

Эконометрический ликбез: динамические стохастические модели общего равновесия

Оценивание динамических стохастических моделей общего равновесия*

Анна Микушева[†]

Массачусеттский Технологический Институт, Кембридж, США

Данное эссе является обзором основных эконометрических методов оценивания динамических стохастических моделей общего равновесия, широко используемых современными центральными банками и федеральными резервными системами. Статья посвящена обсуждению основных эконометрических проблем, возникающих при статистическом оценивании и формировании статистических выводов о параметрах линеаризованных моделей общего равновесия. В статье мы рассматриваем три основных метода оценивания динамических моделей — метод наименьшего расстояния, метод максимального правдоподобия и Байесовский метод. Особое внимание уделено проблемам, которые возникают в связи со скудностью макроэкономических данных. Вопросы макроэкономического моделирования и вопросы решения стохастических динамических систем, хотя и являются исключительно интересными и сложными, находятся за рамками данного эссе.

1 Динамические стохастические модели общего равновесия

Последнее десятилетие принесло большие изменения в эмпирической практике макроэкономических исследований. Произошел качественный сдвиг от использования векторных авторегрессионных моделей (VAR) к массовому использованию динамических стохастических моделей общего равновесия (DSGE). Модели DSGE являются современной попыткой использовать разумные структурные ограничения, накладываемые экономической теорией, при оценивании взаимоотношений между различными ведущими макроиндикаторами. Большим прорывом последнего десятилетия стало развитие численных методов решения динамических систем, которые позволили решать сложные теоретические модели численными методами и тем самым сократили разрыв между теоретическими моделями, которые можно решить, и моделями, которые можно использовать при оценивании.

Основная разница между VAR моделями и DSGE моделями в том, что первые являются достаточно общими линейными моделями, накладывающими минимальные структурные ограничения на взаимодействия между входящими макроэкономическими переменными. В противовес им DSGE модели являются полностью параметризованными моделями с микроэкономическим фундаментом, то есть основанными на полной спецификации действий и поступков всех агентов в экономике (с описанием их функций полезности) а также условий равновесия на всех рынках. Условия максимизации полезности всеми агентами и условия равновесия всех рынков накладывают очень сильные ограничения на возможные взаимодействия между макропеременными и устанавливают функциональные зависимости между

*Цитировать как: Микушева, Анна (2014). «Оценивание динамических стохастических моделей общего равновесия», Квантиль, №12, стр. 1–21. Citation: Mikusheva, Anna (2014). “Estimation of dynamic stochastic general equilibrium models,” *Quantile*, No.12, pp. 1–21.

[†]Адрес: 50 Memorial Drive, Cambridge, MA-02138, USA. Электронная почта: amikushe@mit.edu

коэффициентами приведенной формы. Так например, многие линеаризованные DSGE модели могут быть записаны в форме VAR, однако, коэффициенты соответствующего VAR представления будут не свободными, а строго функционально связанными между собой параметрами. Вопрос о том, когда DSGE модель может быть записана в VAR форме с ограничениями, обсуждается в статье Fernandez-Villaverde, Rubio-Ramirez, Sargent & Watson (2007).

Более жесткая параметризация DSGE моделей имеет свои преимущества. Во-первых, дополнительные ограничения позволяют надеяться на более эффективные оценки всех структурных взаимоотношений. Во-вторых, спецификация функций полезности агентов позволяет изучать эффекты различных типов макроэкономических вмешательств и макроэкономической политики на благосостояние агентов, что невозможно сделать в VAR моделях. И в третьих, дополнительные ограничения на взаимодействия переменных дают надежду на более качественные прогнозы. Основными недостатками DSGE моделей является повышенная вероятность неверной спецификации и сложности с нелинейным оцениванием. DSGE модели легче фальсифицируемы, так как накладывают больше ограничений на данные. Некоторые статьи делающие эмпирические сравнения DSGE и VAR моделей — Chari, Kehoe & McGrattan (2005), Christiano, Eichenbaum & Vigfusson (2006) и Del Negro, Schorfheide, Smets & Wouters (2007).

Статистическое оценивание DSGE модели является многоступенчатым процессом, который включает в себя процесс моделирования (написание динамической макромоделли общего равновесия), решение модели, приведения решения к виду, подходящему для статистического оценивания, и, наконец, эконометрическое оценивание. В этом эссе мы обсудим только последний шаг, а именно, только эконометрическую составляющую этого процесса. Следующая глава дает краткое описание доэконометрических стадий.

На данный момент существует три основных подхода к оцениванию DSGE моделей, два из них классических и один Байесовский. Классические подходы используют либо обобщенный метод моментов и метод наименьшего расстояния, включая случаи, когда для вычисления целевой функции приходится использовать симуляции, либо метод максимального правдоподобия. Байесовское оценивание полагается главным образом на MCMC алгоритм, который является численным методом для получения постериорного распределения. Ниже мы обсудим преимущества различных методов, особенности их применения, а также сложности, с которыми сталкиваются макроэконометристы.

В последнее время было выпущено немало книг и обзорных статей, которые позволяют ознакомиться с теорией и практикой оценивания DSGE моделей. Я бы хотела обратить внимание читателей на книги Canova (2007) и DeJong & Dave (2007), а также лекции Ларри Кристиано,¹ которые являются достаточно полными и интересными источниками по данной теме. Проблеме оценивания DSGE моделей также посвящена обзорная лекция Джима Стока на летней школе NBER в 2008.² Другим интересным источником является лекция Франка Шорфалда (Schorfheide, 2010) на мировом конгрессе, посвященная нерешенным эконометрическим проблемам DSGE моделей.

Для практического использования DSGE моделей я рекомендую ознакомиться с бесплатной софтверной программой Dynare³, которая является надстройкой в пакете MatLab. Программа Dynare позволяет автоматически линеаризовать и решать стандартные DSGE модели, а также стандартно реализовать оценивание методом максимального правдоподобия и Байесовским методом. Dynare поддерживается большим числом академических ученых, а также исследователями, работающими в центральных банках и резервных системах различных стран. Сообщество Dynare поддерживает активный форум, где можно получить ответы на практические вопросы, связанные с оцениванием DSGE моделей.

¹Доступны на веб-сайте faculty.wcas.northwestern.edu/~lchrist/

²www.nber.org/minicourse/_2008.html

³www.dynare.org

2 Стадии, предшествующие эконометрическому анализу

2.1 Моделирование и линеаризация модели

Типичная задача макроэкономического моделирования начинается с описания всех агентов в экономике и их функций полезности или оптимизационных задач. Такими агентами могут быть домохозяйства, которые потребляют конечный продукт, предлагают свои услуги на рынке труда, инвестируют свои сбережения, и, например, фирмы, которые нанимают рабочих, инвестируют в капитал и продают конечную продукцию. Необходимо также описать все бюджетные ограничения, накладываемые на агентов. Делаются также предположения о различных рынках и их несовершенствах. Модель может содержать (и обычно содержит) стохастическую составляющую, которая заключается в описании всех структурных шоков в экономике и каким образом они воздействуют на различные переменные. Построение теоретической макроэкономической модели является важным шагом и полностью находится в ведении макроэкономики. В данной статье мы откажемся от обсуждения различных моделей и их предположений.

На этом этапе наша задача заключается в получении системы уравнений, состоящей из условий первого порядка для оптимизационных задач агентов, уравнений балансов, условий равновесия на всех рынках и уравнений, описывающих динамику всех шоков в экономике. Данная система уравнений описывает динамические зависимости между всеми макропеременными. К сожалению, в подавляющем большинстве случаев мы не можем разрешить данную систему уравнений из-за следующих сложностей: во-первых, многие из уравнений нелинейны, во-вторых, многие уравнения включают условные математические ожидания будущих значений переменных.

Для разрешения первой сложности большинство статей решает не изначальную модель, а ее приближение, полученное в результате лог-линеаризации уравнений вокруг стационарного состояния (steady state). Впрочем, в настоящее время появляются исследования использующие также нелинейные модели; примером статьи, решающей нелинейную DSGE модель, может служить Fernandez-Villaverde & Rubio-Ramirez (2005).

В данном эссе в качестве примера мы будем использовать лог-линеаризованную DSGE модель, предложенную в статье Galí, Lopez-Salido & Valles (2003). Это одна из простейших моделей, на которой построены более сложные современные DSGE модели, включающие в себя большее число макро переменных и позволяющие оценивать большее число структурных параметров.

Пример 1. Предположим, что переменная x_t обозначает логарифм совокупной продукции (и совокупного потребления) в экономике в момент t , переменная π_t обозначает логарифм инфляции, а переменная r_t — реальный процент. Эволюция этих трех макропеременных описывается следующей лог-линеаризованной системой уравнений:

$$\begin{cases} -(r_t - \mathbb{E}_t \pi_{t+1} - \rho \Delta a_t) + \mathbb{E}_t x_{t+1} - x_t = 0, \\ b \mathbb{E}_t \pi_{t+1} + \kappa x_t - \pi_t + \varepsilon_t = 0, \\ \lambda r_{t-1} + (1 - \lambda) \varphi_\pi \pi_t + (1 - \lambda) \varphi_x x_t + u_t = r_t. \end{cases} \quad (1)$$

Первое уравнение в этой системе является линеаризованным уравнением Эйлера, или условием первого порядка в задачи оптимизации полезности потребителем. Второе уравнение в системе (1) — это так называемая кривая Филлипса (Phillips curve). Это уравнение получается как решение модели Калво (Calvo model), в которой фирмы не могут свободно устанавливать цены на свою продукцию каждый период, а меняют цены только в экзогенно заданные случайные моменты времени. Третье уравнение имеет название «правило Тейлора» (Taylor Rule) и является характеристикой монетарной политики, проводимой Федеральной резервной системой США. Символ \mathbb{E}_t обозначает условное математическое ожидание,

использующее всю информацию, известную в момент t ; часто такую парадигму называют «рациональными ожиданиями» (rational expectations).

