VARIA

А. А. Кудрявцев
докт. экон. наук, профессор кафедры статистики и эконометрики Санкт-Петербургского государственного экономического университета

О ДОВЕРИТЕЛЬНОМ ИНТЕРВАЛЕ ОЦЕНОК ТЕМПОВ РОСТА ВВП НА ОСНОВЕ НОРМАЛЬНОЙ АППРОКСИМАЦИИ

Как известно, любые статистические оценки предполагают возможные ошибки и отклонения вследствие ограниченности выборки, неточностей в определении ключевых факторов и других причин. В результате точечные оценки замещаются интервальными. При этом на практике соответствующие доверительные интервалы часто строятся на основе нормальной (гауссовской) аппроксимации, которая в большинстве случаев требует, чтобы изучаемый экономический процесс был массовым со слабой зависимостью между наблюдениями или даже отсутствием таковой.

Такой подход к построению доверительных интервалов довольно широко используется в статистической практике вследствие своей простоты. Тем не менее в реальной экономике нередко встречаются ситуации и процессы, выходящие за границы применения данного подхода, в результате чего выводы, получаемые на его основе, будут неверными и приведут к ошибкам при принятии управленческих решений. В частности, сделанные предположения (которые в большинстве случаев позволяют упростить ход расчетов) могут приводить к противоречиям при построении оценок доверительных интервалов.

Подобная ситуация возникает и при анализе темпов роста. Обычно подходящие предположения о виде распределения делаются для валового внутреннего продукта (ВВП) (Widmann, 2011) либо для самих значений темпов его роста — см., например, (Ianchovichina, Kacker, 2005) и многие другие публикации. В первом случае такое предположение хорошо соответствует реальности (объем транзакций в экономике велик) и моделям, применяемым для анализа данного показателя (предпосылка о «гауссовском белом шуме» в моделях временных рядов), но при высокой вариации ВВП может вызывать дополнительные сложности оценивания. Во втором случае указанное предположение скорее направлено на снижение технических трудностей в процессе конструирования модели, но вызывает вопросы об ее адекватности и, следовательно, об обоснованности прогнозов на основе такой модели. В лучшем случае выполняется логарифмическое преобразование темпов роста (что соответствует предпосылке об их логнормальном распределении, которая, однако, почти никогда не проверяется на соответствие реальному поведению наблюдаемых величин). В итоге прогнозы темпов роста ВВП могут быть поставлены под сомнение либо по техническим причинам, связанным с построением модели, использованной для прогнозирования, либо из-за подозрений в неадекватности такой модели в целом.

Очевидно, более верным является первый из представленных подходов, когда анализ темпов роста базируется на модели ВВП, а их оценка является вторичной. Однако даже в таком случае последний подход, связанный с дополнительными предпосылками о виде распределения, может быть применим в рамках подходящей аппроксимации, хотя иногда и при достаточно жестких условиях. В данной

DOI 10.31085/1814-4802-2022-18-4-156-110-115

статье обсуждаются различные варианты построения нормальной аппроксимации с указанием границ их применения.

Если отталкиваться от модели временных рядов, использованной для описания динамики ВВП, то случайный компонент вариации включается как в числитель, так и в знаменатель показателя темпов роста (или прироста). Иными словами, рассматриваются случайные величины X_0 и X_1 размера ВВП в базовый и прогнозируемый периоды. Случайность является следствием ошибок наблюдения и т. п. источников неопределенности. Математические ожидания указанных случайных величин будут равны μ_0 и μ_1, соответственно. Их дисперсии составят σ_0^2 и σ_1^2. Также оценивается коэффициент (авто)корреляции ρ между ними. Тогда случайная величина $R = X_1 / X_0$ представляет собой темп роста ВВП. Его ошибки оценивания будут подчинены вероятностному распределению сложной формы. Для стационарных временных рядов отклонения сводятся к «гауссовскому белому шуму», так что тогда для темпов роста речь пойдет о вероятностном распределении, порожденном отношением нормально распределенных случайных величин.

Если в знаменателе предполагается наличие нормально распределенной случайной величины, то данный знаменатель может принимать (по крайней мере теоретически) нулевое значение, т. е. рассматриваемое отношение становится бесконечным. Иными словами, все моменты его распределения, включая математическое ожидание и дисперсию, являются бесконечными. Конечно, на практике наблюдаемые величины ВВП будут отличны от нуля, так что этот результат представляется следствием предположения о виде ошибки в модели временного ряда. Если в самой модели динамики ВВП это предположение было полезным и упрощало теоретическое описание макроэкономического процесса, то при рассмотрении темпов роста ВВП оно уже может стать серьезным препятствием для получения оценок.

