
**МАТЕМАТИЧЕСКИЙ АНАЛИЗ
ЭКОНОМИЧЕСКИХ МОДЕЛЕЙ**

**Вычислительная эффективность байесовских эконометрических методов
для «неудобных» плотностей**

© 2021 г. С.М. Иващенко

С.М. Иващенко,

СЗГУ Банка России, Санкт-Петербург; Институт проблем региональной экономики РАН, Санкт-Петербург; Научно-исследовательский финансовый институт Минфина РФ, Москва; Факультет экономики СПбГУ, Санкт-Петербург; e-mail: sergey.ivashchenko.ru@gmail.com; glucke_ru@pisem.net

Поступила в редакцию 05.10.2020

Исследование выполнено при частичной финансовой поддержке Российского фонда фундаментальных исследований (проект 18-010-01185) «Структурные изменения в экономике России: роль человеческого капитала и инвестиций».

Настоящая статья выражает личную позицию автора, которая может не совпадать с официальной позицией Банка России. Банк России не несет ответственности за содержание статьи.

Аннотация. Оценка моделей динамического стохастического общего экономического равновесия (ДСОЭР) методами байесовской эконометрики подразумевает построение марковских цепей Монте-Карло (MCMC). Проведен анализ MCMC-алгоритмов для функций плотности, с неблагоприятными свойствами, характерными для ДСОЭР-моделей (ограниченный носитель функции, тяжелые хвосты, острые пики, невыпуклость логарифма плотности или самой плотности). Рассматривается три группы алгоритмов: случайное блуждание (RW), алгоритм MALA и предложенный автором, алгоритм LTG (local truncated Gauss). Для последних двух алгоритмов рассматривались три версии: с использованием информации о градиенте и гессии логарифма функции плотности в каждой точке; только о градиенте и версия, использующая лишь информацию о моде. Результаты MALA и LTG близки в большинстве случаев, с небольшим преимуществом LTG (включая тест на ДСОЭР-модели). Алгоритм RW проигрывает оставшимся двум, особенно сильно в случае малой размерности. Причем версии, использующие аппроксимацию градиента и гессии, не связаны с существенными дополнительными вычислительными затратами. Наличие тяжелых хвостов ведет к снижению эффективности алгоритмов MALA и LTG. А уровень принятия, обеспечивающий наилучшую эффективность выборки, варьируется в широких пределах, заметно отклоняясь от принятых значений.

Ключевые слова: MCMC; Монте-Карло с марковскими цепями; алгоритм Метрополиса–Гастингса; методы байесовской эконометрики.

Классификация JEL: C11.

DOI: 10.31857/S042473880014916-6

1. ВВЕДЕНИЕ

ДСОЭР-модели — один из основных макроэкономических инструментов (Tovar, 2009). Они базируются на микроэкономических основаниях, т.е. предпочтениях и технологиях. Это решает проблему критики Лукаса (Lucas, 1976). В сочетании с формальной эконометрической оценкой это позволяет получать высокое качество прогнозов (Smets, Wouters, 2007, Tovar, 2009). В результате этот класс моделей активно применяется центральными банками и другими организациями во многих странах (Tovar, 2009). Однако сложная функциональная связь параметров модели и ее поведения ведет к снижению скорости сходимости распределения оценок к асимптотическим, являющимся основой классической эконометрики. Вдобавок микроэкономические основания способствуют естественному формированию априорных представлений о значениях параметров. В результате стали популярны методы байесовской эконометрики, использующие априорную информацию и менее чувствительные к относительно маленьким выборкам, в условиях медленной сходимости к асимптотике. Обратная сторона этих методов — высокие вычислительные затраты, которые еще больше возрастают для распределений, далеких от стандартных.

Методы байесовской эконометрики активно применяются для расчетов по ДСОЭР-моделям. В частности, применяется Metropolis Hating МСМС (МН-МСМС), а точнее Random Walk (RW) МСМС (Smets, Wouters, 2007; Christiano, Eichenbaum, Trabandt, 2016; Herbst, Schorfheide, 2015). Причем в качестве ориентиров приводятся цифры по уровню принятия (acceptance rate), рекомендуемые для усредненных моделей 35% (Smets, Wouters, 2007), 24% (Christiano et al., 2016), 20–40% (Herbst, Schorfheide, 2015). Однако уровень принятия порядка 23,4% не оптимален для множества распределений. Например, для гауссовой радиальной плотности оптимальный уровень принятия составляет 10%, а для экспоненциальной радиальной плотности — 6% (Sherlock, Roberts, 2009). Более того, 23,4% — это асимптотически оптимальный уровень (по размерности), и сходимость к нему может быть крайне медленной (и требовать размерности более 200 000) (Bedard, 2007).

Для апостериорной плотности параметров, характерной для ДСОЭР-моделей, можно встретить ряд неблагоприятных для сходимости МСМС-алгоритмов свойств. Во-первых, плотность часто определена не на всем пространстве R^n . Это связано с условиями Бланшара–Кана, необходимыми для существования единственного невзрывного решения задачи с рациональными ожиданиями (Blanchard, Kahn, 1985; Schmitt-Grohe, Uribe, 2004). Соответственно, при нарушении этого условия плотность принимает штрафное значение. Во-вторых, распространена проблема слабой идентификации отдельных параметров (Iskrev, 2010; Ivashchenko, Mutschler, 2020), т.е. функция апостериорной плотности может быть достаточно плоской в отдельных направлениях. В-третьих, нелинейность зависимости по параметрам и жесткие априорные распределения могут приводить (при принятых объемах выборки) к значительному отклонению формы функции плотности от гауссовой, а в предельных случаях — к невыпуклости логарифма функции правдоподобия.

Известны алгоритмы, демонстрирующие высокую эффективность генерации выборки для ДСОЭР-моделей (Chib, Ramamurthy, 2010). Однако для достижения этой эффективности требуется многократное вычисление функции правдоподобия на каждой итерации (этап максимизации), что связано с большими вычислительными затратами. Исходя из этого, будут рассматриваться алгоритмы МСМС без многочисленных дополнительных расчетов функции правдоподобия.

Таким образом, целью данной работы является анализ методов байесовской эконометрики, используемых для оценки ДСОЭР-моделей. А именно анализ свойств МСМС-алгоритмов для мономодальных функций плотности, обладающих неблагоприятными свойствами, встречающимися в апостериорных плотностях, характерных для ДСОЭР-моделей. Мы приведем описания нескольких алгоритмов МСМС (разд. 2) и их тестов (разд. 3).

2. АЛГОРИТМЫ МСМС

2.1. Общая схема Metropolis Hating (МН)

Алгоритм МН позволяет генерировать МСМС с желаемым асимптотическим распределением $p(x)$. Для этого необходимо уметь генерировать значения из произвольной условной плотности $Q(x_i|x_{i-1})$, а также рассчитать значения плотностей $p(x_i)$ и $Q(x_i|x_{i-1})$ в произвольной точке. После инициализации алгоритма (т.е. какого-то выбора начальной точки x_0) на итерации i вначале генерируется точка x_{try} из плотности $Q(x_{try}|x_{i-1})$. Затем по формуле

$$\alpha = \min \left(\frac{p(x_{try})Q(x_{i-1}|x_{try})}{p(x_{i-1})Q(x_{try}|x_{i-1})}; 1 \right) \quad (1)$$

рассчитывается величина α , а далее с вероятностью α значение принимается ($x_i = x_{try}$), а с вероятностью $1-\alpha$ — отвергается ($x_i = x_{i-1}$). Различия алгоритмов заключаются в использовании различных плотностей $Q(x_i|x_{i-1})$.

