Выбор модели и парадоксы прогнозирования*

Олег Ицхоки†

Гарвардский Университет, Кэмбридж, США
Центральный Экономико-Математический Институт, Москва, Россия

В настоящем эссе мы высказываем ряд теоретических гипотез, позволяющих в той или иной мере разрешить два парадокса прогнозирования: (1) почему простые линейные модели зачастую обладают преимуществом в предсказательной силе над более сложными нелинейными моделями, которые, в свою очередь, позволяют получить более точную внутривыборочную подгонку данных; (2) почему комбинации прогнозов нередко повышают предсказательную силу индивидуальных прогнозов. В работе также приводится численный пример, иллюстрирующий выдвигаемые теоретические положения.

1 Введение

Настоящее эссе посвящено анализу двух широко известных эмпирических фактов в области прогнозирования экономических временных рядов, не имеющих общепризнанного теоретического обоснования. Первый факт, или, скорее, эмпирическое правило, состоит в том, что простые линейные модели, как правило, позволяют получить более качественные вневыборочные прогнозы, несмотря на то, что более сложные нелинейные модели обеспечивают более точную внутривыборочную подгонку данных. Некоторые исследователи полагают, что простота и компактность линейных моделей делает их более устойчивыми к неверной спецификации, что является важным при долгосрочном прогнозировании.

Второе эмпирическое правило состоит в том, что комбинирование прогнозов нередко улучшает качество индивидуальных прогнозов. В серии работ в 1960-е и 1970-е гг. Грэнжер и его соавторы[1] развили технику получения оптимальной комбинации прогнозов, когда все индивидуальные модели являются лишь аппроксимацией процесса, генерирующего данные. Тем не менее, эти работы не дают теоретического объяснения, почему подобные ситуации часто возникают на практике.

Оба эмпирических факта были подробно исследованы на примере большого числа экономических временных рядов из США в работе Стока и Уотсона (Stock & Watson, 1999). Тем не менее, теоретическое обоснование этих фактов до сих пор отсутствует. В данной работе делается небольшой шаг к объяснению данных двух закономерностей.

2 Парадокс

В первую очередь следует подчеркнуть, что описанные эмпирические факты представляют собой некий парадокс с точки зрения стандартного эконометрического подхода к прогнозированию. Действительно, любой временной ряд может быть разложен на предсказуемую и непредсказуемую компоненты: $\exists \mu_{t-1} \in I_{t-1}$, такая, что $y_t = \mu_{t-1} + \varepsilon_t$ и $\mathbb{E}_{t-1}[\varepsilon_t] = 0$, где μ_{t-1} является предсказуемой компонентой (т.е. принадлежит информационному множеству

*Цитировать как: Ицхоки, Олег (2006) «Выбор модели и парадоксы прогнозирования», Квантиль, №1, стр. 43–51. Citation: Itskhoki, Oleg (2006) "Model selection and paradoxes of prediction," Quantile, No.1, pp. 43–51.

†Адрес: Littauer Center, Harvard University, 1875 Cambridge st., Cambridge, MA 02138, USA. Электронная почта: *itskhoki@fas.harvard.edu*

[1]Например, Bates & Granger (1969) и Granger & Newbold (1977).

$I_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$), а ε_t – непредсказуемой компонентой (мартингальным приращением относительно I_{t-1}). В результате, $\mathbb{E}_{t-1}[y_t] \equiv \mathbb{E}[y_t|I_{t-1}] = \mu_{t-1} + \mathbb{E}[\varepsilon_t|I_{t-1}] = \mu_{t-1}$.

Из этого следует, что наилучшим как внутривыборочным, так и вневыборочным прогнозом[2] для y_t является μ_{t-1}. Формально, $\forall g_{t-1} \in I_{t-1}$

$$MSPE(g_{t-1}) \equiv \mathbb{E}\{y_t - g_{t-1}\}^2 = MSPE(\mu_{t-1}) + \underbrace{\mathbb{E}\{\mu_{t-1} - g_{t-1}\}^2}_{\geq 0}.$$

Следовательно, $MSPE(g_{t-1}) \geq MSPE(\mu_{t-1})$ $\forall g_{t-1} \in I_{t-1}$. Данный результат можно обобщить на случай прогноза на h шагов вперед. Согласно закону повторного математического ожидания, оптимальным прогнозом на h шагов вперед является $\mathbb{E}_t y_{t+h} = \mathbb{E}_t \mu_{t+h-1}$.