Даже в этих условиях нормальная аппроксимация (а следовательно, и традиционный метод построения симметричного доверительного интервала на ее основе) еще возможны. Рассмотрим основные подходы к подобным приближениям и зафиксируем границы их применения.

Самый простой (и самый неточный) способ, который можно назвать «*наивным*», состоит в предположении о том, что распределение случайной величины R само может рассматриваться как приближенно нормальное с математическим ожиданием

$$E\left[R\right] \approx \frac{\mu_1}{\mu_0} + \frac{\sigma_0^2\mu_1}{\mu_0^3} - \frac{\rho\sigma_0\sigma_1}{\mu_0^2} \tag{1}$$

и дисперсией

$$D\left[R\right] \approx \frac{\sigma_0^2\mu_1^2}{\mu_0^4} + \frac{\sigma_1^2}{\mu_0^2} - \frac{2\rho\sigma_0\sigma_1\mu_1}{\mu_0^3}. \tag{2}$$

Здесь явно возникает противоречие в предпосылках, хотя в достаточно узких рамках оно может быть оправданно. К сожалению, подобное приближение, часто предполагаемое в практических расчетах, почти никогда не проверяется на адекватность и границы применения. Более того, на практике нередко используют еще более грубые оценки ожидаемого темпа роста, чем представлено в формуле (1), которая получена за счет разложения в ряд Тейлора. Чаще всего используют отношение математических ожиданий $\hat{R} = \mu_1 / \mu_0$, которое в силу неравенства Йенсена будет заведомо отличаться от $E[R]$.

Границы применения такого подхода, в частности, обсуждались в работе (Hayya et al., 1975). Согласно им такая аппроксимация имеет смысл при коэффициенте корреляции числителя и знаменателя, не превышающих по модулю значения 1/2, если коэффициент вариации для числителя достаточно высок, а для знаменателя довольно мал. Для моделей динамики ВВП следует ожидать высоких величин коэффициента автокорреляции соседних значений ряда, так как подобный ряд в реальности меняется довольно медленно (прирост ВВП в несколько процентных пунктов обычно считается хорошим результатом развития народного хозяйства). Именно этот факт представляет собой главное ограничение «наивного» метода. Другое ограничение состоит в том, что коэффициент вариации примерно постоянен. Но последнее условие не так критично: согласно указанной работе соответствующая аппроксимация возможна с уровнем значимости 5% при упомянутом ранее ограничении на величину коэффициента автокорреляции, когда величины коэффициента вариации числителя и знаменателя в показателе темпа роста примерно равны и не превышают 6%, что в целом отвечает наблюдениям динамики ВВП.

Тем не менее имеется подход, обеспечивающий намного более широкие условия для нормальной аппроксимации распределения отношения, который носит название *преобразования Гири—Хинкли* (Geary, 1930; Hinkley, 1969; 1970). Эти авторы показали, что при некоторых условиях случайная величина

$$Z = \frac{\mu_1 - R\mu_0}{\sqrt{\sigma_1^2 - 2R\rho\sigma_0\sigma_1 + R^2\sigma_0^2}} \tag{3}$$

может быть аппроксимирована стандартным нормальным распределением. Методика состоит в том, что фиксируется подходящая квантиль такого распределения, например $z_{1-a/2}$ для подходящего уровня значимости α (при двустороннем доверительном интервале). Затем решается (квадратное) уравнение относительно R, решения которого и дают границы доверительного интервала.

Это приближение считается хорошим, если вероятность нулевого значения в знаменателе мала — в частности, это будет иметь место при $\mu_X > 3\sigma_X$. Именно это условие позволяет перейти к усеченной случайной величине в знаменателе и избежать теоретического появления нулевого значения, а также вызванных им бесконечных значений моментов распределения.

Кроме того, в статье (Hayya et al., 1975) дана более точная оценка соответствующих условий: при любом коэффициенте корреляции коэффициент вариации случайных величин ВВП X_0 и X_1 не должен превышать 0,39. Последние ограничения достаточно широки и зачастую соответствуют реальности (по крайней мере для динамики ВВП при стабильных макроэкономических условиях). По этой причине данный метод следует считать наиболее адекватным.

