Размышления об инструментальных переменных*

Кристофер Симс†

Принстонский Университет, Принстон, США

Рассматривается вопрос об использовании метода инструментальных переменных и метода моментов с точки зрения теории принятия решений. Априорные веры играют чрезвычайно важную роль, поэтому в случаях, когда инструменты, возможно, «слабые», или их количество велико относительно количества наблюдений, важно описать особенности правдоподобия вне рамок ИП- или ММП-оценивания и соответствующих асимптотических (т.е. локально квадратичных) стандартных ошибок. ИП и ОММ привлекательны из-за удобства вычислений в большинстве случаев и (ложного) убеждения, что они применимы при малом числе «предположений». Здесь обсуждаются подходы, позволяющие сделать последнее утверждение более обоснованным.

Ключевые слова: байесовский подход, обобщенный метод моментов, инструментальные переменные, слабые инструменты, выбор инструментов, энтропия

Классификация JEL: C11, C13, C44

1 Введение

Ситуации, где применяются ИП и ОММ, хорошо иллюстрируют ценность байесовского подхода, поскольку в этих случаях, в отличие от большинства ситуаций, при больших выборках трактовка байесовским анализом правдоподобия как апостериорной функции распределения (при плоском априорном распределении) оказывается не всегда эквивалентна классической (небайесовской) трактовке вероятных спецификаций до получения выборки, как если бы они были полноценными вероятными спецификациями после получения выборки.

Байесовский подход к выдаче научных результатов рассматривает весь анализ данных как отчет о форме функции правдоподобия в виде, наиболее удобном для читателя. В этом эссе мы рассматриваем инференцию, основанную на методе моментов, с этой точки зрения.

2 Литература

Описание механики байесовского анализа инструментальных переменных недавно пополнилось новыми результатам, в частности, хорошо отраженными в Geweke (1996) и Kleibergen & Zivot (2000). В ряде недавних работ, куда входят наряду с прочими Kitamura & Stutzer (1997), Zellner, Tobias & Ryu (1997) и Kim (2002), авторы пытаются «вывести» ИП, в некоторых случаях вместе с послевыборочными апостериорными распределениями, используя «автоматические» методы информационной теории с целью избежать использования явных априорных распределений. В одной из работ Филлипс и Чао рекомендуют использовать априорные распределения Джеффриза – другой класс «автоматически» генерируемых априорных распределений.

Исследования с использованием байесовской техники развивают понимание необычных особенностей функций правдоподобия, возникающих в моделях с одновременностью из-за

*Перевод С. Анатольева и А. Белякова. Цитировать как: Симс, Кристофер (2007) «Размышления об инструментальных переменных», Квантиль, №2, стр. 83–94. Citation: Sims, Christopher A. (2007) “Thinking About Instrumental Variables,” Quantile, No.2, pp. 83–94.

†Адрес: Department of Economics, Princeton University, 104 Fisher Hall, Princeton, NJ 08544-1021, USA. Электронная почта: *sims@princeton.edu*

полюсов и нулей у отображения между приведенной формой (ПФ) и структурными параметрами. Часть этого эссе посвящена примерам, показывающим важность таких особенностей на практике. Исследования «автоматических» априорных и апостериорных распределений мотивированы привлекательностью утверждения, что ОММ и ИП требуют немного предположений, и одновременно неудовлетворенностью этим постулатом на практике. Данное эссе начинается с обсуждения слабых предположений.

3 Уклонение от предположений

Какие бы предположения не делались, все они не бесспорные, поэтому эконометристу всегда удобно иметь возможность заявить на критику профессиональной аудитории, что его выводы практически не зависят от каких-либо предположений. Модели, в которых сделано мало предположений, часто называют «непараметрическими», в отличие от перегруженных предположениями параметрических моделей. Некоторое время назад, когда спектральные методы впервые появились в эконометрике, вначале считалось, что те позволяют анализировать временные ряды, используя намного более слабые предположения, чем параметрические ARIMA-модели, которые были и все еще остаются общепринятыми в эконометрике. В обзорной статье в 1974 году (Sims, 1974) мной было показано, что это впечатление обманчиво. Спектральный метод приводит к *асимптотической* теории с малым количеством предположений, но при этом статистики частотной области сглаживаются посредством окон или ядер, сужающихся по мере роста выборки, а также предполагается гладкость представляющих интерес объектов частотных областей. Понятно, что существуют функции в классе интересующих нас объектов, которые допустимы при слабых предположениях, необходимых для вывода асимптотической теории, но они становится недопустимыми, если мы всерьез воспринимаем утверждения о неопределенности (о доверительных интервалах, о тестировании гипотез и т.п.) в имеющейся выборке. При любой ширине окна ядра будут существовать «гладкие» функции, которые осциллируют слишком сильно, чтобы их точно оценить ядром с этой шириной окна. Мы интуитивно применяем эти ограничения при оценке результатов, даже если они не обсуждаются в исследовательских отчетах. Эти проблемы применительно к тестированию гипотез обсуждаются в работе Faust (1999).

В сфере временных рядов быстро стало понятно, что использующие ядра методы спектрального анализа не обладают большей общностью. ARIMA-модели, в которых количество параметров систематически растет с увеличением размера выборки или как функция от данных, могут привести к асимптотической теории настолько же общей, как и методы спектрального анализа. Класс моделей временных рядов, которые можно хорошо аппроксимировать последовательностью параметрических моделей, не больше, в топологическом смысле, чем класс моделей, удовлетворяющих условиям гладкости, подразумевающимся в методах спектрального анализа. Гладкость функций в функциональном пространстве – предположение того же рода, что и норма затухания последовательностей в пространстве последовательностей.

