УДК 621.317.35

***М.Р. Марков*[1], *В.В. Моттль*[2], *И.Б. Мучник*[3], *О.В. Красоткина*[4]**
[1] Markov Processes International, USA, michael.markov@markovprocesses.com
[2] Вычислительный центр им. Дородницына РАН, Россия, vmottl@yahoo.com
[3] Rutgers University, USA, muchnik@dimacs.rutgers.edu
[4] Тульский государственный университет, Россия, krasotkina@uic.tula.ru

Вложенные классы моделей нестационарности сигнала в динамическом анализе состава инвестиционного портфеля

Принцип минимизации структурного риска вложенных классов моделей данных часто применяется в задачах распознавания образов и восстановления регрессионной зависимости в качестве основного инструмента выбора подходящего класса модели. Для подсчета структурного риска мы используем метод скользящего контроля с помощью его инкорпорирования в стандартную схему процедуры динамического программирования. Предложенную технику мы используем для слежения за составом инвестиционного портфеля и анализа стратегии инвестиционной компании.

Введение

В данной работе исследуется проблема построения нестационарных моделей сигналов, представимых в виде модели нестационарной линейной регрессии. Задача оценивания нестационарной регрессии неизбежно связана с необходимостью выбора подходящего уровня нестационарности модели, который изменяется от полной стационарности мгновенных моделей до их полной независимости друг от друга. Мы показываем, что увеличение степени нестационарности определяет последовательность «почти вложенных» классов моделей анализируемого временного ряда. В качестве основного инструмента выбора уровня нестационарности или, другими словами, выбора наиболее подходящего класса моделей мы применяем принцип минимизации структурного риска [1], [2], изначально предложенный для анализа стационарных данных. Ярким примером задачи восстановления нестационарной регрессии является задача анализа состава портфеля инвестиционной компании, играющая огромную роль в современном анализе инвестиций. Мы рассматриваем здесь динамическое обобщение этой задачи, изначальная статическая постановка которой дана нобелевским лауреатом 1990 года по экономике Уильямом Шарпом. Степень нестационарности модели в данной задаче определяет характер управления инвестиционным портфелем. Понятие вложенных моделей нестационарности и предложенный метод оценивания степени нестационарности модели позволяют осуществлять более детальный и точный, в сравнении с традиционной методологией, анализ стратегии инвестиционного портфеля.

Задача анализа состава инвестиционного портфеля

Портфелем инвестиционной компании называется совокупность финансовых активов, которыми владеет инвестиционная компания в пропорциях, обозначенных коэффициентами $(\beta^{(0)},\beta^{(1)},...,\beta^{(n)})$, $\sum_{i=0}^{n}\beta^{(i)}=1$. Долевой состав портфеля обычно неизвестен и такая информация была бы чрезвычайно интересна внешним наблюдателям, в частности, инвесторам в этот портфель, предупреждая о возможной опасности для их капитала. В данной задаче имеется доступ только к информации о доходностях (относительном изменении стоимости) портфеля и ценных бумаг или их классов, его составляющих. Уильям Шарп показал, что если за рассматриваемый период никакие средства не поступали в портфель извне и не изымались из него, то доходность портфеля $r^{(p)}$ определяется как линейная комбинация $r^{(p)} \cong \sum_{i=0}^{n} \beta^{(i)} r^{(i)} + e$ доходностей $r^{(i)}$ составляющих его классов ценных бумаг. Подобные модели Шарп называл многофакторными. Отсюда следует принцип оценивания состава портфеля по Шарпу, известный в современном инвестиционном анализе под названием Returns Based Style Analysis и приводящий к задаче квадратичного программирования:

$$(\hat{\beta}^{(0)},...,\hat{\beta}^{(n)}) : \sum_{t=1}^{N} \left(r_t^{(p)} - \sum_{i=0}^{n} \beta^{(i)} r_t^{(i)} \right)^2 \to \min, \beta^{(i)} \geq 0, \ \sum_{i=0}^{n} \beta^{(i)} = 1 .$$

Модель Шарпа имеет по крайней мере одно существенное ограничение – малореалистичное предположение о неизменности состава портфеля инвестора в течение всего периода владения. Для преодоления этого недостатка Шарп использовал скользящее окно некоторой длины, однако этот прием основан на вынужденном предположении, что состав портфеля не изменялся внутри окна.