При более сильной вариации нельзя использовать приближение нормальным распределением. Выход состоит в *прямой оценке* доверительного интервала *по точному распределению*. Известно, что отношение двух нормально распределенных случайных величин с нулевыми математическими ожиданиями подчинено распределению Коши. В серии статей (Marsaglia, 1965; 2006) даны формулы для общего случая отношения нормально распределенных случайных величин (с произвольными математическими ожиданиями). В частности, показано, что распределение отношения представляет собой смесь распределения Коши и специального бимодального (двухвершинного) распределения, отражающего эффект ненулевых математических ожиданий числителя и знаменателя. Квантили для распределения Коши выражаются достаточно простой формулой, но для указанного специального бимодального распределения подобную оценку следует проводить численно (включая применение методов численного интегрирования), что делает точный расчет неудобным для практического использования.

Из-за наличия двух мод (вершин) в одном из компонентов смеси распределений можно ожидать, что нижний предел приближенного доверительного интервала, полученного на основе нормальной аппроксимации, будет завышен по сравнению с истинным нижним пределом, а сам интервал, соответственно, заужен. В результате оценка на основе преобразования Гири—Хинкли способствует небольшому завышению прогноза темпов роста (обычно это сотые, реже десятые доли процентных пунктов).

В качестве иллюстрации рассмотрим оценки темпов роста ВВП России за период с 2003 по 2007 г. (включительно). Квартальные данные с исключенной сезонностью взяты на сайте Госкомстата. Точечные и интервальные оценки темпов роста будут получены для первого квартала 2008 г., что позволит сравнить их с фактическими (наблюдаемыми) темпами роста. Эти оценки будут вычисляться по отношению к предыдущему кварталу и не пересчитываться на годовой основе.

Указанный период выбран из-за отсутствия кризисов и сломов тенденции, так что процесс экономического роста в рассматриваемом примере следует рассматривать как достаточно стабильный, что снимает часть проблем анализа экономической динамики, важных для исследования темпов роста, но выходящих за границы обсуждения данной статьи. Кроме того, данные были предварительно сглажены Госкомстатом (ныне — Росстат), что позволяет вынести за рамки обсуждения вопросы циклических колебаний, которые существенны для конструирования модели, но искажают (как считается) оценки темпов роста. В результате рассмотренная в примере модель динамики ВВП будет довольно простой и ограничится рассмотрением трендовой и случайной компонент.

При этом трендовая компонента для указанных данных моделируется полиномом третьего порядка, а случайная компонента хорошо согласуется с нормальным распределением. Число положительных и отрицательных отклонений, а также их распределение по группам достаточно, чтобы считать полученные отклонения полностью случайными в соответствии с критерием знаков. Колебания дисперсии во времени также следует считать случайными (имеет место гомоскедастичность). Таким образом, данный пример позволяет полностью сосредоточиться на особенностях построения доверительного интервала показателя темпа роста ВВП.

Формула тренда задает динамику математических ожиданий, среди прочих величин значения для 4 квартала 2007 г. $\mu_0 = 19911{,}65$ и для 1 квартала 2008 г. $\mu_0 = 20441{,}57$. Дисперсия отклонений (ошибок) σ_ε^2 составляет 2028,0630 для обеих случайных величин. Оценка математического ожидания темпа роста ВВП, полученная по формуле (1), составит 1,02662 (2,662% за квартал). В силу того, что коэффициенты вариации малы (они равны 0,0022), эта оценка мало отличается от простого отношения математических ожиданий, хотя последняя и завышена (искажения начинаются с четвертой цифры после запятой, т. е. с сотой доли процентного пункта).

Тем не менее наблюдаемое значение темпа роста (отношение фактических значений ВВП) было ниже: оно составило 1,02653 (2,653% за квартал). Хотя ошибка мала, из-за ее присутствия необходимо рассмотреть вопрос об интервальной оценке (доверительном интервале). «Наивный» метод прямого приближения нормальным распределением здесь не работает из-за высокого значения коэффициента корреляции (выше 0,99). Преобразование Гири—Хинкли дает более хороший результат, т. к. его сфера применения не зависит от значения коэффициента корреляции.

