Парные тесты на одинаковую точность прогнозов⋆

Майкл Маккракен†

Федеральное резервное управление, Вашингтон, США

В настоящем эссе обсуждаются последние достижения, связанные с тестированием на равную точность прогнозов между невложенными или вложенными моделями. Наряду с некоторыми техническими деталями даны рекомендации по практической реализации тестирования.

1 Введение

Иногда в практической работе встречаются ситуации, когда в наличии имеются более одной прогнозирующей модели. В экономике и финансах такая ситуация возникает часто, ибо разные теории о поведении экономических агентов приводят к разным выводам о том, какие переменные и модели имеют предсказательный смысл для ряда экономических переменных. Имея прогнозы из разных моделей, интересно знать, предсказывает ли одна из моделей более точно, чем другая, и если да, является ли разница статистически значимой.

Имеется множество способов это осуществить. Наиболее популярным является внутривыборочная диагностика (F-тест и тест отношения правдоподобия) с целью определить, какая из моделей более точно ложится на данные. В последнее время, однако, становится более популярным оценивать предсказательный смысл моделей, используя вневыборочные методы. В их рамках прогнозы генерируются из «вневыборочной» порции наблюдаемых данных, а результирующие ошибки прогнозов оцениваются с использованием статистических методов.

Для ситуаций, когда две модели являются невложенными, а параметры моделей известны заранее, Diebold & Mariano (1995) предложили способы построения тестов на равную точность прогнозов между невложенными моделями при различных функциях потерь. Когда параметры моделей заранее не известны, а должны оцениваться, West (1996) предоставил аналитический инструментарий, который можно использовать и для построения тестов на равную точность прогнозов между невложенными моделями. Его результаты близки к результатам Diebold & Mariano (1995), но явным образом учитывают дополнительную вариацию в ошибках прогнозов, вызванную оцениванием параметров. West (1996), правда, требует, чтобы функция потерь, используемая для измерения точности прогнозов, была непрерывно дифференцируемой. McCracken (2000) обобщил результаты West (1996) на случаи, когда функция потерь не является непрерывно дифференцируемой.

Если же две рассматриваемые модели являются вложенными, результаты в Diebold & Mariano (1995), West (1996) и McCracken (2000) неверны. А именно, в то время как типичная тестовая статистика на равную точность прогнозов между невложенными моделями имеет нормальное асимптотическое распределение, в Clark & McCracken (2001, 2005) и McCracken (2006) показано, что эти же тестовые статистики при вложенности моделей имеют асимптотические нестандартные распределения, представимые как функционалы от Броуновских движений.

В этом эссе мы обсуждаем результаты перечисленных выше статей. Хотя мы обсуждаем и некоторые технические предположения, по большей части мы концентрируем усилия на описании практических аспектов построения тестовых статистик и получения асимптотически

⋆Перевод С. Анатольева. Цитировать как: Маккракен, Майкл (2006) «Парные тесты на одинаковую точность прогнозов», Квантиль, №1, стр. 53–62. Citation: McCracken, Michael (2006) "Pairwise tests of equal forecast accuracy," Quantile, No.1, pp. 53–62.

†Адрес: Federal Reserve Board of Governors, 20th and Constitution N.W., Washington D.C. 20551, USA. Электронная почта: *michael.w.mccracken@frb.gov*

Таблица 1: Функции потерь

Название	Функция потерь для модели i
Квадратичная	$(y_{t+\tau} - x'_{i,t}\beta_i)^2$
Абсолютная	$\lvert y_{t+\tau} - x'_{i,t}\beta_i\rvert$
Абсолютная α-процентная	$\lvert y_{t+\tau} - x'_{i,t}\beta_i\rvert((1-\alpha)\mathbb{I}(y_{t+\tau} - x'_{i,t}\beta_i < 0) + \alpha\mathbb{I}(y_{t+\tau} - x'_{i,t}\beta_i \geqslant 0))$
Линейно-экспоненциальная	$\exp(\alpha(y_{t+\tau} - x'_{i,t}\beta_i)) - \alpha(y_{t+\tau} - x'_{i,t}\beta_i) - 1$
Скоринговая	$\mathbb{I}(L(y_{t+\tau} - x'_{1,t}\beta_1) \leqslant L(y_{t+\tau} - x'_{2,t}\beta_2)) - 0.5$

верных критических значений, необходимых для реализации тестирования. Ради простоты изложения мы используем линейные многогоризонтные регрессионные модели на стационарных временных рядах, оцениваемые с помощью метода наименьших квадратов (МНК).

Эссе организовано следующим образом. В разделе 2 мы опишем основы вневыборочной методологии и введем минимально необходимые обозначения. В разделе 3 мы обсудим, как построить тесты на равную точность прогнозов между двумя невложенными моделями, а в разделе 4 перейдем к вложенным моделям. Последний раздел содержит некоторые выводы.