В отличие от статической модели Шарпа, мы рассматриваем здесь динамическую многофакторную модель с изменяющимся долевым составом портфеля $\mathbf{в}_t = (\beta_t^{(1)}, ..., \beta_t^{(n)})$. Можно доказать, что для каждого элементарного интервала времени справедливо уравнение, аналогичное статическому уравнению при тех же условиях:

$$y_t = (r_t^{(p)} - r^{(0)}) \cong \sum_{i=1}^{n} \beta_{t-1}^{(i)} (r_t^{(i)} - r^{(0)}) \ = \boldsymbol{\beta}_{t-1}^T \mathbf{x}_t .$$

Ключевым моментом предлагаемого динамического анализа инвестиционного портфеля является понимание коэффициентов регрессии как скрытого процесса, обладающего марковским свойством

$$\boldsymbol{\beta}_t = \mathbf{V}_t \boldsymbol{\beta}_{t-1} + \boldsymbol{\xi}_t , \qquad (1)$$

где матрица $\mathbf{V}_t$ определяет предполагаемую скрытую динамику состава портфеля, а $\boldsymbol{\xi}_t$ вектор белого шума.

Для оценивания нестационарной модели такого вида мы используем метод, получивший в англоязычной литературе название Flexible Least Squares [3], который применительно к анализу инвестиционного портфеля выглядит следующим образом

$$\begin{cases} \hat{\boldsymbol{\beta}}(Y, X, \lambda) = \arg\min J(\boldsymbol{\beta}_t, t = 1,...,N \mid Y, X), \\ J(\boldsymbol{\beta}_t, t = 1,...,N \mid Y, X) = \sum_{t=1}^{N} (y_t - \boldsymbol{\beta}_{t-1}^T \mathbf{x}_t)^2 + \lambda \sum_{t=2}^{N} (\boldsymbol{\beta}_t - \mathbf{V}_t \boldsymbol{\beta}_{t-1})^T \mathbf{U}_t (\boldsymbol{\beta}_t - \mathbf{V}_t \boldsymbol{\beta}_{t-1}). \end{cases} \qquad (2)$$

Вложенные классы нестационарности модели сигнала

Предполагается, что параметры критерия (2) задаются априори. Это предположение оказывается справедливым только для матриц $\mathbf{U}_t$ и $\mathbf{V}_t$ модели перехода (1), которые фактически определяют «вид» скрытой динамики нестационарных коэффициентов регрессии. Что же касается коэффициента λ, который отвечает за «интенсивность» сглаживания, то оказывается весьма проблематичным выбрать его значение априори. При возрастании параметра λ от 0 до ∞ характер мгновенных оценок $\hat{\boldsymbol{\beta}}_t(Y, F, \lambda)$ постепенно последовательно меняется от полной независимости друг от друга при $\lambda \to 0$ до их полной независимости от времени при $\lambda \to \infty$. В терминах статистической теории распознавания образов [2] возрастающие значения λ определяют последовательность «почти» вложенных классов моделей анализируемого временного ряда, в одном из которых и выбирается конкретная модель. Если $\lambda''>\lambda'$, то класс более детерминированных нестационарных моделей, определяемый λ'', «почти содержится» внутри класса более подвижных моделей, соответствующего λ'.

Одним из наиболее известных способов выбора наиболее подходящего класса моделей из множества вложенных классов моделей является принцип минимизации структурного риска [1], [3]. Именно его мы используем здесь для выбора параметра нестационарности λ.

В данной работе мы предполагаем, что последовательность регрессоров $X=(\mathbf{x}_t, t=1,...,N)$ фиксирована. Но выходная переменная $Y=(y_t, t=1,...,N)$ из-за присутствия в модели шума наблюдения ξ_t представляет собой случайный процесс. Для подлежащей анализу временной последовательности $Y=(y_t, t=1,...,N)$, являющейся одной из реализаций этого процесса, эффективной мерой качества модели может служить сумма квадратов остаточной ошибки восстановления выходной переменной в каждой точке оси сигнала. В соответствии с терминологией Вапника мы называем этот функционал эмпирическим риском класса моделей, определенного конкретным значением параметра λ $Emp(Y,\lambda \mid X) = (1/N)\sum_{t=1}^{N}\left(y_t - \mathbf{x}_t^T\hat{\boldsymbol{\beta}}_t(Y,X,\lambda)\right)^2$. Полной характеристикой класса, связанного с конкретным значением параметра $Emp(Y,\lambda \mid X) = (1/N)\sum_{t=1}^{N}\left(y_t - \mathbf{x}_t^T\hat{\boldsymbol{\beta}}_t(Y,X,\lambda)\right)^2$, является условное математическое ожидание эмпирического риска класса моделей

$$\overline{Risk}(\lambda|X)=\mathrm{E}_{Y|X}\left[\mathrm{E}_{Y'|X}\left((1/N)\sum_{t=1}^{N}\left(y'_t-\mathbf{x}_t^T\hat{\boldsymbol{\beta}}_t(Y,X,\lambda)\right)^2\right)\right]. \quad (3)$$

С этой точки зрения значение параметра нестационарности $\lambda^* = \arg\min_\lambda \overline{Risk}(\lambda \mid X)$ было бы наилучшим, но т.к. нам не известно ни условное распределение $p(Y \mid X)$, ни точное математическое выражение для оператора $\hat{\boldsymbol{\beta}}_t(Y,X,\lambda)$, подсчитать средний риск оказывается очень трудной задачей.