Помимо эндогенно меняющихся макропеременных, чья эволюция описана в системе (1), в систему также входят экзогенные переменные, являющиеся шоками. Таковыми в данной системе являются технологический шок Δa_t , монетарный шок u_t и шок потребления ε_t . Дополнительно к системе (1) мы должны описать динамическую эволюцию шоков. Предположим, что шоки удовлетворяют следующим допущениям:

$$\Delta a_t = \rho \Delta a_{t-1} + \varepsilon_{a,t}; \quad u_t = \delta u_{t-1} + \varepsilon_{u,t}; \quad (\varepsilon_t, \varepsilon_{a,t}, \varepsilon_{u,t})' \sim i.i.d. N(0, \Sigma); \quad \Sigma = \text{diag}(\sigma^2, \sigma_a^2, \sigma_u^2).$$

В рассматриваемой модели имеется 10 параметров, обычно называемых структурными параметрами. Часть из них имеет экономическое значение, такие например, как фактор дисконтирования (discount rate) b , параметер Калво κ , и параметры правила Тейлора φ_x, φ_π и λ . Остальные параметры, $\delta, \rho, \sigma, \sigma_a, \sigma_u$, статистически описывают эволюцию экзогенных переменных. Обозначим набор всех неизвестных структурных параметров β . Задача эконометрического оценивания и тестирования состоит в том, чтобы, имея данные $\mathcal{Y}_T = \{x_t, \pi_t, r_t, t = 1, \dots, T\}$, оценить, построить доверительные интервалы и протестировать интересные нас гипотезы о структурных параметрах модели β .

2.2 Решение системы рациональных ожиданий

Модель (1) является линейной, но при этом содержит условные математические ожидания будущих значений переменных. Упомянутые условные ожидания ненаблюдаемы, и с этой точки зрения модель (1) неудобна для оценивания. Следующим шагом является решение линейной системы рациональных ожиданий. В общем виде модель, стандартно получаемую в результате лог-линеаризации DSGE модели, можно записать как

$$\Gamma_0(\beta)Z_t = \Gamma_1(\beta)\mathbb{E}_t Z_{t+1} + \Gamma_2(\beta)Z_{t-1} + \Gamma_3(\beta)\eta_t,$$

где Z_t содержит все эндогенные переменные (как наблюдаемые, так и не наблюдаемые), а η_t содержит все экзогенные шоки. Матрицы $\Gamma_i(\beta)$ являются известными функциями структурных параметров β .

Решить линейную модель рациональных ожиданий означает записать ее в виде стандартной линейной модели пространства состояний (state-space representation), состоящей из двух уравнений

$$\begin{cases} Y_t = \Psi_0(\beta) + \Psi_1(\beta)t + \Psi_2(\beta)S_t, \\ S_t = \Phi_1(\beta)S_{t-1} + \Phi_2(\beta)\eta_t. \end{cases} \quad (2)$$

Здесь S_t — это вектор переменных, описывающих состояние динамической системы (state variable), а Y_t — набор переменных, наблюдаемых эконометристом в момент t . Первое из уравнений системы (2) называется уравнением измерения (measurement equation), а второе — уравнением перехода (transition equation). В случае Примера 1, $S_t = (\pi_t, x_t, r_t, \Delta a_t, u_t, \varepsilon_t)$.

Существует целый ряд различных алгоритмов, позволяющих численно решать линейные системы с рациональными ожиданиями; наиболее известны из них методы, предложенные в статьях Blanchard & Kahn (1980), Anderson & Moore (1985), Klein (2000) и Sims (2001). Большинство из вышеупомянутых методов запрограммированы и являются частью стандартного пакета Dynare.⁴ Следует сказать, что решается система численно, а не аналитически. Это значит, что для каждого значения параметра β можно численно посчитать соответствующие матрицы $\Gamma_i(\beta)$ и решить систему рациональных ожиданий, т.е. численно найти значение матриц $\Psi_i(\beta)$ и $\Phi_i(\beta)$.

⁴www.dynare.org

Проблема решения системы рациональных ожиданий очень интересна и сложна. Для некоторых значений параметров система не имеет решения, а в некоторых случаях решений несколько, что соответствует нескольким потенциальным равновесиям. Стандартная на данный момент методология состоит в ограничении области определения параметра только теми значениями, при которых существует только единственное устойчивое решение линейной системы ожиданий. В данной статье мы не будем обсуждать как решать системы рациональных ожиданий, а предположим, что DSGE модель уже линеаризована и решена в виде (2).

3 Методы оценивания: метод наименьшего расстояния

Наиболее ранним является классический метод оценивания DSGE моделей, основанный на методе наименьшего расстояния и впервые в данном контексте примененный в статье Rotemberg & Woodford (1997); примеры более современного использования метода можно найти в статьях Voivin & Giannoni (2006), Christiano, Eichenbaum & Evans (2005) и Schorfheide (2000). Альтернативными названиями в эконометрической литературе являются английские термины *matching moments* и *indirect inference*.

Метод состоит в минимизации расстояния между некоторыми эмпирически значимыми характеристиками процесса оцененными на данных и полученными из теоретической модели. Чаще всего такими характеристиками являются функция импульсного отклика (*impulse response function*), автоковариационная и спектральная функции процесса.

Пусть β_0 является истинным (неизвестным) значением параметра β , которое мы пытаемся оценить. Предположим, что характеристика процесса, которую мы будем приближать, скажем $\delta = \delta(\beta)$, зависит от структурных параметров модели β . Пусть $\delta_0 = \delta(\beta_0)$ является истинным значением характеристики для наблюдаемого процесса, а $\hat{\delta} = \hat{\delta}(\mathcal{Y}_T)$ является состоятельной оценкой этой характеристики, основанной на доступных данных \mathcal{Y}_T . Предположим, что данная оценка удовлетворяет условию

$$\sqrt{T}(\hat{\delta} - \delta_0) \Rightarrow N(0, \Omega).$$

Оценка $\hat{\beta}$ по методу наименьшего расстояния определяется как

$$\hat{\beta} = \arg \min_{\beta} (\hat{\delta} - \delta(\beta))' \Xi (\hat{\delta} - \delta(\beta)),$$

где Ξ — положительно определенная симметричная матрица. В классических предположениях, что структурный параметр β идентифицируем, значение β_0 является внутренней точкой пространства параметров, что матрица первых производных $R = \partial\delta(\beta_0)/\partial\beta$ имеет полный ранг, равный размерности параметра β , и что матрица $R'\Xi R$ имеет обратную, классический результат гарантирует, что оценка $\hat{\beta}$ состоятельна и

$$\sqrt{T}(\hat{\beta} - \beta_0) \Rightarrow N(0, (R'\Xi R)^{-1} R'\Xi \Omega \Xi R (R'\Xi R)^{-1}).$$

Оптимальным выбором матрицы взвешивания разных характеристик является $\Xi = \Omega^{-1}$, при котором асимптотическая матрица ковариаций равна $(R'\Omega^{-1}R)^{-1}$.

Целевая функция, используемая для оценивания, может также использоваться для тестирования спецификации в случае, когда размерность характеристики δ выше размерности структурного параметра β . В частности в предположении правильной спецификации мы можем показать, что статистика $\min_{\beta} (\hat{\delta} - \delta(\beta))' \Omega^{-1} (\hat{\delta} - \delta(\beta))$ асимптотически сходится к χ^2 -распределению с числом степеней свободы, равным разности между размерностями параметров δ и β .

Метод наименьшего расстояния хорошо изучен в эконометрике. Здесь мы обсудим только особенности его применения для оценивания DSGE моделей. Метод наименьшего расстояния использует для оценивания только часть возможных статистических связей между переменными, вытекающих из модели (только те связи, которые описаны характеристиками δ , используемыми для оценивания). В этом смысле метод наименьшего расстояния является процедурой с ограниченной информацией (limited information procedure). Из этого факта проистекают как преимущества, так и недостатки. Основным недостатком является потеря эффективности по сравнению с методом максимального правдоподобия, который использует всю совокупность статистических связей наиболее эффективным образом. Принимая во внимание, что даже более эффективный метод максимального правдоподобия страдает от проблемы слабой идентификации, которую мы обсудим в деталях в следующей секции, метод наименьшего расстояния еще более уязвим к этой проблеме. С другой стороны, для задания функции правдоподобия необходимы предположения о том, как распределены шоки в системе, такие как предположения о гауссовости в Примере 1. Для метода наименьшего расстояния предположения о распределениях не нужны, и таким образом, такой метод оценивания более робастен к неверной спецификации модели.

