словами, для BVAR построение прогнозной плотности происходит по следующей схеме (Karlsson (2012), с. 8, 18):

1. Сгенерировать параметры апостериорного распределения. Для априорного распределения Миннесоты можно использовать алгоритм на стр. 12, для сопряженного нормального-обратного Уишарта — алгоритм на стр. 16, для независимого нормального-обратного Уишарта — алгоритм на стр. 18.
2. Сгенерировать $\varepsilon_{T+1}^{[s]}, \dots, \varepsilon_{T+H}^{[s]}$ из $\varepsilon_t \sim N(0, \Sigma^{[s]})$ (в случае априорного распределения Миннесоты $\Sigma^{[s]} = \Sigma$) и посчитать рекурсивно:

$$\tilde{y}_{T+h}^{[s]} = \Phi_{ex}^{[s]} + \sum_{i=1}^{h-1} \Phi_i^{[s]} \tilde{y}_{T+h-i}^{[s]} + \sum_{i=h}^p \Phi_i^{[s]} y_{T+h-i}^{[s]} + \varepsilon_{T+h}^{[s]} \quad (51)$$

3. Увеличить s на единицу и перейти к шагу 1.

Далее следует «забыть» о том, что разные прогнозы были сделаны для разных значений параметров. Для априорного распределения Миннесоты и сопряженного нормального-обратного Уишарта следует рассматривать $\{\tilde{y}_{T+1}^{[s]}, \dots, \tilde{y}_{T+H}^{[s]}\}_{s=1}^S$ 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 $s > B$: $\{\tilde{y}_{T+1}^{[s]}, \dots, \tilde{y}_{T+H}^{[s]}\}_{s=B+1}^S$.

Стоит заметить, что в случае сопряженного нормального априорного распределения процесс генерации параметров можно ускорить. Для этого матрицу параметров Φ можно посчитать как:

$$\Phi = \bar{\Phi} + \text{chol}(\bar{\Omega}) \times V \times \text{chol}(\Sigma)' \quad (52)$$

При этом прогноз на один шаг вперёд представляет собой линейную функцию параметров, и поэтому апостериорная прогнозная плотность для $h = 1$ может быть задана аналитически. Она принимает вид матричного матричного t -распределения (MT), параметры которого зависят от конкретного используемого априорного распределения. Например, для случая сопряженного нормального распределения (Carriero, Clark, and Marcellino (2015) с. 54) прогноз имеет многомерное t -распределение с параметрами:

$$y'_{T+1}|x'_{T+1} \sim MT\left(x'_{T+1}\bar{\Phi}, (x'_{T+1}\bar{\Omega}x_{T+1})^{-1}, \bar{S}, \bar{\nu}\right) \quad (53)$$

Зачем снова уточняется, что это для сопр норм? Ведь уже выше сказано, что это только в этом случае? Какая тогда свобода в используемом априорном распределении?

Прогноз на большее количество периодов требует численной процедуры, описанной выше.

После построения случайной выборки из апостериорного прогнозного распределения можно рассчитывать точечные и интервальные прогнозы. Выбор конкретного типа точечного прогноза (например, мода или медиана прогнозной плотности) может быть сделан на основании функции потерь. Можно считать, что существует функция потерь $\mathbb{L}(a, y_{T+1:T+H})$, определяющая, какой вектор значений a будет выбран в качестве точечного прогноза. Вектор значений a выбирается таким образом, чтобы минимизировать ожидаемые потери при условии доступных данных Y_T (Karls-son (2012):р. 3):

$$\mathbb{E}[\mathbb{L}(a, y_{T+1:T+H})|Y_T] = \int \mathbb{L}(a, y_{T+1:T+H})p(y_{T+1:T+H}|Y_T)dy_{T+1:T+H} \quad (54)$$

При заданных функции потерь и прогнозной плотности решение задачи минимизации есть функция только от доступных данных, $a(Y_T)$. Для частных случаев функции потерь решение принимает простую форму. Например, для квадратичной функции потерь решение принимает вид условного математического ожидания: $a(Y_T) = \mathbb{E}(y_{T+1:T+H}|Y_T)$, а для функции потерь в виде абсолютного значения решение представляет собой моду прогнозного распределения.

