УДК 681.513.685

ФИЛЬТРАЦИЯ ВРЕМЕННЫХ РЯДОВ С ПРОПУСКАМИ НАБЛЮДЕНИЙ

Е.Ю. Алексеева, А.А. Беседин

Построен алгоритм фильтрации случайных процессов с пропусками наблюдений. Численные эксперименты показали среднеквадратическую ошибку прогнозирования реального процесса валютного курса USD / RUB на один шаг (день) порядка 1 %.

Ключевые слова: идентификация, оценивание параметров, линейные системы, нормальное распределение, максимальное правдоподобие, фильтр Калмана, численная максимизация.

Введение. Любой процесс в малом диапазоне колебаний всегда линеаризуется. Когда изменения процесса не слишком существенны (процесс работает в локальной окрестности рабочей точки), то он приближенно может быть линеаризован. Модель получается линейной. Локальная линеаризация – стандартный прием теории управления. Глобальная модель невозможна в какой-то форме.

Есть массив данных. На каждом интервале массива данных, который выглядит как стационарный, мы можем построить стационарную модель и определить коэффициенты. Коэффициенты модели зависят от области изменения процесса и некоторых возмущающих факторов.

Отсюда – задача. Построить совокупность моделей для различных интервалов изменений и проанализировать численно, как меняются параметры моделей от диапазонов изменений.

Исходным материалом служит последовательность наблюдений за реализацией случайного процесса на валютном рынке.

$$Z_1^n = [Z_1\,, Z_2\,, Z_3\ \dots\ Z_n\,].$$

Следует отметить, что механизм формирования цен на валютном рынке чрезвычайно сложен, он включает в себя большое количество факторов, трудных для выявления. Временные ряды такого рода содержат ряд статистических закономерностей: таких как тяжелые хвосты и острый пик распределения, нестационарность, волатильности. Исходя из этого, авторами [1–2] была выявлена ограниченная предсказуемость курса валют, но, тем не менее, авторами был сделан вывод о возможности их прогнозирования в краткосрочном периоде.

Ряд исследователей [3–6], считают, что прогнозировать на валютном рынке можно только преимущественно с помощью нелинейных моделей.

Но в нашей работе проведенные исследования, позволяют сделать вывод об отсутствии преимуществ нелинейных моделей перед классическими линейными, что было и ранее отмечено в работе [7]. Так, для построения нелинейной модели необходимо определить и задать специфические особенности и характеристики конкретного валютного рынка на конкретном периоде времени. Учитывая подобную субъективность, традиционные линейные модели являются наиболее общим и качественным инструментом прогнозирования.

Прогнозирование значения процесса Z_{n+1} состоит в вычислении условной плотности вероятности $f(Z_{n+1} \,/\, Z_1^n)$. По полученной плотности строится либо точечный прогноз величины Z_{n+1}, либо интервальный. Точечный прогноз обычно дается условным математическим ожиданиям $M(Z_{n+1} \,/\, Z_1^n)$. Интервальный прогноз строится по плотности вероятности и значению доверительной вероятности. Основой для решения этих задача служит математическая модель процесса, содержащая структуру и параметры модели. Параметры оцениваются по наблюдениям обычно методом максимального правдоподобия. После этого модель позволяет строить прогноз. Если модель признается адекватной, можно оценить точность прогнозирования, например, по среднему квадрату ошибки прогнозирования и использовать результаты прогнозирования в принятии решений относительно изучаемого процесса. Средний квадрат ошибки прогнозирования может служить мерой адекватности модели и процесса.

Математическая модель. Используется популярная модель АРМА для построения процедуры прогноза. Принимаем гипотезу, что изучаемый процесс описывается уравнением движения.

$$Y_{k+1} = a_1 y_k + a_2 y_{k-1} + v_k. \quad (1)$$

где y_k – значение процесса, v_k – внутренний шум модели, типа белого шума – последовательности независимых величин с некоторым математическим ожиданием mv и дисперсией dv. Уравнение наблюдений

$$Z_k = y_k + E_k\,. \quad (2)$$

где E_k – ошибка измерения, независимый белый шум с математическим ожиданием me и дисперсией de. Распределения величин v_k и E_k предполагаются нормальными. Параметры модели a_1, a_2, mv, dv, me, de предполагаются на интервале прогноза постоянными и оцениваются по результатам наблюдений Z_k. Очевидно, что и начальные значения последовательности y_1 и y_2 должны быть включены в параметры модели и оцениваться вместе. При желании размерность модели может быть увеличена.

Преобразуем уравнение (1) в систему уравнений первого порядка.

