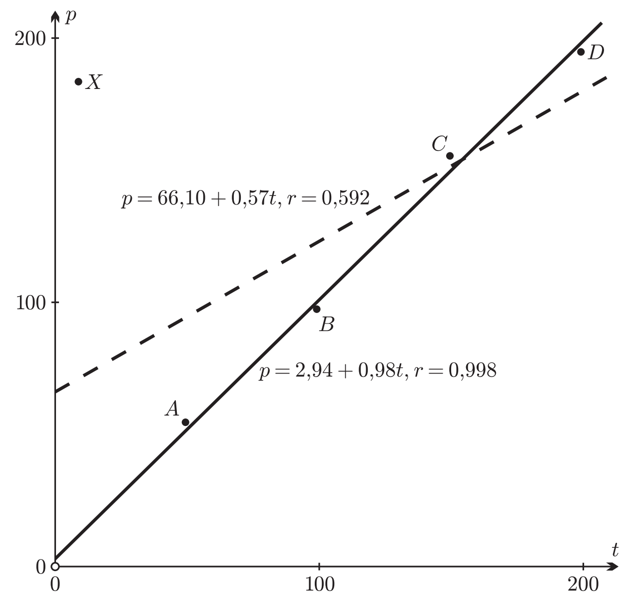
Jest to zapis odczytu wygłoszonego na XL Szkole Matematyki Poglądowej, *Matematyczne Obrazki*, styczeń 2008.

Bąble, moda i inni

Anna ŻYLICZ, Warszawa

Postawiony przed koniecznością przeanalizowania jakiegoś szeregu czasowego (tj. sekwencji informacji o jednym zjawisku, uporządkowanych w czasie) ekonomista próbował będzie zwykle dopasować do obserwowanej ścieżki rozwoju badanej zmiennej właściwy model teoretyczny. Dobrze dobrany i skalibrowany model teoretyczny pozwala bowiem nie tylko na rzetelne opisanie rzeczywistości, ale również na dokonywanie prognoz przyszłych wartości rozważanego zjawiska (na przykład: prognoz stopy bezrobocia, zapotrzebowania rynkowego na jakiś produkt czy giełdowej ceny akcji). Dobranie (lub wręcz stworzenie) właściwej teorii bywa czasem niezwykle skomplikowane, ale w przeważającej części przypadków nie na tym kończy się najtrudniejsza praca. W rzeczywistości w gospodarce rzadko kiedy mamy bowiem do czynienia z procesami „idealnymi", w których na wartość danej zmiennej nie mają wpływu żadne losowe zjawiska ani (tymczasowe) odstępstwa od reguł. W praktyce zawsze ekonomista-praktyk narażony jest na to, że badany przez niego szereg zawierał będzie obserwacje, które mniej lub bardziej odstają od zakładanego (teoretycznie prawdziwego) schematu. Małe, losowe zaburzenia nie stanowią przy obecnym rozwoju technik badawczych dużego problemu. Jednak gdy zaburzenie jest większe, albo systematyczne, sprawia zwykle więcej kłopotów. Dwóm typom takich odstępstw: *obserwacjom nietypowym* i *bąblom spekulacyjnym* warto przyjrzeć się bliżej; poświęcono im wiele miejsca w literaturze przedmiotu.

Obserwacje nietypowe to (według jednej z wielu przyjmowanych definicji) takie, które nie odpowiadają „ogólnej prawidłowości" i/lub zdecydowanie odstają wielkością od reszty obserwacji. Oczywiście, żeby stwierdzić, czy jakaś obserwacja jest nietypowa, należy najpierw określić, co jest tą „ogólną prawidłowością", i jakie od niej odstępstwo uznamy za niepokojące. Wyłapanie z szeregu czasowego takich „dewiantów" jest tyleż trudne, co ważne, gdyż poprawne zidentyfikowanie takich obserwacji pozwala uniknąć groźnych pułapek.



