

Mariusz Hamulczuk

Katedra Ekonomiki Rolnictwa i Międzynarodowych Stosunków Gospodarczych
SGGW

Częstość danych statystycznych a wyniki prognozowania

Wstęp

Rola prognoz w sferze gospodarki jest ogromna, choć nie zawsze doceniana. Większość osób zdaje sobie sprawę z faktu, że tylko trafne odgadnięcie tego, co napotkamy w przyszłości pozwala na poprawne zaprogramowanie i zaplanowanie przyszłych działań. Oznacza to, że głównym zadaniem prognozy jest stworzenie dodatkowych przesłanek w procesie podejmowania decyzji. Sporządzane prognozy charakteryzują się często określonym błędem. Łącznych przyczyn ich powstawania jest wiele. Jedną z nich jest konstruowanie prognoz w tak zwanych punktach zwrotnych, w których następuje zmiana dotychczasowego trendu. Prognozowanie za pomocą większości metod w takich okresach jest obciążone bardzo dużym ryzykiem, które maleje wraz ukształtowaniem się nowego trendu.

W niniejszym opracowaniu podjęto próbę sprawdzenia, w jakim stopniu dokładność prognozowania krótkookresowego za pomocą wybranych modeli adaptacyjnych zależy od częstości danych statystycznych stanowiących szereg czasowy. Prognozy zostały skonstruowane na podstawie szeregu charakteryzującego się częstymi zmianami kierunku trendu.

W artykule dokonano oceny błędów prognoz przy wykorzystaniu szeregów czasowych o różnej częstości danych. Do oceny wpływu częstości danych statystycznych na wyniki prognozowania wykorzystano materiał statystyczny stanowiący średnie miesięczne notowania cukru na giełdzie londyńskiej i obejmujący informacje od stycznia 1994 r. do grudnia 2001 r. Dane empiryczne (rys. 1) zostały przedstawione w postaci trzech szeregów czasowych: miesięcznych (96 obserwacji), kwartalnych (32 obserwacje) i półrocznych (16 obserwacji). Przeciętna cena cukru w badanym okresie wyniosła 293,64 USD/t, a współczynnik zmienności kształtował się w przedziale 24,4–24,9% w zależności od częstości danych. Analiza graficzna szeregów czasowych pozwala na stwierdzenie, że zjawisko charakteryzuje się częstymi wzrostami o różnym tempie, jak i spadkami o różnej szybkości.

Metodyka prognozowania i oceny modeli

W procesie prognozowania zjawisk i procesów gospodarczych mogą być wykorzystywane różne metody. Wyboru techniki prognozowania dokonuje się na podstawie przesłanek prognostycznych dotyczących badanego zjawiska oraz właściwości metod użytych do predykcji [Cieślak 2001]. Do badania wybrano metody adaptacyjne, mające praktyczne zastosowanie w przypadku szeregów czasowych charakteryzujących się nieregularnym przebiegiem. „Modele adaptacyjne konstruowane są w taki sposób, by omijać (rozluźniać) założenia klasycznej teorii predykcji, uwzględniając między innymi niestałość struktury ekonomicznej i możliwość zmian parametrów w czasie. Duża elastyczność modeli adaptacyjnych i ich zdolność dostosowawcza, w przypadku nieregularnych zmian kierunku lub szybkości trendu (...) czynią z nich wygodne narzędzie krótkookresowych (operatywnych) prognoz statystycznych” [Zeliaś 1997, s. 220–221]. „Postępowanie prognostyczne oparte na podstawie tych modeli zakłada, że rozwój zjawiska gospodarczego w czasie może być segmentowy, tzn. »gładki« tylko w pewnych przedziałach czasu” [Zeliaś 1997, s. 38]. Prognoza budowana za pomocą tych metod wyznaczana jest w głównej mierze na podstawie informacji najbliższych okresowi prognozowania.

W przypadku występowania w szeregu czasowym trendu monotonicznego lub niemonotonicznego zastosowanie znajduje **model wygładzania wykładniczego Browna drugiego rzędu** (liniowy). Idea prognozowania za pomocą tego modelu jest następująca: dany szereg wygładza się według wzoru (1), a następnie otrzymany szereg wygładza się ponownie zgodnie ze wzorem (2) [Farnum, Stanton 1989, Stańko 1999]:

$$\hat{Y}_t = \alpha Y_{t-1} + (1 - \alpha) \hat{Y}_{t-1} \quad (1)$$

$$\hat{Y}_t^* = \alpha \hat{Y}_t + (1 - \alpha) \hat{Y}_{t-1}^* \quad (2)$$

gdzie: Y_t – wygładzona wartość zmiennej prognozowanej w momencie t ,

\hat{Y}_t^* – wygładzona wartość zmiennej prognozowanej w momencie t metodą wygładzania wykładniczego rzędu drugiego,

α – stała wygładzania ($0 \leq \alpha < 1$).