2.2. Random Walk (RW) МСМС

Один из самых простых алгоритмов, называемый «случайное блуждание», это применение в качестве $Q(x_i|x_{i-1})$ — плотности гаусса с дисперсией h^2V и средним x_{i-1} . Обычно в качестве дисперсии V берется обратный гессиан в моде и подбирается масштабирующий множитель h для желаемого уровня принятия. Именно этот алгоритм стоит по умолчанию в пакете расчетов по ДСОЭР-моделям — dynare (Adjemian et al., 2011).

2.3. Metropolis-adjusted Langevin algorithm (MALA)

Альтернативный подход предполагает использование локальных данных о форме плотности, а именно градиента логарифма функции плотности. Такой подход связан с большими вычислительными издержками (в основном на расчет градиента), но обеспечивает существенно более быструю сходимость (Titsias, Dellaportas, 2019; Durmus et al., 2017). Генерирование происходит по формуле

$$x_{iry} = x_{i-1} + (h^2 / 2)V \frac{\partial \log p(x_{i-1})}{\partial x_{i-1}} + h(V)^{1/2} \varepsilon_i, \tag{2}$$

где ε_i сгенерировано из стандартного нормального распределения.

В некоторых работах применяется формулировка с единичной матрицей вместо V (Jarner, Roberts, 2007). Но более естественный вариант — матрица, аналогичная RW . Помимо этого есть естественное обобщение (2) до вида

$$x_{iry} = x_{i-1} + (h^2 / 2) \left(\frac{\partial^2 \log p(x_{i-1})}{\partial x_{i-1} \otimes x_{i-1}} \right)^{-1} \frac{\partial \log p(x_{i-1})}{\partial x_{i-1}} + h \left(\frac{\partial^2 \log p(x_{i-1})}{\partial x_{i-1} \otimes x_{i-1}} \right)^{-1/2} \varepsilon_i \tag{3}$$

с использованием информации о гессиане. В таком случае (2) оказывается частным случаем (3) с аппроксимацией для гессиана.

Также можно применять аппроксимацию и для градиента. Подробнее о том, как строится аппроксимация, а также о том, как считать корень из матрицы, будет описано ниже.

Соответственно, получается 5 версий MALA-алгоритма: с градиентом и гессианом; с градиентом и аппроксимацией гессиана (гессиан в моде); с аппроксимацией градиента и гессиана; а так же версии с единичной матрицей в качестве гессиана.

2.4. Local truncated Gauss (LTG)

Имея данные о градиенте и гессиане логарифма плотности в каждой точке, естественно возникает квадратичная аппроксимация (квадратичная функция логарифма плотности у распределения Гаусса). Однако точность аппроксимации зависит от расстояния от точки разложения, отсюда возникает идея использовать урезанное нормальное распределение в качестве распределения, из которого генерируется пробная точка. То есть градиент и гессиан (или их аппроксимации) преобразуются в математическое ожидание μ_i и дисперсию V_i в соответствии с формулами:

$$\left(\frac{\partial^2 \log p(x_{i-1})}{\partial x_{i-1} \otimes x_{i-1}} \right)^{-1} = V_i, \tag{4}$$

$$x_{i-1} - \left(\frac{\partial^2 \log p(x_{i-1})}{\partial x_{i-1} \otimes x_{i-1}} \right)^{-1} \left(\frac{\partial \log p(x_{i-1})}{\partial x_{i-1}} \right) = \mu_i, \tag{5}$$

где \otimes — произведение Кронекера.

Существует множество способов урезать многомерное нормальное распределение. Нами был выбран вариант наиболее простой с точки зрения расчета, а именно урезание независимых стандартных нормальных распределений. Для масштабирования дистанции используются стандартные отклонения в моде, т.е. согласно формуле

$$\left| (V_i)^{-1/2} (x_i - x_{i-1}) \right| \leq (V_i)^{-1/2} (V_{mode})^{1/2} 1_{n_x} r_{std} \tag{6}$$

в случае совпадения дисперсий в моде (V_{mode}) и текущей точке (V_i) абсолютное отклонение (после перехода к независимому распределению согласно текущей матрице дисперсии) поэлементно не превосходит r_{std} . Таким образом, плотность, из которой генерируются наблюдения $Q(x_i|x_{i-1})$, — это урезанное нормальное распределение со средним μ_i , дисперсией V_i и урезанное вокруг точки x_{i-1} с радиусом r_{std} .

Генерирование наблюдений и расчет плотности (нормирующей константы) для одномерного урезанного распределения Гаусса легко произвести благодаря аналитическому интегрированию нормальной плотности. Однако при слишком большом удалении от моды возникают численные проблемы. В таких случаях используется алгоритм (Chopin, 2011).

2.5. Замечания об аппроксимации

Аппроксимацией гессиана является гессиан в моде. Аппроксимация градиента строится на основе значений функции плотности в текущей точке и в моде (точнее, нормальная аппроксимация плотности). Для этого определяется математическое ожидание на отрезке между текущей точкой и модой плотности, т.е. по двум значениям плотности находятся два параметра (c_0, c_1) аппроксимации в форме

$$\log(p(x)) = c_0 - \left(x - (x_{i-1}c_1 + (1-c_1)x_{mode}) \right)' V^{-1} \left(x - (x_{i-1}c_1 + (1-c_1)x_{mode}) \right) / 2. \quad (7)$$

У матрицы дисперсии нет индекса, так как используется гессиан в моде. Соответственно, вес рассчитывается по формуле

$$c_1 = \frac{2(\log(p(x_{i-1})) - \log(p(x_{mode}))) + (x_{i-1} - x_{mode})' V^{-1} (x_{i-1} - x_{mode})}{2(x_{i-1} - x_{mode})' V^{-1} (x_{i-1} - x_{mode})}. \quad (8)$$

Получающееся математическое ожидание для генерации точки согласно MALA-алгоритму (9) напоминает Preconditioned Crank–Nicolson-алгоритм (Cotter et al., 2013). Отличие заключается в том, что смещение идет не в направлении начала координат, а в направлении мода. Так же взаимосвязь дисперсии и смещения имеет другую параметризацию, плюс автоматическая корректировка этой взаимосвязи, в зависимости от того насколько плотность отклоняется от нормальной аппроксимации в моде:

$$\begin{aligned} \mu_i = x_{i-1} + (h^2 / 2) \left(\frac{\partial^2 \log p(x_{i-1})}{\partial x_{i-1} \otimes x_{i-1}} \right)^{-1} \frac{\partial \log p(x_{i-1})}{\partial x_{i-1}} = x_{i-1} + (h^2 / 2) V \left(-V^{-1} (x_{i-1} - \right. \\ \left. - (x_{i-1}c_{1,i} + (1-c_{1,i})x_{mode}) \right) = x_{i-1} - 0,5(x_{mode} - x_{i-1})h^2(1-c_{1,i}). \end{aligned} \quad (9)$$

Еще одна важная деталь алгоритма касается нахождения корня из матрицы. В случае когда матрица положительно определенная, стандартным подходом является разложение Холецкого. Однако в отдельных точках гессиан логарифма плотности может не быть положительно определенным. Для защиты от такой проблемы используется LDL-разложение, которое для положительно определенных матриц ведет себя аналогично разложению Холецкого. В случае неположительно определенного гессиана матрица D заменяется диагональной матрицей, в которой устанавливается минимальное значение диагональных элементов (1000^{-2}). Поскольку нарушение положительно определенности гессиана логарифма плотности не в фокусе данного исследования, подобной эвристики достаточно.