Хм, несмотря на коррелированность срабатывает? я знаю этот результат для скалярного точечного прогноза, а для целого вектора у меня сомнения

В прикладных работах, использующих BVAR, строятся точечные прогнозы и прогнозы плотности. При этом использование альтернативных моделей для построения прогноза позволяет оценивать указанные модели с точки зрения их точности.

3.2 Оценка точности точечного прогноза: одномерный случай

Для оценки точности точечного прогноза чаще всего используют среднеквадратическую ошибку прогноза (MSFE, mean squared forecasting error) или квадратный корень из нее (RMSFE, root mean squared forecasting error):

$$MSFE_{var,h}^M = \frac{1}{N_h} \sum_T (y_{var,T+h|T}^M - y_{var,T+h|T})^2, \quad (55)$$

$$RMSFE_{var,h}^M = \sqrt{MSFE_{var,h}^M}, \quad (56)$$

где $y_{var,T+h|T}^M$ — прогноз переменной var , сделанный в момент T на h шагов вперёд на основе модели M . Число N_h отражает количество таких прогнозов для горизонта h , сделанных для разных T .

Альтернативной мерой точности прогноза служит значение средней абсолютной ошибки прогноза (MAE, mean absolute forecast error):

$$MAFE_{var,h}^M = \frac{1}{N_h} \sum_T |y_{var,T+h|T}^M - y_{var,T+h|T}| \quad (57)$$

тильды у прогноза? здесь пропали :)

3.3 Оценка точности точечного прогноза: многомерный случай

В многомерном случае, когда с помощью одного индикатора требуется описать точность прогноза сразу по нескольким переменным, используют две статистики, предложенные в работе Adolfson, Lindé, and Villani (2007). Это след и логарифм определителя матрицы среднеквадратических ошибок $\Sigma_M(h)$, рассчитываемой как:

$$\Sigma_A(h) = \frac{1}{N_h} \sum_{t=T}^{T+N_h-1} \tilde{\varepsilon}_{t+h|t} \tilde{\varepsilon}_{t+h|t}', \quad (58)$$

где $\tilde{\varepsilon}_{t+h|t} = A^{-1/2} \varepsilon_{t+h|t}$, а $\varepsilon_{t+h|t}$ — ошибка прогноза на горизонте h и A — произвольная положительно определенная матрица. Ранжирование прогнозов с помощью логарифма определителя не зависит от выбора матрицы A , а с помощью следа — зависит, так как $\ln |\Sigma_A(h)| = \ln |\Sigma_I(h)| - \ln |A|$, а $\text{tr} |\Sigma_A(h)| = \text{tr} |A^{-1} \Sigma_I(h)|$.

В работе Adolfson, Lindé, and Villani (2007) в качестве матрицы A была взята диагональная матрица выборочных дисперсий прогнозируемых переменных. В этом случае след матрицы равен взвешенному среднему среднеквадратичных ошибок прогнозов индивидуальных временных рядов. Однако в указанной работе было утверждается, что ранжирование точечных прогнозов с помощью многомерных индикаторов может вводить в заблуждение. Дело в том, что на значение этих статистик особенно сильно влияет возможность прогнозирования переменных, которые сложнее всего поддаются прогнозированию. Плохо прогнозируемые переменные обычно включают в себя большую часть последней главной компоненты. При этом эти переменные не обязательно представляют особый интерес для исследователя (Adolfson, Lindé, and Villani (2007)).

3.4 Оценка точности апостериорной прогнозной плотности

3.4.1 Использование апостериорной функции распределения

Точечный прогноз представляет собой вероятное будущее значение переменной и не несёт никакой информации о степени неопределенности прогноза. В отличие от точечного прогноза, прогноз плотности отражает всю меру неопределенности прогноза. Прогноз плотности можно изобразить, например, с помощью веерных диаграмм (fan charts). Веерные диаграммы часто используются центральными банками для анонсирования своих прогнозов по инфляции или другим ключевым макроиндикаторам. В отличие от стандартных доверительных интервалов, прогноз плотности в общем случае не является симметричным относительно точечного прогноза. Основная сложность при оценивании прогноза плотности является то, что истинная плотность является ненаблюдаемой, а наблюдаемыми являются лишь отдельные фактические значения прогнозируемых переменных, по одному на каждый момент времени.