Вышеописанные проблемы выявляются в той же форме в теории непараметрического оценивания регрессионных функций или функций плотности. По моему впечатлению, все-таки практики вне области временных рядов менее осведомлены о том, что явно непараметрическая модель – это всего лишь еще одна версия того, какая именно модель будет хорошо работать в конкретном приложении, а не способ получения результата с меньшим числом предположений.[1]

[1] Возможно, это не из-за того, или не только из-за того, что эконометристы, работающие со временными рядами, умнее. Спектральные модели во временных рядах очень неуклюжи, если требуется спрогнозировать будущее по данным на текущий момент. Так как это стандартная задача в прикладном анализе временных рядов, для практиков было облегчением получить теоретическое доказательство того, что непараметриче-

Проблема также возникает и в стандартных утверждениях, что анализ, основанный на МНК, ИП и ОММ, свободен от предположений о виде распределения возмущений. Эти заявления обычно используются для утверждений о неопределенности результатов, которые (для данных модели и выборки) оправданы только при более жестком ограничении класса распределений возмущений, чем это требуется в асимптотической теории. Во многих случаях выводы о распределениях, основанные на асимптотической теории, справедливы в точности при предположениях о нормальности ошибок и справедливы в приемлемом приближении для данной выборки, если возмущения близки к нормальным.

По сути, в конечном счете мы обосновываем предположение о нормальности, прибегая к асимптотической теории. Это наводит на мысль, что есть шансы, что это предположение робастно. Рассуждения с помощью энтропии говорят о том, что это предположение, тем не менее, может быть в некотором смысле и консервативным.

4 Модель

Рассмотрим модель

$$\underset{T\times 1}{y} = x\beta + \varepsilon = \underset{T\times k}{Z}\gamma\beta + \nu\beta + \varepsilon, \tag{1}$$

$$\underset{T\times 1}{x} = Z\gamma + \nu, \tag{2}$$

$$\mathbb{V}([\nu\beta + \varepsilon \;\; \nu]) = \Sigma. \tag{3}$$

Заметим, что ту же модель можно параметризовать «наоборот» без изменения класса возможных вероятностных моделей данных просто переменой мест y и x в этих формулах. Эту симметрию можно расширить, если записать $Y = [y \; x]$, $\Pi = [\gamma\beta \; \gamma]$ и $\eta = [\nu\beta + \varepsilon \; \nu]$, а модель – как

$$Y = Z\Pi + \eta. \tag{4}$$

В модели требуется, чтобы $k \times 2$ матрица Π имела ранг 1, и как только мы накладываем такое ограничение, модель в форме (4) пробегает то же множество вероятностных моделей для данных.

При использовании параметризации (1–2) правдоподобие не стремится к нулю при $\beta \to \infty$ и постоянном Σ, какой бы ни был размер выборки. Это происходит из-за того, что при очень больших β наилучшая подгонка достигается при очень малых $\|\gamma\|$, при этом γ выбирается так, чтобы $\beta\gamma$ обеспечивало бы наилучшую возможную подгонку в уравнении (1) для y. Это, конечно же, приводит в большинстве случаев к плохой подгонке в уравнении (2) для x, но в пределе при больших β и очень малых $\|\gamma\|$ правая часть уравнения для x будет приблизительно нулем. Таким образом, при $\beta \to \infty$ степень подгонки ухудшается, но только до определенного значения, которое не загоняет правдоподобие в нуль.

Это приводит к наличию ловушек для наивных байесовских подходов. Динамические Методы Монте-Карло (ДММК), применяемые непосредственно к правдоподобию, как если бы оно было ненормированной функцией плотности распределения, не сходятся, и аналитическое интегрирование параметров из правдоподобия может не дать результата или выдать несобственные маржинальные «апостериорные распределения».

Плоские априорные распределения для Π в (4) без ограничений на ее ранг (неограниченная приведенная форма, НПФ), напротив, дают интегрируемые апостериорные распределения,

ские модели врожденно не являются более общими. В непараметрических регрессиях, с другой стороны, локальное свойство ограничений гладкости, неявно присутствующее в ядерных методах, более обоснованно, чем ограничения, налагаемые, скажем, ортогональными полиномиальными, параметрическими моделями.

если размер выборки не слишком мал. Таким образом, разумно использовать Евклидово расстояние для введения метрики на подмногообразие матриц Π с уменьшенным рангом, после чего преобразовать плоские априорные распределения (по мере Лебега) на этом подмногообразии к координатам β и γ. Этот вывод сам по себе не гарантирует, что апостериорные распределения при этих несобственных априорных распределениях будут собственными, но это многообещающий подход. Несобственное априорное распределение для β и γ, возникающее в этом подходе, есть

$$\left|\frac{\partial\Pi}{\partial(\beta,\gamma)}\left(\frac{\partial\Pi}{\partial(\beta,\gamma)}\right)'\right|^{\frac{1}{2}} = \|\gamma\|\left(1+\beta^2\right)^{\frac{1}{2}}. \tag{5}$$

Это действительно приводит к собственным апостериорным распределениям.

Даже при таких априорных распределениях хвосты апостериорных распределений стремятся к нулю лишь с полиномиальной скоростью, причем степень этого полинома не увеличивается с ростом выборки. В линейных регрессионных моделях же с плоскими априорными распределениями, наоборот, хвосты уменьшаются с полиномиальной скоростью, которая увеличивается линейно с размером выборки.

Обсудим теперь, к чему это приводит при комбинировании априорных распределений с Гауссовскими хвостами или с хвостами, убывающими как полином высокого порядка, с правдоподобием, взвешенным с помощью (5). Пусть вначале априорное среднее и пик правдоподобия совпадают, а затем пик правдоподобия удаляется от априорного среднего, при этом формы правдоподобия и априорной плотности распределения остаются теми же. Тогда апостериорные среднее, медиана и мода сначала будут удаляться от априорного среднего (как ожидалось), но затем поменяют направление и в конце концов совпадут с априорным средним, когда пик правдоподобия окажется далеко от априорного среднего. Это ситуация изображена на Рис. 1.