3. СХЕМА ТЕСТИРОВАНИЯ

МСМС-подход ведет к созданию коррелированных наблюдений, имеющих заданное асимптотическое распределение. Для оценки качества получающейся выборки применяют IF (inefficiency factor), который показывает, во сколько раз выборка должна быть больше по сравнению с независимой для получения одинаковой дисперсии выборочного среднего:

$$IF_k = 1 + 2 \sum_{i=1}^{\infty} \text{correl}(x_{k,t+i}, x_{k,t}). \quad (10)$$

Для чистых случаев МН-алгоритма (не имеющих элементов адаптации и т.п.) условное распределение зависит только от вектора на предыдущем шаге. Соответственно, сумму ковариаций можно рассчитать при помощи VAR(1)-представления с матрицей A по формуле

$$\sum_{i=1}^{\infty} \text{cov}(x_{t+i}, x_t) = \sum_{i=1}^{\infty} A^i \text{var}(x_t) = A(I - A)^{-1} \text{var}(x_t). \quad (11)$$

При оценке VAR(1) методом наименьших квадратов, когда истинные собственные числа матрицы A близки к 1, можно получить оценку A с собственными числами большими 1 с достаточно большой вероятностью (Hammad, 2014). Если собственные числа A больше 1, то это взрывной процесс и формула (11) перестает работать. Используя знание теоретического среднего (чтобы уменьшить число оцениваемых параметров), а также то, что дисперсия начальной точки и после шага алгоритма МН совпадают (т.е. удвоим число наблюдений для расчета обратной матрицы дисперсии регрессоров), можно уточнить оценки. В ряде случаев (при относительно большом числе регрессоров) все равно получаются взрывные траектории, в таком случае

для сокращения числа регрессоров и получения невзрывного VAR(1)-представления будет использоваться метод главных компонент.

Поскольку мы работаем с распределениями, генерирование наблюдений из которых является стандартным, можно применять цепи длиной в одно наблюдение с целью расчета IF. То есть генерируется начальная точка непосредственно из интересующего распределения, а далее генерируется один цикл МСМС-алгоритма, что позволяет существенно увеличить эффективность вычислений (особенно в случаях, когда IF очень велик) без каких-либо потерь. Исходя из этого, число цепей установлено на уровне 1000 (расчет для каждого рассматриваемого случая строится на основе 1000 независимых наблюдений).

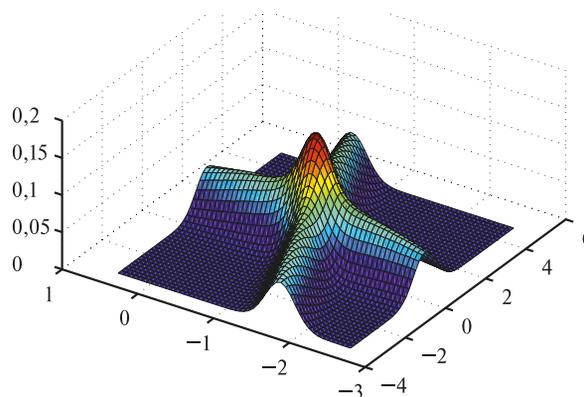


Рис. 1. Плотность вида X

Опишем преобразование базовых одномерных распределений в многомерные, используемые для тестов. Каждое одномерное распределение центрируется вокруг его моды, т.е. вычитается мода, и у модифицированного распределения мода оказывается в 0. Таким образом, если z — вектор таких модифицированных базовых одномерных распределений, то вектор интересующего распределения x будет рассчитываться по формуле

$$x = Qz + \mu, \tag{12}$$

где вектор моды μ и матрица Q генерируются (один раз на тест) поэлементно из стандартного нормального распределения.

Следующие базовые плотности будут использоваться для тестов:

- 1) стандартное нормальное распределение (среднее 0, дисперсия 1);
- 2) гамма-распределение с параметрами 9 и 1/3, дающими среднее 3, дисперсию 1 и моду, равную 8/3. Это распределение комбинирует более медленное убывание плотности с наличием барьерного ограничения;
- 3) распределение Вейбула с параметрами 3 и $10^{-1/2}$, дающими среднее 2,685, дисперсию 0,867 и моду, равную 2,66. Это распределение комбинирует более быстрое убывание плотности с наличием барьерного ограничения;
- 4) урезанное стандартное нормальное распределение с областью поддержки от $-2,5$ до $2,5$. Среднее и мода при этом остаются в 0, а дисперсия уменьшается до 0,9, т.е. это распределение дает лишь эффект наличия барьерного ограничения;
- 5) распределение Стьюдента с 3 степенями свободы — это распределение берется для демонстрации свойств алгоритмов в условии крайне медленного убывания функции плотности, с нарушением выпуклости логарифма плотности.

Дополнительный тест будет произведен для плотности вида X в качестве базовой. Это частный случай смеси нормальных распределений. С вероятностью 1/2 используется распределение Гаусса со средним 0 и стандартным отклонением 3 для четных компонент и 1/3 — для нечетных; с вероятностью 1/2 используется аналогичное нормальное распределение, в котором стандартные отклонения четных и нечетных компонент меняются местами. Данная базовая плотность обладает средним, равным 0, и дисперсией — около 4,56 (по каждой компоненте). Внешний вид такой плотности (после добавления константы и вращения) для двумерного случая представлен на рис. 1. Это случай нарушения выпуклости функции плотности, при сохранении единственности моды.

Второй дополнительный тест — это смесь с равными вероятностями случаев 1–4. Внешний вид базовой плотности представлен на рис. 2. Там же приведены плотности трех компонент (гамма, Вейбул и Гаусс). Поведение смеси на одном хвосте распределения определяется в основном гамма-распределением, а на другом — в основном плотностью Гаусса.