W powyższych wzorach szacowane wartości wygładzone dla okresu t są średnią ważoną z ostatniej informacji oraz wartości wygładzonej dla okresu poprzedniego. Głównym zadaniem prognosty jest właściwe określenie parametru wygładzania α . Wartości parametru bliskie jedności oznaczają nadawanie większej wagi obserwacjom najnowszym, natomiast wartości bliskie zero

oznaczają, że efekt obserwacji poprzedzających będzie trwał dłużej [Cieślak 2001, Pawłowski 1982, Stańko 1999]. Wartości stałej α ustala się zazwyczaj eksperymentalnie na takim poziomie, przy którym np. średni błąd prognoz wygasłych jest najniższy.

Informacje liczbowe z otrzymanych szeregów czasowych stanowią podstawę do obliczenia poziomu trendu liniowego $\hat{T}_t(t)$ i jego zmian $\hat{\beta}_1(t)$ [Stańko 1999]:

$$\hat{T}_t(t) = 2\hat{Y}_t - \hat{Y}_t^* \quad (3)$$

$$\hat{\beta}_1(t) = \frac{\alpha}{1-\alpha} (\hat{Y}_t - \hat{Y}_t^*) \quad (4)$$

Zbudowana prognoza w okresie t dla p okresów jest następująca [Stańko 1999]:

$$\tilde{y}_{t+p}^* = \hat{T}_t + p\hat{\beta}_1(t) = 2\hat{Y}_t - \hat{Y}_t^* + p\frac{\alpha}{1-\alpha} (\hat{Y}_t - \hat{Y}_t^*) \quad (5)$$

Jeżeli dane z próby wskazują na istnienie trendu nieliniowego, to możemy do prognozowania użyć **modelu Browna rzędu III** (kwadratowego). Podstawą wyrównywania są wówczas wzory (1) i (2) oraz trzeci w postaci:

$$\hat{Y}_t^{**} = \alpha\hat{Y}_t^* + (1-\alpha)\hat{Y}_{t-1}^{**} \quad (6)$$

gdzie: \hat{Y}_t^{**} – wygładzona wartość szeregu czasowego metodą potrójnego wygładzania wykładniczego w okresie t .

W wyniku otrzymujemy trzy wygładzone szeregi czasowe, które stanowią podstawę do obliczenia parametrów funkcji kwadratowej:

$$\hat{\beta}_0(t) = 3\hat{Y}_t - \hat{Y}_t^* + \hat{Y}_t^{**} \quad (7)$$

$$\hat{\beta}_1(t) = \frac{\alpha}{2(1-\alpha)^2} [(6-5\alpha)\hat{Y}_t - 2(5-4\alpha)\hat{Y}_t^* + (4-3\alpha)\hat{Y}_t^{**}] \quad (8)$$

$$\hat{\beta}_2(t) = \frac{\alpha^2}{(1-\alpha)^2} (\hat{Y}_t - 2\hat{Y}_t^* + \hat{Y}_t^{**}) \quad (9)$$

Prognozę dla p okresów budujemy zgodnie z następującym wzorem [Pawłowski 1982, Stańko 1999]:

$$\hat{Y}_{t+p} = \hat{\beta}_0(t) + p\hat{\beta}_1(t) + p^2\hat{\beta}_2(t) \quad (10)$$

Jako kolejnej procedury użyto modelu wyrównywania **liniowo-wykładniczego Holta**. Zasada prognozowania za pomocą tego modelu jest

podobna do prognozowania za pomocą modelu Browna rzędu drugiego, z tym że występują w nim dwie różne stałe wygładzania. Wygładzanie za pomocą dwóch niezależnych parametrów α i γ , odpowiednio: do poziomego trendu i jego zmian, daje większą elastyczność wyboru tempa, w którym dąży trend i jego zmiany [Farnum, Stanton 1989]. Prognozując tą metodą, najpierw wygładza się poziom trendu $\hat{T}_t(t)$ zgodnie z wzorem (11), a potem jego zmiany $\hat{\beta}_1(t)$ według wzoru (12).