Идея оценивания путем минимизации расстояния между важными характеристиками наблюдаемыми в данных и предсказанными моделью была предложена в статье Kydland & Prescott (1982); в частности и за это авторы были награждены Нобелевской премией по экономике. Kydland & Prescott (1982) были против методов, основанных на правдоподобии из соображений что в случае максимального правдоподобия наше понимание того, хорошо ли описывает модель данные, основывается на статистических критериях, которые не всегда совпадают с практическими и эмпирически значимыми целевыми функциями. Например, если некая модель построена для объяснения эффектов монетарной политики на макроиндикаторы, то разумно использовать в качестве теста спецификации целевую функцию, измеряющую расстояние между функциями импульсного отклика оцененными в данных и предсказанными моделью. Этот тест может противоречить тесту, основанному на функции правдоподобия, так как на последнюю могут влиять много факторов, плохо объясняемых моделью, большая часть из которых нам может быть неинтересна.

3.1 Оценивание характеристики δ на данных: метод

Заметим, что если DSGE модель решена в виде системы уравнений (2), то вычисление для любого значения параметра β теоретических (предписанных моделью) автоковариационной и спектральной функций для наблюдаемых переменных, а также функций импульсного отклика является относительно простой задачей. Действительно, например, чтобы получить теоретическое значение импульсного отклика Y_{t+h} на единичный шок η_t при значении структурного параметра β , достаточно положить $S_{t-1} = 0$, $\eta_{t+s} = 0$ для всех $s > 0$ и сравнить значения Y_{t+h} , получаемые из системы (2) при $\eta_t = 1$ и $\eta_t = 0$.

Основная сложность метода состоит в оценивании соответствующих характеристик δ на данных \mathcal{U}_T . Остановимся на наиболее частом случае, когда характеристика δ , на которой основано оценивание, есть функция импульсного отклика на структурные шоки в экономике. В таком случае оценка $\hat{\delta}$ обычно получается как результат оценивания структурной VAR модели. Оценивание структурной VAR модели состоит из двух шагов. На первом шаге оценивается приведенная VAR модель, т.е. производится регрессия по методу наименьших квадратов переменных $Y_t = (x_t, \pi_t, r_t)$ на их лаги:

$$Y_t = A_0 + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + e_t. \quad (3)$$

Используя данную регрессию, мы можем получить функцию импульсного отклика для при-

веденной модели:

$$\theta_{i,h} = \frac{\partial Y_{t+h}}{\partial e_{t,i}} = A_1 \theta_{i,h-1} + \dots + A_p \theta_{i,h-p}.$$

В данном случае мы используем предположение что $\theta_{i,h} = 0$ для всех $h < 0$ и $\theta_{i,h} = 1$ для $h = 0$. Пусть Θ_h обозначает матрицу, содержащую все импульсные отклики всех наблюдаемых переменных Y_{t+h} на все шоки e_t приведенной модели (на горизонте h). Проблема заключается в том, что шоки e_t не являются структурными шоками. Вторым шагом в оценивании структурной VAR модели является оценивание матрицы C , связывающей шоки приведенной модели со структурными шоками $\eta_t = Ce_t$. Функция импульсного отклика структурной модели δ_h связана с функцией импульсного отклика приведенной модели θ_h линейно: $\delta_h = \theta_h C^{-1}$.

Для оценивания импульсного отклика на структурные шоки требуется наложить дополнительные ограничения, позволяющие идентифицировать матрицу C . Данные ограничения должны быть согласованы со структурной DSGE моделью. Как оказывается, существует достаточная свобода в выборе условий идентификации, и конечный результат, а именно оценка функции импульсного отклика, очень чувствителен к этому выбору. Для ознакомления с теорией структурных VAR моделей я советую читателю ознакомиться со статьей Нобелевского лауреата Sims (1980). Пример наложения краткосрочных идентификационных условий можно найти в King, Plosser, Stock & Watson (1991), а пример долгосрочных идентификационных условий — в Blanchard & Quah (1989).

3.2 Оценивание характеристики δ на данных: проблемы

Существует несколько теоретических и практических сложностей при использовании метода наименьшего расстояния путем мэтчинга функции импульсного отклика, что делает этот метод проблематичным на практике.

Первая сложность заключается в том, что стандартная DSGE модель, записанная в виде системы (2), вообще говоря не является VAR моделью конечного порядка, она является моделью типа VARMA, т.е. у нее есть компонента типа скользящего среднего. Таким образом, стандартная теоретическая DSGE модель может быть записана как VAR модель бесконечного порядка (включающая в себя бесконечное число лагов). В этом смысле оценивание структурной VAR модели, описанное выше, является оцениванием неверно специфицированной модели. Как показали Chari, Kehoe & McGrattan (2005), погрешность вносимая неверной спецификацией динамики процесса, может быть огромной, и оценка импульсного отклика, полученная в результате оценивания структурной VAR модели, будет до неузнаваемости отличаться от истинной оцениваемой функции импульсного отклика.

Вторая сложность состоит в том что оценки функции импульсного отклика являются смещенными в конечной выборке, даже если абстрагироваться от факта неверной спецификации VAR модели. Смещенность оценки имеет несколько истоков, один из них — смещенность в конечной выборке оценки VAR коэффициентов редуцированной VAR модели. Оценивание методом наименьших квадратов системы временных рядов типа (3) обычно дает смещенную оценку, впрочем смещение для оценок VAR коэффициентов не очень велико. Однако, если хотя бы один из характеристических корней близок к единичному корню, то это небольшое смещение оценок коэффициентов транслируется в существенное смещение оценки функции импульсного отклика (см. Mikusheva, 2012). Другой источник смещения оценки функции импульсного отклика заключается в том, что импульсный отклик является очень сильно нелинейной функцией VAR коэффициентов, а любое нелинейное преобразование несмещенных оценок неминуемо влечет появление смещения в конечных выборках. Попыткой уменьшить влияние смещения оценки импульсного отклика на структурную оценку β , получаемую ме-

тодом наименьшего расстояния, является коррекция Когли–Нейсона, предложенная в статье Cogley & Nason (1995).

Наконец, дополнительной практической сложностью является выбор характеристик процесса δ , на основе которых строится оценка наименьшего расстояния. Предположим, что мы решили минимизировать расстояние между теоретической функцией импульсного отклика и ее оценкой в данных. Возникает вопрос, какие именно импульсные отклики использовать: отклики каких переменных на какие шоки и на каком горизонте. Использование каждого дополнительного импульсного отклика может нести дополнительную информацию о структурном коэффициенте β , но в то же время это вносит дополнительную неопределенность так как оценка этого отклика в данных имеет ошибку. Оптимальный выбор характеристик для оценки по методу наименьшего расстояния должен балансировать эти два фактора. В статье Hall, Inoue, Nason & Rossi (2012) предложен информационный критерий, который помогает сделать данный выбор.

4 Методы оценивания: Метод максимального правдоподобия

Второй часто используемый классический метод оценивания DSGE моделей — метод максимального правдоподобия. Примеры использования метода максимального правдоподобия для оценивания DSGE моделей можно найти в статьях Ingram, Kocherlakota & Savin (1994), Ireland (2004) и Lindé (2005).

Метод состоит в нахождении функции правдоподобия $\mathcal{L}(\mathcal{Y}_T|\beta)$ и последующей ее максимизации по отношению к параметру β . Значение $\hat{\beta}_{MLE}$ параметра β , в котором достигается максимум, называется оценкой максимального правдоподобия. При некоторых условиях регулярности модели оценка максимального правдоподобия является состоятельной и асимптотически нормальной:

$$\sqrt{T}(\hat{\beta}_{MLE} - \beta_0) \Rightarrow N(0, \mathcal{J}^{-1}), \quad (4)$$

здесь \mathcal{J} обозначает так называемую информацию Фишера. В регулярных случаях состоятельной оценкой информации Фишера является наблюдаемая информация, равная отрицательной второй производной логарифма функции правдоподобия:

$$\hat{J} = -\frac{1}{T} \frac{\partial^2}{\partial \beta \partial \beta'} \log \mathcal{L}(\mathcal{Y}_T|\hat{\beta}_{MLE}). \quad (5)$$

Теория максимального правдоподобия также позволяет тестировать различные гипотезы, используя один из трех классических тестов — тест Вальда, тест основанный на мультипликаторе Лагранжа (Lagrange multiplier test или score test) и тест отношения правдоподобия (likelihood ratio test).

Условия регулярности, при которых верны указанные выше асимптотические утверждения в случае временных рядов, менее тривиальны, чем классическая теория максимального правдоподобия, верная для независимых одинаково распределенных наблюдений. Хорошим учебником по данной теме является Hall & Heyde (1980).

Если DSGE модель линеаризована и решена в виде модели пространства состояний (2), то функция правдоподобия может быть получена при помощи применения фильтра Калмана. Хорошим учебником, освещающим теорию фильтрации является Harvey (1989). Фильтр Калмана в применении к линеаризованной DSGE модели реализован в программе Dynare. В случае, когда модель нелинейна, но решена в виде модели в пространстве состояний, проблема получения функции правдоподобия гораздо более сложна, обзор возможных методов может быть найден в статье Цыплакова (2011). В частности, один из часто употребляемых методов для получения функции правдоподобия — это использование так называемых корпускулярных фильтров (particle filters).