Метод скользящего контроля для выбора класса моделей

Один из способов преодоления этого ограничения состоит в нахождении подходящей оценки среднего риска с помощью замены его на эмпирический риск, полученный с помощью процедуры скользящего контроля. Основным отличием классической процедуры скользящего контроля, описанной в литературе для случая стационарной регрессии и распознавания образов [2], и предлагаемой здесь процедурой является то, что коэффициенты регрессии различны в различных точках обучающей выборки и связаны между собой скрытой моделью переходов (1).

Мы удаляем одну пару $(y_t, \mathbf{x}_t)$ из полной временной последовательности $(Y_N,X_N)=\left((y_t,\mathbf{x}_t), t=1,...,N\right)$ всеми возможными способами, подсчитываем N различных оценок последовательности коэффициентов регрессии

$$\hat{B}(Y_N^{(t)},X_N^{(t)},\lambda)=\left(\hat{\boldsymbol{\beta}}_t(Y_N^{(t)},X_N^{(t)},\lambda), t=1,...,N\right)$$

и формируем эмпирический риск – оценку среднего риска $\overline{Emp}(\lambda, X_N, Y_N) = =(1/N)\sum_{t=1}^{N}\left(y_t - \mathbf{x}_t^T\hat{\boldsymbol{\beta}}_t(Y_N^{(t)}, X_N^{(t)}, \lambda)\right)^2$. Так как мы здесь рассматриваем только условное математическое ожидание этой оценки для фиксированной последовательности коэффициентов регрессии, то в результате получим следующее уравнение $\mathrm{E}_{Y_N|X_N}\left[\overline{Emp}(\lambda, X_N, Y_N)\right] = (1/N)\sum_{t=1}^{N}\mathrm{E}_{Y_N^{(t)}|X_N}\left\{\mathrm{E}_{y_t'|X_N}\left[\left(y_t' - \mathbf{x}_t^T\hat{\boldsymbol{\beta}}_t(Y_N^{(t)}, X_N^{(t)}, \lambda)\right)^2\right]\right\}$.

Внутреннее математическое ожидание под знаком суммирования есть условный средний риск конкретного значения параметра λ при оценивании последовательности нестационарности коэффициентов регрессии с выколотой точкой: $Risk\left(\hat{B}(Y_N^{(t)}, X_N^{(t)}, \lambda) | X_N\right) = \mathrm{E}_{y_t|X_N}\left[\left(y_t - \mathbf{x}_t^T\hat{\boldsymbol{\beta}}_t(Y_N^{(t)}, X_N^{(t)}, \lambda)\right)^2\right]$, что приводит нас к уравнению $\mathrm{E}_{Y_N|X_N}\left[\overline{Emp}(\lambda, X_N, Y_N)\right] = (1/N)\sum_{t=1}^{N}\mathrm{E}_{Y_N^{(t)}|X_N}\left[Risk\left(\hat{B}(Y_N^{(t)}, X_N^{(t)}, \lambda) | X_N\right)\right]$.

Назовем математическое ожидание под знаком суммирования частичным условным средним риском оценивания коэффициентов регрессии для некоторого параметра λ: $\overline{Risk}^{(t)}(\lambda | X_N) = \mathrm{E}_{Y_N^{(t)}|X_N}\left[Risk\left(\hat{B}(Y_N^{(t)}, X_N^{(t)}, \lambda) | X_N\right)\right]$. Сравнение правой части этого уравнения с (3) показывает, что мы получили арифметическое среднее частичного условного среднего риска последовательности с выколотой точкой вместо полного условного среднего риска. По аналогии с классической задачей оценивания регрессии мы можем сказать, что эмпирический риск $\overline{Emp}(\lambda, X_N, Y_N)$, полученный с помощью процедуры скользящего контроля, является «почти» несмещенной оценкой условного среднего риска $\overline{Risk}(\lambda | X_N)$ для некоторого значения параметра нестабильности λ. Этот факт позволяет нам использовать принцип скользящего контроля вместо действительного среднего риска для выбора наиболее подходящего значения λ. Для того, чтобы можно было сравнивать значения эмпирического риска, полученные для различных инвестиционных портфелей, мы преобразуем его к стандартному финансовому показателю, который носит в англоязычной литературе наименование *Predicted R-squared*:

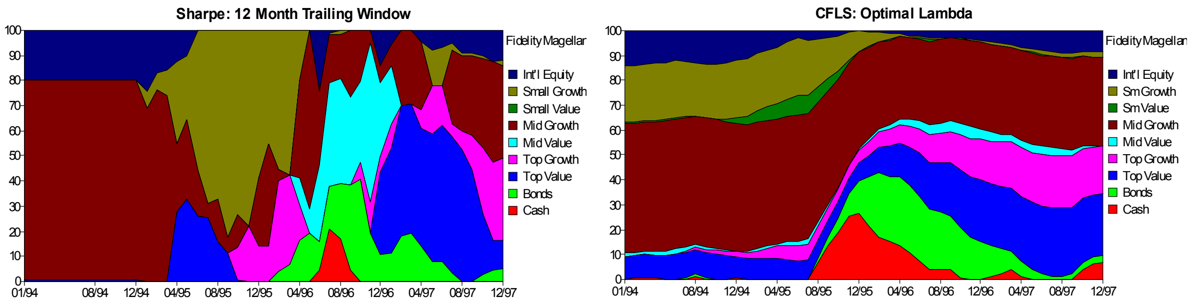
$$\lambda^* = \arg\max_{\lambda} PR^2(\lambda), \quad PR^2(\lambda) = 1 - \overline{Emp}(\lambda, X, Y) \Big/ \sum_{t=1}^{N} y_t^2 .$$

Способ подсчета эмпирического риска с помощью метода скользящего контроля был предложен авторами в [4] и представляет собой один проход процедуры динамического программирования с линейной вычислительной сложностью по отношению к длине временной последовательности.

Пример анализа стратегии инвестиционного портфеля

В качестве примера мы рассмотрим инвестиционную политику фонда Фиделити Магеллан, одного из крупнейших и наиболее успешных открытых инвестиционных фондов в США. В октябре 1995 года управляющий фондом, опасаясь, что в мировой экономике произойдет серьезный кризис, резко сократил объем средств фонда, вложенных в акции, и перевел освободившиеся средства в более надежные, но намного менее доходные государственные облигации. Вопреки его ожиданиям кризиса не произошло и фонд потерпел убытки вследствие слишком осторожной политики управляющего. В дальнейшем фонд большую часть государственных облигаций погасил и до конца 1996 г. вернул почти 90 процентов всех принадлежащих ему акций. Мы проанализировали поведение фонда Магеллан на временном отрезке с начала

1994 года по конец 1997 сначала с помощью статической многофакторной модели Шарпа, а затем с помощью динамической многофакторной модели, в обоих случаях используя ограничение о неотрицательности долевых составляющих портфеля. На рис. 1а видно, что распределение портфеля согласно модели Шарпа не дает удовлетворительного качества анализа. На втором этапе мы обработали те же данные с использованием динамической многофакторной модели с подбором параметра сглаживания посредством процедуры скользящего контроля. Как можно видеть из рис. 2а, результаты восстановления динамики портфеля с помощью динамической многофакторной модели адекватно отражают все этапы изменения инвестиционной политики фонда.



а) б)

Рисунок 1 – Восстановление скрытой динамики портфеля фонда Fidelity Magellan: а) с помощью многофакторной модели Шарпа со скользящим окном в 12 месяцев; б) с помощью динамической многофакторной модели

Литература

1. Vapnik V. Principles of risk minimization for learning theory // Touretzky, editors, Advances in Neural Information Processing Systems 4 / D.S. Lippman, J.E. Moody, and D.S. Morgan Kaufmann, 1992. – P. 831-838.
2. Vapnik V. Statistical Learning Theory. – John-Wiley & Sons, Inc., 1998.
3. Kalaba R., Tesfatsion L. Time-varying linear regression via flexible least squares // International Journal on Computers and Mathematics with Applications. – 1989. – Vol. 17. – P. 1215-1245.
4. Kostin A., Krasotkina O., Markov M., Mottl V., Muchnik I. Dynamic programming algorithms for analysis of nonstationary signals // Computational Mathematics and Mathematical Physics. – 2004. – Vol. 44, № 1. – P. 62-77.

М.Р. Марков, В.В. Моттль, І.Б. Мучник, О.В. Красоткіна

Вкладені класи моделей нестаціонарності сигналу у динамічному аналізі складу інвестиційного портфеля

Принцип мінімізації структурного ризику вкладених класів моделей даних часто застосовується у задачах розпізнавання образів і відновлення регресійної залежності як основного інструменту вибору відповідного класу моделі. Для підрахунку структурного ризику ми використовуємо метод ковзного контролю за допомогою його інкорпорування до стандартної схеми процедури динамічного програмування. Запропоновану техніку ми використовуємо для спостереження за складом інвестиційного портфеля і аналізу стратегії інвестиційної компанії.

Статья поступила в редакцию 26.04.2006.