Следует отметить, что алгоритмы MALA и RW используют масштабирующий параметр h , в то время как в LTG применяется r_{std} . Необходимо соотнести эти два параметра. Возьмем следующий набор значений для r_{std} : 0,02; 0,1; 0,2; 0,5; 1; 1,5; 2; 3; 5; 10; 20; в качестве h — стандартные

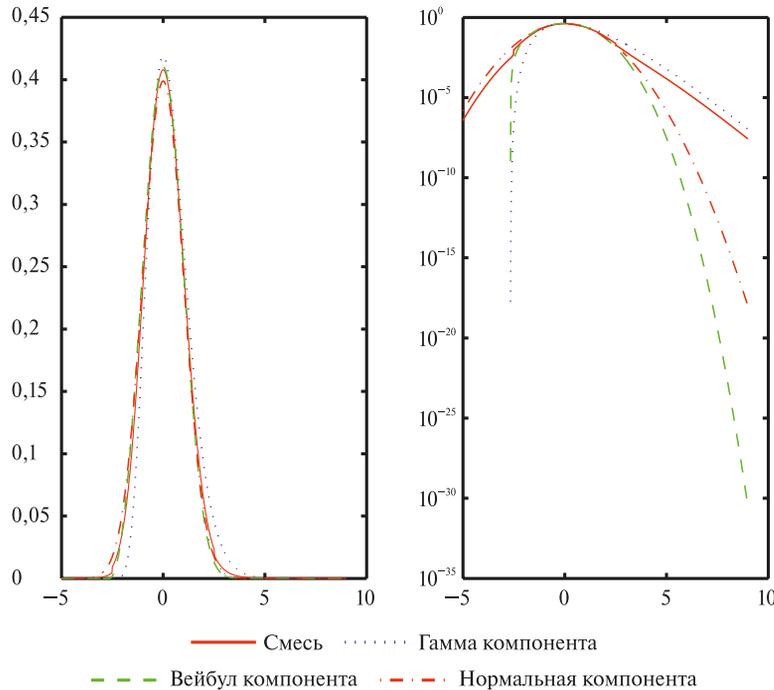


Рис. 2. Плотность смеси распределений гамма, Вейбула, Гаусса и урезанной Гаусса

отклонения урезанного (на интервал от $-r_{std}$ до r_{std}) стандартного нормального распределения, т.е. значения 0,146; 1,63; 4,59; 17,6; 0,446, 0,691; 0,859; 0,985; 1,0; 1,0; 1,0.

Введем сетку размерности для тестов (целая часть 1,5 в степени от 2 до 13), покрывающую диапазон размерности, характерный для числа параметров от сверхмалых ДСОЭР-моделей до крупномасштабных.

4. РЕЗУЛЬТАТЫ

Результаты для распределения Гаусса представлены в табл. 1 (Н обозначает гессиан, G — градиент, 0 — аппроксимацию градиента и гессиана). Использование метода MALA-diag, т.е. единичной матрицы в качестве ковариационной, ведет к крайне слабым результатам, несмотря на имеющиеся данные о градиенте. RW характеризуется растущей неэффективностью с ростом размерности, но этот рост не очень быстр. Для рассматриваемых размеров выборки и решетки масштабирующего множителя оптимальный уровень принятия заметно отличается от асимптотически оптимальных 23,4%. Для остальных алгоритмов нормальная плотность полностью воссоздается, и, соответственно, уровень неэффективности оказывается близок к 1.

Результаты для гамма-распределения представлены в табл. 2. Показатели для случайного блуждания меняются несильно. Результаты MALA-diag остаются неконкурентоспособно слабыми. Достаточно слабыми оказываются результаты методов, использующих значение гессиана в каждой точке, особенно в методе LTG. Вероятно, это связано с тем, что для более плоской функции плотности оптимальным оказывается относительно большой шаг и локальная квадратичная аппроксимация не очень хорошо описывает поведение в области в целом. Алгоритмы MALA оказываются сопоставимы с RW для гамма-распределения при большой размерности пространства, а при малой и средней — явное преимущество за MALA. Версия LTG(G), использующая градиент, кратно превосходит RW для всех рассматриваемых размерностей, а версия LTG(0) дает сопоставимые результаты при большой размерности пространства и кратно преимущество при малой и средней размерности.

Уровень принятия соответствующий минимальной неэффективности выборки часто оказывается ниже 23,4%, причем это наблюдается даже для средней размерности пространства. Подобные случаи есть и для RW, и для MALA, для которого часто рекомендуют более высокий уровень принятия.

Результаты для распределения Вейбула представлены в табл. 3. Ситуация для RW и MALA-diag аналогична предыдущим случаям. Использование данных о гессиане не дает заметного улучшения

Таблица 1. Нормальное распределение. IF, уровень принятия

Размерность	LTG (H, G)	LTG (G)	LTG (0)	MALA (H, G)	MALA (G)	MALA (0)	MALA-diag (G)	MALA-diag (0)	RW
2	0,93 100%	1,09 100%	0,91 100%	1,03 100%	0,95 100%	1,00 100%	14,7 45%	17,6 47%	8,11 55%
3	0,99 100%	0,98 100%	0,96 100%	0,96 100%	1,01 100%	1,12 100%	26,5 76%	26,4 74%	10,9 44%
5	1,01 100%	1,08 100%	1,12 100%	0,91 100%	1,00 100%	1,04 100%	187,8 8%	211,1 9%	14,7 34%
7	1,02 100%	0,85 100%	0,92 100%	0,97 100%	0,83 100%	0,87 100%	1772,8 33%	479,9 4%	15,6 39%
11	1,04 100%	1,29 100%	1,09 100%	1,03 100%	1,08 100%	1,18 100%	4055,2 16%	3513,6 15%	33,8 15%
17	1,37 100%	1,11 100%	1,02 100%	1,10 100%	1,06 100%	1,17 100%	1841,1 1%	6301,2 1%	53,2 18%
25	1,14 100%	1,24 99%	1,25 100%	1,33 94%	1,10 100%	1,07 100%	221816 3%	12362 2%	59,3 27%
38	1,22 100%	1,21 100%	1,15 100%	1,18 100%	1,21 100%	1,05 100%	∞	∞	86,9 20%
57	1,26 100%	1,32 100%	1,24 100%	1,20 100%	1,05 100%	1,26 100%	∞	∞	125,1 51%
86	1,14 100%	1,23 100%	1,11 100%	1,07 100%	0,92 100%	1,17 100%	∞	∞	116,3 39%
129	1,24 100%	1,32 100%	1,18 100%	1,16 100%	1,10 100%	1,29 100%	∞	∞	164,8 32%
194	1,19 97%	1,14 100%	1,16 100%	1,11 100%	1,25 100%	1,14 100%	∞	∞	220,2 22%

результатов, а для LTG(H, G) происходит заметное ухудшение при большой размерности пространства. Результаты остальных версий LTG и MALA дают заметно лучшие значения, чем RW. Исключением является лишь максимальная размерность при использовании градиента. Версии