Похожесть апостериорного прогнозного распределения на фактические данные называется калибровкой («соответствие между вероятностными прогнозами и фактическим поведением прогнозируемой величины», Цыплаков (2012), стр. 118). Оценивание точности калибровки рассматривается, например, в работах Gneiting, Balabdaoui, and Raftery (2007), Mitchell and Wallis (2011). Естественно, возможные несоответствия между прогнозом и фактическими реализациями могут по-разному восприниматься исследователем. Формально говоря, это определяется его функцией потерь, о которой мы писали выше. Распространенный метод оценки калибровки функции потерь, PIT, состоит в том, чтобы подставить фактические значения переменных в апостериорную прогнозную функцию распределения¹⁷ (см., например, Diebold, Gunther, and Tay (1998), Gerdrup et al. (2009), Gonzalez-Rivera and Sun (2015)).

Предположим, что прогноз на момент τ имеет апостериорную функцию распределения F , а фактическое значение прогнозируемой переменной равно y_τ . Если фактические реализации действительно представляют собой выборку из указанного распределения, то значения функции распределения $F(y_\tau)$ имеют равномерное распределение на промежутке $[0, 1]$. Визуальная оценка качества прогноза происходит с помощью гистограммы. При совпадении фактического и предполагаемого апостериорного распределения гистограмма должна соответствовать равномерному распределению, то есть иметь столбики примерно одинаковой высоты.

В отличие от анализа функции плотности прогноза (predictive density), метод PIT является частотным, т.к. сравнивает распределение фактических значений с распределением, которое они имели бы, если бы процесс генерации данных совпадал с тем, что используется моделью. Некоторые формальные тесты на основе PIT можно найти в Geweke and Amisano (2010).

Можно выделить две проблемы, связанные с применением PIT. Во-первых, на практике условие равномерности распределения PIT сложно проверить: даже если рассматриваемая модель действительно является процессом генерации данных для некоторой переменной, прогнозная плотность не обязательно будет с ней совпадать из-за неопределенности относительно параметров.

не понял мысль, мы же всё выинтегрировали, что мешало?

Во-вторых, в работе Gneiting, Balabdaoui, and Raftery (2007) с помощью симуляций было показано, что на основе PIT невозможно однозначно выбрать истинную модель среди нескольких конкурирующих. Поэтому в этой же работе было предложено максимизировать «остроту прогнозной плотности при заданной калибровке», где под остротой понимается концентрация распределений вокруг фактических реализаций, что, в свою очередь, можно оценить с помощью диаграмм размаха (box plots) или скоринговых правил.

¹⁷По английски этот метод называется вероятностным интегральным преобразованием, (probability integral transform, PIT), в силу того, что апостериорная функция распределения получается взятием интеграла от апостериорной функции плотности. В русском языке подобная терминология не используется.

Ежели мы чаго максимизируем, то как это оценивает точность прогноза? см. название раздела? Да и по каким параметрам? может добавить слово иногда невозможно выбрать?

3.4.2 Скоринговые правила

Скоринговые правила представляют собой функции, в соответствии с которыми прогнозу начисляется определенное количество условных баллов в зависимости от реализации прогнозируемой величины. Подробный обзор скоринговых правил можно найти в (Цыплаков (2012)). Если речь идет о серии прогнозов, полученных по разным моделям, то наиболее точной признается та модель, которая получила наивысший скоринговый балл.

Наиболее часто в литературе используется скоринговое правило для логарифма прогнозной плотности (log predictive density scores), предложенное в работе Good (1952) и описанное в Geweke and Amisano (2010). Свежие примеры его использования можно найти в Adolfson, Lindé, and Villani (2007), Christoffel, Warne, and Coenen (2010) и Carriero, Clark, and Marcellino (2015).