Данное свойство – это результат того, что непредсказуемая компонента $\varepsilon_t = y_t - \mu_{t-1}$ не коррелирует с любым элементом из информационного множества I_{t-1}:

$$\mathbb{E}\left[(y_t - \mu_{t-1})(\mu_{t-1} - g_{t-1})\right] = \mathbb{E}\left[\mathbb{E}_{t-1}[\varepsilon_t] \cdot (\mu_{t-1} - g_{t-1})\right] = 0.$$

Другими словами, непредсказуемая компонента ε_t не коррелирует с разностью между ошибкой $\nu_t = y_t - g_{t-1}$ произвольного прогноза g_{t-1} и самой собой:

$$\mathbb{E}\left[\varepsilon_t \cdot (\nu_t - \varepsilon_t)\right] = \mathbb{E}\left[\varepsilon_t \cdot (\mu_{t-1} - g_{t-1})\right] = 0 \quad \Leftrightarrow \quad \mathbb{C}[\varepsilon_t, \nu_t] = \mathbb{V}[\varepsilon_t].$$

Это эквивалентно тому, что ошибку ν_t произвольного прогноза g_{t-1} можно разложить на ошибку оптимального прогноза ε_t и полностью прогнозируемый «шум» $\xi_{t-1} \in I_{t-1}$:

$$\nu_t = \varepsilon_t + \xi_{t-1}.$$

Именно поэтому комбинирование прогнозов не может улучшить индивидуальный прогноз μ_{t-1}:

$$\begin{aligned} MSPE\left[\alpha\mu_{t-1} + (1-\alpha)g_{t-1}\right] &= \mathbb{E}\left[(\alpha\varepsilon_t + (1-\alpha)\nu_t)^2\right] \\ &= \mathbb{V}\left[\varepsilon_t\right] + (1-\alpha)^2\mathbb{V}\left[\nu_t - \varepsilon_t\right] \geq \mathbb{V}\left[\varepsilon_t\right] \quad \forall\alpha \in \mathbb{R}. \end{aligned}$$

Отметим, что здесь мы предположили сбалансированность комбинированного прогноза, поскольку сумма весов комбинируемых индивидуальных прогнозов равна единице. Аналогичный аргумент остается верным и в общем случае несбалансированных комбинаций с произвольными весами.

3 Разрешение парадокса

Возможно ли примирить теорию с наблюдаемыми эмпирическими закономерностями? Очевидным и практически тривиальным наблюдением является тот факт, что в реальных экономических приложениях мы никогда не знаем μ_{t-1} и почти никогда не можем ее точно оценить. Заметим также, что с теоретической точки зрения данная ситуация является единственным потенциальным разрешением парадокса.

Точкой отсчета для большинства эконометрических моделей является предположение о том, что μ_{t-1} принадлежит некоторому классу параметрических[3] функций аргумента $z_{t-1} \in I_{t-1}$ и параметра $\beta \in B \subset \mathbb{R}^p$, которые (функции) мы будем обозначать как $f(z_{t-1}; \beta)$. Данная предпосылка позволяет свести сложную задачу оценивания объекта μ_{t-1} к более простой

[2] Критерием качества прогноза является квадратичная функция потерь — стандартный критерий в эконометрике и прогнозировании в частности. В условиях квадратичной функции потерь лучший прогноз минимизирует среднеквадратическую ошибку прогноза – выборочный аналог среднеквадратического отклонения.

[3] Альтернативным подходом является более гибкое непараметрическое оценивание. Есть основание полагать, что непараметрические методы могут иметь лучшие прогнозирующие качества по сравнению с нелинейными параметрическими моделями.

задаче оценивания вектора параметров β. Однако в этом случае нам будет известна лишь оценка $\hat{\mu}_{t-1} = f(z_{t-1};\hat{\beta}) \neq \mu_{t-1}$, а, следовательно, вышеизложенные теоретические результаты для μ_{t-1} уже не должны выполняться для $\hat{\mu}_{t-1}$. Тем не менее, если мы можем состоятельно оценить β, а функция $f(z_{t-1};\cdot)$ является непрерывной, то при достаточно большом размере выборки наши теоретические результаты должны быть приблизительно верными и для $\hat{\mu}_{t-1}$.