Вектор параметров модели

$$\theta = (y_1, y_2, a_1, a_2, mv, dv, me, de).$$

Вводим вектор состояния

$$X_k = (y_k, y_{k+1})^T \qquad X_1 = (y_1, y_2)^T.$$

Тогда уравнение (1) может быть записано в виде

$$X_{k+1} = AX_k + W_k, \qquad A = \begin{bmatrix} 0 & 1 \\ a_2 & a_1 \end{bmatrix}, \tag{3}$$

где

$$X_{k+1} = (y_{k+1}, y_{k+2})^T,\ W_k = (0 \quad V_{k+1}).$$

Уравнение наблюдения

$$Z_k = HX_k + E_k \quad H = [1 \quad 0]. \tag{4}$$

Первой задачей является задача вычисления оценок параметров модели θ. Для её решения используем метод максимального правдоподобия

$$\hat{\theta} = argmax_\theta\, f(Z_1^n \,/\, \theta).$$

Для упрощения записи аргумент θ будем опускать. Как известно

$$f(Z_1^n) = f(Z_1, Z_2 \dots Z_n) = f(Z_1) \prod_{k=2}^n f(Z_k \,/\, Z_1^{k-1}). \tag{5}$$

Допустим, что уже вычислена плотность вероятности вектора состояния X_k при наличии наблюдений Z_1^{k-1}. Как известно, в условиях задачи $f\left(X_k / Z_1^{k-1}\right)$ будет нормальной с некоторым средним значением m_k и матрицей ковариации D_k

$$f(X_k \,/\, Z_1^{k-1}) = N(m_k, D_k). \tag{6}$$

Тогда из уравнения наблюдения (4) следует, что $f\left(Z_k / Z_1^{k-1}\right)$ будет нормальной с математическим ожиданием $Hm_k + me$ и дисперсией $HD_kH^T + de$.

$$f\left(Z_k / Z_1^{k-1}\right) = N(Hm_k + me, HD_kH^T + de). \tag{7}$$

Соотношения (5), (7) дают алгоритм вычисления функции правдоподобия, которая при известных измерениях Z_1^n является функцией вектора параметров θ. Численное определение экстремума функции правдоподобия любым известным методом по вектору θ дает оценки $\hat{\theta}$. Последовательность m_k, D_k вычисляется по рекуррентным соотношениям (П7).

$$m_{k+1} = Am_k + \frac{AD_kH^T}{de + HD_kH^T}(Z_k - Hm_k - me)$$

$$D_{k+1} = AD_kA^T - \frac{AD_kH^THD_kA^T}{de + HD_kH^T} + DW,\ \ DW = \begin{pmatrix} 0 & 0 \\ 0 & dv \end{pmatrix}.$$

Функция правдоподобия $f(Z_1^n)$ согласно (5), (7) имеет вид:

$$f(Z_1^n) = \frac{1}{\sqrt{2\pi de}} e^{\frac{-\frac{1}{2}(Z_1 - me - Y_1)^2}{de}} \prod_{k=2}^n \frac{1}{2\pi\sqrt{HD_kH^T + de}} \exp\left(-\frac{1}{2}\, \frac{(Z_k - Hm_k - me)^2}{HD_kH^T + de}\right).$$

Удобнее рассматривать логарифмическую функцию правдоподобия (константы 2π отбросим).

$$L(Z_1^n, \theta) = -2\ln f(Z_1^n) = \ln(de) + \frac{(Z_1 - me - Y_1)^2}{de} +$$

$$+ \sum_{k=2}^n \ln(HD_kH^T + de) + \sum_{k=2}^n \frac{(Z_k - Hm_k - me)^2}{HD_kH^T + de}. \tag{8}$$

Минимум выражения (8) по θ соответствует максимуму функции правдоподобия.

Заключение. По соотношениям (7), (8) разработана программа для прогнозирования курса доллара в пакете MATLAB для различных периодов данных. Используемые в работе временные ряды являются последовательной выборкой (объема) за период с 04 января 2000 года по 05 марта 2020 года из динамики курса валют, доступной на сайте http://www.cbr.ru/currency_base/dynamics/. Численные эксперименты показали точность прогнозирования курса доллара на 1 шаг около 1 % (ошибка прогноза).

Приложение. Построим рекуррентные уравнения, связывающие последовательные значения параметров условных распределений $f(X_k\,/\,Z_1^{k-1})$ и $(X_{k+1}\,/\,Z_1^{k})$ (фильтр Калмана) [8].