Rys. 1. Wpływ obserwacji nietypowych na regresję liniową

Konsekwencje niefrasobliwego uwzględniania w modelu obserwacji nietypowych dobrze ilustruje następujący przykład. Wyobraźmy sobie, że obserwujemy dwie zmienne: t (np. czas, liczony w dniach) i p (np. cena metra kwadratowego w coraz bardziej popularnej i modnej lokalizacji). Przypuśćmy, że w „idealnej" gospodarce relacja między tymi zmiennymi kształtuje się w następujący sposób:

$$p = c + \alpha t, \tag{1}$$

gdzie c jest stałą (w praktyce nieznaną *a priori* badaczowi, np. rzeczywistym kosztem wybudowania metra kwadratowego mieszkania), tak jak i parametr α. Taka postać modelu (1) oznacza, że cena rośnie liniowo; jeśli dodatkowo $\alpha = 1$, to w takim samym tempie, jak upływa czas (cena wykazuje liniowy trend rosnący). Wyobraźmy sobie teraz, że w ciągu 200 dni od początku sprzedaży na rynku sprawdziliśmy cenę cztery razy u wybranego dewelopera. Zaobserwowane wartości cen oznaczone są jako punkty A, B, C, D na wykresie 1 obok.

Jeśli wiemy (np. tak jak w tym przypadku, z zakładanej teorii), że relacja między zmiennymi t i p jest liniowa, możemy dopasować do zebranych czterech obserwacji model typu (1), tj. na podstawie danych wyznaczyć „najbardziej pasujące" wartości parametrów c i α. Najprościej zadanie to można wykonać przeprowadzając regresję liniową (tj. dopasowując do danych prostą, która najlepiej opisuje między nimi zależność tego typu). W naszym przypadku, dopasowanie takiej prostej metodą najmniejszych kwadratów (tj. tak, żeby zminimalizować sumę kwadratów odchyleń wartości obserwacji od prostej) pozwala na stwierdzenie, że relacja między zmiennymi t i p kształtuje się w następujący sposób: $p = 2,94 + 0,98t$ (zaznaczona na rysunku ciągłą linią), a współczynnik korelacji zmiennych p i t wynosi $r = 0,998$. Oznacza to, że model liniowy jest prawie idealnie dopasowany, gdyż współczynnik korelacji co do modułu przyjmuje wartości między 0 a 1, gdzie 0 oznacza brak zależności liniowej, a 1 – doskonałą zależność liniową. Oczywiście w praktyce nie można się spodziewać, że model będzie idealnie wyjaśniał obserwowane zjawisko, gdyż w otaczającej nas rzeczywistości zawsze mamy do czynienia z małymi losowymi zaburzeniami (sprzedawca zapomniał zmienić etykiety z cenami z poprzedniego dnia, potrzebował dodatkowego zarobku, więc marginalnie zwiększył cenę, miał dobry humor i ogłosił promocję itd.).

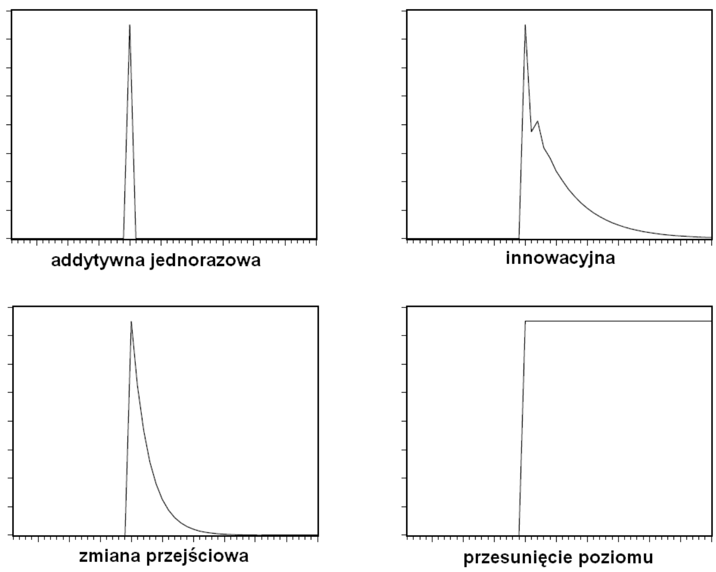
Przypuśćmy teraz, że mamy do dyspozycji piątą obserwację (oznaczoną na wykresie jako punkt X) – np. pochodzącą od innego dewelopera, albo podaną przez nowoprzyjętego do pracy, niezorientowanego sprzedawcę, albo też niechlujnie (omyłkowo) zapisaną, albo wynikającą z jakiegoś eksperymentu zarządu. Jeśli nie uwzględnimy tego, że dodatkowa obserwacja ta jest nietypowa, i przeprowadzimy jak poprzednio analizę regresji liniowej metodą najmniejszych kwadratów, ale dla wszystkich pięciu obserwacji, wyniki będą następujące: najlepiej do danych dopasowana jest prosta $p = 66{,}10 + 0{,}57t$ (na rysunku linia przerywana), przy współczynniku korelacji $r = 0{,}592$. Oznacza to, że, po pierwsze, model liniowy jest słabo dopasowany do danych, więc można by było wysnuć błędny wniosek, że teoria nie sprawdza się w praktyce; po wtóre, gdybyśmy próbowali na tej podstawie wyciągać wnioski na temat kosztu budowy metra kwadratowego mieszkania (tj. stałej c w modelu), to pomylilibyśmy się o rząd wielkości; po trzecie, ponieważ wszystkie parametry opisujące zależność wyestymowane są błędnie, niepoprawnie prognozowalibyśmy cenę, jakiej możemy się spodziewać w przyszłości. A więc występowanie w danych jakiejkolwiek obserwacji nietypowej – wynikającej z błędu w zapisie, pochodzącej z innej struktury, czy związanej z jakimś innym, losowym zaburzeniem – może prowadzić do błędnych wniosków i prognoz, co oczywiście może nieść ze sobą poważne konsekwencje (także finansowe).