4.1 Проблемы метода максимального правдоподобия

Теперь мы обсудим особенности и проблемы применения метода максимального правдоподобия к оцениванию DSGE моделей.

Первой проблемой, с которой чаще всего сталкивается прикладной исследователь, является проблема численного поиска максимума функции правдоподобия. Большинство известных и решенных на данный момент DSGE моделей обладают крайне нерегулярной функцией правдоподобия. Типичная функция правдоподобия для DSGE модели имеет множество локальных максимумов, участки типа «плато», где функция почти не меняется, а также области резкого изменения производной («обрывы»). Все это, а также тот факт, что типичная DSGE модель обычно имеет большое число структурных параметров (наша простейшая модель в Примере 1 имеет 10 неизвестных параметров), осложняет численную задачу поиска оценки максимального правдоподобия.

В большинстве случаев использование численных оптимизаторов, основанных на градиентных методах, не позволяют найти максимум функции правдоподобия. Практические рекомендации включают в себя предложение использовать неградиентные методы (например, симплекс-метод) и использовать оптимизационный алгоритм несколько раз, стартуя из разных начальных точек, в надежде найти глобальный (а не локальный) максимум.

Другой частой проблемой является то, что исследователь редко догматично верит в модель, и скорее рассматривает ее как аппроксимацию к действительности. В этом смысле оцениваемая модель зачастую неверна буквально, т.е. неверно специфицирована. Таким образом, метод максимального правдоподобия более чувствителен к проблемам неверной спецификации по сравнению с обсуждаемым выше методом наименьшего расстояния. Метод максимального правдоподобия использует абсолютно все связи, налагаемые моделью на наблюдаемые переменные, а также требует спецификацию распределения всех шоков (гауссовость в Примере 1). В этом смысле правильнее рассматривать функцию правдоподобия не как истинную, а как функцию квази-правдоподобия. Эта идея была введена в статье White (1982). При некоторых условиях регулярности оценка максимального правдоподобия является состоятельной оценкой псевдоистинного параметра β^* . К сожалению, в сложных моделях связь между псевдоистинным параметром β^* и изначальным параметром β неочевидна, что усложняет интерпретацию результатов такого оценивания.

White (1982) также указывает, что если рассматривать оценку максимального правдоподобия с точки зрения потенциально неверной спецификации, то следует использовать стандартные ошибки, робастные к неверной спецификации. Робастная формула для оценки асимптотической матрицы вариаций \mathcal{J}^{-1} в утверждении типа (4) дана формулой $\hat{J}^{-1}\hat{I}\hat{J}^{-1}$, где

$$\hat{I} = -\frac{1}{T} \sum_{t=1}^T \left(\frac{\partial}{\partial \beta} \log f(Y_t | \mathcal{Y}_{t-1}, \hat{\beta}_{MLE}) \right) \left(\frac{\partial}{\partial \beta} \log f(Y_t | \mathcal{Y}_{t-1}, \hat{\beta}_{MLE}) \right)', \quad (6)$$

а \hat{J} задана формулой (5).

Дополнительной проблемой оценивания DSGE моделей методом максимального правдоподобия является проблема стохастической сингулярности. Если представить себе DSGE модель, в которой число случайных ненаблюдаемых шоков меньше числа наблюдаемых переменных, то часто одним выводом модели может быть наличие некой детерминистической (то есть выполняемой во всех реализациях данных) связи между наблюдаемыми переменными. Статья Ingram, Kocherlakota & Savin (1994) дает очень простой пример такой ситуации. Однако, подобная детерминистическая связь между наблюдаемыми переменными практически никогда не верна в реальных данных, и как результат, модель является неверно специфицированной и будет отвергнута в большой выборке. Из данного рассуждения следует частая практическая рекомендация (которая сама по себе все же не гарантирует отсутствие

стохастической сингулярности), что число шоков в DSGE модели должно быть не меньше числа наблюдаемых переменных.

Существует два подхода к разрешению проблемы сингулярности. Один из них состоит в введении в систему дополнительных системных шоков; этот подход предложен в статье Ingram, Kocherlakota & Savin (1994). Другим подходом является введение ошибок измерения. Например, предположим, что некая модель устанавливает детерминистическую зависимость вида $c_t = \alpha y_t + \beta c_{t-1}$, где c_t и y_t — наблюдаемые переменные потребления и производства, а α и β — параметры. В реализации данных, очевидно, такая зависимость не может быть удовлетворена при всех t ни для каких фиксированных значений α и β . Введение ошибок измерения состоит в том, чтобы не менять модель (и ее выводы), а предположить что на самом деле настоящие данные о потреблении c_t недоступны, вместо этого эконометрист наблюдает переменную c_t^* , которая является измерением c_t с ошибкой $c_t^* = c_t + e_t$.

5 Эконометрические проблемы: слабая идентификация

В данном разделе мы поговорим о вопросах идентификации и слабой идентификации, которые очень часто возникают при оценивании DSGE моделей, и создают проблему для классического оценивания (так называемого подхода фреквентиста). Вопросы слабой идентификации являются на данный момент активной областью эконометрических исследований, и решения для многих из них пока неизвестны.

5.1 Идентификация

Параметр β считается (глобально) идентифицируемым в точке $\beta = \beta_0$, если для любого другого значения параметра $\beta^* \neq \beta_0$ распределение данных \mathcal{Y}_T , сгенерированных из модели с параметром β_0 , отличается от распределения данных, сгенерированных из модели, где параметр равен β^* . Эконометрическая литература, посвященная вопросам идентификации, очень быстро растет в последние годы. Однако, большая ее часть посвящена кросс-секционным данным, в то время как идентификация в моделях временных рядов имеет свои особенности.

Примером сложности идентификации в моделях временных рядов может служить, например, идентификация простейшего процесса скользящего среднего.

Пример 2. Рассмотрим процесс $x_t = e_t + \theta e_{t-1}$, где e_t являются независимыми одинаково распределенными гауссовскими случайными величинами $e_t \sim N(0, \sigma^2)$. Предположим, что эконометрист наблюдает данные $\mathcal{X}_T = \{x_1, \dots, x_T\}$. Несложно показать, что вектор \mathcal{X}_T является гауссовским и параметры $\theta = \delta, \sigma^2 = \sigma_1^2$ и $\theta = 1/\delta, \sigma^2 = \delta^2 \sigma_1^2$ дают одинаковое распределение вектора \mathcal{X}_T . Таким образом, без дополнительных ограничений параметры процесса скользящего среднего являются неидентифицируемыми.

Другим примером исчезновения идентификации может служить пример сокращения корней авторегрессионной части и скользящего среднего.

Пример 3. Предположим, что нас интересует процесс ARMA(1,1): $x_t = \alpha x_{t-1} + e_t - \beta e_{t-1}$, где $e_t \sim N(0, 1)$ — независимые и одинаково распределенные шоки, α и β — неизвестные параметры. Мы наблюдаем вектор $\mathcal{X}_T = \{x_1, \dots, x_T\}$. Легко показать, что если $\alpha = \beta$, то $x_t = e_t$, и параметры неидентифицируемы.

Оба примера выявляют сложности с идентификацией динамики процесса в моделях временных рядов. Если DSGE модель линейна и разрешена в виде (2), то наблюдаемый процесс является процессом типа VARMA с ограничениями на коэффициенты. Статья Komunjer &

Ng (2011) устанавливает критерии идентификации таких моделей. Альтернативный подход представлен в статье Qu & Tkachenko (2012), которые формулируют условия идентификации линейных DSGE моделей в терминах спектральной плотности. Iskrev (2008, 2010) связывает условия идентификации с невырожденностью матрицы информации Фишера и указывает на простой способ вычисления ее для логлинеаризованных DSGE моделей.

Следует отметить, что все эти подходы устанавливают условия для локальной (а не глобальной) идентификации. Под локальной идентификацией понимается следующее свойство: модель локально идентифицируема при значении параметра β_0 , если существует открытая окрестность параметра β_0 такая, что модель, ограниченная на данную окрестность, является идентифицируемой. Так, модель в Примере 2 является локально идентифицируемой, так как после ограничения пространства параметров до $|\theta| < 1$ она становится идентифицируемой. В то же время модель в Примере 3 является ни глобально, ни локально неидентифицируемой. Условия для глобальной идентификации DSGE моделей на данный момент неизвестны.

Как говорилось ранее, большинство DSGE моделей являются нелинейными по входящим макропеременным, что осложняет их решение и оценивание. Большая часть DSGE литературы посвящена решению и оцениванию не изначальных моделей, а их линеаризованных версий. Однако вполне возможно, что в процессе линеаризации идентифицируемая модель может потерять идентификацию. Это мотивирует поиск и развитие методов решения и оценивания нелинейных моделей.

5.2 Слабая идентификация

Одна из важных и нерешенных на данный момент проблем оценивания DSGE моделей является проблема слабой идентификации. Термин «слабая идентификация» хотя и применяется часто в эконометрической литературе, не имеет до сей поры строгого и четкого определения. Обычно предполагается, что модель является идентифицированной в смысле, описанном в предыдущем подразделе, однако это само по себе не гарантирует, что оценка неизвестных параметров будет хороша, или что классические тесты будут иметь правильный уровень значимости. Термин «слабая идентификация» чаще всего описывает ситуацию, когда распределения в конечных выборках классических оценок и тестовых статистик далеки от распределений, предписываемых классической асимптотической теорией, часто вследствие того, что данные не информативны в достаточной степени о неизвестном параметре. Иными словами, термин «слабая идентификация» зачастую используется как замена утверждению «асимптотика не работает».