2 Постановка

Стационарные наблюдения $\{y_t, x'_t\}_{t=1}^{T+\tau}$ включают скалярную случайную величину y_t, которую будем предсказывать, и k-мерный вектор потенциальных предикторов $x_t = (x'_{1,t}, x'_{2,t})'$. Выборка разделена на внутривыборочную и вневыборочную порции. Внутривыборочная порция содержит наблюдения с 1 по R. Пусть P обозначает количество τ-шаговых ($1 \leqslant \tau$) прогнозов. Вневыборочная порция содержит наблюдения с $R+\tau$ по $R+P+\tau-1 = T+\tau$. Обозначим $\pi = \lim(P/R)$ при $T \to \infty$.

Прогнозы $y_{t+\tau}$, $t = R, \ldots, T$, генерируются с использованием двух линейных моделей в форме $x'_{i,t}\beta_i^*$ ($i = 1, 2$), каждая из которых оценивается с помощью МНК. Прогнозы могут быть рекурсивными, скользящими или фиксированными. При рекурсивной схеме параметры каждой модели переоцениваются по мере того, как прогнозирование смещается во времени: для $t = R, \ldots, T$ прогноз $y_{t+\tau}$ от i-й модели, $x'_{i,t}\hat{\beta}_{i,t}$, строится с использованием оценок параметров

$$\hat{\beta}_{i,t} = \left(\frac{1}{t}\sum_{s=1}^{t-\tau} x_{i,s}x'_{i,s}\right)^{-1}\left(\frac{1}{t}\sum_{s=1}^{t-\tau} x_{i,s}y_{s+\tau}\right).$$

Для остальных схем прогнозы в момент t строятся похожим образом с использованием либо скользящего окна из последних R наблюдений $s = t-R+1, \ldots, t$, либо начальных R наблюдений $s = 1, \ldots, R$ для скользящей и фиксированной схем соответственно.

Обозначим оцененные τ-шаговые ошибки прогнозов как $\hat{u}_{i,t+\tau} = y_{t+\tau} - x'_{i,t}\hat{\beta}_{i,t}$ ($i = 1, 2$). Эти ошибки, только с использованием популяционных значений параметров, обозначим как $u_{i,t+\tau} = y_{t+\tau} - x'_{i,t}\beta_i^*$ с соответствующими значениями потерь $L(\hat{u}_{i,t+\tau})$ ($i = 1, 2$). Примеры наиболее часто используемых функций потерь приведены в таблице 1.

Выбрав функцию потерь и имея две последовательности потерь, по одной для каждой из моделей, мы хотим построить тест на одинаковую предсказательную точность этих моделей. А именно, мы хотим протестировать нулевую гипотезу

$$H_0 : \mathbb{E}[L(u_{1,t+\tau})] - \mathbb{E}[L(u_{2,t+\tau})] = 0$$

для всех t. Естественной отправной точкой для построения теста на одинаковую точность прогнозов – рассмотреть поведение вневыборочного среднего разницы потерь

$$P^{-1}\sum_{t=R}^{T}[L(\hat{u}_{1,t+\tau})-L(\hat{u}_{2,t+\tau})].$$

Интуитивно ясно, что если это среднее «велико» и положительно в какой-то метрике, мы сделаем вывод, что модель 2 более точная, чем модель 1, и наоборот, если это среднее отрицательно. В противном случае, если это среднее «мало», мы сделаем вывод, что разница статистически не значима, и модели эквивалентны по точности прогнозирования.

К сожалению, уже в данный момент становится важным различать между невложенными и вложенными моделями. В частности, асимптотическое распределение, используемое для определения, насколько «велика» средняя разница потерь, зависит от того, вложены модели или нет.

3 Невложенные модели

Для последовательностей прогнозов от двух невложенных моделей Diebold & Mariano (1995) и West (1996) рассматривают асимптотическое распределение t-статистики в форме

$$P^{1/2}\frac{P^{-1}\sum_{t=R}^{T}[L(\hat{u}_{1,t+\tau})-L(\hat{u}_{2,t+\tau})]}{\hat{\Omega}_T^{1/2}}, \tag{1}$$

где $\hat{\Omega}_T$ обозначает состоятельную оценку асимптотической дисперсии средней разницы потерь $\Omega \equiv \lim Var(P^{-1/2}\sum_{t=R}^{T}[L(\hat{u}_{1,t+\tau})-L(\hat{u}_{2,t+\tau})])$. В обеих статьях авторы показывают, что статистика асимптотически стандартно нормальная. Числитель (1) легко построить. Тяжелее сконструировать подходящую оценку Ω. Ниже мы обсуждаем некоторые асимптотически адекватные методы.

Diebold & Mariano (1995)

Простейшая ситуация – это когда вопреки постановке предыдущего раздела параметры известны заранее и их не нужно оценивать. В контексте параметрического предсказывания такое вряд ли возможно. Тем не менее, если «моделью» является мнение из опроса профессиональных предсказателей или же финансовое трейдинговое правило, в ней действительно нет параметров, которые надо было бы оценивать.