Таким образом, анализируемый парадокс можно описать следующим образом:

— Оценка предсказуемой компоненты $\hat{\mu}_{t-1} = f(\{\cdot\};\hat{\beta})$ для класса сложных нелинейных функций $\{f(\{\cdot\};\cdot)\}$ хорошо описывает процесс, генерирующий данные, внутри выборки, однако при этом имеет низкую вневыборочную прогнозирующую силу;[4]

— Комбинирование прогнозов, полученных на основе различных методов оценивая μ_{t-1} (в том числе из разных параметрических классов), позволяет повысить предсказательную силу прогноза вне выборки.

Ниже представлен ряд теоретических аргументов, которые потенциально могут объяснить эти два наблюдения.

Устойчивость к неверной спецификации модели

Истинная функция $f(z_{t-1};\beta)$ вполне может быть нелинейной по параметрам β. При этом данная функция может оказаться весьма сложной и, как следствие, не принадлежать никакому стандартному параметрическому классу.[5]

Таким образом, модели из стандартных параметрических классов являются лишь приближением истинного процесса, генерирующего данные. Тем не менее, они могут частично отражать некоторые нелинейные характеристики процесса, что позволяет данным моделям иметь описательное преимущество внутри выборки над более простыми линейными моделями.

Несмотря на это, линейные модели обладают большей гибкостью. Более формально, они являются устойчивыми к неправильной спецификации функциональной формы модели. Это, в свою очередь, может приводить к более высокому качеству вневыборочного прогноза.

Отметим также, что согласно разложению Вольда[6], любой стационарный процесс имеет линейное представление в виде бесконечного скользящего среднего ($\mathsf{MA}(\infty)$) с некоррелированными инновациями. Из этого следует, что линейные модели временных рядов являются достаточно разумным приближением истинного процесса. В то же время, нелинейные модели при заданном количестве параметров могут иметь как слишком мало, так и слишком много степеней свободы, что может приводить к низкой вневыборочной предсказательной силе. При этом слишком большое число степеней свободы может быть даже хуже, чем слишком маленькое.

Например, рассмотрим задачу прогнозирования «белого шума». Практически любой параметрический класс включает белый шум, как частный случай. Несмотря на это, нелинейные модели, как правило, способны отыскивать «закономерности» и «предсказывать» белый шум внутри выборки. При этом качество вневыборочного прогноза становится намного хуже, чем у простых линейных моделей, которые правильно делают выбор в пользу непредсказуемого белого шума даже внутри выборки.[7]

[4] Следует отметить, что гипотеза линейности нередко отвергается внутри выборки, но, несмотря на это, линейная модель по-прежнему лучше предсказывает вне выборки.

[5] Данное предположение кажется вполне оправданным, особенно если учесть скудность существующих нелинейных моделей временных рядов: билинейные модели, пороговые авторегрессии, модели с марковскими переключениями, модели с ненаблюдаемыми компонентами и несколько других реже используемых моделей.

[6] См., например, учебник Гамильтона по временным рядам (Hamilton, 1994).

[7] Другим примером может стать случай нестационарного временного ряда с детерминированным трендом

Погрешности и смещенность оценивания

Оценивание нелинейных моделей может представлять собой еще одну проблему. В большинстве случаев нелинейные модели не могут быть оценены с высокой степенью точности, по крайней мере, для макроэкономических данных. Как правило, нелинейные модели по сравнению с линейными с тем же количеством параметров требуют существенно больших размеров выборки для получения аналогичной точности оценивания. Более того, наименее точно оцениваются именно те параметры, которые делают модель нелинейной.

Данный факт может иметь следующее интуитивное пояснение. Многие нелинейные классы включают простую линейную модели в качестве частного случая. Однако, если при этом линейная модель истинна, часть параметров нелинейной модели не могут быть идентифицированы (как, например, это происходит в случае пороговой авторегрессии, см. ниже). Если же линейная модель просто является хорошим приближением, то ряд параметров в нелинейной модели могут оказаться слабо идентифицированными.