Хорошо изученным примером слабой идентификации является случай так называемых слабых инструментов, когда рассматривается регрессия, оцениваемая методом инструментальных переменных, но инструменты очень слабо коррелированы с эндогенным регрессором. В таком случае стандартная классическая асимптотика дает очень плохое приближение распределениям в конечных выборках. Так, например, распределение двухшаговой оценки наименьших квадратов в модели со слабыми инструментами достаточно сильно отличается от гауссовского распределения, оно сильно смещено в сторону несостоятельной оценки наименьших квадратов и имеет более тяжелые хвосты. Распределения классических статистик (t - и F -статистик) также существенно отличаются от своих асимптотических пределов. Использование классических методов инструментальной регрессии в этом случае ведет к неверным статистическим выводам. Вместо этого необходимо использовать методы, робастные к слабым инструментам.

Существует несколько обобщений феномена слабых инструментов на нелинейные модели. Один подход выражен в статье Stock & Wright (2000), которая вводит понятие слабо идентифицированного обобщенного метода моментов (ОММ). В упомянутой статье функция, задающая моментное условие в ОММ, почти не меняется при изменении параметра.

Асимптотически это моделируется так, что функция зависит от размера выборки, в предположении, что в пределе производная моментной функции равна нулю. При этом неизвестный параметр предполагается идентифицируемым в конечных выборках, но информация о нем не накапливается асимптотически. Как результат, ОММ оценка этого параметра не является состоятельной, ее распределение нестандартно, и классические тесты имеют распределения, отличающиеся от классических асимптотических пределов.

Другим обобщением понятия слабой идентификации является подход, предложенный в статье Andrews & Cheng (2012). Этот подход предполагает, что модель идентифицируема за исключением одного специального значения некоторого параметра, при котором другое значение параметра теряет свою идентификацию. Термином слабая идентификация Andrews & Cheng (2012) называют ситуацию, когда настоящее значение параметра приближается к точке потери идентификации. Как пример такой ситуации рассмотрим процесс, описанный в Примере 3. Мы видели что если $\alpha = \beta$, то ни α ни β не идентифицируемы. Предположим вместо этого, что $\beta = \alpha + \delta$, и назовем новой параметризацией пару (α, δ) . Тогда если $\delta = 0$, то параметр α не идентифицируем. Ситуация, когда $\delta \neq 0$, но при этом параметр δ очень близок к нулю, описывает ситуацию, где данные обладают очень малой информацией о параметре α . Асимптотикой слабой идентификации считается набор моделей, где параметр $\delta_n = \text{const}/\sqrt{n}$ меняется с величиной выборки n и асимптотически сходится к нулю. В данной асимптотике параметр α не может быть оценен состоятельно, стандартные оценки (например оценка максимального правдоподобия) не являются асимптотически нормальными, и классические статистики имеют нестандартные предельные распределения.

Общее для этих двух подходов то, что асимптотическая теория дает очень плохое приближение для распределений оценок и статистик в конечных выборках из-за недостаточности информации в данных относительно какого-то параметра или группы параметров. Похожий феномен присутствует в DSGE моделях. Впервые это было замечено в статье Canova & Sala (2009), которые обратили внимание, что типичная функция правдоподобия для типичной DSGE модели имеет области постоянства, где она очень слабо меняется при существенном изменении параметров. В такой области постоянства может находиться несколько локальных (а также и глобальный) максимумов. Эта особенность DSGE моделей очень хорошо известна прикладным макроэкономистам. Так например Fernández-Villaverde (2010) характеризует проблему следующим образом: (1) поверхность функции правдоподобия нерегулярна, обладает многочисленными локальными экстремумами и имеет области постоянства, что затрудняет поиск численного значения оценки; (2) даже когда оценка найдена, возникают проблемы по поиску стандартной ошибки, характеризующей точность данной оценки, так как стандартные формулы дают странные значения; (3) получаемые доверительные интервалы имеют уровень доверия, сильно отличающийся от декларируемого.

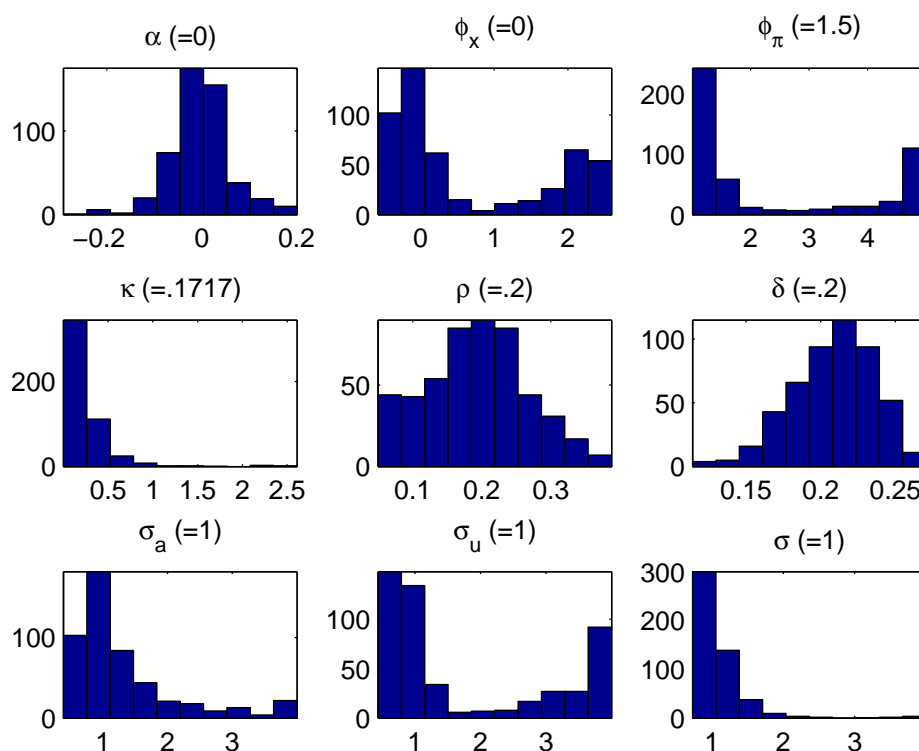
Чтобы продемонстрировать возникающие проблемы, воспроизведем эксперимент Монте-Карло, описанный в статье Andrews & Mikusheva (2011).

Пример 1 (продолжение). Рассмотрим простейшую логлинеаризованную DSGE модель, описанную в Примере 1. Предположим, что настоящие значения экономически значимых параметров равны следующим величинам (широко согласующимся с макроэкономическим консенсусом): коэффициент дисконтирования $b = 0,99$, Калво параметр $\kappa = 0,1717$, параметры правила Тейлора $\varphi_x = 0$, $\varphi_\pi = 1,5$ и $\lambda = 0$. Положим значения параметров статистического описания равным следующим значениям: $\rho = 0,2$ и $\delta = 0,2$ (чтобы внести автокорреляцию шоков, при этом сохранив стационарность) и $\sigma = \sigma_a = \sigma_u = 1$.

Из описанной модели мы генерируем выборки размера 200 периодов, что является типичным размером для макроэкономических данных. На полученных выборках мы будем производить оценивание по методу максимального правдоподобия и тестировать различные гипотезы, которые выполняются в популяции. В данном Монте-Карло эксперименте мы бу-

дем считать значение коэффициента дисконтирования b известным, так как в практике он очень часто оценивается вне модели и калибруется. Оставшиеся 9 неизвестных параметров мы назовем вектором β . Оценивание по методу максимального правдоподобия происходит через численное решение модели, нахождение функции правдоподобия фильтром Кальмана и максимизацией функции правдоподобия. Эксперименты производятся с использованием пакета Dynare.

Рис. 1: Гистограмма распределения оценок по методу максимального правдоподобия. Истинные значения параметров приведены в скобках над каждым графиком. Размер выборки $T = 200$. Число симуляций 500.