Скоринговый балл рассчитывается как:

$$s_h = \sum_{t=T}^{T+N_h-1} \ln p_t(y_{t+h}) \quad (59)$$

что такое $p_t()$? почему нет внутри $|Y_T$?

Скоринговый балл для прогноза на один шаг может быть выражен с помощью функции маргинального правдоподобия (Adolfson, Lindé, and Villani (2007), с. 324-325):

$$\begin{aligned} s_1 &= \ln[p_T(y_{T+1}) \cdots p_{T+N_1-1}(y_{T+N_1})] = \\ &= \ln[p_T(y_{T+1}, \dots, y_{T+N_1})] = \ln m(T + N_1) - \ln m(T), \end{aligned} \quad (60)$$

опять же, что такое $p_t()$ и какие внутри условия явно не написаны?

где $m(t) = p_0(y_1, \dots, y_t) = \int p(y_1, \dots, y_t | \phi) p(\phi) d\phi$ и $m(t)$ — функция маргинального правдоподобия данных вплоть до момента времени t , $p(\theta)$ — априорная плотность. Важно отметить, что внутри интеграла находится только априорная плотность параметров $p(\phi)$ и функция правдоподобия $p(y_1, \dots, y_t | \phi)$. Следовательно, при подсчёте маргинального правдоподобия не используются фактические данные, что позволяет интерпретировать маргинальное правдоподобие как меру точности вневыборочного прогноза плотности, а не меру внутривыборочной подгонки.

Подсчет s_h для $h > 1$ является более сложным, т.к. $p(y_{t+h})$ не имеет аналитического решения. Один подход состоит в оценке $p(y_{t+h})$ на основе реализаций прогнозов с помощью, например, ядерной оценки плотности (kernel density estimator). Практически это реализуемо, только если количество переменных в модели не очень велико.

В обратном случае Adolfson, Lindé, and Villani (2007) предлагают предположить, что $p(y_{t+h})$ — есть плотность многомерного нормального распределения и оценить вектор средних этого распределения и его ковариационную матрицу из прогнозной выборки.

Именно этот второй подход использован, например, в работе Carriero, Clark, and Marcellino (2015). В этом случае применяется следующий алгоритм построения условного балла:

1. Сгенерировать выборку из прогнозов для переменной var на горизонт h
2. Рассчитать оценку логарифма предсказательной плотности по формуле:

$$s_t(y_{t+h}^{var}) = \log p(y_{t+h}|Y_t, m) = \quad (61)$$

$$= -0.5 \left[\ln(2\pi + \ln(V_{\tau+h|\tau}^{var}) + (y_{t+h}^{var} - \bar{y}_{\tau+h|\tau})^2 / V_{\tau+h|\tau}^{var} \right] \quad (62)$$

где $p(y_{t+h}|Y_t, m)$ — маргинальная прогнозная условная плотность для y_{t+h}, \dots, y_{t+h} , зависящая от наблюдаемых данных $Y_t = \{y_1, \dots, y_T\}$, $\bar{y}_{\tau+h|\tau}$ и $V_{\tau+h|\tau}^{var}$ обозначают апостериорное среднее и дисперсию симулированного распределения прогноза для переменной var и прогнозного окна h .

3. Рассчитать условный балл как среднее арифметическое полученных оценок для каждого прогнозного горизонта: $s_t(y_{t+h}^{var})$:

$$\bar{s}_{var,h}^M = \frac{1}{N_h} \sum s_t(y_{t+h}^{var}), \quad (63)$$

где N_h — количество построенных прогнозов на горизонт h .