Wiemy już więc, że z istnienia obserwacji nietypowych musimy sobie zdawać sprawę, by móc poprawnie modelować procesy gospodarcze. Jednak jak można takie obserwacje wykrywać? W analizie szeregów czasowych przeprowadzanej zwykle podczas modelowania tego typu zjawisk (np. takiej, jaką w Polsce przeprowadza się w Narodowym Banku Polskim analizując związki wzrostu gospodarczego, stóp procentowych, inflacji i innych zmiennych makroekonomicznych) wyróżnia się zwykle cztery podstawowe rodzaje obserwacji nietypowych:

1. Addytywne obserwacje jednorazowe (AO, ang. *additive outlier*)
2. Obserwacje innowacyjne (IO, ang. *innovative outlier*)
3. Zmiany przejściowe (TC, ang. *temporary change*)
4. Przesunięcie poziomu (LS, ang. *level shift*)

Rysunek 2 przedstawia wpływ obserwacji nietypowych poszczególnych rodzajów w modelu: nietypowa jednorazowa obserwacja addytywna powoduje, że zaburzenie (duży błąd, odstępstwo od prawidłowości) widoczne jest w modelu w jednym okresie; obserwacja innowacyjna i obserwacja przejściowa mają wygasający w czasie wpływ, przy czym wpływ obserwacji innowacyjnej związany jest ściśle ze strukturą zależności szeregu czasowego; przesunięcie poziomu powoduje zaś trwałe zwiększenie błędu w modelu.

Wykrywanie obserwacji z ustalonego katalogu rodzajów nietypowości przeprowadza się zwykle w następujący sposób: do *stricte* teoretycznej zależności dodaje się zmienne, które mają uchwycić występowanie obserwacji nietypowych, po kolei dla każdej obserwacji (dla każdej wartości zmiennej t) i dla każdego typu obserwacji. Estymuje się z osobna każdy z powstałych w ten sposób modeli, by stwierdzić, czy w którymś przypadku dodanie dodatkowej zmiennej nie poprawia



Rys. 2. Błędy w modelach wywołane obserwacjami nietypowymi różnych rodzaj??w

modelu (w sposób znaczący). Jeśli możemy zidentyfikować jakieś obserwacje nietypowe (identyfikując zmienne, które po dodaniu poprawiają model), to wybieramy najbardziej istotną z nich, korygujemy i powtarzamy całą procedurę, aż nie będzie już istotnych parametrów przy zmiennych oznaczających jakiekolwiek obserwacje nietypowe. „Korektę” nietypowej obserwacji można przeprowadzić na bardzo wiele sposobów, np. zmieniając jej wagę w estymacji, „przycinając” do wartości teoretycznej czy wyrzucając w ogóle z modelu.