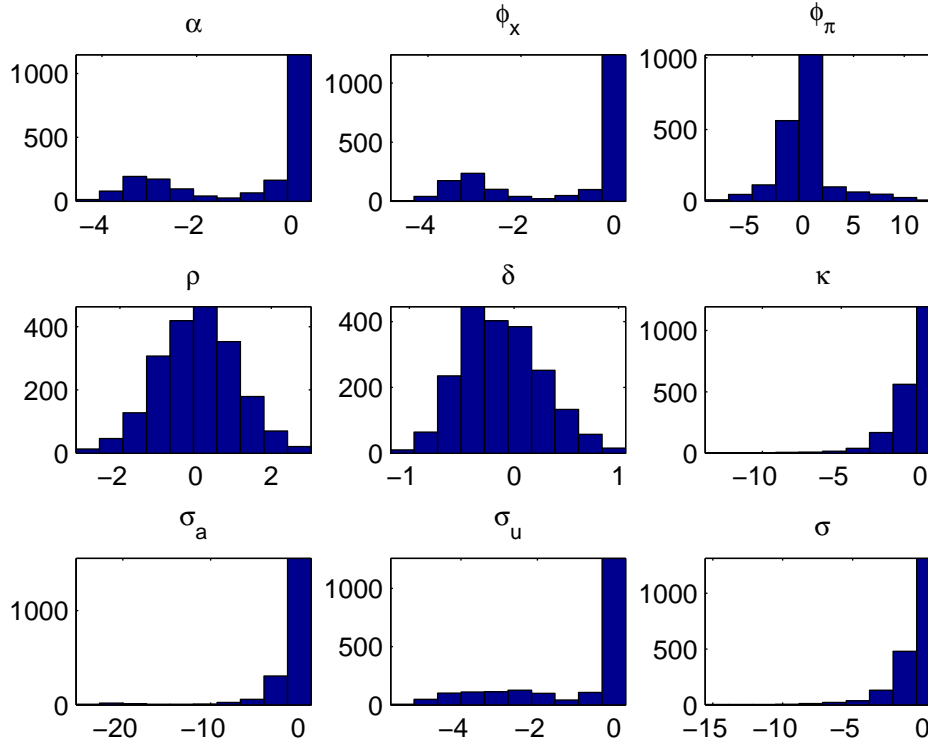


Для начала рассмотрим распределение в конечной выборке (размера 200 периодов) оценки максимального правдоподобия, получающуюся в результате оптимизации функции максимального правдоподобия по 9-мерному параметру β . Гистограммы соответствующих распределений компонент β приведены на Рис. 1. Согласно асимптотической теории, в больших выборках распределение оценки должно быть близко к нормальному. Однако, как мы видим на Рис. 1, для большинства параметров распределение достаточно далеко от нормального. В этом смысле асимптотическая теория дает неверное представление о статистической неопределенности ассоциированной с оценкой наибольшего правдоподобия для данной модели.

Любая оценка должна иметь характеристику статистической погрешности. Часто эту роль выполняет матрица ковариаций оценки. В случае метода максимального правдоподобия асимптотическая теория предписывает, что данная матрица асимптотически равна обратной матрице к матрице информации Фишера. В случае корректной спецификации модели, а в случае нашего эксперимента Монте-Карло модель корректно специфицирована, информация Фишера может быть оценена на практике двумя способами: через квадратную вариацию первой производной (см. формулу (6)) и через вторую производную функции правдоподобия (см. формулу (5)). Однако на практике эти два способа дают существенно разные ответы, и оба способа дают очень странные значения для матрицы Фишера.

На Рис. 2 изображены гистограммы распределения в конечной выборке t -статистик для

Рис. 2: Гистограмма распределения t -статистик для тестирования отдельных параметров. Размер выборки $T = 200$. Число симуляций 500. Информация Фишера подсчитана как вариация первой производной функции правдоподобия.



тестирования верного значения каждого параметра в отдельности. Здесь мы использовали формулу (6) для вычисления матрицы Фишера. Если асимптотическая теория дает хорошее приближение, то данные распределения должны быть близки к стандартному нормальному. Однако, как можно видеть, распределения еще более далеки от нормальности, чем распределения на Рис. 1.

Таблица 1: Фактический уровень значимости классических тестов метода максимального правдоподобия для простой гипотезы $H_0 : \beta = \beta_0$.

	LR	Wald ($\hat{J}(\theta_0)$)	Wald ($\hat{J}(\hat{\theta})$)	Wald ($\hat{I}(\theta_0)$)	Wald ($\hat{I}(\hat{\theta})$)
Размер 5%-го теста	3,20%	65,45%	63,05%	68,05%	68,15%
Размер 10%-го теста	7,05%	67,20%	64,30%	70,80%	71,00%

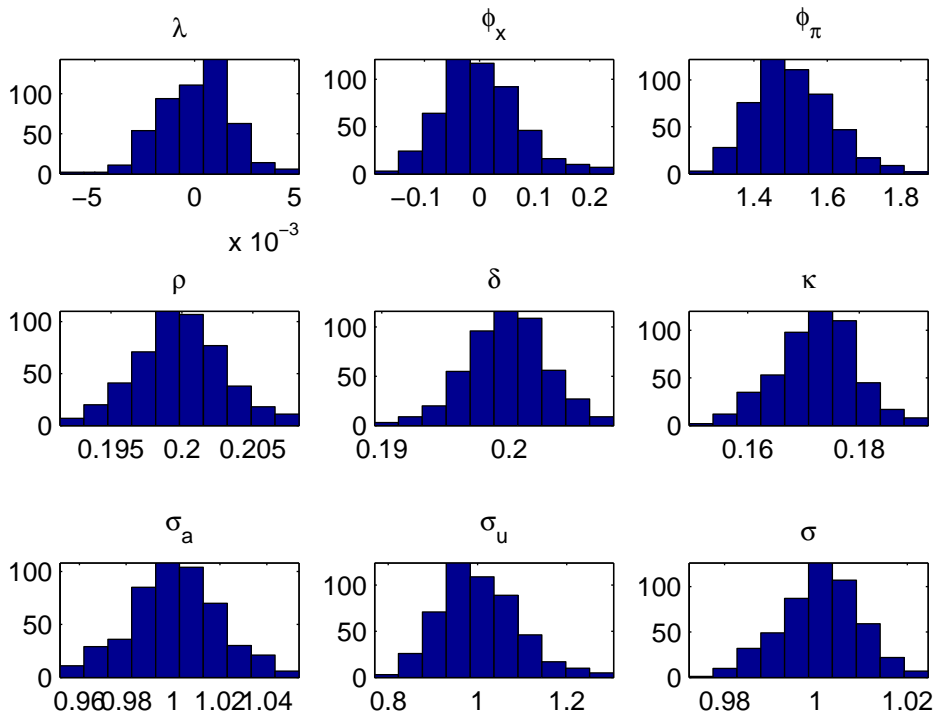
Замечания: Декларируемый уровень значимости 5% и 10%. Число симуляций 2000.

Наконец, обратимся к классическим тестам. Рассмотрим здесь только простейший случай, а именно проблему тестирования простой гипотезы $H_0 : \beta = \beta_0$, где все 9 неизвестных параметров тестируются одновременно. Таблица 1 дает фактический уровень значимости классических теста отношения правдоподобия (LR) и теста Вальда. Последний рассчитан в нескольких видах в зависимости от того, как считается информация Фишера. В данном случае тест, основанный на статистике отношения правдоподобия, является слегка консервативным, а все варианты теста Вальда слишком часто отвергают верную гипотезу и ведут к совершенно неверным статистическим выводам.

Таким образом видим, что результаты классической асимптотической теории, применен-

ные к нашей простой DSGE модели, дают статистически неверные выводы и не могут быть использованы для оценивания и статистических выводов. Может ли наблюдаемый эффект быть назван слабой идентификацией? Скорее да, чем нет. Проблема заключается в том, что на данный момент не существует тестов на слабую идентификацию в общем случае. Но есть несколько соображений, указывающих на то, что ситуация полностью аналогична известным моделям со слабой идентификацией.

Рис. 3: Гистограмма распределения оценок по методу максимального правдоподобия при размере выборки $T = 100000$.



Во-первых, данная модель локально идентифицируема, что следует из применения к ней критериев идентификации, предложенных Komunjer & Ng (2011) и Iskrev (2008, 2010). Во-вторых, наблюдаемая неспособность асимптотики аппроксимировать настоящие распределения относится только к малым выборкам, так например, Рис. 3 изображает гистограммы распределения оценки максимального правдоподобия в случае, когда размер выборки $T = 10^5$. Как видим, в этом случае распределения близки к нормальным. Наконец, Andrews & Mikusheva (2011) показали, что все известные на данный момент модели со слабой идентификацией обладают одним общим интересным свойством, а именно для них две оценки матрицы информации Фишера (данные в формулах (5) и (6)) сильно отличаются друг от друга. Более конкретно, White (1982) показал, что если функция правдоподобия правильно специфицирована, то $\mathbb{E}\hat{I} = \mathbb{E}\hat{J}$. Однако для моделей со слабой идентификацией оценка \hat{I} может быть нормализована так, что она сходится асимптотически к детерминистической невырожденной матрице, в то время как \hat{J} при той же нормализации сходится к случайной матрице, которая не обязана быть положительно определена. Последнее особенно парадоксально, учитывая что \hat{J} является оценкой положительно определенной матрицы информации Фишера. В нашем Монте-Карло эксперименте мы обнаружили, что при размере выборки $T = 200$ в 95% выборок оценка \hat{J} не является положительно определенной матрицей.

5.3 Методы, робастные к слабой идентификации

Метод считается робастным к слабой идентификации, когда он дает тесты с фактическим уровнем значимости, не превышающем (с небольшой погрешностью) декларируемый уровень значимости даже в ситуациях, когда классические методы не работают. Большинство методов так или иначе опираются на асимптотические приближения, однако качество подобных аппроксимаций может сильно варьироваться. Например, в случае слабых инструментов робастные методы опираются на предположение, что центральная предельная теорема, примененная к моментному условию, дает хорошее асимптотическое приближение для распределения нормализованной частичной суммы. То есть считается, что выборка достаточно велика, чтобы простейшие асимптотические законы работали, в то время как выборка не достаточна для того, чтобы были верны асимптотические приближения более сложных нелинейных функций от частичных сумм.