4 Заключение

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

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

Список литературы

- Adolfson, Malin, Jesper Lindé, and Mattias Villani (2007). “Forecasting Performance of an Open Economy DSGE Model”. In: *Econometric Reviews* 26(2-4), pp. 289–328.
- Alvarez, Ignacio (2014). “Bayesian inference for a covariance matrix”. In: *arXiv preprint arXiv:1408.4050*.
- Bañbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2010). “Large Bayesian vector auto regressions”. In: *Journal of Applied Econometrics* 25(1), pp. 71–92.
- Bauwens, Luc, Michel Lubrano и Jean-Francois Richard (2000). *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press.
- Berg, Tim Oliver and Steffen Henzel (2013). *Point and Density Forecasts for the Euro Area Using Many Predictors: Are Large BVARs Really Superior?* Tech. rep. ifo Working Paper.
- Blake, Andrew and Haroon Mumtaz (2012). *Applied Bayesian econometrics for central bankers*. 1st ed. Centre for Central Banking Studies, Bank of England.
- Canova, Fabio (2007). *Methods for applied macroeconomic research*. Vol. 13. Princeton University Press.
- Carriero, Andrea, Todd E Clark, and Massimiliano Marcellino (2015). “Bayesian VARs: specification choices and forecast accuracy”. In: *Journal of Applied Econometrics* 30(1), pp. 46–73.
- Carriero, Andrea, George Kapetanios и Massimiliano Marcellino (2012). “Forecasting government bond yields with large Bayesian vector autoregressions”. B: *Journal of Banking & Finance* 36(7), с. 2026–2047.
- Christoffel, Kai, Anders Warne, and Günter Coenen (2010). *Forecasting with DSGE models*. Working Paper Series 1185. European Central Bank.
- De Mol, Christine, Domenico Giannone, and Lucrezia Reichlin (2008). “Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?” In: *Journal of Econometrics* 146(2), pp. 318–328.
- Del Negro, Marco and Frank Schorfheide (2011). “Bayesian macroeconometrics”. In: *The Oxford handbook of Bayesian econometrics*, pp. 293–389.
- Diebold, Francis, Todd A Gunther, and Anthony S Tay (1998). “Evaluating Density Forecasts with Applications to Financial Risk Management”. In: *International Economic Review* 39(4), pp. 863–83.
- Doan, Thomas, Litterman, and Sims (1984). “Forecasting and conditional projection using realistic prior distributions”. In: *Econometric reviews* 3(1), pp. 1–100.
- Favero, Carlo A (2001). *Applied macroeconometrics*. Oxford University Press.
- Gerdrup, Karsten R. et al. (2009). *Evaluating ensemble density combination - forecasting GDP and inflation*. Working Paper 2009/19. Norges Bank.
- Geweke, John and Gianni Amisano (2010). “Comparing and evaluating Bayesian predictive distributions of asset returns”. In: *International Journal of Forecasting* 26(2), pp. 216–230.

- Gneiting, Tilmann, Fadoua Balabdaoui, and Adrian E. Raftery (2007). “Probabilistic forecasts, calibration and sharpness”. In: *Journal of the Royal Statistical Society Series B* 69(2), pp. 243–268.
- Gonzalez-Rivera, Gloria and Yingying Sun (2015). “Generalized autocontours: Evaluation of multivariate density models”. In: *International Journal of Forecasting* 31(3), pp. 799–814.
- Good, I. J. (1952). “Rational Decisions”. In: *Journal of the Royal Statistical Society Ser. B*,
- Kadiyala, K and Sune Karlsson (1997). “Numerical Methods for Estimation and Inference in Baesian VAR-Models”. In: *Journal of Applied Econometrics* 12(2), pp. 99–132.
- Karlsson, Sune (2012). *Forecasting with Bayesian Vector Autoregressions*. Working Papers 2012:12. Örebro University, School of Business.
- Koop, Gary and Dimitris Korobilis (2010). “Bayesian Multivariate Time Series Methods for Empirical Macroeconomics”. In: *Foundations and Trends(R) in Econometrics* 3(4), pp. 267–358.
- Litterman (1979). *Techniques of forecasting using vector autoregressions*. Working Papers 115. Federal Reserve Bank of Minneapolis.
- Litterman (1986). “Forecasting with Bayesian vector autoregressions—five years of experience”. In: *Journal of Business & Economic Statistics* 4(1), pp. 25–38.
- Mitchell, James and Kenneth Wallis (2011). “Evaluating density forecasts: forecast combinations, model mixtures, calibration and sharpness”. In: *Journal of Applied Econometrics* 26(6), pp. 1023–1040.
- Robertson, John C and Ellis W Tallman (1999). “Vector autoregressions: forecasting and reality”. In: *Economic Review-Federal Reserve Bank of Atlanta* 84(1), p. 4.
- Sims (1980). “Macroeconomics and Reality”. In: *Econometrica* 48(1), pp. 1–48.
- Sims (1993). “A nine-variable probabilistic macroeconomic forecasting model”. In: *Business Cycles, Indicators and Forecasting*. University of Chicago Press, pp. 179–212.
- Sims, Christopher A and Tao Zha (1998). “Bayesian methods for dynamic multivariate models”. In: *International Economic Review*, pp. 949–968.
- Zellner, A. (1996). *An Introduction to Bayesian Inference in Econometrics*. Wiley Classics Library. Wiley. ISBN: 9780471169376.
- Малаховская, О.А. и С. Э. Пекарский (2012). «Исследования причинно-следственных взаимосвязей в макроэкономике: нобелевская премия по экономике 2011 года». В: *Экономический журнал Высшей школы экономики* 16(1), с. 3–30.
- Цыплаков, А. А. (2012). «Оценка качества вероятностных прогнозов: корректные скоринговые правила и моменты». В: *Прикладная эконометрика* 3, с. 115–132.