W omawianym przykładzie procedura przeprowadzona byłaby tak: do podstawowego modelu (1) dodawane byłyby zmienne $z_\tau^o(t)$ dla τ przyjmującego (wszystkie pięć) wartości zmiennej t, a o oznaczającego typy obserwacji nietypowych ($o = AO, IO, TC, LS$), przemnożone przez (nieznaną *a priori*) stałą ω, oznaczającą siłę wpływu obserwacji. Estymowane byłoby więc po kolei dwadzieścia modeli postaci

$$p = c + \alpha t + \omega z_\tau^o.$$

W szczególności – jako jeden z dwudziestki pomocniczych – estymowany byłby model z nietypową jednorazową zmianą addytywną w punkcie X:

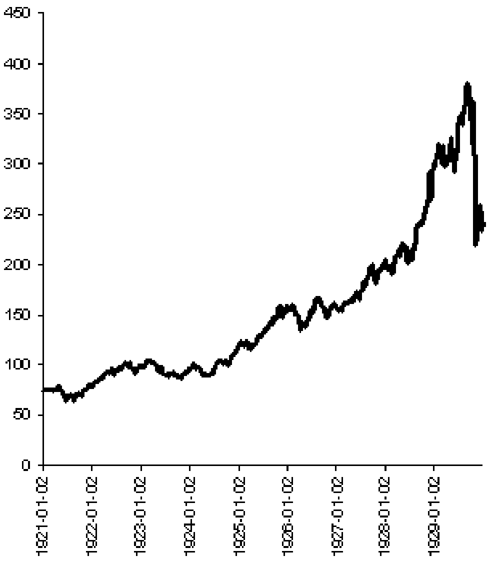
$p = c + \alpha t + \omega z_8^{AI}(t)$, w którym zmienna $z_8^{AI}(t)$ ma postać $z_8^{AI}(t) = \begin{cases} 1 & t = 8 \\ 0 & \text{w p.p.} \end{cases}.$

W naszym przypadku okazuje się, że taki model daje najlepsze (takie samo jak wyjściowy model dla czterech obserwacji) dopasowanie ze wszystkich dwudziestu jeden rozważanych modeli, z istotnym (tj. zdecydowanie różniącym się od zera) parametrem stojącym przy zmiennej opisującej nietypową obserwację typu addytywnego w czasie $\tau = 8$:

$$p = 2,94 + 0,98t + 175,12 z_8^{AI}(t).$$

W związku z tym, badający zjawisko ekonomista powiedziałby, że dodatkowa (piąta) obserwacja jest jednorazową addytywną obserwacją nietypową, i postępował dalej w zgodzie z przyjętymi zasadami estymacji (najprawdopodobniej, w tym przypadku zdecydowałby się na wyrzucenie obserwacji z rozważanego zbioru danych, jako pochodzącej z zupełnie „innej bajki”). Powtórzenie opisanej procedury dla zbioru czterech (podstawowych) obserwacji nie powoduje już zidentyfikowania żadnej kolejnej jako odstępstwa.

Przypuśćmy jednak, że zamiast obserwować wygasający wpływ jakichś zewnętrznych bodźców na badaną zmienną, ekonomista obserwuje zjawisko wręcz przeciwne, a mianowicie „wybuchania” szeregu czasowego. Oznaczać to może, że w badanym procesie ukryty jest *bąbel spekulacyjny*.



Rys. 3. Indeks giełdy nowojorskiej Dow Jones w latach 1921–1929

Tak jak i w przypadku obserwacji nietypowych, ogólna definicja bąbli spekulacyjnych nie istnieje. Intuicyjnie, pojęcie to odnosi się do takich procesów, jak holenderska histeria tulipanowa (obserwowana w latach 1634-1637, kiedy ceny pojedynczych cebulek urosły do niebotycznych wielkości – rzędu wartości domu w Amsterdamie – by spaść praktycznie do zera) czy krach giełdowy obserwowany w Stanach Zjednoczonych w 1929 roku (po kilkuletnim okresie nabierającego tempa wzrostu, w ciągu dwóch tygodni wartość indeksu spadła o około 35%), czyli procesów charakteryzujących się wystąpieniem gwałtownego załamania cen poprzedzonego dłuższym okresem intensywnego wzrostu (rys. 3). Najczęściej przyjmowana przez ekonomistów formalna definicja bąbla spekulacyjnego to odstępstwo od przyjętego *modelu fundamentalnego*, czyli modelu podstawowego, uwzględniającego wszystkie istotne z punktu widzenia przyjętej teorii zmienne. Oczywiście nie chcielibyśmy nazywać bąblem każdego odstępstwa od takiego modelu – oprócz małych, czysto losowych zaburzeń występują też przecież, jak już wiemy, „zwykłe” obserwacje nietypowe; bąbel spełniać więc musi kilka dodatkowych założeń.