Для уровня значимости 5% зафиксируем квантиль для двустороннего доверительного интервала $z_{0,975} = 1{,}96$, подставим ее в равенство (3) и решим соответствующее квадратное уравнение относительно R. В результате получим доверительный интервал (1,02605; 1,02718), который включает в себя наблюдаемое значение для 1 квартала 2008 г. Центр этого интервала близок к величине, полученной по формуле (1).

Что касается точного доверительного интервала, то в данном случае он будет бимодальным (двухвершинным), так как вес распределения Коши крайне мал, а соответственно вес специального бимодального распределения близок к единице. Это является следствием низких значений коэффициентов вариации оценок ВВП. В результате нижний предел приближенного доверительного интервала, полученного ранее, по-видимому, несколько завышен, что объясняет, почему истинное значение наблюдается в «нижней половине» построенного доверительного интервала.

Подводя итоги, можно сделать следующие выводы и рекомендации.

- При конструировании доверительного интервала для показателя темпов роста ВВП могут возникать технические сложности, вызванные противоречиями в предположениях модели оценивания.
- «Наивный» метод, связанный с прямым использованием нормальной аппроксимации, часто дает сбои из-за высокой автокорреляции.
- Преобразование Гири—Хинкли, как правило, позволяет решить проблему и обеспечивает приемлемую оценку доверительного интервала на основе нормальной аппроксимации. Тем не менее можно ожидать некоторого завышения нижнего предела такого доверительного интервала, а следовательно, и небольшого завышения прогноза темпов роста ВВП (обычно речь идет о сотых долях процентных пунктов).
- Большинство практических ситуаций для оценки доверительного интервала темпов роста ВВП соответствуют границам применения преобразования Гири—Хинкли, так что соответствующий метод должен рекомендоваться как основной.
- При высокой вариации ВВП следует отказываться от применения нормальной аппроксимации при построении доверительного интервала темпов роста ВВП и переходить к прямым расчетам, которые проводятся численно, что делает такой подход не очень удобным для практических расчетов.

Источники

Geary R. C. The Frequency Distribution of the Quotient of Two Normal Variables // Journal of the Royal Statistical Society. 1930. Vol. 93. P. 442—446.

Hayya J., Armstrong D., Gressis N. A Note on the Ratio of Two Normally Distributed Variables // Management Science. 1975. Vol. 21. N 11. P. 1338—1341.

Hinkley D. V. On the Ratio of Two Correlated Normal Random Variables // Biometrika. 1969. Vol. 56. P. 635—639.

Hinkley D. V. Correction: On the Ratio of Two Correlated Normal Random Variables // Biometrika. 1970. Vol. 57. P. 635—639.

Ianchovichina E., Kacker P. Growth Trends in Developing World: Country Forecasts and Determinants // World Bank Policy Research Working Paper 3775. November 2005.

Marsaglia G. Ratios of Normal Variables and Ratios of Sums of Uniform Variables // Journal of the American Statistical Association. 1965. Vol. 60. P. 193—204.

Marsaglia G. Ratios of Normal Variables // Journal of Statistical Software. 2006. Vol. 16. Issue 4. P. 1—10.

Wiedmann M. Money, Stock Prices and Central Banks: a Coinegrated VAR Analysis. Berlin, 2011.

References

Geary R. C. The Frequency Distribution of the Quotient of Two Normal Variables. *Journal of the Royal Statistical Society*, 1930, vol. 93, pp. 442–446.

Hayya J., Armstrong D., Gressis N. A Note on the Ratio of Two Normally Distributed Variables. *Management Science*, 1975, vol. 21, N 11, pp. 1338–1341.

Hinkley D. V. On the Ratio of Two Correlated Normal Random Variables. *Biometrika*, 1969, vol. 56, pp. 635–639.

Hinkley D. V. Correction: On the Ratio of Two Correlated Normal Random Variables. *Biometrika*, 1970, vol. 57, pp. 635–639.

Ianchovichina E., Kacker P. Growth Trends in Developing World: Country Forecasts and Determinants. *World Bank Policy Research Working Paper* 3775, November 2005.

Marsaglia G. Ratios of Normal Variables and Ratios of Sums of Uniform Variables. *Journal of the American Statistical Association*, 1965, vol. 60, pp. 193–204.

Marsaglia G. Ratios of Normal Variables. *Journal of Statistical Software*, 2006, vol. 16, iss. 4, pp. 1–10.

Wiedmann M. Money, Stock Prices and Central Banks: a Coinegrated VAR Analysis. Berlin, 2011.