Этот особый случай является простейшим, поскольку он означает, что мы наблюдаем, а не оцениваем, популяционные значения ошибок прогнозов. Это означает не только то, что $L(\hat{u}_{1,t+\tau})-L(\hat{u}_{2,t+\tau}) = L(u_{1,t+\tau})-L(u_{2,t+\tau})$, но и то, что на долгосрочную дисперсию $L(\hat{u}_{1,t+\tau})-L(\hat{u}_{2,t+\tau})$ не влияет неопределенность, вызванная ошибкой оценивания параметров. Правильно оценить Ω просто, если учесть возможную серийную корреляцию в разнице потерь. Например, строя τ-шаговые прогнозы, мы ожидаем серийную корреляцию порядка по крайней мере $\tau-1$ в $L(u_{1,t+\tau})-L(u_{2,t+\tau})$. В такой ситуации можно использовать стандартные непараметрические ядерные оценки, с ядрами типа Бартлетовского или квадратично-спектрального, для состоятельного оценивания долгосрочной дисперсии.

West (1996)

Оценивать Ω становится сложнее, когда прогнозы строятся на основе параметрических моделей с оцениваемыми параметрами. Как отмечено выше, ключевым отличием является то, что при оцениваемых параметрах мы не наблюдаем популяционные значения ошибок прогнозов $u_{i,t+\tau}$, а наблюдаем их оценки $\hat{u}_{i,t+\tau}$. Эта маленькая разница означает, что

$$\Omega \equiv \lim \mathbb{V}\left(P^{-1/2}\sum_{t=R}^{T}\left[L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})\right]\right)$$

необязательно равна

$$\lim \mathbb{V}\left(P^{-1/2}\sum_{t=R}^{T}[L(u_{1,t+\tau}) - L(u_{2,t+\tau})]\right).$$

Эта мысль подчеркивалась явным образом в West (1996) и была развита далее в McCracken (2000). А именно, West (1996) показал, как подобающим образом построить Ω, когда функция потерь дважды непрерывно дифференцируема, как в случае квадратичных и линейно-экспоненциальных потерь из таблицы 1. McCracken (2000) расширил теорию для случая остальных функций потерь из таблицы 1.

Ключевым элементом обоих результатов является асимптотическое разложение с двумя раздельными компонентами: одна компонента включает ошибки прогнозов популяционного уровня, а вторая обусловлена тем, что параметры неизвестны и оцениваются. Чтобы понять смысл этого разложения, сначала рассмотрим нормализованную среднюю разницу потерь

$$P^{-1/2}\sum_{t=R}^{T}\left[L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})\right],$$

и вспомним обсуждение из раздела 2, где мы ограничились рассмотрением прогнозов, порожденных линейными многогоризонтными моделями, оцененными с помощью МНК. Определим $k \times 1$ вектор, $k \times k$ матрицу и $k \times 1$ вектор-функцию

$$F = \left(\partial\mathbb{E}[L(y_{t+\tau} - x'_{1,t}\beta_1)]/\partial\beta'_1|_{\beta_1=\beta_1^*}, -\partial\mathbb{E}[L(y_{t+\tau} - x'_{2,t}\beta_2)]/\partial\beta'_2|_{\beta_2=\beta_2^*}\right)',$$

$$B = \begin{pmatrix} \mathbb{E}[x_{1,t}x'_{1,t}]^{-1} & 0_{k_1\times k_2} \\ 0_{k_2\times k_1} & \mathbb{E}[x_{2,t}x'_{2,t}]^{-1} \end{pmatrix},$$

$$H(t) = \begin{cases} \left(t^{-1}\sum_{s=1}^{t-\tau} h'_{1,s+\tau}, t^{-1}\sum_{s=1}^{t-\tau} h'_{2,s+\tau}\right)' & \text{для рекурсивной схемы,} \\ \left(R^{-1}\sum_{s=t-R+1}^{t-\tau} h'_{1,s+\tau}, R^{-1}\sum_{s=t-R+1}^{t-\tau} h'_{2,s+\tau}\right)' & \text{для скользящей схемы,} \\ \left(R^{-1}\sum_{s=1}^{R-\tau} h'_{1,s+\tau}, R^{-1}\sum_{s=1}^{R-\tau} h'_{2,s+\tau}\right)' & \text{для фиксированной схемы,} \end{cases}$$

где $h_{i,s+\tau} = u_{i,s+\tau}x_{i,s}$ $(i = 1, 2)$. Используя эти обозначения, West (1996) и McCracken (2000) показывают, что

$$\begin{aligned} P^{-1/2}\textstyle\sum_{t=R}^{T}\left[L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})\right] &= P^{-1/2}\textstyle\sum_{t=R}^{T}[L(u_{1,t+\tau}) - L(u_{2,t+\tau})] \\ &+ FBP^{-1/2}\textstyle\sum_{t=R}^{T} H(t) + o_p(1). \end{aligned} \tag{2}$$