Przypuśćmy, że ceny na badanym rynku kształtują się zgodnie z zadaną (teorią) funkcją f według wzoru

$$p_t = f(x, t),$$

gdzie t – czas, p_t – cena w chwili t, a x – zestaw zmiennych obserwowanych lub przewidzianych, które według przyjętej teorii mają wpływ na p_t (np. opóźnione lub przyszłe oczekiwane wartości p_t). Można wówczas zadać pytanie, czy dla ustalonych wartości zmiennych x, innych niż ceny, proces $\{p_t\}$ jest wyznaczony jednoznacznie przez f? Jeśli tak, to nie możemy mieć do czynienia z bąblem spekulacyjnym, a wszelkie odstępstwa od modelu obserwowane w rzeczywistości mają inny charakter (np. zaburzenia losowe, czy obserwacje nietypowe). Jednak jeśli proces $\{p_t\}$ nie jest wyznaczony jednoznacznie, to jest możliwość występowania bąbla jako odstępstwa od modelu fundamentalnego („właściwej” ścieżki procesu).

W standardowym modelu rynków efektywnych (stosowanym na przykład do wyznaczania cen akcji giełdowych) zakłada się, że ceny na rynku kształtowane są zgodnie ze wzorem

$$p_t = \gamma E_t(p_{t+1} + d_{t+1}), \tag{2}$$

gdzie d_{t+1} – dywidenda wypłacana w chwili $t+1$, $0 < \gamma < 1$ – stały czynnik dyskontujący, E_t – warunkowa wartość oczekiwana (pod warunkiem zbioru informacji dostępnego w chwili t; tu: zawierającego informację o strukturze modelu i obecnych i przeszłych wartościach rozważanych szeregów czasowych); a więc wartość aktywu w chwili t jest zdyskontowaną spodziewaną wartością akcji w chwili $t+1$, powiększoną o spodziewaną dywidendę.

Przepisując równanie (2) dla $t+1$ i podstawiając do równania dla t, otrzymujemy

$$p_t = \gamma E_t(\gamma E_{t+1}(p_{t+2} + d_{t+2}) + d_{t+1}).$$

Korzystając teraz z *prawa iterowanych oczekiwań* (najlepszą prognozą przyszłej prognozy wartości jakiejś zmiennej jest obecna prognoza tej zmiennej), a ściśle:

$$E_t(E_{t+j}(x_{t+j+k})) = E_t(x_{t+j+k}),$$

dla dowolnej zmiennej x i nieujemnych j i k, dostajemy:

$$p_t = \gamma^2 E_t(p_{t+2}) + \gamma^2 E_t(d_{t+2}) + \gamma E_t(d_{t+1}).$$

Wynika to z probabilistycznego twierdzenia, że wartość oczekiwana warunkowej wartości oczekiwanej pewnej zmiennej jest wartością oczekiwaną tej zmiennej, o ile tylko zbiór warunków jest „porządny”.