Приложение 1. Доступные реализации кода

Источник	Среда	Min	Conj N-IW	Ind N-IW	SoC	IO
Carriero	Matlab	-	?	-	+	+
Blake Mumtaz	Matlab	-	+	+	+	+
Koop Korobilis	Matlab	+	+	+	-	-
Zha	Matlab		?		+	+
Le Sage	Matlab	?				
Sims	Matlab		?	?	+	+
Canova	Matlab		?			
BMR	R	+	-	+	-	-
MSBVAR	R	-	+	-	+	+
bvarr	R	+	+	+	+	+
Sims	R		?	?	+	+
Встроенная функция	EViews	+	+	-	+	+
Встроенная функция	Dynare	?	+	?	+	+

1. Carriero: Дамми-наблюдения вводятся как для conj NIW, при этом делает Gibbs sampling, 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 псевдо-обратная. Имеется странный хак, повторно генерирующий VAR коэффициенты, если собственные числа за пределами единичного круга. <http://cremfi.econ.qmul.ac.uk/efp/info.php>
2. Blake Mumtaz: Называет Миннесотой Ind NIW. Код для conj NIW построен так же, как у Carriero через Gibbs Sampling: разница — добавляет две одинаковых строки для дамми при определении коэффициента при константе. http://www.bankofengland.co.uk/education/Pages/ccbs/technical_handbooks/techbook4.aspx
3. Koop Korobilis: Код крайне негибкий. Без правки кода нет возможности прогнозировать больше чем на один шаг, базовые conj N-IW и ind N-IW априорное распределения не содержит гиперпараметров и задаются через фиксированные матрицы. <https://sites.google.com/site/dimitriskorobilis/matlab>
4. Zha: в соответствии со статьей ограничения накладываются на структурную форму VAR, соответственно дается другая интерпретация для некоторых гиперпараметров
5. Sims: Недостаточно подробное описание. Нужно прочитать весь код, чтобы что-то модифицировать. <http://sims.princeton.edu/yftp/VARtools/>
6. BMR: Симуляции реализованы в C++. оценивает DSGE. Реализует TVPVAR. Отличная документация. <http://bayes.squarespace.com/bmr/>

7. MSBVAR: Симуляции реализованы на фортране и C++. Реализует также марковские BVAR с переключением. <https://cran.r-project.org/web/packages/MSBVAR/>
8. bvarr: Сопряжённое нормальное-Уишарта реализовано максимально гибко. При плохо обусловленной матрице $X'X$ используется псевдо-обратная. Код для Миннесоты и независимого нормального Уишарта является переводом кода Koops-Korobilis и потому крайне негибкий. <https://github.com/bdemeshev/bvarr>
9. Eviews: Код игнорирует тот факт, что коэффициенты были оценены байесовскими методами, прогнозы делаются точно так же, как вобычной модели. Коэффициент λ_{kron} равен по умолчанию 0,99 и не может быть изменен. Большую свободу представляет прямое задание ковариационной матрицы. Матожидания коэффициентов при первых лагах могут быть заданы только одинаковыми для всех переменных.
10. Dynare: Функция позиционируется как BVAR à la Sims. Оценка возможна только «комплектно», но можно поменять априорное распределение для априорного распределения ковариационной матрицы <http://www.dynare.org/>