Данная проблема, как правило, не приводит к снижению качества внутривыборочной подгонки модели. Тем не менее, низкая точность оценки параметров может привести к очень низким вневыборочным прогнозирующим характеристикам модели. Как было показано в эмпирической работе Стоком и Уотсоном (Stock & Watson, 1999), нелинейные модели нередко дают абсолютно неадекватные прогнозы, что достаточно редко случается с линейными моделями. Данное наблюдение, скорее всего, объясняется низкой точностью оценки параметров нелинейной модели, в результате чего нелинейные модели могут выдавать «дикие» прогнозы, особенно если текущая ситуация была нетипичной для выборки.

Таким образом, даже при полностью верной спецификации нелинейной модели из-за неточности оценивания параметров она может давать прогнозы более низкого качества, чем простая линейная модель, параметры которой (являющиеся коэффициентами линейной проекции) оценены точно. Кроме того, линейные модели, как правило, являются более устойчивыми к ошибкам оценивания и выбросам в данных.

Аналогичную проблему представляет собой смещение оцененных параметров нелинейных моделей. Обычно это смещение является значительным в небольших выборках. Все вычисления в предыдущем разделе, очевидно, опирались на то, что μ_{t-1} является несмещенным прогнозом; в противном случае свойства оптимальности μ_{t-1} не выполняются. Даже в условиях верной спецификации нелинейной модели, значительные смещения в конечных выборках могут приводить к бо́льшим $MSPE$ нелинейных моделей по сравнению с линейными.

Следует отметить, что если линейная модель является лишь приближением, ее коэффициенты зачастую не имеют существенной экономической интерпретации. Тем не менее, они имеют точные аналоги в популяции – коэффициенты теоретической линейной проекции. Эти коэффициенты, как правило, могут быть оценены с высокой точностью и без значительных смещений даже в небольших выборках. Таким образом, прогнозы линейных моделей могут быть очень близкими к популяционным прогнозам, основанным на линейной проекции.

Также важно отметить, что присутствие значительных смещений при оценивании параметров может являться одной из важных причин, приводящих к значительным преимуществам комбинированных прогнозов над индивидуальными. В условиях, когда различные модели дают не слишком сильно коррелированные смещения, комбинирование прогнозов может уменьшать $MSPE$ просто за счет снижения этих смещений.

сложной функциональной формы. Очевидно, что простая модель линейного тренда будет доминироваться более сложными полиномами внутри выборки. При этом с большой вероятностью линейная модель побьет любой полином во вневыборочном прогнозировании.

Структурные сдвиги и дрейф

Другим свойством экономических временных рядов, потенциально приводящим к наблюдаемым закономерностям, могут являться структурные сдвиги или дрейф в параметрах модели. Даже незначительные сдвиги в параметрах истинного процесса могут приводить к большим погрешностям при прогнозировании с помощью нелинейных моделей, оцененных по выборке до структурного сдвига.

Линейные модели нередко оказываются более устойчивыми к различным стуктурным сдвигам в параметрах истинного процесса. Именно этот факт, вероятно, объясняет, почему качество прогнозов нелинейных моделей резко ухудшается с ростом горизонта прогнозирования. Так, Сток и Уотсон (Stock & Watson, 1999) демонстрируют, что нелинейные модели дают неплохие прогнозы, сравнимые с прогнозами линейных моделей или даже лучше, на один-два шага вперед, однако прогнозы резко ухудшаются с ростом горизонта и становятся существенно хуже прогнозов линейных моделей.

Комбинирование прогнозов

Любые оцениваемые модели являются лишь приближением к истинному процессу, генерирующему данные. Даже в том случае, когда нелинейная модель правильно специфицирована, неточности оценивания и смещение в конечных выборках, наряду с потенциальными структурными сдвигами в параметрах модели, приводят к существенным отклонениям оцененного прогноза $\hat{\mu}_{t-1}$ от оптимального прогноза μ_{t-1}. В подобных условиях комбинирование прогнозов может иметь значительный потенциал в повышении качества прогнозирования.

Рассмотрим следующий пример, который во многом может мотивировать использование комбинированных прогнозов. Допустим, перед исследователем стоит проблема пронозирования процесса, который по своей структуре является агрегатом некоторого числа микропроцессов. Например, необходимо спрогнозировать совокупное производство страны, которое складывается из региональных выпусков отдельных субъектов. В этом случае гораздо более эффективно прогнозировать каждый индивидуальный ряд по отдельности, а затем агрегировать прогнозы оптимальным образом. Теоретически это можно обосновать с помощью закона больших чисел, поскольку агрегирование индивидуальных прогнозов позволяет «диверсифицировать» ошибки и таким образом снизить погрешность.