Dalej, postępując analogicznie, dostajemy

$$p_t = \sum_{i=1}^{\infty} \gamma^i E_t(d_{t+i}) + \lim_{n\to\infty} \gamma^n E_t(p_{t+n}).$$

Przy dodatkowym *warunku transwersalności*

$$\lim_{n\to\infty} \gamma^n E_t(p_{t+n}) = 0 \tag{3}$$

Warunek transwersalności jest często stosowaną regułą przy wyznaczaniu rozwiązań równań Eulera (a właśnie z takim mamy do czynienia w przypadku modelu rynków efektywnych), pozwalającą zawęzić klasę rozważanych rozwiązań.

rozwiązanie jest jedno i zadane jest wzorem

$$p_t^f = \sum_{i=1}^{\infty} \gamma^i E_t(d_{t+i}).$$

W analizie ekonomicznej, takiej jak prezentowana tutaj, warunek transwersalności jest konsekwencją przyjęcia założenia, że rynek daje się opisywać za pomocą analizy problemu maksymalizacyjnego dla kompetentnego, reprezentatywnego i nieskończenie długo żyjącego inwestora. Nieskończenie długo żyjący inwestor może bowiem przyjąć następującą (nieobciążoną żadnym ryzykiem związanym z fluktuacjami cen giełdowych) strategię: kupić akcję i nigdy jej nie sprzedawać, tzn. trzymać ją w nieskończoność i czerpać zyski wyłącznie z tytułu dywidend (wówczas spodziewany zysk to p_t^f, zdyskontowany strumień dywidend). Inwestorowi żyjącemu wiecznie powyższa strategia będzie się opłacała, jeśli tylko $p_t^f > p_t$. Gdyby cena rynkowa p_t była niższa od wartości p_t^f, to nadwyżka popytu na akcję spowodowałaby wzrost p_t; analogicznie, gdyby $p_t > p_t^f$, to nadwyżka podaży spowodowałaby obniżkę ceny akcji. W związku z tym, jeśli inwestorzy żyją nieskończenie długo, a rynek jest w równowadze, to cena p_t nie może odbiegać od wartości fundamentalnej.

Jednak gdy od rozwiązania równania (2) nie wymaga się spełnienia warunku (3) – a niespełnienie tego warunku nie musi wcale oznaczać nieracjonalności inwestorów giełdowych – to dowolny proces spełniający warunki

$$\begin{aligned} p_t &= p_t^f + b_t, \\ E_t(b_{t+1}) &= \gamma^{-1} b^t, \end{aligned} \tag{4}$$

spełnia wyjściowy model rynku. Składnik b_t ceny, czyli odstępstwo od modelu fundamentalnego, może być bąblem spekulacyjnym!

Najprostszym przykładem bąbla spekulacyjnego jest bąbel deterministyczny, wyrażony wzorem $b_t = b_0\gamma^{-t}$, dla $b_0 > 0$. Nietrudno stwierdzić, że spełnia on warunek (4), jest również stosunkowo łatwy do wykrycia w praktyce. Jedyny szkopuł polega na tym, że proces ten rośnie wykładniczo w nieskończoność. . . , a z takimi przypadkami w praktyce ekonomiści nie mają do czynienia.

Przykładem bardziej „życiowego” bąbla jest bąbel pękający okresowo:

$$b_t = \begin{cases} \dfrac{b_{t-1} - \bar{b}}{\pi_t \gamma} & \text{z prawdopodobieństwem } \pi_t, \\ \dfrac{\bar{b}}{(1-\pi_t)\gamma} & \text{z prawdopodobieństwem } 1 - \pi_t, \end{cases}$$

dla pewnej wartości $\bar{b} > 0$ (racjonalny bąbel okresowy, po pęknięciu, nie może przyjąć wartości 0, gdyż wtedy z (4) i faktu, że bąbel spekulacyjny nie może być ujemny – oznaczałoby to bowiem, że inwestorzy mogą oczekiwać, że na giełdzie w którymś momencie ceny spadną poniżej zera – już zawsze musiałby być równy 0). Bąbel ten nie jest już wprawdzie deterministyczny, ani nie rośnie w nieskończoność, ale kolejne wzrosty bąbla związane są zawsze z taką samą ścieżką.