Таблица 2. Гамма-распределение. IF, уровень принятия

Размерность	LTG (H, G)	LTG (G)	LTG (0)	MALA (H, G)	MALA (G)	MALA (0)	MALA-diag (G)	MALA-diag (0)	RW
2	9,05 52%	3,18 72%	3,02 75%	6,59 54%	2,95 73%	2,64 77%	5,68 72%	4,06 80%	11,0 54%
3	10,5 41%	3,78 62%	3,29 67%	11,6 41%	3,78 62%	3,27 68%	20,8 43%	22,2 40%	14,1 45%
5	21,4 23%	5,03 51%	4,33 56%	20,2 31%	5,42 49%	3,82 55%	111,9 19%	92,2 23%	16,4 31%
7	38,8 12%	5,75 41%	4,91 46%	27,9 12%	5,98 40%	4,90 45%	217,5 98%	958,5 4%	24,2 24%
11	100,2 4%	7,96 36%	7,66 33%	56,4 15%	9,95 38%	8,06 35%	3108,5 40%	1734,0 1%	38,0 18%
17	270,4 13%	15,1 20%	13,6 24%	105,8 19%	15,4 27%	12,4 24%	8613,5 14%	4990,6 12%	58,4 37%
25	511,1 22%	19,3 23%	22,1 16%	132,3 50%	24,2 18%	21,3 13%	619761 2%	130646 1%	75,6 25%
38	457,7 11%	32,2 10%	59,0 6%	134,3 40%	41,6 30%	50,0 6%	∞	∞	71,6 16%
57	1358,7 16%	54,8 19%	110,1 22%	159,7 25%	66,8 20%	143,8 41%	∞	∞	101,3 47%
86	1830,3 6%	46,4 10%	180,7 11%	285,7 15%	121,5 44%	151,6 32%	∞	∞	238,6 39%
129	5857,5 35%	51,3 8%	224,8 42%	303,9 8%	182,9 7%	211,7 21%	∞	∞	234,7 26%
194	3868,0 0%	29,5 11%	282,2 37%	706,0 60%	291,1 22%	454,6 12%	∞	∞	275,7 21%

Таблица 3. Распределение Вейбула. IF, уровень принятия

Размерность	LTG (H, G)	LTG (G)	LTG (0)	MALA (H, G)	MALA (G)	MALA (0)	MALA-diag (G)	MALA-diag (0)	RW
2	2,12 85%	0,93 89%	1,10 94%	2,21 83%	0,95 89%	0,97 92%	11,4 72%	11,2 78%	7,77 55%
3	2,23 80%	1,06 84%	1,06 90%	2,45 79%	1,03 84%	1,05 92%	33,7 53%	26,9 28%	10,7 43%
5	2,65 71%	1,22 78%	1,25 86%	2,66 71%	1,30 79%	1,32 87%	398,6 4%	257,1 7%	15,1 27%
7	3,23 65%	1,71 72%	1,47 90%	3,43 62%	1,68 66%	1,29 81%	1271,4 2%	781,4 6%	19,8 38%
11	4,08 50%	1,95 57%	1,44 86%	3,93 53%	1,91 59%	1,48 76%	858,2 50%	1421,3 16%	30,7 16%
17	6,35 37%	3,21 48%	1,97 69%	6,83 38%	3,00 47%	1,87 72%	2870,2 0%	923,5 91%	55,9 32%
25	8,31 28%	3,22 49%	2,06 74%	8,32 28%	3,96 39%	1,89 58%	300370 1%	179307 3%	46,7 25%
38	24,4 14%	5,59 36%	2,63 63%	18,6 17%	7,47 26%	2,83 52%	∞	∞	111,8 54%
57	39,9 9%	8,03 26%	3,52 53%	26,4 14%	11,8 24%	4,36 40%	∞	∞	92,2 44%
86	511,0 33%	16,8 16%	5,74 41%	56,1 21%	23,4 16%	6,56 42%	∞	∞	143,1 37%
129	957,3 20%	37,4 10%	8,82 30%	175,2 47%	65,0 8%	8,68 32%	∞	∞	252,3 25%
194	1944,4 9%	231,1 6%	16,9 18%	207,8 39%	347,9 21%	18,3 22%	∞	∞	250,0 16%

с градиентом дают худшие результаты, чем без него. С точки зрения уровня принятия наблюдаются случаи за пределами интервала 23,4–40%, причем с отклонениями в обе стороны.

Результаты для урезанного нормального распределения представлены в табл. 4. Показатели LTG и MALA оказываются намного лучше и кратно обыгрывают RW. В условиях нормальной

Таблица 4. Урезанное нормальное распределение. IF, уровень принятия

Размерность	LTG (H, G)	LTG (G)	LTG (0)	MALA (H, G)	MALA (G)	MALA (0)	MALA-diag (G)	MALA-diag (0)	RW
2	0,97 97%	1,15 98%	1,06 98%	1,07 98%	0,99 98%	0,96 98%	3,20 49%	3,56 51%	8,76 54%
3	1,04 97%	1,05 97%	1,09 96%	1,04 97%	1,06 97%	1,12 97%	34,3 45%	38,3 44%	10,7 46%
5	1,32 92%	1,15 93%	1,24 96%	1,26 93%	1,12 93%	1,26 95%	537,1 3%	506,4 3%	15,8 30%
7	1,11 92%	1,22 94%	1,19 93%	1,22 92%	1,09 92%	1,07 93%	832,1 28%	861,4 3%	22,3 19%
11	1,17 87%	1,13 88%	1,30 87%	1,02 90%	1,11 86%	1,24 88%	160,7 0%	726,8 6%	30,0 45%
17	1,53 81%	1,64 80%	1,59 83%	1,58 81%	1,45 81%	1,51 84%	1277,7 13%	3554,4 87%	47,1 34%
25	1,95 72%	1,81 72%	1,95 73%	2,10 73%	1,89 75%	1,86 75%	8498,8 2%	358136 2%	71,9 25%
38	2,36 64%	2,46 63%	2,34 64%	2,58 65%	2,64 61%	2,28 64%	∞	∞	104,6 14%
57	4,33 47%	3,73 49%	3,68 51%	3,86 49%	3,87 50%	3,52 49%	∞	∞	154,6 44%
86	5,39 36%	5,31 37%	5,13 36%	6,14 34%	5,33 38%	5,89 35%	∞	∞	173,2 32%
129	11,7 22%	10,9 20%	12,6 20%	10,9 22%	12,9 20%	12,4 20%	∞	∞	293,9 21%
194	45,2 8%	34,7 10%	36,1 10%	48,8 11%	28,1 11%	39,8 10%	∞	∞	322,8 14%