Согласно (6)

$$f(X_k\,/\,Z_1^{k-1}) = \frac{1}{2\pi\sqrt{|D_k|}}\exp\left(-\frac{1}{2}(x_k-m_k)^T D_k^{-1}(x_k-m_k)\right). \qquad (П1)$$

По формуле Байеса

$$f(X_k\,/\,Z_1^{k}) = \frac{f(X_k\,/\,Z_1^{k-1}) * f(Z_k\,/\,X_k Z_1^{k-1})}{f(Z_k\,/\,Z_1^{k-1})}. \qquad (П2)$$

При фиксированном X_k из (4) следует, что распределение Z_k нормально с математическим ожиданием $HX_k + me$ и дисперсией de. Следовательно

$$f(X_k\,/\,Z_1^{k}) = const\,\exp\left(-\frac{1}{2}(x_k-m_k)^T D_k^{-1}(x_k-m_k)\right) \times$$

$$\times \exp\left(-\frac{1}{2}\frac{(Z_k-HX_k-me)^2}{de}\right). \qquad (П3)$$

Знаменатель (П2) при известных наблюденных есть константа и включен в const. Правая часть (П3) есть выражение, подобное (П1), следовательно представляет собой нормальную плотность вероятности с некоторым параметрами, математическим ожиданием M_k и матрицей ковариации Γ_k.

$$f(X_k\,/\,Z_1^{k}) = \frac{1}{2\pi\sqrt{|\Gamma_k|}}\exp\left(-\frac{1}{2}(x_k-m_k)^T \Gamma_k^{-1}(x_k-m_k)\right). \qquad (П4)$$

Приравнивая коэффициенты при одинаковых степенях X_k в (П3) и (П4) получаем соотношения для вычисления M_k, Γ_k.

$$\begin{cases} \Gamma_k^{-1} = D_k^{-1} + \dfrac{H^T H}{de} \\ T_k^{-1} M_k = D_k^{-1} m_k + \dfrac{H^T}{de}(Z_k - me) \end{cases} \qquad (П5)$$

Соотношения (П5) можно представить в более удобном виде

$$\begin{cases} \Gamma_k = D_k - \dfrac{D_k H^T H D_k}{de + H D_k H^T} \\ M_k = m_k + \dfrac{D_k H^T}{de + H D_k H^T}(Z_k - H m_k - me) \end{cases}. \qquad (П6)$$

Из соотношений (3) и (П6) следует что

$$m_{k+1} = M\{X_{k+1}\,/\,Z_1^{k}\} = A * M_k = A m_k + \frac{A D_k H^T}{de + H D_k H^T}(Z_k - H m_k - me) \qquad (П7)$$

$$D_{k+1} = cov\{X_{k+1} \,/\, Z_1^k\} = AD_kA^T - \frac{AD_kH^THD_kA^T}{de+HD_kH^T} + DW,$$

здесь $DW = cov\{W_k\} = \begin{pmatrix} 0 & 0 \\ 0 & dv \end{pmatrix}$.

При наличиях пропусков измерений соотношения (П7) примут вид

$$\begin{cases} m_{k+1} = Am_k \\ D_{k+1} = AD_kA^T + DW \end{cases}. \qquad (П8)$$

Это эквивалентно условию $de = \infty$.

Начальные условия для системы (П7)

$$m_1 = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = X_1, \; D_1 = 0.$$

Библиографический список

1. Балошникова, В.А. методы хаоса в анализе обменных курсов валют по центробанку россии / В.А. Балошникова // Вестник инжэкона. Серия: экономика. – 2009. – № 3(30). – С. 251–252.

2. Балошников, А.М. Моделирование динамики обменных курсов основных валют / А.М. Балошников, В.А.Балошникова // Прикладная информатика. – 2010. – № 1(25). – С. 15–20.

3. Мусаев, А.А. Quod est veritas. Трансформация взглядов на системную составляющую наблюдаемого процесса / А.А. Мусаев // Труды СПИИРАН. – 2010. – Вып. 15. – С. 53–74.

4. Brooks, C. Testing for non-linearity in daily sterling exchange rates / C. Brooks // Applied Financial Economics. – 1996. – P. 307–317.

5. De Jong. The likelihood for a state space model / De Jong // Biometrika. – 1988. – P. 165–169.

6. Engle, R.F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom inflation / R.F. Engle // Econometrica. – 1982. – P. 987–1007.

7. Kohn, R. Estimation, prediction, and interpolation for ARFMA models with missing data / R. Kohn, C.F. Ansley // Journal of the American Statistical Association. – 1986. – P. 751–761.

8. Калман, Р. Очерки по математической теории систем / Р. Калман, П. Фалб, М. Арбиб; пер. с англ. под ред. Я.З. Цыпкина. – М.: Мир, 1971. – 400 с.

К содержанию