$$\hat{T}_t(t) = \alpha Y_t + (1 - \alpha)[\hat{T}_t(t-1) + \hat{\beta}_1(t-1)] \quad (11)$$

$$\hat{\beta}_1(t) = \gamma[\hat{T}_t(t) - \hat{T}_t(t-1)] + (1 - \gamma)\hat{\beta}_1(t-1) \quad (12)$$

gdzie: $\hat{T}_t(t)$ – ocena wartości średniej w momencie (okresie) t ,

$\hat{\beta}_1(t)$ – ocena przyrostu trendu w momencie (okresie) t ,

Y_t – rzeczywista wartość zmiennej w momencie (okresie) t ,

$\hat{T}_t(t-1)$ – ocena wartości średniej w momencie (okresie) $t-1$,

$\hat{\beta}_1(t-1)$ – wygładzona wartość przyrostu trendu w momencie $t-1$,

γ – stała wygładzania dla zmian trendu.

Równanie prognozy na moment (okres) $t > n$ zapisujemy następująco [Farnum, Stanton 1989]:

$$\hat{Y}_{t+p} = \hat{T}_t(t) + p \hat{\beta}_1(t) \quad (13)$$

Do budowy modelu potrzebne są początkowe wartości komponentów $\hat{T}_t(1)$ i $\hat{\beta}_1(1)$. Najczęściej przyjmuje się uznanie za $\hat{T}_t(1)$ pierwszej wartości zmiennej prognozowanej y_1 , a za $\hat{\beta}_1(1)$ różnicy między drugą a pierwszą wartością szeregu czasowego. Parametry modelu szacujemy tak, aby przy ich kombinacji odpowiedni błąd prognoz wygasłych był najmniejszy [Farnum, Stanton 1989].

Analiza i prognozowanie składają się z dwóch faz. Pierwszą jest wybór odpowiedniego modelu i jego parametrów, druga obejmuje weryfikację otrzymanych wyników na podstawie wybranych miar dokładności prognoz. W opracowaniu wybór parametrów poszczególnych modeli został dokonany z punktu widzenia minimalizacji średniego bezwzględnego błędu procentowego MAPE, określającego stopień dopasowania wartości szeregu czasowego uzyskanych z modelu \hat{y}_t do wartości empirycznych y_t [Cieślak 2001, Farnum, Stanton 1989]:

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| 100\%, \quad (14)$$

gdzie: y_t – wartości rzeczywiste w momencie (okresie) t ,
 \hat{y}_t – wartości uzyskane z modelu prognostycznego,
 n – liczba porównywanych okresów.

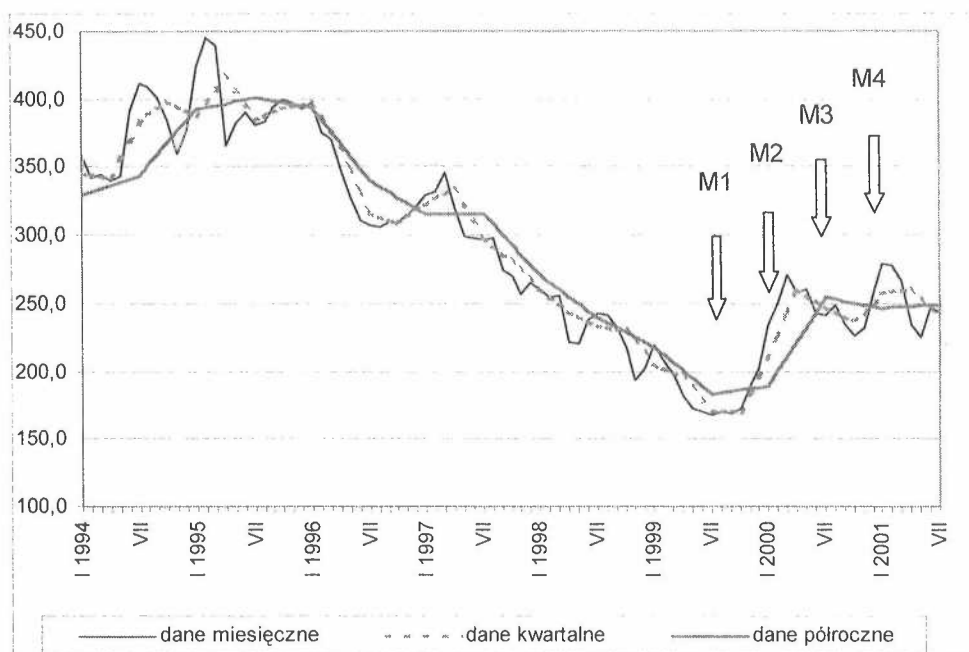
Korzystając z procedury prognozowania za pomocą wygładzania wykładniczego, nie można obliczyć mierników dokładności predykcji *ex ante*. Dlatego dokonamy porównania metod i wyników prognozowania na podstawie błędów prognoz *ex post* obliczanych po upływie czasu, na który prognoza była wyznaczona. Przedział, w którym prognoza jest sprawdzana nazywamy przedziałem empirycznej weryfikacji prognoz. Wyznaczenie błędów *ex post* pozwoli nam uzyskać odpowiedź na pytanie, dla jakiej formy danych oraz których metod wyniki prognozowania są dokładniejsze. Miernikiem oceny będzie średni bezwzględny błąd procentowy MAPE w przedziale weryfikacji.