Таблица 5. Распределение Стьюдента. IF, уровень принятия

Размерность	LTG (H,G)	LTG (G)	LTG (0)	MALA (H,G)	MALA (G)	MALA (0)	MALA-diag (G)	MALA-diag (0)	RW
2	1,81 46%	9,96 81%	8,47 70%	32,0 32%	11,2 82%	7,98 72%	20,6 50%	31,1 39%	23,8 69%
3	1,31 41%	12,1 77%	10,8 60%	55,1 27%	12,9 74%	12,1 61%	118,5 22%	289,8 24%	31,4 58%
5	2,57 24%	15,1 67%	18,8 45%	81,2 12%	13,8 66%	14,2 47%	445,6 13%	3033,8 5%	40,1 53%
7	4,51 16%	15,8 60%	21,1 35%	277,6 5%	15,8 59%	18,9 39%	7367,4 40%	17191,1 19%	45,9 39%
11	16,6 7%	17,8 54%	36,2 26%	3354,9 2%	19,2 49%	42,6 21%	2057,7 0%	19304,0 24%	55,0 28%
17	1443,0 1%	22,8 40%	56,2 12%	2474,9 0%	22,8 40%	52,9 11%	539,2 19%	3395,5 4%	63,3 19%
25	97401,8 0%	32,5 34%	245,7 6%	∞	34,5 32%	187,5 6%	∞	∞	115,8 13%
38	∞	39,7 24%	701,4 46%	∞	41,2 23%	544,0 28%	∞	∞	110,7 14%
57	∞	59,8 15%	762,8 63%	∞	56,7 15%	604,3 16%	∞	∞	145,7 65%
86	∞	152,4 7%	1006,1 24%	∞	139,8 21%	874,1 67%	∞	∞	162,7 57%
129	∞	261,0 28%	1504,2 14%	∞	257,6 12%	881,8 54%	∞	∞	170,2 50%
194	∞	230,4 20%	1351,7 39%	∞	428,5 40%	2389,9 48%	∞	∞	306,4 39%

плотности (пусть и урезанной) аппроксимации показывают себя очень хорошо и, соответственно, дают результаты, аналогичные использующим информации о градиенте и гессиане. Тем не менее с ростом размерности пространства растет вероятность оказаться вблизи границы носителя функции, в результате чего снижается оптимальный уровень принятия. Результаты MALA и LTG также очень близки друг к другу.

Результаты для распределения Стьюдента представлены в табл. 5 (0 — указывает на аппроксимации градиента и гессиана). Еще более плоское, чем предыдущие, распределение с более тяжелыми хвостами усугубило особенности, наблюдавшиеся для гамма-распределения. Привлечение данных о гессиане ведет к провальным результатам, за исключением малых размерностей. Это неудивительно, учитывая нарушение глобальной выпуклости логарифма функции плотности. Результаты RW ухудшаются, но в меньшей степени, чем у других алгоритмов (особенно в случае большой размерности). Алгоритмы, использующие данные о градиенте (или его аппроксимации), демонстрируют хорошие результаты при малой и средней размерности пространства, но при большой сравниваются (для LTG) или начинают уступать (для MALA) подходу RW.

Минимальная неэффективность выборки достигается при уровне принятия, заметно выходящем за пределы интервала 23,4–40%, причем это наблюдается и на большой, и на малой размерности пространства.

Результаты для распределения вида X представлены в табл. 6. Показатели RW существенно ухудшаются, а для алгоритмов с информацией о гессиане оказываются намного лучше. Алгоритмы с данными о градиенте также многократно обыгрывают RW и сильно уступают методам с гессианом. А вот методы, использующие аппроксимацию градиента, демонстрируют крайне слабые результаты, что, по всей видимости, связано с неудачностью аппроксимации гессиана (гессиан вдали от мода сильно отличается от гессиана в моде). Можно отметить, что MALA(H, G) дает практически единичную эффективность, обыгрывая LTG(H, G), но в случае отсутствия данных о гессиане уже LTG начинает обыгрывать MALA-алгоритм.

Результаты для смеси распределения представлены в табл. 7. Результаты RW несильно ухудшаются по сравнению с результатами компонентов смеси. Результаты MALA и LTG во всех версиях для малой и средней размерности заметно лучше, чем у RW. Однако для большой размерности резко растет неэффективность при использовании гессиана и чуть медленнее — при градиенте.

Таблица 6. Распределение X. IF, уровень принятия

Размерность	LTG (H, G)	LTG (G)	LTG (0)	MALA (H, G)	MALA (G)	MALA (0)	MALA-diag (G)	MALA-diag (0)	RW
2	1,96 65%	79,9 87%	173,7 26%	2,01 45%	97,3 63%	180,0 26%	352,3 88%	887,3 68%	140,9 82%
3	2,70 74%	117,4 62%	250,5 25%	1,32 66%	140,5 48%	295,9 28%	3837,8 24%	6925,8 2%	169,3 46%
5	1,73 87%	87,8 65%	493,8 14%	1,10 91%	123,8 43%	570,0 15%	30451,7 4%	1757,0 0%	475,4 56%
7	1,97 90%	59,5 79%	609,8 8%	1,00 98%	136,6 40%	413,9 7%	11648,5 0%	13947,6 13%	381,9 22%
11	6,56 91%	59,1 72%	3014,7 44%	1,04 100%	126,5 35%	3224,6 2%	∞	∞	587,9 16%
17	7,67 95%	50,5 64%	5333,2 10%	1,15 100%	142,9 33%	1921,2 90%	∞	∞	916,4 65%
25	1,89 92%	49,4 74%	3840,4 50%	1,12 100%	157,5 30%	3444,0 65%	∞	∞	1202,2 20%
38	5,58 74%	92,9 49%	9001,7 42%	1,19 100%	91,1 80%	9951,8 47%	∞	∞	451,9 19%
57	2,17 79%	36,1 58%	3059,3 34%	1,12 100%	166,4 23%	2959,2 46%	∞	∞	1301,6 38%
86	2,92 75%	73,4 68%	9325,0 21%	1,18 100%	62,8 69%	5337,5 33%	∞	∞	888,4 45%
129	2,30 72%	18,2 62%	20349,5 23%	1,12 100%	616,8 11%	6748,9 32%	∞	∞	2077,6 23%
194	3,91 51%	70,0 50%	35326,6 6%	1,13 100%	73,1 55%	26331,2 60%	∞	∞	1467,8 21%

По всей видимости, это связано с ростом вероятности, что одна из компонент оказывается в зоне тяжелых хвостов. Сравнивая LTG и MALA, можно наблюдать очень близкие результаты.

Уровень принятия оказывается в еще более широких диапазонах, чем в рассмотренных ранее случаях. Также есть общая динамика по снижению уровня принятия с ростом размерности пространства.

Итак, оказалось, что MALA-diag дает ужасные результаты для всех распределений. Результаты RW близки почти для всех рассматриваемых случаев (исключение — нарушение выпуклости функции плотности). Для малой (а часто и средней) размерности методы MALA и LTG оказываются эффективнее, чем RW. Отклонение от лог-квадратичного вида плотности (наличие тяжелых хвостов или острых пиков) ведет к снижению эффективности использования информации о гессиане. Данные об аппроксимации градиента ведут к ухудшению результатов в случае тяжелых хвостов. Результаты LTG очень близки к MALA, но часто немного их превосходят (в среднем отношение IF MALA к IF LTG составляет 5,33 для гессиана, 1,75 — для градиента и 0,98 — для аппроксимации градиента). Что касается уровня принятия, то оно заметно отклоняется от асимптотически оптимального уровня во многих случаях.