Так, согласно Hall & Heyde (1980), качество асимптотической аппроксимации в классической теории максимального правдоподобия основывается на следующих предположениях: (1) функция правдоподобия хорошо аппроксимируется квадратичной функцией, (2) центральная предельная теорема дает хорошее приближение для некоторых частичных сумм (часто речь идет о первой производной функции правдоподобия), и (3) нормированная вторая производная функции правдоподобия близка к неслучайной константе (информации Фишера). Из предыдущих дискуссий видно, что условия (1) и (3) вряд ли выполняются для DSGE моделей в небольших выборках. Существует несколько статей, пытающихся найти методы для построения статистических выводов, не использующих данные предположения, которые тем не менее опираются на предположения типа (2).

На данный момент существует несколько методов тестирования простых гипотез (гипотез вида $H_0 : \beta = \beta_0$), робастных к слабой идентификации. Andrews & Mikusheva (2011) показали, что один из классических тестов, а именно тест, основанный на мультипликаторе Лагранжа (LM), является робастным к слабой идентификации, если в качестве оценки матрицы Фишера используется \hat{I} . Qu (2011) предложил использовать приближенный тест Лагранжа, основанный на приближении спектральной плотности процесса. Guerгон-Quintana, Inoue & Kilian (2012) предложили использовать тест отношения правдоподобия для приведенной модели, которая пренебрегает многими структурными связями, накладываемыми DSGE моделью ради робастности к слабой идентификации. Наконец, Dufour, Khalaf & Kichian (2009) предложили робастный тест, который отходит от принципа максимального правдоподобия и использует метод наименьшего расстояния. Таблица 2 приводит фактический уровень значимости для всех упомянутых тестов в модели из Примера 1 при размере выборки 200 периодов. Как можно видеть, в то время как метод максимального правдоподобия дает неверные результаты, все четыре робастных теста имеют уровень значимости, очень близкий к декларируемому.

Таблица 2: Фактический уровень значимости робастных тестов для тестирования простой гипотезы в Примере 1.

Декларируемый уровень значимости	AM (2011)	Qu (2011)	GQIK (2012)	DKK (2009)
5%	4,8%	5,1%	6,6%	6,9%
10%	8,9%	10,4%	13,1%	12,4%

Замечания: Число симуляций 1000. Размер выборки 200 периодов.

К сожалению, прогресс в создании робастных тестов относится только к тестированию простых гипотез, т.е. гипотез, которые одновременно тестируют значения всех неизвестных параметров (девятимерный параметр в нашем примере). На практике мы чаще заинтересованы в тестировании гипотез о каждом параметре в отдельности, что является составной ги-

потезой (composite hypothesis). Возможность тестирования каждого одномерного параметра отдельно необходимо также для построения отдельных доверительных интервалов для каждого параметра. К сожалению, задача робастного тестирования сложных гипотез на данный момент неразрешена. Предположим, что в Примере 1 мы хотим протестировать гипотезу $H_0 : \kappa = 0,1717$ (которая в нашем примере верна). В таком случае все остальные восемь параметров задачи являются неизвестными и не оговариваются гипотезой. Они называются шумовыми параметрами. Если предположить что все эти шумовые параметры являются сильно идентифицируемыми, то существуют модификации всех упомянутых выше робастных тестов для тестирования сложной гипотезы. К сожалению, предположение о сильной идентификации шумовых параметров, скорее всего, неверно для типичной DSGE модели, и к тому же не существует на данный момент тестов, которые могли бы протестировать данное предположение. Таким образом, вопрос о том, как производить статистические выводы в случае слабой идентификации модели, остается открытой эконометрической проблемой.

6 Методы оценивания: Байесовский подход

Частично из-за провала классических методов (слабая идентификация), частично из-за появления новых численных алгоритмов, самым популярным методом оценивания DSGE моделей на данный момент является Байесовский метод. Хорошим изложением методологии является статья An & Schorfheide (2007). Наиболее цитируемый пример оценивания DSGE модели Байесовским способом приведен в статье Smets & Wouters (2007).

Входными компонентами Байесовского анализа являются корректно специфицированная функция правдоподобия $\mathcal{L}(\mathcal{Y}_T|\beta)$ и априорное распределение (или вера) $p(\beta)$. На основе их получается постериорное распределение параметра

$$p(\beta|\mathcal{Y}_T) = \frac{\mathcal{L}(\mathcal{Y}_T|\beta)p(\beta)}{\int \mathcal{L}(\mathcal{Y}_T|\tilde{\beta})p(\tilde{\beta})d\tilde{\beta}}. \quad (7)$$

В качестве Байесовской оценки обычно берется либо среднее либо медиана постериорного распределения, а в качестве доверительного интервала- множество, чья постериорная вероятность равна уровню доверия.

Долгое время основным препятствием к использованию Байесовского метода была техническая сложность в вычислении многомерных интегралов, стоящих в знаменателе формулы (7). Постериорное распределение было известно только в ограниченном числе случаев при очень специализированных априорных распределениях. В последнее десятилетие был сделан огромный прорыв, позволяющий применять Байесовский подход в широком наборе случаев. Этот прорыв заключается в разработке алгоритма MCMC (Markov Chain Monte-Carlo). Алгоритм состоит в построении Марковской цепи β_1, β_2, \dots таким образом чтобы для некоего большого числа N (в пределе) случайная величина β_N имела распределение совпадающее с постериорным распределением. Таким образом, вместо того чтобы считать нормализующую константу в уравнении (7), MCMC алгоритм ставит своей целью получить случайную выборку (любого необходимого размера) из постериорного распределения. Имея случайную выборку из постериорного распределения легко получить Байесовские оценки и сделать Байесовские статистические выводы. Например, среднее такой выборки приблизительно равно среднему постериорного распределения, при этом последнее может быть подсчитано с любой точностью.

Существует несколько версий MCMC алгоритма, наиболее часто употребляющиеся для оценки DSGE моделей это алгоритм Метрополис-Хастинга и сэмплер Гиббса. Первый реализован в стандартном пакете Dynare. Для ознакомления с MCMC алгоритмами я советую прочитать статьи Chib (2001) и Chib & Greenberg (1995, 1996), которые дают очень доступное изложение материала.

6.1 Откуда берутся априорные распределения?

Основной сложностью на данный момент в применении Байесовского метода для оценивания DSGE моделей является ответ на вопрос, откуда берутся априорные распределения. С философской точки зрения априорное распределение или вера является отображением текущего уровня знаний или ожиданий ученого относительно вероятностей, что параметр принимает то или иное значение, до того как ученый увидел данные. Постериорное распределение является коррекцией априорной веры с учетом выборки, имеющейся у исследователя. К сожалению, большинство прикладных макроэкономистов признают, что у них чаще всего нет содержательных кандидатов для априорной веры.

На практике многие статьи, оценивающие DSGE модели Байесовским методом, используют априорные распределения, предложенные в статье Smets & Wouters (2007), или аналогичные им. Принцип построения таких априорных распределений следующий: во-первых, априорно все одномерные параметры модели считаются независимо распределены (т.е. многомерная априорная плотность равна произведению априорных плотностей всех параметров). Во-вторых, одномерные распределения выбраны из стандартных семейств распределений, по возможности сопряженных (conjugate) с моделью, т.е. например коэффициенты линейной регрессии считаются априорно нормально распределенными, дисперсии берутся из обратного Гамма-распределения, авторегрессионные коэффициенты считаются априори распределенными по Бета-закону. В-третьих, если есть какие-то оценки параметров, по которым достигнут консенсус среди макроэкономистов, то априорное распределение выбирается центрированным на данной оценке. И наконец, по возможности выбирается распределение, позволяющее широкий разброс значений (или большую дисперсию).

Классическим результатом Байесовской теории считается теорема Бернштейна–фон Мизеса, которая показывает, что асимптотически выбор априорного распределения неважен, так как с точки зрения фреквентистского подхода при широких условиях Байесовское постериорное распределение сходится (когда размер выборки стремится к бесконечности) к гауссовскому распределению вокруг оценки максимального правдоподобия. Эта теорема часто служит утверждением, что Байесовский метод может применяться и сторонником фреквентистского (классического) подхода как удобный численный механизм, коль скоро результаты не сильно отличаются от результатов метода максимального правдоподобия.

К сожалению, эта логика не вполне применима к моделям со слабой идентификацией, а значит и к DSGE моделям в частности. Теорема Бернштейна–фон Мизеса существенным образом использует асимптотическую теорию максимального правдоподобия, которая дает очень неверное представление фактических распределений в конечных выборках для типичных DSGE моделей. В Байесовском методе информация агрегируется из двух источников: выборки и априорного распределения. Теорема Бернштейна–фон Мизеса говорит, что когда выборка растет, информация, приходящая из априорного распределения, асимптотически доминируется информацией из выборки и становится несущественной. В моделях со слабой идентификацией выборка является малоинформативной, и значение априорного распределения велико.

Тот факт, что в DSGE моделях априорное распределение влияет на результат оценивания, заостряет проблему выбора априорного распределения и делает ее центральной. Распространенным заблуждением является утверждение, что если при построении априорного распределения взять одномерные распределения параметров, имеющие большую дисперсию, то оно будет неинформативным, и все результаты будут зависеть главным образом от данных, а не от априорного распределения. Müller (2012) показал, что это не вполне верно. Так например, само предположение, что априори различные одномерные параметры независимы друг от друга, является очень информативным.