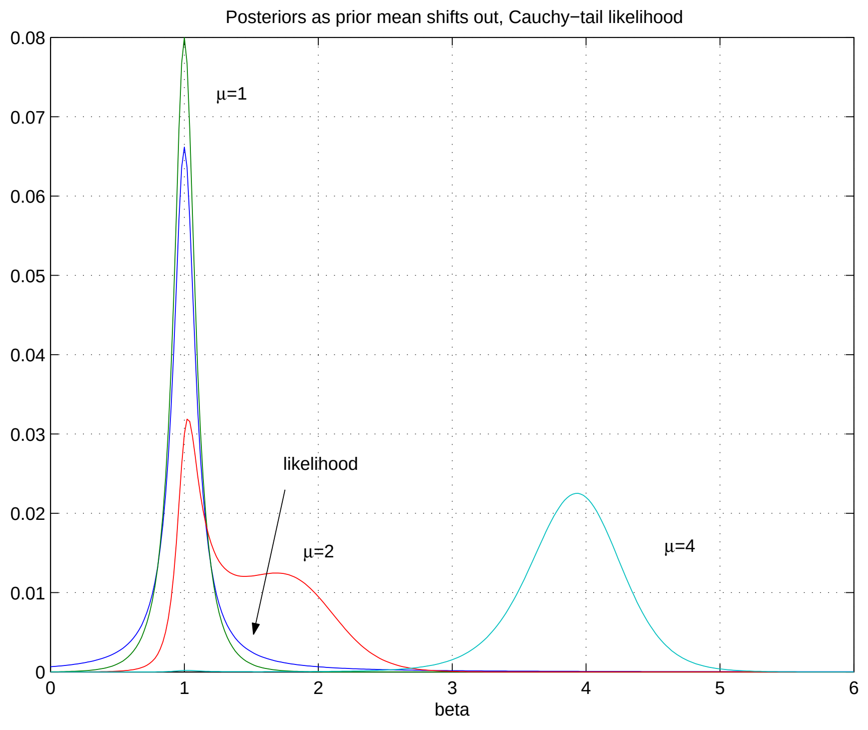

Рис. 1: Апостериорное распределение при увеличивающемся расстоянии между ММП-оценкой и априорным средним. Априорное t-распределение с 9-ю степенями свободы и $\sigma = 1/3$. Функция плотности – Коши с $\sigma = 0{,}1$, центрирована в 1.

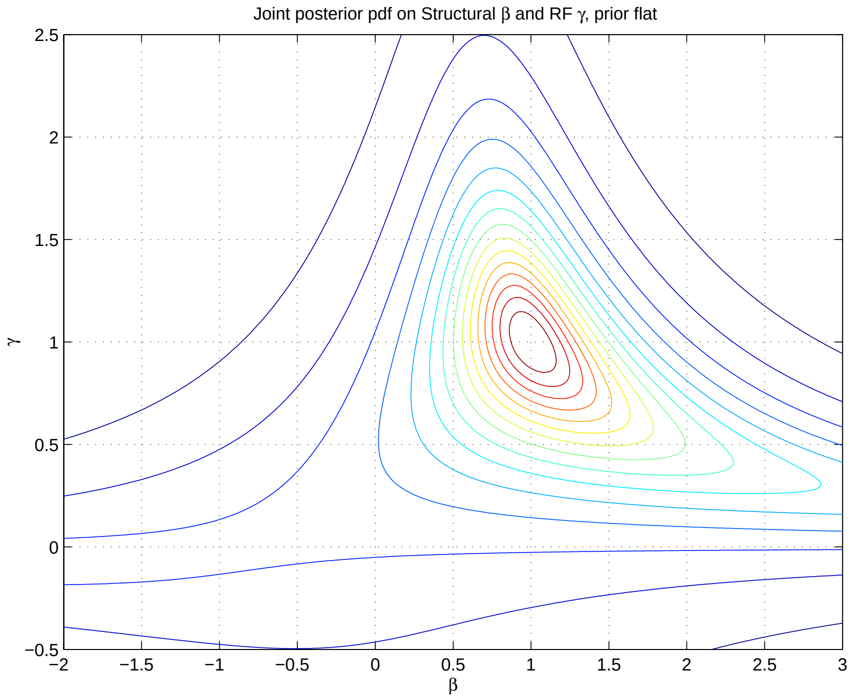

Рис. 2: Совместная плотность распределения структурного параметра β и ПФ-параметра γ. Плоское априорное распределение. Кроме того, $Z'Z = 4$, $\hat{\Sigma} = \begin{pmatrix} 0{,}67 & 0{,}33 \\ 0{,}33 & 0{,}67 \end{pmatrix}$, $\hat{\beta} = \hat{\gamma} = 1$.