Первое слагаемое в правой части (2) – это разница потерь, которая возникла бы, если бы параметры были известны заранее и не оценивались бы. Второе слагаемое в правой части (2) возникло из-за того, что параметры неизвестны и оцениваются. Каждое из этих двух слагаемых может внести свой вклад в предельную дисперсию в (1). Пусть $h_{t+\tau} = (h'_{1,t+\tau}, h'_{2,t+\tau})'$, и обозначим за S долгосрочную дисперсию $(L(u_{1,t+\tau}) - L(u_{2,t+\tau}), h'_{t+\tau})'$ с блочно-диагональными элементами S_{dd} и S_{hh} и внедиагональным элементом S_{dh}. При дополнительных мягких предположениях

$$P^{-1/2}\sum_{t=R}^{T}\left[L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})\right]$$

сходится по распределению к нормальной случайной величине с предельной дисперсией

$$\Omega = S_{dd} + 2\lambda_{dh}(FBS'_{dh}) + \lambda_{dd}(FBS_{hh}B'F'), \tag{3}$$

где

Схема	λ_{dh}	λ_{hh}
Рекурсивная	$1 - \pi^{-1}\ln(1+\pi)$	$2[1 - \pi^{-1}\ln(1+\pi)]$
Скользящая с $\pi \leqslant 1$	$\pi/2$	$\pi - \pi^2/3$
Скользящая с $1 < \pi < \infty$	$1 - 1/2\pi$	$1 - 1/3\pi$
Фиксированная	0	π

Из равенства (3) ясно, что оценивание параметров может повлиять на предельную дисперсию вневыборочных тестовых статистик. Первая компонента в правой части Ω, S_{dd}, – это как раз предельная дисперсия, предложенная в Diebold & Mariano (1995) для тестирования на одинаковую точность прогнозов, когда параметры известны заранее. Если остальные слагаемые сокращаются и $\Omega = S_{dd}$, мы говорим, что ошибка оценивания параметров асимптотически нерелевантна.

В общем случае, однако, ошибка оценивания релевантна. Вторая и третья компоненты (3) появляются именно благодаря тому, что параметры неизвестны и оцениваются. Дополнительная вариация из-за ошибки оценивания – это слагаемое $FBP^{-1/2}\sum_{t=R}^{T} H(t)$ в равенстве (2). Эта вариация порождает дополнительную дисперсионную компоненту $\lambda_{hh}(FBS_{hh}B'F')$ и дополнительную ковариационную компоненту $2\lambda_{dh}(FBS'_{dh})$ в Ω. Эти две компоненты и являются источником беспокойства. Если их игнорировать, тесты нулевой гипотезы будут иметь неверный размер и пониженную мощность.

Теперь мы очертим четыре различных способа, с помощью которых можно учесть ошибку оценивания параметров.

Случай 1: $F = 0$

Временами можно показать, что дополнительная вариация из-за ошибки оценивания параметров асимптотически нерелевантна. Так как в данном обсуждении мы предполагаем, что параметры модели состоятельно оцениваются МНК, это означает, что $F = 0$, если мы оцениваем точность прогнозов, используя квадратичные потери.

Случай 2: $\pi = 0$

Другой важный случай, когда ошибки оценивания параметров асимптотически нерелевантны, встречается при $\pi = 0$. Заметим, что когда $\pi = 0$, и λ_{dh}, и λ_{hh} – нули. Если это имеет место, второе и третье слагаемые в правой части (3) равны нулю. На практике такая ситуация никогда не встречается, т.к. P и R – конечные положительные целые числа. Но иногда может быть разумным считать отношение P/R достаточно близким к нулю, так что результат будет следовать по непрерывности. Этот аргумент приводится в явном виде в Chong & Hendry (1986).

Случай 3: рекурсивная схема и $FBS'_{dh} = -FBS_{hh}BF'$

Третья и наименее очевидная ситуация, когда ошибки оценивания параметров могут быть асимптотически нерелевантны, встречается, когда второе и третье слагаемые в правой части (3) в сумме дают нуль при том, что каждое из них не равно нулю. Заметим, что $2\lambda_{dh} = \lambda_{hh}$ для рекурсивной схемы. Из-за этого в тестах, где $FBS'_{dh} = -FBS_{hh}BF'$, ошибки оценивания параметров асимптотически нерелевантны, когда используется рекурсивная схема.

Эта ситуация иногда встречается и в других вневыборочных тестах на прогнозирующую способность, таких как на равенство нулю средней ошибки прогноза или серийной корреляции одношаговых ошибок прогнозов, но реже встречается при тестировании на равную предсказательную точность. В любом случае такое сокращение происходит только для определенных процессов, порождающих данные, и никогда не происходит при использовании скользящих и фиксированных схем.