Интересным вопросом с этой точки зрения является вопрос о том, сколько информации приходит в результат из данных, а сколько из априорного распределения, и насколько резуль-

тат чувствителен к выбору априорного распределения. Попыткой ответить на этот вопрос является статья Müller (2012). Müller (2012) применил эти методы к модели из Примера 1 и типичным данным и показал, что вся информация о Калво параметре приходит из априорного распределения, а данные главным образом не информативны о нем. Как результат, Байесовская оценка параметра Калво очень чувствительна к выбору априорной веры.

7 Заключение

Большие структурные модели становятся все более и более привлекательными для макроэкономистов. Однако, макроэкономические данные не очень информативны и фактически не позволяют сколь либо хорошо оценивать сколь нибудь сложные модели. Это порождает очень сложные и интересные эконометрические проблемы, которые на данный момент не решены и открывают большую область для деятельности эконометристов.

На мой взгляд, интересными проблемами являются задача поиска статистических методов, робастных к слабой идентификации, вопрос измерения информативности данных и априорной веры в Байесовском подходе, изучение фреквентистских свойств Байесовских процедур а также задача построения неинформативных априорных распределений. Прорыв в решении любой из этих задач способен изменить практику оценивания DSGE моделей и оказать существенное влияние на развитие прикладной макроэконометрики.

Список литературы

- Цыплаков, А. (2011). Введение в моделирование в пространстве состояний. *Квантиль* 9, 1–24.
- An, S. & F. Schorfheide (2007). Bayesian analysis of DSGE models. *Econometric Reviews* 26, 113–172.
- Anderson, G. & G. Moore (1985). A linear algebraic procedure for solving linear perfect foresight models. *Economics Letters* 17, 247–252.
- Andrews, D.W.K. & X. Cheng (2012). Estimation and inference with weak, semi-strong and strong identification. *Econometrica* 80, 2153–2211.
- Andrews, I. & A. Mikusheva (2011). Maximum likelihood inference in weakly identified DSGE models. Unpublished manuscript, MIT.
- Blanchard, O.J. & C.M. Kahn (1980). The solution of linear difference models under rational expectations. *Econometrica* 48, 305–311.
- Blanchard, O.J. & D. Quah (1989). Dynamic effects of aggregate demand and supply disturbances. *American Economic Review* 79, 655–673.
- Boivin, J. & M. Giannoni (2006). Has monetary policy become more effective? *The Review of Economics and Statistics* 88, 445–462.
- Canova, F. (2007). *Methods for Applied Macroeconomic Research*. Princeton University Press.
- Canova, F. & L. Sala (2009). Back to square one: identification issues in DSGE models. *Journal of Monetary Economics* 56, 431–449.
- Chib, S. (2001). Markov chain Monte Carlo methods: computation and inference. Глава 5 в *Handbook of Econometrics* том 5 под редакцией J.J. Heckman & E. Leamer, 3564–3634. Amsterdam: North-Holland.
- Chib, S. & E. Greenberg (1995). Understanding the Metropolis–Hastings algorithm. *American Statistician* 49, 327–335.
- Chib, S. & E. Greenberg (1996). Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory* 12, 409–431.
- Chari, V., P. Kehoe & E. McGrattan (2005). A critique of structural VARs using business cycle theory. Research Department Staff Report 364, Federal Reserve Bank of Minneapolis.
- Christiano, L.J. (2007). A short course on estimation, solution and policy analysis using equilibrium monetary models. Lecture Notes, Northwestern University.

- Christiano, L., Eichenbaum, M. & C. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy* 113, 1–45.
- Christiano, L., M. Eichenbaum & R. Vigfusson (2006). Assessing structural VARs. *NBER Macroeconomics Annual* 21, 1–72.
- Cogley, T. & J. Nason (1995). Output dynamics in real-business-cycle models. *American Economic Review* 85, 492–511.
- DeJong, D. N. & C. Dave (2007). *Structural Macroeconometrics*, Princeton University Press.
- Del Negro, M. & F. Schorfheide (2004). Priors from general equilibrium models for VARs. *International Economic Review* 45, 643–673.
- Del Negro, M., F. Schorfheide, F. Smets, & R. Wouters (2007). On the fit of new Keynesian models. *Journal of Business & Economic Statistics* 25, 123–143.
- Dufour, J.M., L. Khalaf & M. Kichian (2009). Structural multi-equation macroeconomic models: identification-robust estimation and fit. Working Paper, Bank of Canada.
- Fernández-Villaverde, J. (2010). The econometrics of DSGE models. *SERIES: Journal of the Spanish Economic Association* 1, 3–49.
- Fernandez-Villaverde, J. & J. Rubio-Ramirez (2005). Estimating dynamic equilibrium economies: linear versus nonlinear likelihood. *Journal of Applied Econometrics* 20, 891–910.
- Fernandez-Villaverde, J., J. Rubio-Ramirez, T. Sargent & M. Watson (2007). The ABC and (D's) to understand VARs. *American Economic Review* 97, 1021–1026.
- Gali, J., J.D. Lopez-Salido & J. Valles (2003). Technology shocks and monetary policy: assessing the Fed's performance. *Journal of Monetary Economics* 50, 723–743.
- Guerron-Quintana, P., A. Inoue & L. Kilian (2012). Frequentist inference in weakly identified DSGE models. *Quantitative Economics*, в печати.
- Hall, A., A. Inoue, J. Nason & B. Rossi (2012). Information criteria for impulse response function matching estimation of DSGE models. *Journal of Econometrics* 170, 499–518.
- Hall, P. & C.C. Heyde (1980). *Martingale Limit Theory and its Application*. Academic Press.
- Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and Kalman Filter*. Cambridge University Press.
- Ingram, B.F., N. Kocherlakota & N.E. Savin (1994). Explaining business cycles: a multi-shock approach. *Journal of Monetary Economics* 34, 415–428.
- Ireland, P. (2004). A method for taking models to data. *Journal of Economic Dynamics and Control* 28, 1205–26.
- Iskrev, N. (2008). Evaluating the information matrix in linearized DSGE models. *Economics Letters* 99, 607–610.
- Iskrev, N. (2010). Evaluating the strength of identification in DSGE models: an a priori approach. Working paper, Bank of Portugal.
- King, R.G., C.I. Plosser, J.H. Stock & M.W. Watson (1991). Stochastic trends and economic fluctuations. *American Economic Review* 81, 819–840.
- Klein, P. (2000). Using the generalized Schur form to solve a multivariate linear rational expectations model. *Journal of Economic Dynamics and Control* 24, 1405–1423.
- Komunjer, I. & S. Ng (2011). Dynamic identification of dynamic stochastic general equilibrium models. *Econometrica* 79, 1995–2032.
- Kydland, F. & E. Prescott (1982). Time to build and aggregate fluctuations. *Econometrica* 50, 1345–1370.
- Lindé, J. (2005). Estimating new-Keynesian Phillips curves: a full information maximum likelihood approach. *Journal of Monetary Economics* 52, 1135–1149.
- Mikusheva, A. (2012). One-dimensional inferences in autoregressive models in potential presence of unit root. *Econometrica* 80, 173–212.
- Müller, U. (2012). Measuring prior sensitivity and prior informativeness in large Bayesian models. *Journal of Monetary Economics* 59, 581–597.
- Qu, Z. (2011). Inference and specification testing in DSGE models with possible weak identification. Unpublished manuscript, Boston University.

- Qu, Z. & D. Tkachenko (2012). Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models. *Quantitative Economics* 3, 95–132.
- Rabanal, P. & J. Rubio-Ramirez (2005). Comparing new Keynesian models of the business cycle: a Bayesian approach. *Journal of Monetary Economics* 52, 1151–1166.
- Rotemberg, J. & M. Woodford (1997). An optimization-based econometric framework for the evaluation of monetary policy. *NBER Macroeconomics Annual* 12, 297.
- Schorfheide, F. (2000). Loss function-based evaluation of DSGE models. *Journal of Applied Econometrics* 15, 645–670.
- Schorfheide, F. (2010). Estimation and evaluation of DSGE models: progress and challenges. Invited Lecture at 2010 Econometric Society World Congress in Shanghai.
- Sims, C.A. (1980). Macroeconomics and reality. *Econometrica* 48, 1–48.
- Sims, C.A. (2001). Solving linear rational expectations models. *Computational Economics* 20, 1–20.
- Smets, F. & R. Wouters (2007). Shocks and frictions in US business cycles: a Bayesian DSGE approach. *American Economic Review* 97, 586–606.
- Stock, J.H. & J.H. Wright (2000). GMM with weak identification. *Econometrica* 68, 1055–1096.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25.

Estimation of dynamic stochastic general equilibrium models

Anna Mikusheva

Massachusetts Institute of Technology, Cambridge, USA

This essay is a survey of the main econometric approaches to estimation of the dynamic stochastic general equilibrium (DSGE) models widely used by central banks and federal reserves. The paper discusses in detail the main econometric problems arising in inferences about the parameters of the log-linearized DSGE models. We examine three main estimation approaches: the minimum distance method, the maximum likelihood method and the Bayesian approach. We focus on the problems of weak identification that are due to scarcity of macro data. The issues of economic modelling and methods of solving dynamic models are beyond the scope of the current essay.