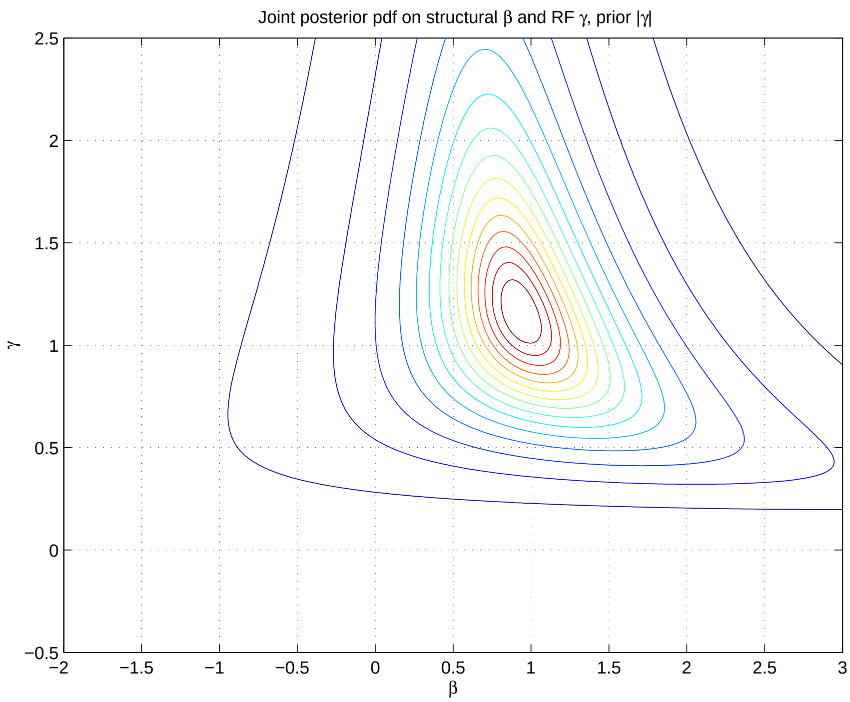

Рис. 3: Совместная плотность распределения структурного параметра β и ПФ-параметра γ. Априорное распределение для $|\gamma|$ (см. Рис. 2).

Другими словами, в модели инструментальных переменных тяжелые хвосты правдоподобия означают, что когда правдоподобие значительно отличается от априорного распределения, априорное распределение доминирует, даже если правдоподобие имеет пик достаточно резкий, чтобы заглушить априорную информацию вблизи него. Это имеет важные практические последствия для интерпретации ИП-оценок, которые отражены в общем признании того, что существует проблема «слабых инструментов». Но с байесовской точки зрения слабые инструменты – это не свойство популяции или даже выборки, а скорее свойство взаимосвязи выборки и априорных вер. ИП иногда дают «изменчивые» оценки, и практики это знают. Не существует, однако, способа «протестировать» оценки на изменчивость; надо учитывать априорные веры и факт, является ли выборочное правдоподобие в окрестности разумных величин для β действительно почти логарифмически плоским.

5 Совместные и маржинальные апостериорные распределения в случае точной идентификации

На Рис. 2 и 3 изображены линии уровня апостериорного распределения в случае точной идентификации, когда совместное апостериорное распределение легко выводится аналитически. Заметим его явно не эллиптическую форму, что означает, что априорная информация о γ будет сильно влиять на точность и форму апостериорного распределения β и наоборот. Также заметим, что в случае плоских априорных распределений для β и γ на графике видна довольно сложная картина вблизи $\|\gamma\| = 0$.

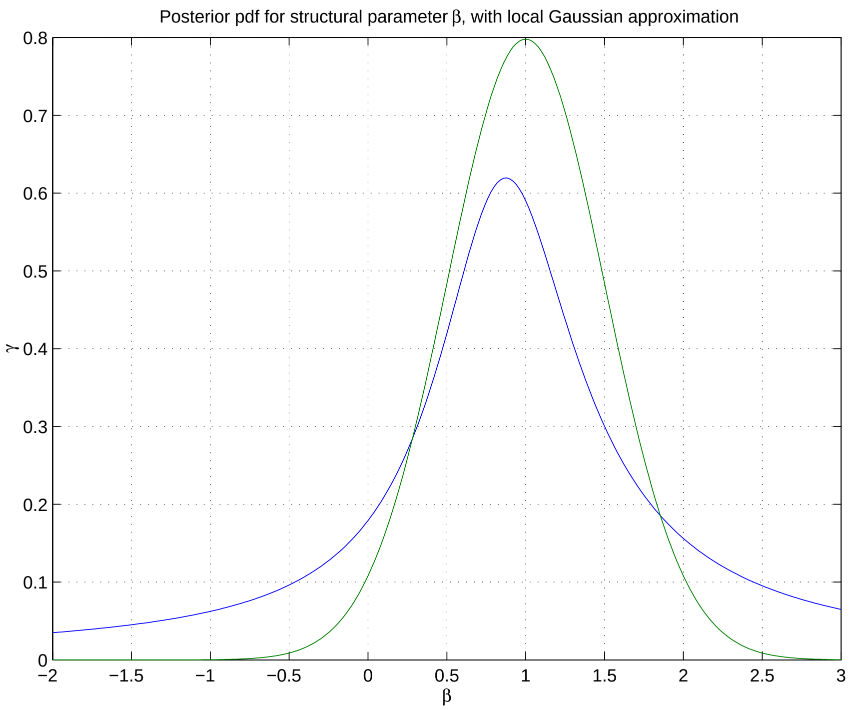

Рис. 4: Апостериорная функция плотности распределения для структурного параметра β. Локальная гауссовская аппроксимация (см. Рис. 2).

На Рис. 4 изображено маржинальное апостериорное распределение β, соответствующее Рис. 3. Заметим, насколько обманчива в этом случае локальная гауссовская аппроксимация.