Случай 4: оценивание Ω

Самый непосредственный способ скорректировать на ошибки оценивания параметров – попросту построить состоятельную оценку Ω и использовать ее в знаменателе t-статистики в (1). Все компоненты Ω оцениваются напрямую. Например, $\hat{B} = \operatorname{diag}((P^{-1}\sum_{t=R}^{T} x_{1,t}x'_{1,t})^{-1}$, $(P^{-1}\sum_{t=R}^{T} x_{2,t}x'_{2,t})^{-1})$ формируют состоятельную оценку B. Так как λ_{dh} и λ_{hh}– непрерывные функции от $\pi = \lim(P/R)$, их можно оценить, подставляя P/R вместо π.

Члены S_{dd}, S_{dh} и S_{hh} требуют знания того, какие используются данные, но в остальном это непосредственно долгосрочные ковариации. Например, если мы строим τ-шаговые прогнозы, мы ожидаем серийную корреляцию порядка $\tau - 1$ и в $L(u_{1,t+\tau}) - L(u_{2,t+\tau})$, и в $h_{t+\tau}$. В такой ситуации стандартные непараметрические ядерные оценки, с ядрами типа Бартлетовского или квадратично-спектрального, могут быть использованы для состоятельного оценивания долгосрочных ковариаций S_{dd}, S_{dh} и S_{hh}. В частности, West (1996), West & McCracken (1998) и McCracken (2000) показывают, как последовательность оцененных разниц потерь $L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})$, $t = R, ..., T$, и последовательность оцененных условий ортогональности $\hat{h}_{t+1} = (\hat{u}_{1,t+\tau}x'_{1,t}, \hat{u}_{2,t+\tau}x'_{2,t})'$, можно использовать как если бы они являлись соответствующими популяционными значениями. Поэтому состоятельные оценки для S_{dd}, S_{dh} и S_{hh} легко получаются при помощи стандартных процедур (например, Newey & West, 1987).

Наиболее сложный аспект оценивания Ω с использованием равенства (3) – это, возможно, построение состоятельной оценки F. В случаях, когда функция потерь дифференцируема, это просто. Например, положим, что потери измеряются линейно-экспоненциальной функцией. В этом случае можно использовать

$$\hat{F} = \left(P^{-1}\sum_{t=R}^{T} \alpha(1 - \exp(\alpha\hat{u}_{1,t+\tau}))x'_{1,t}, -P^{-1}\sum_{t=R}^{T} \alpha(1 - \exp(\alpha\hat{u}_{2,t+\tau}))x'_{2,t} \right)'$$

в качестве оценки F.

Когда функция потерь недифференцируема, оценивание F может усложниться. Не очень сложным случаем являются абсолютные потери. Когда абсолютное значение ошибки является мерой потерь, McCracken (2000) показывает, что $F = (\mathbb{E}[\operatorname{sgn}(u_{1,t+\tau})x'_{1,t}], -\mathbb{E}[\operatorname{sgn}(u_{2,t+\tau})x'_{2,t}])$, и более того, что

$$\hat{F} = \left(P^{-1}\sum_{t=R}^{T} \operatorname{sgn}(\hat{u}_{1,t+\tau})x'_{1,t}, -P^{-1}\sum_{t=R}^{T} \operatorname{sgn}(\hat{u}_{2,t+\tau})x'_{2,t} \right)'$$

является состоятельной оценкой F.

4 Вложенные модели

Для последовательностей прогнозов от двух вложенных моделей, Clark & McCracken (2001, 2005) рассматривают асимптотическое распределение двух различных статистик, которые можно использовать для тестирования равной точности прогнозов. Первая статистика – та же, что использовалась для невложенного случая, а вторая – новая:

$$P^{1/2}\frac{P^{-1}\sum_{t=R}^{T}[L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})]}{\hat{\Omega}_T^{1/2}}, \quad 2P\frac{P^{-1}\sum_{t=R}^{T}[L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})]}{\hat{c}} \tag{4}$$

для скалярного $\hat{c}$, зависящего от функции потерь $L(\cdot)$.

Поскольку первая статистика в (3) идентична той, что использовалась в для невложенного случая, можно подумать, что она асимптотически нормальна. Однако это не так. Ненормальность асимптотического распределения происходит из того, что при нулевой гипотезе не только $\mathbb{E}[L(u_{1,t+\tau}) - L(u_{2,t+\tau})] = 0$, но в больших выборках $u_{1,t+\tau} = u_{2,t+\tau}$, и поэтому $L(u_{1,t+\tau}) - L(u_{2,t+\tau}) = 0$. Отсюда следует, что $\Omega = 0$, и поэтому асимптотическая теория, разработанная для невложенного случая, требующего положительности Ω, не переносится на вложенный случай.