Несмотря на то, что большинство экономических временных рядов являются агрегатами, их индивидуальные компоненты зачастую недоступны. Тем не менее, различные модели временных рядов, являясь разными приближениями истинного процесса, могут лучше или хуже описывать различные характеристики истинного процесса. В этом случае ошибки прогнозирования различных моделей могут оказаться слабо коррелированными,[8] а это, в свою очередь, позволяет комбинациям прогнозов превышать предсказательную силу индивидуальных прогнозов.

Напомним, что качество комбинации двух произвольных прогнозов g'_{t-1} и g''_{t-1} с ошибками ν'_t и ν''_t и весами α и $1-\alpha$ соответственно может быть охарактеризовано как

$$MSPE\big[\alpha g'_{t-1} + (1-\alpha) g''_{t-1}\big] = \alpha^2 \mathbb{V}[\nu'_t] + (1-\alpha)^2 \mathbb{V}[\nu''_t] + 2\alpha(1-\alpha)\mathbb{C}[\nu'_t, \nu''_t].$$

Как видно из этой формулы, комбинация прогнозов может иметь меньший $MSPE$ по сравнению с индивидуальными прогнозами, если ошибки двух индивидуальных прогнозов слабо коррелированы.[9]

[8] Что подтверждается в новой эмпирической работе Стока и Уотсона (Stock & Watson, 2006).

[9] Отрицательная корреляция ошибок прогнозов встречается крайне редко, и этот случай можно исключить из анализа практически без потери общности с точки зрения прикладного анализа.

4 Численный пример

В качестве иллюстрации некоторых приведенных выше аргументов разберем один частный случай на численном примере. В качестве примера возьмем стандартную гладкопороговую авторегрессию (STAR) с меняющимся во времени авторегрессионным параметром. Пусть истинный процесс задается следующим образом:

$$y_t = (\alpha + \beta e^{-\gamma y_{t-1}^2}) y_{t-1} + \varepsilon_t, \qquad \alpha, \beta, \gamma \geq 0, \tag{1}$$

где ε_t является мартингальным приращением с постоянной условной дисперсией σ_ε^2.

Данный процесс не пользуется особой популярностью среди эмпирических эконометристов, однако он имеет простое содержательное обоснование и может вполне адекватно описывать некоторые эмпирические закономерности. Данный процесс позволяет временному ряду быть более связным в окрестности маленьких по абсолютной величине значений y_t, при этом связность уменьшается по мере того, как y_t растет по абсолютной величине и отклоняется от своего долгосрочного равновесного значения – безусловного математического ожидания $\mathbb{E}[y_t] = 0$.

Альтернативной, линейной, спецификацией является обычная авторегрессия первого порядка AR(1):

$$y_t = \rho y_{t-1} + \nu_t, \qquad \rho \geq 0, \tag{2}$$

где ν_t является ошибкой линейной проекции с некоторой условной дисперсией σ_ν^2. Из теоретических соображений следует, что $\sigma_\nu^2 \geq \sigma_\varepsilon^2$ с равенством только в случае справедливости нулевой гипотезы линейности, $\mathbb{H}_0 : \beta = 0$.

Ошибки прогнозирования для истинной нелинейной модели и линейного AR-приближения равны, соответственно,

$$\begin{aligned} \hat{\varepsilon}_{t+1|t} &\equiv y_{t+1} - (\hat{\alpha} + \hat{\beta} e^{-\hat{\gamma} y_t^2}) y_t \\ &= \varepsilon_{t+1} + \left[(\alpha - \hat{\alpha}) + e^{-\gamma y_t^2} (\beta - \hat{\beta}) + \hat{\beta} (e^{-\gamma y_t^2} - e^{-\hat{\gamma} y_t^2}) \right] y_t, \\ \hat{\nu}_{t+1|t} &\equiv y_{t+1} - \hat{\rho} y_t = \nu_{t+1} + (\rho - \hat{\rho}) y_t. \end{aligned}$$

Таким образом, приблизительная (асимптотическая) среднеквадратическая ошибка прогнозирования для нелинейной модели STAR равна

$$MSPE_{T+1|T}^{\text{STAR}} = \sigma_\varepsilon^2 \left[1 + g'_{\theta,T+1} \left(\sum_{t=1}^{T-1} g_{\theta,t} g'_{\theta,t} \right)^{-1} g_{\theta,T+1} \right],$$