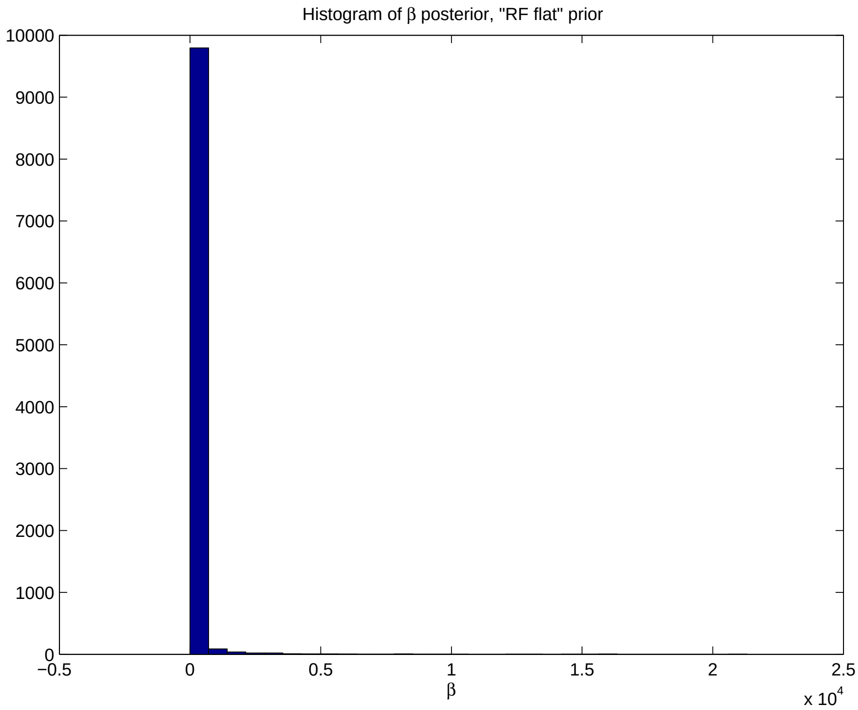

Рис. 5: Гистограмма апостериорного распределения β. Плоское априорное распределение для ПФ-параметра. Кроме того, $x = Z\begin{pmatrix}1\\0{,}1\end{pmatrix} + \varepsilon$, $y = Z\begin{pmatrix}1\\0{,}1\end{pmatrix} + v$, $\underset{20\times 2}{Z}, \varepsilon, v$ – все $N(0,1)$. Значение ИП-оценки $\hat{\beta} = 1{,}5132$, асимптотическая стандартная ошибка 0,8412.

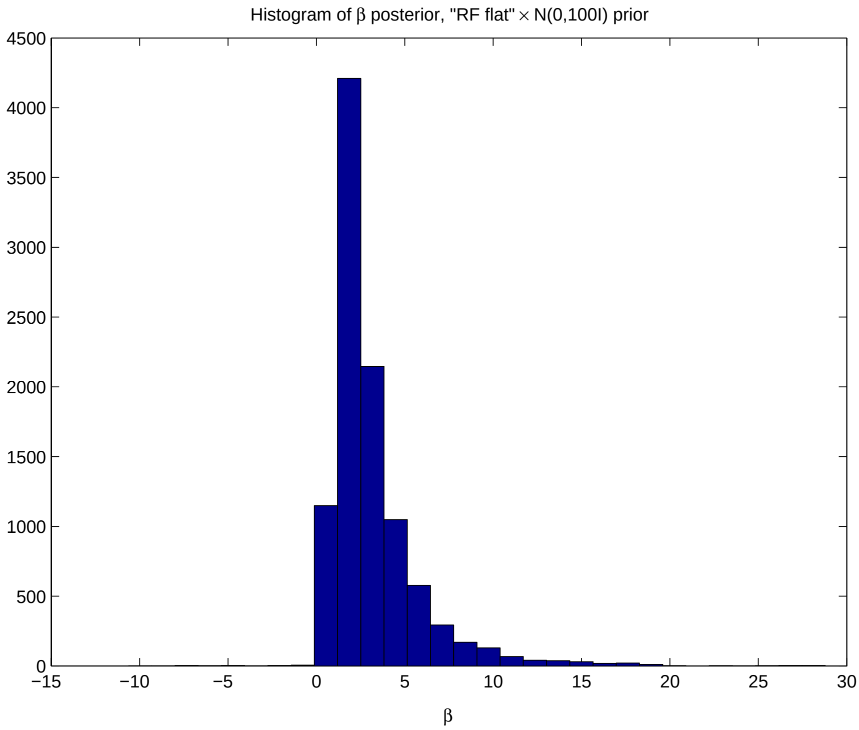

Рис. 6: Гистограмма апостериорного распределения β. Плоское априорное распределение $\times\, N(1, 100I)$ для ПФ-параметра. Модель и выборка как для Рис. 5.

6 Случай слабых инструментов

В выборке, где апостериорные распределения сильно не гауссовские со значительными и медленно спадающими хвостами, даже явная очень слабая априорная информация может существенно влиять на результаты анализа. На представленных далее графиках (Рис. 5–11) изображены эффекты наложения априорного распределения $N(0, 100I)$ на β и γ и в модели, и в выборке, что подразумевает оцененное с существенной неопределенностью значение β около 1. Даже такое слабое априорное распределение сильно влияет на апостериорное, в основном за счет устранения чрезвычайно вытянутых хвостов. Конечно же, апостериорные средние сильно зависят от добавления такой слабой априорной информации.

Для построения этих графиков я сгенерировал выборки размера 10000 при помощи ДММК, выбирая последовательно из условных правдоподобий для $\{\beta \,|\, \gamma, \Sigma\}$, $\{\gamma \,|\, \beta, \Sigma\}$ и $\{\Sigma \,|\, \gamma, \beta\}$, имеющих стандартную форму, и осуществляя один шаг алгоритма Метрополиса-Гастингса для отражения влияния априорного распределения. Когда априорное распределение не используется, в выборке возникают большие значения β, и когда β становиться большим, зависимость β от $\|\gamma\|$ очень сильная. Как известно, сильная зависимость приводит к медленной сходимости алгоритма Гиббса. Графики 10 и 11 показывают, как использование априорного распределения улучшает свойства сходимости алгоритма Гиббса путем устранения чрезвычайно больших экземпляров β.