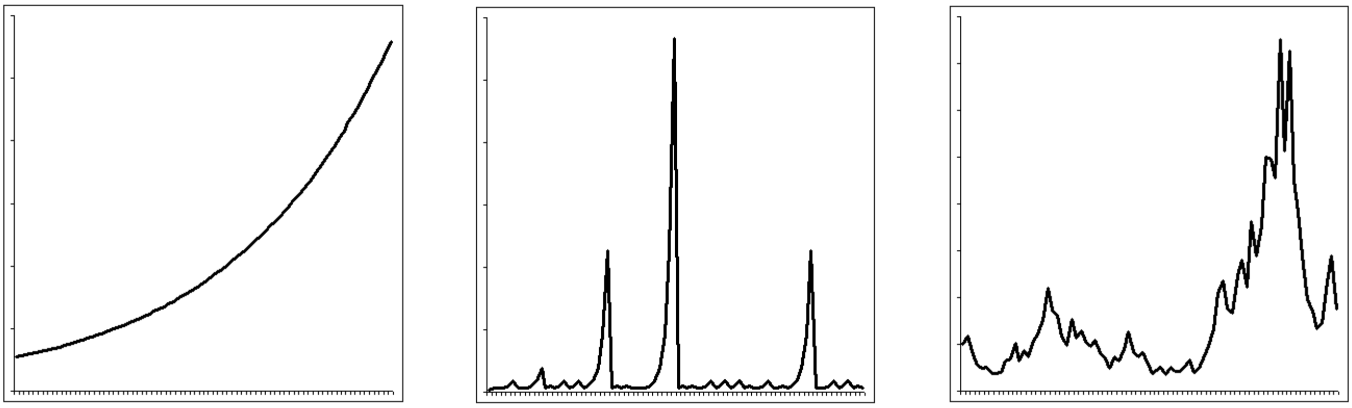
Już na pierwszy rzut oka widać, że na bąblach spekulacyjnych inwestorzy mogą nieźle zarabiać (ale oczywiście tylko ci, którzy wiedzą, kiedy jest najwłaściwszy moment do wycofania się z rynku!), więc literatura przedmiotu rozwija się bardzo dynamicznie. Pogoń za coraz bardziej elastycznym modelem bąbla prowadzi do rozważania coraz bardziej skomplikowanych postaci funkcyjnych, które niosą ze sobą cały szereg problemów metodologicznych, są coraz trudniejsze do zaimplementowania i wykrywania w praktyce. Przykładem może być bąbel

$$b_{t+1} = \begin{cases} \gamma^{-1} b_t u_{t+1} & \text{jeśli } b_t \leqslant \alpha, \\ \big(\delta + \pi^{-1}\gamma^{-1}\theta_{t+1}(b_t - \gamma\delta)\big)u_t & \text{jeśli } b_t > \alpha, \end{cases}$$

gdzie δ i α są nieujemnymi parametrami spełniającymi $0 < \delta < \gamma^{-1}\alpha$ (warunek nałożony jest po to, by bąbel miał szansę wzrastać), u_t są niezależnymi nieujemnymi zmiennymi losowymi o identycznych rozkładach spełniającymi

$E_t u_{t+1} = 1$, zaś θ_t jest procesem Bernoulliego (kolejne wartości to wyniki prób w schemacie Bernoulliego), niezależnym od u, o prawdopodobieństwie sukcesu w jednej próbie równym π.

Proces taki opisuje bąbel, który rośnie ze średnim tempem γ^{-1} do czasu, aż przekroczy wartość α; od tego momentu bąbel rośnie szybciej, bo w tempie $\gamma^{-1}\pi^{-1}$, ale w każdym okresie z prawdopodobieństwem $1 - \pi$ załamuje się, by spaść do poziomu δ, po czym sytuacja powtarza się. W przypadku takiego bąbla, w przeciwieństwie do zwykłego bąbla pękającego okresowo, za każdym razem przebieg może być zupełnie inny, a to ze względu na obecność zmiennej losowej u_t. Za pomocą parametrów α, δ i θ można sterować częstością wybuchów bąbli, średnim czasem ich trwania i skalą zjawiska. Jednak stwierdzenie za pomocą ekonometrycznej analizy, czy tego typu mechanizm występuje w praktyce, graniczy z niemożliwością...



Bąbel deterministyczny **Bąbel pękający okresowo** **Bąbel stochastyczny**

Rys. 4 Przykłady realizacji różnych procesów bąbli

Nic więc dziwnego, że doniesienia poszczególnych autorów na temat występowania różnego rodzaju bąbli na różnych rynkach (w szczególności, na giełdzie Nowojorskiej, czy Warszawskiej Giełdzie Papierów Wartościowych) są ze sobą sprzeczne. Ale gdyby tak udało się właściwie przewidzieć przyszłość...