Для дважды непрерывно дифференцируемых функций потерь и одношагового прогнозирования на основе потенциально нелинейных моделей McCracken (2006) выводит асимптотическое распределение этих двух статистик. Эти распределения не нормальны, но представимы как функционалы от Броуновского движения. Более того, они зависят от используемой функции потерь в том смысле, что нормирующая константа $\hat{c}$ от нее зависит.

В дальнейшем мы подробно рассмотрим эти статистики в сочетании с использованием популярной квадратичной функции потерь $L(u) = u^2$. В частности, Clark & McCracken (2005) выводят асимптотические распределения статистик типа (4) при квадратичных потерях. Если мы положим $\hat{d}_{t+\tau} = \hat{u}^2_{1,t+\tau} - \hat{u}^2_{2,t+\tau}$ и определим $\mathrm{MSE}_i = (P-\tau+1)^{-1}\sum_{t=R}^{T-\tau}\hat{u}^2_{i,t+\tau}$ $(i = 1,2)$, $\bar{d} = (P-\tau+1)^{-1}\sum_{t=R}^{T-\tau}\hat{d}_{t+\tau} = \mathrm{MSE}_1 - \mathrm{MSE}_2$ и

$$\hat{\Gamma}_{dd}(j) = (P-\tau+1)^{-1}\sum_{t=R+j}^{T-\tau}(\hat{d}_{t+\tau} - \bar{d})(\hat{d}_{t+\tau-j} - \bar{d}), \quad \hat{\Gamma}_{dd}(-j) = \hat{\Gamma}_{dd}(j),$$

то для стандартной ядерной оценки долгосрочной дисперсии, $\hat{S}_{dd} = \sum_{j=-\bar{j}}^{\bar{j}} K(j/M)\hat{\Gamma}_{dd}(j)$, статистики принимают вид

$$\text{MSE-t} = (P-\tau+1)^{1/2} \times \frac{\bar{d}}{\sqrt{\hat{S}_{dd}}}.$$

$$\text{MSE-F} = \begin{cases} (P-\tau+1)\dfrac{\bar{d}}{\mathrm{MSE}_2}, & 0 < \pi < \infty, \\ R^{1/2}(P-\tau+1)^{1/2}\dfrac{\bar{d}}{\mathrm{MSE}_2}, & \pi = 0. \end{cases}$$

Как и в случае, рассматриваемом в McCracken (2006), эти статистики обычно имеют асимптотические распределения, которые не являются нормальными, а представимы как функционалы k_2-мерного Броуновского движения $W(\omega)$.

Когда $\lim(P/R) = \pi > 0$, асимптотические распределения обеих статистик зависят от выборочной схемы (рекурсивная, скользящая, фиксированная), от относительных размеров внутривыборочной и вневыборочной порций наблюдений (π и $\lambda = (1+\pi)^{-1}$) и количества внешних предикторов в неограниченной модели (k_2, размерность $W(\omega)$). Они также зависят от взвешивающей матрицы $S_{\tilde{h}\tilde{h}}$, о которой мы поговорим позже.

Когда $\lim(P/R) = \pi = 0$, асимптотические распределения сильно упрощаются. На самом деле, статистика MSE-t асимптотически стандартно нормальна независимо от выборочной схемы, параметра деления выборки π и количества внешних предикторов. Это также справедливо независимо от того, гетероскедастичны ли ошибки прогнозов и скоррелированны ли серийно. Асимптотическое распределение статистики MSE-F также упрощается, хотя и не так радикально. Оно более не зависит от выборочной схемы, но все еще зависит от количества внешних предикторов (k_2) посредством зависимости от размерности двух стандартно нормальных векторов V_0 и V_1 и взвешивающей матрицы $S_{\tilde{h}\tilde{h}}$.

Таблица 2: Асимптотические 95%-ные критические значения для статистик MSE-t и MSE-F