Приложение 2. Таблица обозначений

Буква	Размер	Описание	Формула
p	скаляр	количество лагов	
m	скаляр	количество эндогенных переменных	
d	скаляр	количество экзогенных переменных	
k	скаляр	количество параметров в одном уравнении	$k = mp + d$
T	скаляр	количество наблюдений	
z_t	$d \times 1$	вектор экзогенных переменных (считая константу)	
y_t	$m \times 1$	вектор эндогенных переменных	$y_t = \Phi' x_t + \varepsilon_t$
x_t	$k \times 1$	вектор всех регрессоров	$x_t = [y'_{t-1} \dots y'_{t-p} z'_t]'$
ε_t	$m \times 1$	вектор случайных ошибок	$y_t = \Phi' x_t + \varepsilon_t$
Y	$T \times m$	все эндогенные переменные	$Y = [y_1, y_2, \dots, y_T]'$
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	$mT \times 1$	векторизация Y	$y = \text{vec}(Y)$
ε	$mT \times 1$	векторизация E	$\varepsilon = \text{vec}(E)$
Φ_1, \dots	$m \times m$	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 Φ_1, \dots	$\Phi = [\Phi_1 \dots \Phi_p \Phi_{const}]'$
ϕ	$km \times 1$	вектор из матрицы Φ	$\phi = \text{vec } \Phi$
$\underline{\Phi}$	$k \times m$	априорное математическое ожидание Φ	
$\underline{\phi}$	$km \times 1$	вектор из матрицы $\underline{\Phi}$	$\underline{\phi} = \text{vec } \underline{\Phi}$
$\overline{\Phi}$	$k \times m$	апостериорное математическое ожидание Φ	
$\overline{\phi}$	$km \times 1$	вектор из матрицы $\overline{\Phi}$	$\overline{\phi} = \text{vec } \overline{\Phi}$
Ξ	$km \times km$	Априорная ковариационная матрица Φ	
$\overline{\Xi}$	$km \times km$	Апостериорная ковариационная матрица Φ	$\overline{\Xi} = (\Xi^{-1} + \Sigma^{-1} \otimes X'X)^{-1}$
$\underline{\nu}$	скаляр	априорное количество степеней свободы	
$\overline{\nu}$	скаляр	апостериорное количество степеней свободы	$\overline{\nu} = T + \underline{\nu}$

$\underline{\Omega}$	$k \times k$	Матрица априорных масштабирующих коэффициентов ковариационной матрицы Φ	$\underline{\Xi} = \Sigma \otimes \underline{\Omega}$
$\overline{\Omega}$	$k \times k$	Матрица апостериорных масштабирующих коэффициентов ковариационной матрицы Φ	$\overline{\Omega} = (\underline{\Omega}^{-1} + X'X)^{-1}, \overline{\Xi} = \Sigma \otimes \overline{\Omega}$
Σ	$m \times m$	Ковариационная матрица ошибок	$\mathbb{E} \varepsilon_t \varepsilon_t' = \Sigma$

Приложение 3. Соответствие гиперпараметров в разных работах

DM15	CCM15	BGR10, BH13	KK97
λ_{tight}	λ_1	λ	$\sqrt{\pi_1}$
λ_{kron}	$\lambda_2 = 1$	$\vartheta = 1$	$\sqrt{\pi_2/\pi_1}$
λ_{lag}	1	1	0.5
λ_{const}	λ_0/λ_1	∞	$\sqrt{\pi_3/\pi_1}$
λ_{exo}	NA	NA	NA
λ_{sc}	λ_3	τ	
λ_{io}	λ_4	NA	

DM15 — данная работа, CCM15 — Carriero, Clark, and Marcellino (2015), BGR10 — Bańbura, Giannone, and Reichlin (2010), BH13 — Berg and Henzel (2013), KK97 — Kadiyala and Karlsson (1997)