7 Случай большого количества инструментов

Когда $T \leq k$, то есть когда нет степеней свободы в «регрессии на первом шаге», правдоподобие имеет два бесконечных пика: первый – там, где γ выбрана для идеальной подгонки x в (2), второй – где $\beta\gamma$ выбирается для идеальной подгонки y в (1). Обычно мы не считаем убедительными соответствующие оценки, которые основываются на МНК при прогоне регрессии y на x или при прогоне регрессии x на y, в последнем случае с оцениванием β обращением оценки коэффициента обратной регрессии. Причина в том, что мы на самом деле не верим, что любое из двух уравнений в НПФ имеет идеальную подгонку. Следовательно, имеет смысл попытаться отразить эти веры в априорном распределении, которое могло бы отвлечь от МНК-оценок.

Один из способов сделать это – использовать сопряженные априорные распределения, что можно осуществить добавлением «фиктивных наблюдений» в данные. Для определенности пусть $T = k = 20$, и зададим расширенные данные как

$$x^* = \begin{pmatrix} x \\ \underset{T\times 1}{0} \end{pmatrix}, \quad y^* = \begin{pmatrix} y \\ \underset{T\times 1}{0} \end{pmatrix}, \quad Z^* = \begin{pmatrix} Z \\ \lambda \underset{T\times T}{I} \end{pmatrix}. \tag{6}$$

Вычисления, лежащие в основе Рис. 12, соответствуют $\lambda = 0{,}5$. Естественно считать, что эти фиктивные наблюдения стремятся «прижать» результаты к нулю. Однако из-за того, что они связывают априорные дисперсии γ и $\beta\gamma$ возле нуля с дисперсией возмущения в уравнениях, они также ослабляют «идеальные подгонки». На рисунке 12 видно, что последний эффект доминирует, так как применение априорного распределения сдвигает апостериорное вверх от априорного среднего (и от МНК-оценки). Этот подход перспективен как способ избежать утилизации информации из-за неиспользования несомненно годных и доступных инструментальных переменных.

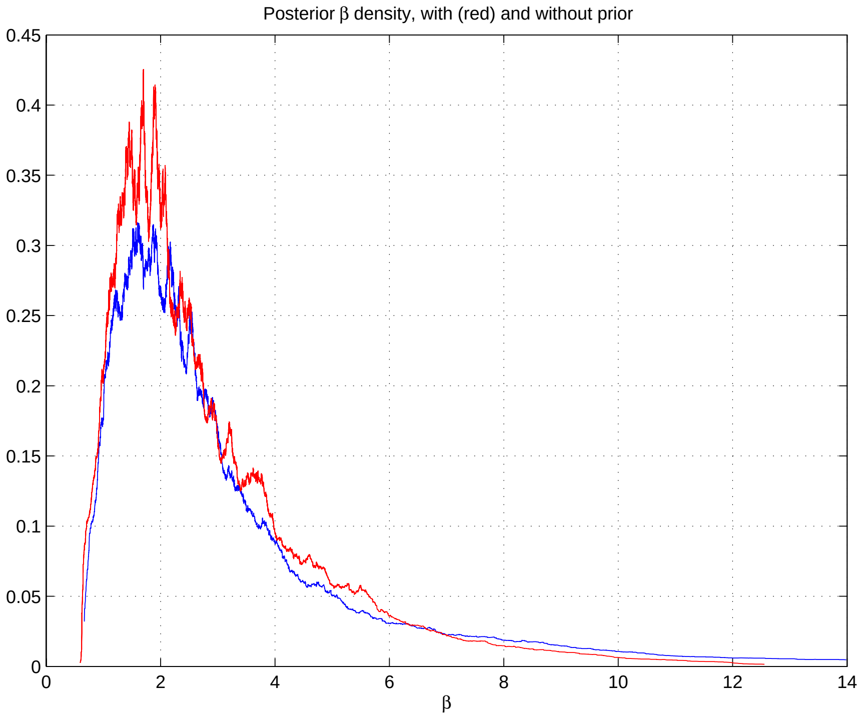

Рис. 7: Апостериорное распределение β при наличии априорного распределения (красным цветом) и без такового. Модель и выборка как для Рис. 5. Плотности оцениваются по методу «ближайших соседей», коих здесь 300.

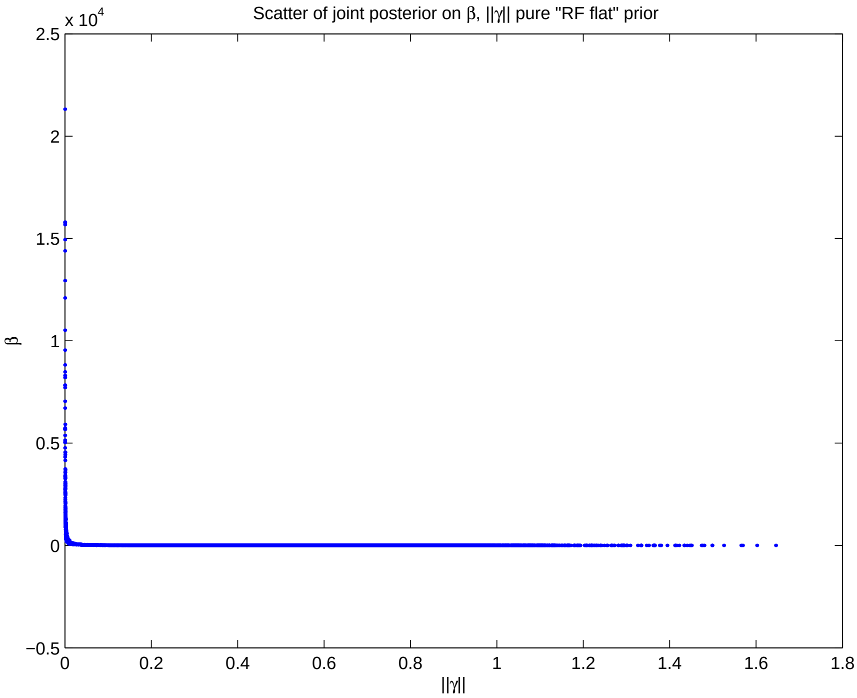

Рис. 8: Точки совместного апостериорного распределения β и $\|\gamma\|$ при плоском априорном распределении для ПФ-параметра. Модель и выборка как для Рис. 5.