Существенным дополнительным фактором при выборе алгоритма MCMC являются вычислительные затраты. Для генерации 1000 наблюдений время расчетов отличается несильно, например, для нормальной плотности (при размерности 194) оно составляет 0,84 мин для версий с гессианом и 0,55–0,57 мин для остальных версий, а для смеси 1,63–1,68 мин для версий с гессианом и 1,19–1,38 мин для остальных. Это связано с тем, что основные вычислительные затраты, особенно для расчетов по ДСОЭР-моделям, связаны с расчетом функции апостериорной плотности. Например, для смеси размерностью 129 выборка LTG(G) в 16,9 раз эффективнее, чем RW. Однако если рассчитывать градиент методом конечных разностей, то время расчетов LTG(G) возрастает в 129 раз и уже RW становится в 7,6 раз эффективнее. Если же расчет градиента аналитически ведет к меньшему увеличению времени расчетов, преимущество LTG(G) сохранится. То есть для сравнения алгоритмов, применяющих гессиан/градиент и их не использующих, необходимо знать вычислительные затраты, связанные с их расчетом.

Отметим, что полученные показатели эффективности выборки для RW, на рассмотренных плотностях, намного лучше, чем величины, получающиеся для ДСОЭР-моделей (Chib, Ramamurthy, 2010). Поэтому посмотрим, какими будут результаты для ДСОЭР-модели: для этого возьмем модель

Таблица 7. Смесь распределений. IF, уровень принятия

Dim	LTG (H, G)	LTG (G)	LTG (0)	MALA (H, G)	MALA (G)	MALA (0)	MALA-diag (G)	MALA-diag (0)	RW
2	1,65 88%	1,32 91%	1,14 94%	1,70 87%	1,32 93%	1,22 94%	1,85 56%	1,70 57%	8,99 56%
3	1,77 81%	1,43 91%	1,26 93%	2,04 79%	1,34 88%	1,18 91%	4,46 67%	3,45 43%	12,9 46%
5	1,54 76%	1,32 85%	1,19 91%	1,99 76%	1,33 83%	1,25 90%	85,3 21%	83,5 18%	13,6 31%
7	2,48 71%	1,53 80%	1,44 89%	2,63 65%	1,54 81%	1,22 89%	680,6 2%	540,4 6%	14,7 30%
11	3,40 57%	2,49 75%	1,65 86%	4,65 55%	2,13 75%	1,53 86%	1358,7 45%	2818,1 9%	57,2 17%
17	6,97 40%	4,37 68%	2,98 83%	7,50 37%	3,96 70%	2,41 83%	2516,1 0%	1471,7 1%	61,1 17%
25	6,83 34%	5,06 62%	3,13 83%	10,5 24%	6,22 64%	3,16 84%	70469 2%	152489 2%	65,8 27%
38	9,70 25%	5,11 55%	2,96 77%	19,8 16%	4,65 55%	3,82 78%	∞	∞	98,6 18%
57	15,1 18%	5,16 48%	2,63 72%	28,5 11%	3,50 48%	2,99 73%	∞	∞	138,0 10%
86	70,9 9%	6,44 37%	2,85 67%	175,6 21%	5,99 41%	2,77 67%	∞	∞	125,5 39%
129	2160,3 0%	15,9 25%	5,68 58%	944,3 11%	21,3 23%	4,62 56%	∞	∞	268,2 30%
194	4680,4 0%	201,2 12%	8,02 50%	942,7 27%	147,7 13%	9,28 45%	∞	∞	384,5 72%

Ивашенко (Ивашенко, 2018) с теми же статистическими данными. Изначально модель оценивалась методом максимального правдоподобия. Однако несколько параметров располагаются очень близко к границе допустимых значений, что усложняет работу МСМС-алгоритмов сверх обычного уровня. Соответственно, априорное распределение параметров было изменено (нормальное распределение

Таблица 8. Характеристики МСМС для ДСОЭР-модели России

Мера	МСМС	Цепь					Среднее по цепям	Медиана по цепям
		1	2	3	4	5		
Уровень принятия	LTG	0,82%	2,33%	1,25%	2,78%	1,25%	1,68%	1,25%
	MALA	21,37%	19,91%	11,46%	13,63%	20,67%	17,41%	19,91%
	RW	18,21%	17,55%	17,35%	17,89%	17,00%	17,60%	17,55%
Максимум (IF) по перемен.	LTG	3,69E+4	8,49E+4	7,43E+4	3,45E+4	1,17E+5	6,95E+4	7,43E+4
	MALA	5,19E+4	1,24E+6	2,29E+6	3,52E+5	3,32E+5	8,54E+5	3,52E+5
	RW	1,41E+6	1,86E+9	2,41E+5	3,85E+4	3,60E+4	3,73E+8	2,41E+5
Среднее (IF) по перемен.	LTG	6,67E+3	7,21E+3	6,25E+3	3,88E+3	8,46E+3	6,49E+3	6,67E+3
	MALA	5,94E+3	3,30E+4	1,03E+5	3,79E+4	1,59E+4	3,92E+4	3,30E+4
	RW	5,21E+4	4,44E+7	7,16E+3	2,21E+3	1,92E+3	8,90E+6	7,16E+3
Медиана (IF) по перемен.	LTG	2,60E+3	3,28E+3	2,63E+3	9,07E+2	2,39E+3	2,36E+3	2,60E+3
	MALA	3,29E+3	1,51E+3	4,68E+3	7,82E+3	1,74E+3	3,81E+3	3,29E+3
	RW	3,54E+3	8,36E+3	3,37E+2	3,83E+2	2,51E+2	2,58E+3	3,83E+2
Минимум (IF) по перемен.	LTG	4,24E+1	1,42E+1	2,89E+1	1,04E+0	2,77E+1	2,28E+1	2,77E+1
	MALA	7,03E+2	2,31E+0	7,83E+2	9,16E+2	2,26E+0	4,81E+2	7,03E+2
	RW	2,86E+2	2,64E+2	1,18E+0	1,00E+0	1,00E+0	1,11E+2	1,18E+0

Примечание. Лучший результат по каждой мере выделяется полужирным шрифтом.

со средним, соответствующим оценке ММП и дисперсией для каждого параметра, соответствующего дисперсии оценок ММП). Вдобавок 10 из 52 параметров было зафиксировано. После этого было построено 5 цепей по 50 000 наблюдений (первые 10 000 наблюдений не учитывались) каждым из трех методов (RW, MALA(0), LTG(0)). Результаты представлены в табл. 8.

Анализируя данные в табл. 8, можно увидеть преимущество LTG-алгоритма, однако оно наблюдается лишь в 4 из 5 цепей. Если сравнивать MALA и RW, ситуация неоднозначная — исходя из неэффективности в среднем, преимущество за MALA-алгоритмом, а если брать неэффективность средней цепи, то небольшое преимущество у RW-алгоритма. Если же смотреть на неэффективности не худшего, а среднего параметра (медиану по параметрам), показатели окажутся ближе друг к другу и может показаться (исходя из медианы по цепям), что RW дает лучшие результаты. Однако сходимость правильнее измерять по худшему из параметров.