$\pi \rightarrow$	$0,0$	$0,2$	$1,0$	$2,0$	$0,0$	$0,2$	$1,0$	$2,0$
$k \downarrow$	MSE-t				MSE-F			
Рекурсивная схема								
1	$1,645$	$1,111$	$0,771$	$0,610$	$3,270$	$1,038$	$1,548$	$1,518$
2	$1,645$	$1,140$	$0,704$	$0,478$	$4,826$	$1,453$	$1,802$	$1,706$
3	$1,645$	$1,120$	$0,610$	$0,386$	$5,946$	$1,710$	$1,909$	$1,612$
4	$1,645$	$1,101$	$0,502$	$0,221$	$6,712$	$1,964$	$1,809$	$1,029$
5	$1,645$	$1,061$	$0,386$	$0,081$	$7,404$	$2,082$	$1,449$	$0,459$
10	$1,645$	$0,890$	$0,043$	$-0,339$	$10,414$	$2,489$	$0,205$	$-2,378$
Скользящая схема								
1	$1,645$	$1,117$	$0,651$	$0,334$	$3,270$	$1,112$	$1,583$	$1,215$
2	$1,645$	$1,105$	$0,484$	$0,103$	$4,826$	$1,481$	$1,695$	$0,504$
3	$1,645$	$1,088$	$0,381$	$-0,084$	$5,946$	$1,752$	$1,532$	$-0,471$
4	$1,645$	$1,087$	$0,274$	$-0,222$	$6,712$	$2,078$	$1,228$	$-1,487$
5	$1,645$	$1,034$	$0,155$	$-0,385$	$7,404$	$2,191$	$0,764$	$-2,765$
10	$1,645$	$0,872$	$-0,258$	$-1,011$	$10,414$	$2,520$	$-1,733$	$-9,863$
Фиксированная схема								
1	$1,645$	$1,416$	$1,252$	$1,218$	$3,270$	$1,015$	$1,667$	$1,862$
2	$1,645$	$1,342$	$1,072$	$0,955$	$4,826$	$1,421$	$2,116$	$2,195$
3	$1,645$	$1,277$	$0,909$	$0,733$	$5,946$	$1,653$	$2,319$	$2,275$
4	$1,645$	$1,281$	$0,755$	$0,509$	$6,712$	$1,947$	$2,238$	$1,784$
5	$1,645$	$1,193$	$0,646$	$0,291$	$7,404$	$2,018$	$2,167$	$1,249$
10	$1,645$	$1,007$	$0,167$	$-0,358$	$10,414$	$2,611$	$0,936$	$-2,404$

Реализация инференции для этих тестов на равную точность прогнозов может быть сложным, по крайней мере оно сложнее, чем в невложенном случае. Конечно, проще всего использовать статистику MSE-t и предполагать, что P/R достаточно близко к нулю, так что асимптотика стандартно нормальная. Хотя следование такому подходу упрощает нахождение критических значений (их можно считать прямо из соответствующих таблиц), показано, что размер теста при такой аппроксимации очень плох, а его мощность очень мала при размерах выборки, обычно имеющихся у макроэкономистов. Главная трудность кроется в матрице шумовых параметров $S_{\tilde{h}\tilde{h}}$. Ниже мы очертим ряд методов проведения инференции, основанных на том, является ли эта матрица единичной или нет.

Случай 1: $S_{\tilde{h}\tilde{h}}$ – единичная матрица

Когда прогнозы одношаговые, а ошибки прогнозов условно гомоскедастичны и серийно нескоррелированы, $S_{\tilde{h}\tilde{h}}$ является единичной матрицей. Такую ситуацию полезно рассмотреть, так как распределения упрощаются еще сильнее. В частности, они теперь зависят только от известных величин, таких как P/R, k_2 и от выборочной схемы. Для этого случая McCracken (2006), используя симуляции, строит асимптотически верные оценки соответствующих критических значений, которые можно использовать для реализации инференции. Некоторые критические значения приведены в таблице 2.

Случай 2: $S_{\tilde{h}\tilde{h}}$ не является единичной матрицей

Когда прогнозы не являются одношаговыми, или ошибки прогнозов условно гетероскедастичны или серийно скоррелированы, $S_{\tilde{h}\tilde{h}}$ не является единичной матрицей. В таком случае предельные распределения зависят от шумовых параметров, и невозможно так затабулировать критические значения, чтобы ими можно было бы пользоваться в разных приложениях. Это не значит, что критические значения невозможно получить, просто это требует больше усилий. Ниже приведены два способы это сделать.

Прямые симуляции

Заметим, что каждое из распределений зависит от интегралов квадратичных форм от Броуновского движения. Используя методы Монте–Карло, каждый из них можно состоятельно оценить: интегралы можно приблизить средними, а Броуновское движение – случайными блужданиями. Если матрицу $S_{\tilde{h}\tilde{h}}$ можно состоятельно оценить, то квадратичные формы тоже можно. Соответственно, нужные критические значения можно построить, используя состоятельную оценку $S_{\tilde{h}\tilde{h}}$ и численные методы следующим образом:

1. Подогнать ограниченную (чтобы наложить нулевую гипотезу) прогнозирующую модель ко всей имеющейся выборке и сохранить остатки $\hat{u}_{t+\tau}$; оценить $\hat{S}_{hh} = \mathrm{LR}\mathbb{V}(X_{t,2}\hat{u}_{t+\tau})$ по методу Newey & West (1987) с шириной окна $1.5 \cdot \tau$ при $\tau > 1$ и 0 при $\tau = 1$.