Wyniki obliczeń

Dla wybranych modeli oszacowano ich parametry (stałe wygładzania) i zbudowano prognozy na pół roku dla trzech szeregów czasowych o różnej częstotliwości danych. Następnie porównano je z wartościami rzeczywistymi, obliczając błędy MAPE (wzór 14). Prognozy wykonano w czterech momentach: na podstawie danych do końca 1999 – M1 i 2000 r. – M3 oraz na podstawie danych do końca pierwszego półrocza 2000 r. – M2 i 2001 r. – M4 (rys. 1). Są to okresy leżące w bezpośrednim sąsiedztwie zmiany kierunku trendu.

Dopasowanie szeregu

Parametry modeli oszacowano przez poszukiwanie takich charakterystyk, dla których średni bezwzględny błąd procentowy między wartościami empirycznymi a prognozowanymi na jeden okres naprzód był najmniejszy (wzór 14). Optymalne stałe wygładzania obliczano oddzielnie w każdym momencie. Ich wartości liczbowe dla poszczególnych modeli przedstawia tabela 1. Możemy zauważyć pewne różnice pomiędzy parametrami poszczególnych modeli w zależności od częstotliwości danych statystycznych. Zaobserwowano jedynie niewielki spadek stałej α modelu liniowego Browna oraz duży wzrost stałej β modelu Holta, służącej do zmian trendu. Oznacza to, że w modelu Browna wraz ze



Rysunek 1

Ceny cukru białego na giełdzie w Londynie w okresie od stycznia 1994 r. do grudnia 2001 r.; strzałki oznaczają momenty wykonania prognoz

Źródło: Opracowanie własne na podstawie danych IERiGŻ.

zmniejszaniem częstości danych statystycznych w okresie retrospektywnym coraz większego znaczenia nabierają informacje starsze, natomiast w modelu Holta jest odwrotnie. Wynikać to może z faktu występowania różnej liczby stałych wygładzania i większej możliwości wyboru zmian poziomu i kierunku trendu tego modelu, w którym jest ich więcej (Holta). Potwierdzają to lepsze wyniki uzyskanego dopasowania modelu oraz błędy prognoz na pół roku.

Tabela 1

Stałe wygładzania dla szeregów o różnej częstości danych

Model	Dane miesięczne	Dane kwartalne	Dane półroczne
Model liniowy Browna	$\alpha = 0,7$	$0,5 \leq \alpha \leq 0,6$	$0,5 \leq \alpha \leq 0,6$
Model kwadratowy Browna	$0,3 \leq \alpha \leq 0,5$	$\alpha = 0,4$	$0,3 \leq \alpha \leq 0,4$
Model Holta	$\alpha = 0,9$ $\beta = 0,1$	$\alpha = 0,9$ $0,1 \leq \beta \leq 0,2$	$\alpha = 0,9$ $0,3 \leq \beta \leq 0,6$

Źródło: Obliczenia własne.

W celu porównania zbudowanych modeli dopasowanie zostało wyrażone za pomocą bezwzględnego błędu procentowego MAPE. Wyniki dopasowania wartości wynikających z modelu do danych rzeczywistych przedstawia tabela 2.

Tabela 2

Ocena dopasowania modeli do danych empirycznych wyrażona za pomocą MAPE w %

Częstość danych	Model liniowy Brownna	Model kwadratowy Brownna	Model Holta
Dane miesięczne	4,83	5,41	4,43
Dane kwartalne	6,86	7,68	6,12
Dane półroczne	8,59	9,52	8,09

Źródło: Obliczenia własne.