5. ЗАКЛЮЧЕНИЕ

В данной работе проводится анализ методов байесовской эконометрики, использующихся для оценки ДСОЭР-моделей. А точнее, свойств МСМС-алгоритмов для мономодальных функций плотности, обладающих неблагоприятными свойствами, встречающимися в апостериорных плотностях, характерных для ДСОЭР-моделей. Такими свойствами являются: носитель плотности отличается от R^n (наличие штрафного барьера); тяжелые хвосты; острые пики; нарушение выпуклости логарифма плотности или функции плотности. Рассматриваются версии МСМС: стандартный алгоритм случайного блуждания (RW); алгоритм MALA с версиями, использующими локальный гессиан, градиент или аппроксимацию градиента; предложенный алгоритм LTG с версиями, аналогичными MALA.

Было выявлено, что оптимальный уровень принятия, т.е. соответствующий наименьшей неэффективности выборки, отклоняется от асимптотически оптимального для случайного блуждания (23,4%) уровня в достаточно широких пределах. Эффективность алгоритма RW является наиболее стабильной, что объясняет стандартность такого выбора. Однако алгоритмы MALA и LTG демонстрируют большую эффективность практически во всех версиях для малой и средней размерности пространства. Для пространства большой размерности при наличии тяжелых хвостов версии с гессианом демонстрируют плохие результаты, но версии только с градиентом или его аппроксимацией превосходят RW (хотя масштаб преимущества может сильно различаться). В случае использования аппроксимации градиента (или быстрого аналитического его расчета) переход на алгоритмы MALA или LTG не связан с дополнительными вычислительными затратами.

Предложенный алгоритм LTG дает в среднем чуть более эффективную выборку по сравнению с аналогичными версиями MALA для рассмотренных плотностей и их размерностей (в большинстве случаев результаты очень близки). Проверка на ДСОЭР-модели также демонстрирует преимущество LTG.

Данные результаты могут быть полезны применительно не только к ДСОЭР-моделям, но и другим областям, где встречаются мономодальные плотности с аналогичными неудобными свойствами.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- Ивашенко С.М.** (2018). Экономическая политика России: модель с дискреционной политикой или с инструментальными правилами // *Вестник СПбГУ. Экономика*. Т. 34. Вып. 1. С. 149–172. [**Ivashchenko S.M.** (2018). Russian economic policy: A model with discretionary policy or policy rules. *St. Petersburg University Journal of Economic Studies*, 34, 1, 149–172 (in Russian).]
- Adjemian S., Bastani H., Juillard M., Karame F., Mihoubi F., Perendia G., Pfeifer J., Ratto M., Villemot S.** (2011). Dynare: Reference manual, Version 4. *Dynare Working Papers*, 1, CEPREMAP.
- Bedard M.** (2007). Weak convergence of Metropolis algorithms for non-i.i.d. target distributions. *Annals of Applied Probability*, 17, 4, 1222–1244.
- Blanchard O., Kahn C.** (1985). The solution of Linear Difference Models under rational expectations. *Econometrica*, 45, July, 1305–1311.

- Chib S., Ramamurthy S.** (2010). Tailored randomized block MCMC methods with application to DSGE models. *Journal of Econometrics*, 155, 1, 19–38.
- Chopin N.** (2011). Fast simulation of truncated Gaussian distributions. *Statistics and Computing*, 21, 2, 275–288.
- Christiano L.J., Eichenbaum M., Trabandt M.** (2016). Unemployment and business cycles. *Econometrica*, 84, 1523–1569.
- Cotter S.L., Roberts G.O., Stuart A.M., White D.** (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statistical Science*, 28, 3, 424–446.
- Durmus A., Roberts G.O., Vilmart G., Zygalkis K.C.** (2017). Fast Langevin based algorithm for MCMC in high dimensions. *Annals of Applied Probability*, 27, 4, 2195–2237.
- Hammad Q.** (2014). Explosive roots in level vector autoregressive models and vector error correction models. *SSRN Electronic Journal*, 10.2139/ssrn.2652306.
- Herbst E., Schorfheide F.** (2015). *Bayesian estimation of DSGE models*. Princeton: Princeton University Press.
- Iskrev N.** (2010). Local identification in DSGE models. *Journal of Monetary Economics*, 57, 2, 189–202.
- Ivashchenko S., Mutschler W.** (2020). The effect of observables, functional specifications, model features and shocks on identification in linearized DSGE models. *Economic Modelling*, 88, June, 280–292.
- Jarner S., Roberts G.** (2007). Convergence of heavy-tailed Monte Carlo Markov Chain Algorithms. *Scandinavian Journal of Statistics*, 34 (4), 781–815.
- Lucas R.E.** (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1, 1, 19–46.
- Schmitt-Grohe S., Uribe M.** (2004). Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics and Control*, 28, 4, 755–775.
- Smets F., Wouters R.** (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review*, 97 (3), 586–606.
- Sherlock C., Roberts G.** (2009). Optimal scaling of the random walk Metropolis on elliptically symmetric unimodal targets. *Bernoulli*, 15, 3 (August), 774–798.
- Titsias M.K., Dellaportas P.** (2019). *Gradient-based adaptive Markov chain Monte Carlo*. arXiv:1911.01373
- Tovar C.E.** (2009). DSGE models and central banks. *Economics-The Open-Access, Open-Assessment E-Journal*, 3 (16), 1–31.

Computational efficiency of Bayesian estimation techniques for “unfavorable” density

© 2021 S.M. Ivashchenko

S.M. Ivashchenko,

The Institute of Regional Economy Studies (Russian Academy of Sciences), Saint Petersburg; Financial Research Institute, Ministry of Finance, Moscow; Saint-Petersburg State University, Saint Petersburg, Russia; e-mail: sergey.ivashchenko.ru@gmail.com; glucke_ru@pisem.net

Received 05.10.2020

The reported study was partly funded by RFBR according to the research (project 18-010-01185) «Structural changes in Russia: The role of human capital and investments».

The author expresses personal ideas, which may not correlate with the official position of the Bank of Russia. The Bank of Russia carries no responsibility for the content of the article.

Abstract. The Bayesian estimation of dynamic stochastic general equilibrium (DSGE) models implies usage of Monte Carlo Markov Chain (MCMC) algorithms. The analysis of MCMC algorithms is made for densities with unfavorable properties common in DSGE models (restricted density support, heavy tails, sharp peaks, non-convex log-density, and non-convex density). Three groups of algorithms are investigated: random walk (RW), MALA and suggested LTG (local truncated Gauss). Three versions of MALA and LTG are investigated: the version using local hessian and gradient of log-density, the version using only local gradient and the version that uses only information about the mode. Performance of MALA and LTG are close to each other. There is some advantage of LTG in average and in test with DSGE model. RW performance is worse than MALA or LTG (especially for small-scale cases). The computational costs are almost the similar to RW and approximation based versions of MALA or LTG. Existence of heavy tails leads to decrease of advantage of MALA and LTG algorithms. Acceptance rate (corresponding to the lowest inefficiency of sample) can be quite different from conventional values.

Keywords: MCMC, Markov chain Monte Carlo, Metropolis–Hastings algorithm, Bayesian econometrics techniques.

JEL Classification: C11.

DOI: 10.31857/S042473880014916-6