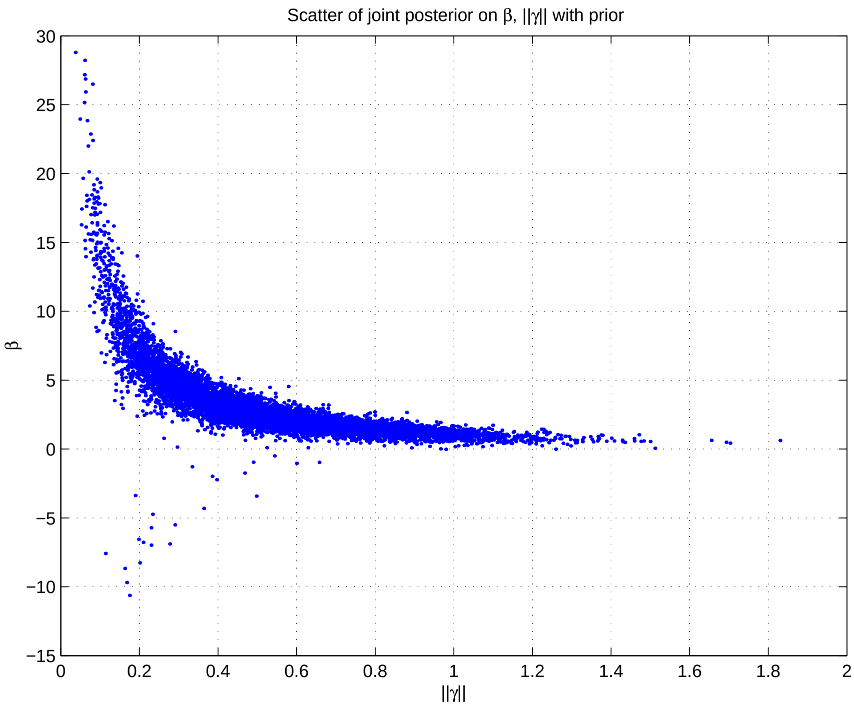

Рис. 9: Точки совместного апостериорного распределения β, $\|\gamma\|$ при априорном распределении для ПФ-параметра. Модель и выборка как для Рис. 5.

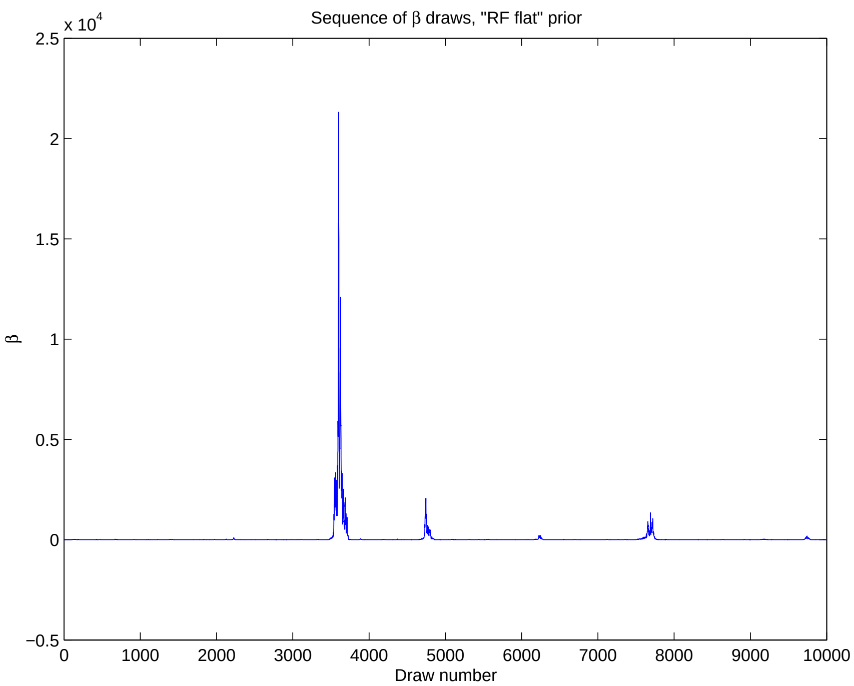

Рис. 10: Последовательность вытягиваний β. Плоское априорное распределение для ПФ-параметра. Модель и выборка как для Рис. 5.