где $g_t(\theta) \equiv (\alpha + \beta e^{-\gamma y_{t-1}^2}) y_{t-1}$, $\theta \equiv (\alpha, \beta, \gamma)'$ и $g_{\theta,t} \equiv \partial g_t(\theta)/\partial \theta'$. Асимптотическая среднеквадратическая ошибка для AR-модели равна, соответственно,

$$MSPE_{T+1|T}^{\text{AR}} = \sigma_\nu^2 \left[1 + y_{T+1}^2 / \sum_{t=1}^{T-1} y_t^2 \right].$$

Отметим, что из априорных соображений невозможно упорядочить эти две среднеквадратические ошибки прогнозов. Ошибка проекции ν_t имеет большую дисперсию, чем ошибка лучшего прогноза ε_t, однако второй элемент в формуле $MSPE$ в квадратных скобках, связанный с погрешностями при оценивании параметров, может вполне быть больше для нелинейной модели.

Для дальнейшего анализа мы используем числовой пример, где процессом, генерирующим данные, является (1) со следующими параметрами: $\alpha = 0,7$, $\beta = 0,3$, $\gamma = 0,5$ и $\sigma_\varepsilon^2 = 0,25$, что соответствует достаточно связному, но все же стационарному процессу.[10]

[10]При этом большинство полученных результатов являются качественно устойчивыми при значительном изменении исходных параметров.

Таблица 1: Асимптотический $MSPE$ для разных размеров выборки

Размер выборки	σ_ε^2	STAR	AR
100	0,250	0,549	0,257
200	0,250	0,350	0,255
500	0,250	0,265	0,254
1000	0,250	0,252	0,254

Таблица 2: Качество прогноза в конечных выборках

Модель	STAR		AR	
Размер выборки	Внутри выборки	Вне выборки	Внутри выборки	Вне выборки
100	0,244	0,359	0,266	0,279
200	0,247	0,295	0,258	0,266
500	0,249	0,255	0,257	0,256

Мы проведем два эксперимента. Во-первых, мы оцениваем приблизительные (асимптотические) $MSPE$ двух моделей для размеров выборки в 100, 200, 500 и 1000 наблюдений, в соответствие с приведенными выше формулами. Отметим, что эти показатели не учитывают смещенность оценок параметров в конечных выборках. Результаты первого эксперимента приведены в таблице 1. Во-вторых, мы оцениваем эмпирическую предсказательную силу двух моделей в конечных выборках: мы подсчитываем характеристики внутривыборочной подгонки данных (дисперсию ошибки внутри выборки) и характеристики ошибки вневыборочных предсказаний (дисперсию ошибки предсказания на один шаг вперед) для различного числа исходных наблюдений, используя метод Монте-Карло. Результаты второго эксперимента представлены в таблице 2. Таблица 1 наглядно демонстрирует основной вывод первого эксперимента: при небольших размерах выборки линейная модель бьет истинную нелинейную модель. Это результат неточности оценивания параметров нелинейной модели – данная нелинейная модель требует более 500 наблюдений для достаточно точной оценки параметров, чтобы превзойти линейную модель по показателю $MSPE$. Параметры же линейной модели оцениваются с очень высокой точностью даже при выборках в 100 наблюдений. Из таблицы 2 можно сделать два важных наблюдения. Во-первых, внутривыборочная подгонка истинной нелинейной модели всегда лучше подгонки линейной проекции. Это очевидный факт, поскольку линейная проекция является ограниченной версией исходной модели. Более того, внутривыборочная подгонка для оцененной нелинейной модели превосходит даже истинный процесс, генерирующий данные (другими словами, нелинейную модель с истинными коэффициентами), для которого ошибка подгонки равна $\sigma_\varepsilon^2 = 0,25$. Этот любопытный факт сродни описанному выше примеру, в котором нелинейные модели способны отыскивать закономерности при описании белого шума. Также отметим, что данное отличие во внутривыборочной подгонке данных приводит к отклонению гипотезы линейности в пользу принятия истинной модели.[11]

Во-вторых, вневыборочная предсказательная сила (среднее квадрата ошибки прогноза на один шаг вперед по большому количеству симуляций) нелинейной модели ниже предсказательной силы линейной модели при умеренных размерах выборки (менее 500 наблюдений).