2. Используя оценки $\hat{B}_i = (\sum_{t=1}^{R+P} x_{i,t}x'_{i,t})^{-1}$ и $\hat{S}_{12} = (\sum_{t=1}^{R+P} x_{22,t}x'_{1,t})$, где $\hat{\sigma}^2$ обозначает остаточную дисперсию модели, оцененной на шаге 1, подсчитать $\hat{D} = \hat{B}_{22}^{-1} - \hat{S}_{12}\hat{B}_1\hat{S}'_{12}$ и сформировать

$$\hat{S}_{\tilde{h}\tilde{h}} = \hat{\sigma}^{-2} \begin{pmatrix} 0_{k_2\times k_1} & \hat{D}^{.5} \end{pmatrix} \hat{B}_2\hat{S}_{hh}\hat{B}_2 \begin{pmatrix} 0_{k_1\times k_2} \\ \hat{D}^{.5} \end{pmatrix},$$

где $\hat{D}^{.5}$ есть разложение Холеского для $\hat{D}$.

3. Подсчитать характеристические числа матрицы $\hat{S}_{\tilde{h}\tilde{h}}$.

4. Вытянуть 5000 независимых «наблюдений» из асимптотического распределения каждой тестовой статистики при данных k_2 и $\widehat{\pi} = P/R$. Генерируя эти «наблюдения», необходимые k_2 Броуновских движений симулируются как случайные блуждания, каждое из которых использует независимую последовательность из 10000 приращений, распределенных как $N(0, 10000^{-.5})$. Интегралы заменяются суммами взвешенными квадратичными формами от случайных блужданий, используя характеристические числа матрицы $\hat{S}_{\tilde{h}\tilde{h}}$ в качестве весов. 10%-ное критическое значение рассчитывается как 90%-ный квантиль результирующих статистик.

Бутстрап

Альтернативным методом построения критических значений является бутстрап. При линейности структур всех участвующих моделей, Clark & McCracken (2005) рассматривают параметрический бутстрап, использованный в Kilian (1999), следующим образом:

1. Оценить векторную авторегрессию для y_t и $x_{2,t}$ с помощью МНК, используя всю выборку и налагая нулевое ограничение, что в $x_{22,t}$ нет прогнозирующей силы для y_t, и сохранить остатки. Заметим, что генерирующее уравнение для y_t принимает точно такую же форму, что и ограниченная прогнозирующая модель при $\tau = 1$ (но оцененная для всех имеющихся данных). Порядки авторегрессии для y_t и $x_{2,t}$ определяются с помощью AIC, допуская разные длины лагов (скажем, от 0 до 8) при каждой из переменных.

2. Бутстрапированные временные ряды для y_t и $x_{2,t}$ генерируются вытягиванием с возвращением из набора сохраненных остатков, используя авторегрессионные структуры моделей для итеративного построения данных. Начальные наблюдения, т.е. наблюдения, предшествующие выборке данных, участвовавших в оценивании моделей, генерируются вытягиванием из первоначальных данных. В частности, начальные наблюдения выбираются взятием случайной даты, а затем взятием начальных наблюдений, начиная с этой даты, но в обратном порядке.

3. В каждом из 999 бутстраповских повторов используются бутстрапированные данные для оценивания ограниченной и неограниченной прогнозирующих моделей. Из результирующих прогнозов затем строятся тестовые статистики. Критические значения рассчитываются просто как квантили бутстрапированных тестовых статистик.

5 Заключение

В этом эссе мы обсудили тестирование на равную точность прогнозов между двумя либо невложенными, либо вложенными моделями. Мы сконцентрировались на подробностях такого тестирования, когда прогнозы строятся на основе оцененных параметров модели. И в невложенном, и во вложенном случаях имеются детали, усложняющие построение асимптотически верного теста. В свете этих сложностей мы внесли предложения по построению тестов, имеющих правильный размер в больших выборках.

Список литературы

Chong, Y.Y. & D.F. Hendry (1986). Econometric evaluation of linear macro-economic models. *Review of Economic Studies* 53, 671–690.

Clark, T. E. & M.W. McCracken (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics* 105, 85–110.

Clark, T. E. & M.W. McCracken (2005). Evaluating direct multi-step forecasts. *Econometric Reviews* 24, 369–404.

Diebold, F.X. & R.S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13, 253–263.

Kilian, L. (1999). Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions? *Journal of Applied Econometrics* 14, 491–510.

McCracken, M.W. (2000). Robust out-of-sample inference. *Journal of Econometrics* 99, 195–223.

McCracken, M.W. (2006). Asymptotics for out-of-sample tests of Granger causality. *Journal of Econometrics*, в печати.

Newey, W.K. & K.D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–08.

West, K.D. (1996). Asymptotic inference about predictive ability. *Econometrica* 64, 1067–1084.

West, K.D. & M.W. McCracken (1998). Regression-based tests of predictive ability. *International Economic Review* 39, 817–40.

Pairwise tests of equal forecast accuracy

Michael McCracken

Federal Reserve Board of Governors, Washington D.C., USA

This essay reviews recent work regarding pairwise tests of equal forecast accuracy between nested and non-nested models. While some technical details are given, special emphasis is placed on the practical implementation of the tests.