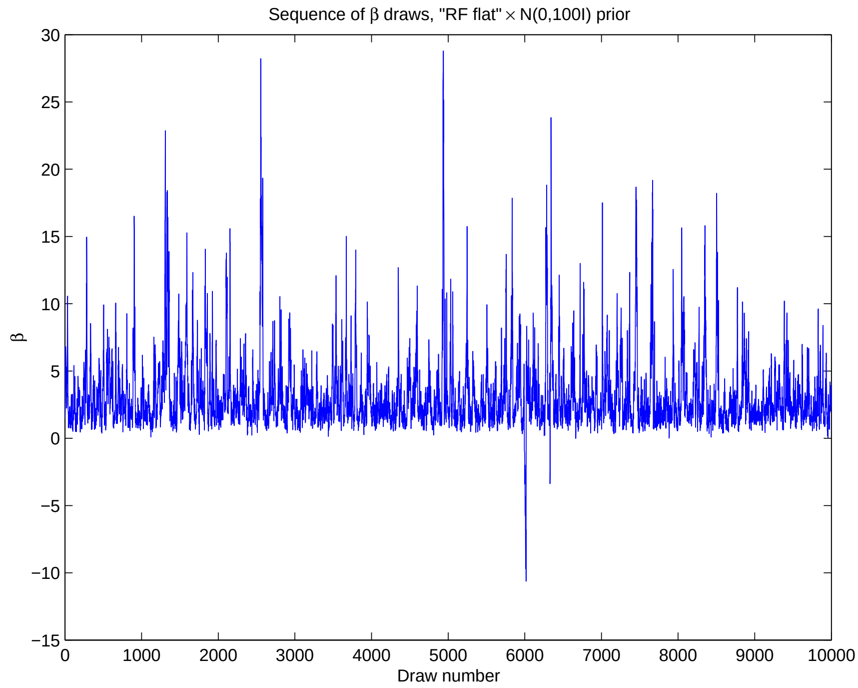

Рис. 11: Последовательность вытягиваний β. Плоское априорное распределение $\times N(1, 100I)$ для ПФ-параметра. Модель и выборка как для Рис. 5.

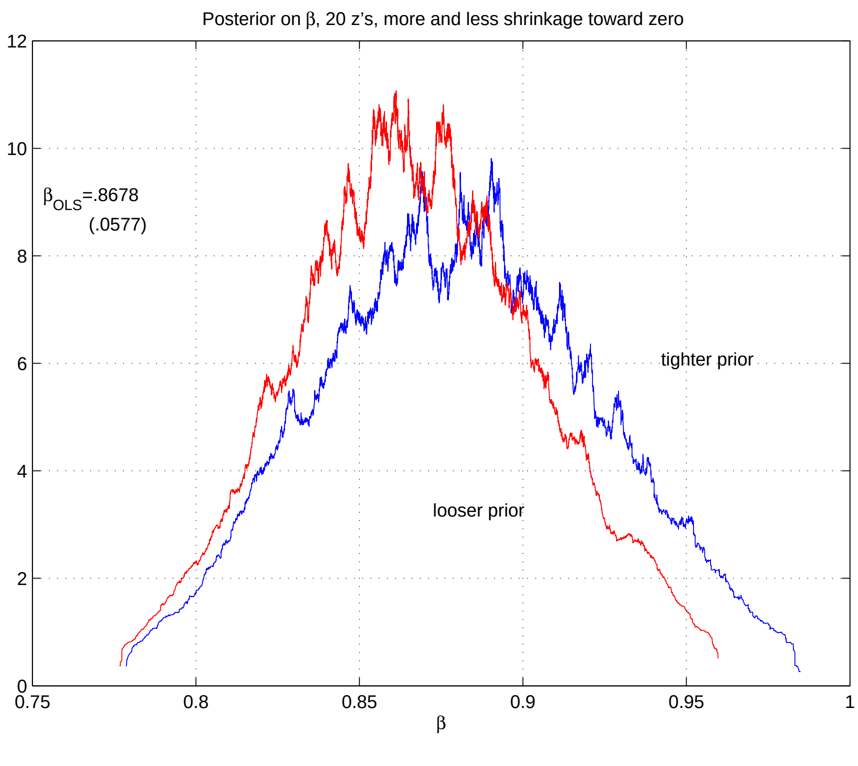

Рис. 12: Матрица Z размером 20×20 тянется из N(0,1). Коэффициенты приведенной формы при Z все равны 1. Выборка такова, что $\hat{\beta}_{OLS} = 0{,}8678$ с обычной стандартной ошибкой 0,0577. Полученная из обратной регрессии $\hat{\beta} = 0{,}9446$.

Список литературы

Faust, J. (1999). Conventional confidence intervals for points on spectrum have confidence level zero. *Econometrica* 67, 629–37.

Geweke, J. (1996). Bayesian reduced rank regression in econometrics. *Journal of Econometrics* 75, 121–146.

Kim, J.-Y. (2002). Limited information likelihood and Bayesian analysis. *Journal of Econometrics* 107, 175–193.

Kitamura, Y. & M. Stutzer (1997). An information-theoretic alternative to generalized method of moments estimation. *Econometrica* 65, 861–874.

Kleibergen, F. & E. Zivot (2000). Bayesian and classical approaches to instrumental variable regression. Technical report, Econometric Institute, Rotterdam.

Sims, C.A. (1974). Distributed lags. In M. Intrilligator & D. Kendrick (eds.). *Frontiers of Quantitative Economics II.* Amsterdam: North-Holland.

Zellner, A., J. Tobias & H.K. Ryu (1997). Bayesian method of moments (BMOM) analysis of parametric and semiparametric regression models. Technical report, University of Chicago.

Thinking about instrumental variables

Christopher A. Sims

Princeton University, Princeton, USA

We take a decision-theoretic view on the question of how to use instrumental variables and method of moments. Since prior beliefs play an inevitably strong role when instruments are possibly “weak”, or when the number of instruments is large relative to the number of observations, it is important in these cases to report characteristics of the likelihood beyond the usual IV or ML estimates and their asymptotic (i.e. second-order local) approximate standard errors. IV and GMM appeal because of their legitimate claim to be convenient to compute in many cases, and a (spurious) claim that they can be justified with few “assumptions”. We discuss some approaches to making such a claim more legitimately.

Keywords: Bayesian approach, GMM, instrumental variables, weak instruments, instrument selection, entropy

JEL Classification: C11, C13, C44