[11]Формальный тест на линейность имеет вид $\sup LR$ статистики, распределение которой принадлежит к семейству χ^2-функционалов (Hansen, 1996). Значение этой статистики равно $91,7$ уже при 100 наблюдениях, что существенно превосходит критические значения.

Это уже не такой очевидный факт. Заметим также, что качество прогноза нелинейной модели существенно снижается при размерах выборки менее 200 наблюдений, в то время как качество линейной модели практически не меняется. Кроме того, примечательно, что линейная модель дает примерно одинаковое качество как внутривыборочной, так и вневыборочной подгонки данных, в то время как для нелинейной модели расхождение качества внутривыборочной и вневыборочной подгонки является очень значительным.

Также заметим, что корреляция между ошибками прогнозов двух рассматриваемых моделей составляет более $0,8$, что практически не позволяет улучшить итоговый прогноз при помощи комбинирования двух моделей. Это и неудивительно, поскольку обе модели являются очень хорошими и достаточно точными приближениями к истинному процессу.

Таким образом, данный простой пример позволяет наглядно проиллюстрировать многие из вышеизложенных теоретических гипотез.

5 Заключение

Данная работа является небольшим шагом к объяснению двух парадоксальных эмпирических закономерностей в области прогнозирования экономических временных рядов. Мы попытались изложить ряд теоретических гипотез, объясняющих, почему простые линейные модели зачастую обладают преимуществом в предсказательной силе над более сложными нелинейными моделями, позволяющими, в свою очередь, получить более точную внутривыборочную подгонку данных. Оказывается, многие из этих гипотез позволяют также объяснить, почему комбинирование индивидуальных прогнозов может улучшать их предсказательную силу.

Обобщая все вышеизложенное, можно подвести итог, что преимущество линейных моделей происходит из-за их общей «робастности» – устойчивости к неверной спецификации модели, устойчивости к смещению и неточности при оценивании, устойчивости к структурным сдвигам и дрейфу параметров модели.

Комбинирование прогнозов может быть оптимальным в условиях, когда все индивидуальные модели являются несовершенным приближением истинного процесса. Кроме того, вероятно, во многих ситуациях сам прогнозируемый процесс является комбинацией более простых микропроцессов.

Предложенный численный пример иллюстрирует теоретические гипотезы. Результаты численного анализа внушают оптимизм по поводу высказанных теоретических идей. Дальнейшие исследования данной проблемы должны выявить конкретные механизмы, которые могут стоять за предложенными гипотезами, а также оценить важность этих механизмов эмпирически.

Благодарности

Автор благодарен Джеймсу Стоку за постановку проблемы, анализируемой в работе, и Станиславу Анатольеву за ознакомление с основами современной эконометрики, а также за помощь при переводе и редактировании текста.

Список литературы

Bates, J. & C. Granger (1969). The combination of forecasts. *Operations Research Quarterly* 20, 451–468.

Chan, K. (1990). Testing for threshold autoregression. *Annals of Statistics* 18, 1886–1894.

Granger, C. & P. Newbold (1977). Forecasting Economic Time Series. New York: Academic Press.

Hamilton, J. (1994). Time Series Analysis. Princeton: Princeton University Press.

Hansen, B. (1996). Inference when a nuisance parameter is not identified under the null. *Econometrica* 64, 413–430.

Stock, J. & M. Watson (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. Глава 1 в R. Engle & H. White (eds.). *Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger*. Oxford: Oxford University Press, 1–44.

Stock, J. & M. Watson (2006). Forecasting with many predictors. Глава в G. Elliott, C. Granger, A. Timmermann, K. Arrow & M. Intriligator (eds.). *Handbook of economic forecasting*. Elsevier: North Holland.

Model selection and paradoxes of prediction

Oleg Itskhoki

Harvard University, Cambridge, USA
Central Economics & Mathematics Institute, Moscow, Russia

In this essay we postulate a number of theoretical hypotheses allowing one to resolve in some degree the following two prediction paradoxes: (1) why simple linear models often have an advantage in predictive power over more complex nonlinear models that lead to a better in-sample fit; (2) why combinations of forecasts often increase the predictive power of individual forecasts. We also give a numerical example illustrating our theoretical statements.