Mierniki dopasowania zawierają cenne informacje o wartości prognozy poszczególnych metod dla budowy prognoz. Najlepsze dopasowanie danych miesięcznych otrzymano dla modelu Holta (4,43%), a najwyższe dla modelu kwadratowego Brownna (5,41%). Jeżeli chodzi o dwie pozostałe formy danych, to model Holta okazał się również najbardziej skuteczny do wyznaczania prognoz na jeden okres naprzód. Należy zaznaczyć, że średnie odchylenia wartości teoretycznych od rzeczywistych tylko w przypadku danych miesięcznych (model liniowy i Holta) modeli mieszczą się w granicy 5% wartości oszacowanych prognoz. Jednocześnie zaobserwowano, w przypadku wszystkich metod, tendencję do wzrostu dopasowania modelu do danych empirycznych wraz ze zwiększaniem częstości danych statystycznych. Błędy dla danych półrocznych okazały się średnio o 80% wyższe niż w przypadku szeregów czasowych w formie miesięcznej, co jest zrozumiałe z uwagi na różny horyzont prognoz wygasłych.

Ocena *ex post* prognoz wygasłych

Analiza *ex post* prognoz wygasłych zbudowanych na pół roku wskazuje, że błędy prognoz zależą od momentu ich wykonania. Wygładzenie szeregu czasowego przez zmniejszenie częstości danych statystycznych spowodowało, że wartości błędów prognoz mają różny „rozkład” ze względu na moment prognozowania. W przypadku danych miesięcznych najwyższe ich wartości zaobserwowano w pierwszym i ostatnim okresie prognozowania. W tych miejscach (lub bezpośrednio po nich) dochodzi do zmiany kierunku przebiegu zjawiska. W przypadku danych kwartalnych i półrocznych wystąpiła sytuacja odwrotna i największe wartości MAPE są związane z prognozami wykonanymi w dwóch

okresach środkowych – M2 i M3. Dotyczy to wszystkich badanych modeli (tab. 3). Zmiana częstości danych spowodowała wygładzenie szeregu czasowego, a co za tym idzie – w innych okresach następuje większa zmiana kierunku trendu, co ma bezpośredni wpływ na uzyskane prognozy.

Najlepszą metodą okazał się model Holta, w którym zanotowano przeciętne wartości MAPE na poziomie 12,6% dla danych miesięcznych, 12,68% dla danych kwartalnych i 16,29% dla danych półrocznych. Błędy kolejnego modelu liniowego są wyższe o ponad 50% niż w modelu Holta, przy każdej częstości danych. Wartości prognoz modelu kwadratowego są ponaddwukrotnie wyższe niż najlepszego modelu Holta.

Tabela 3

Wartości błędów MAPE w przedziale weryfikacji pół roku w zależności od częstości danych i momentu wykonania prognozy

Forma danych	Moment prognozowania	Model liniowy Browna	Model kwadratowy Browna	Model Holta
Miesięczne	M1	24,83	31,20	19,79
	M2	9,37	24,98	6,24
	M3	15,01	48,06	6,02
	M4	28,01	51,46	18,34
	Średnio	19,30	38,92	12,60
Kwartalne	M1	13,43	14,20	20,21
	M2	47,21	46,51	21,86
	M3	15,90	42,95	4,29
	M4	5,69	13,77	4,35
	Średnio	20,56	29,36	12,68
Półroczne	M1	12,19	25,86	18,00
	M2	48,97	59,32	33,69
	M3	42,58	54,34	12,69
	M4	7,95	10,47	0,80
	Średnio	27,92	37,50	16,29

Źródło: Obliczenia własne.

Jeżeli chodzi o wpływ częstości danych na jakość wykonanych prognoz, to w przypadku modelu Holta zaobserwowano wzrost wartości MAPE wraz ze zmniejszaniem częstości danych. Najniższe wartości zaobserwowano dla danych miesięcznych – 12,6% oraz kwartalnych – 12,68%. Największe rozbieżności między wartościami prognozowanymi na pół roku a wartościami empirycznymi zaobserwowano dla danych półrocznych. Wynoszą one 16,29%. Po-

dobną tendencję zaobserwowano w przypadku modelu liniowego, gdzie wartości błędu MAPE dla danych miesięcznych, kwartalnych i półrocznych wynosiły odpowiednio: 19,30, 20,56 i 27,92%.

W przypadku modelu kwadratowego Browna nie ma wyraźnej zależności dokładności prognozowania od częstości danych statystycznych. Najniższą wartość błędu MAPE zaobserwowano dla szeregu czasowego zawierającego obserwacje kwartalne (29,36%), co jest wynikiem gorszym ponad 2,3 razy od najlepszej wartości modelu Holta. Znacznie gorsze rezultaty zostały osiągnięte dla danych półrocznych – 37,5% i miesięcznych – 38,92%. Prognozowanie za pomocą tego modelu, w szeregu charakteryzującym się częstymi zmianami kierunku trendu lub dużymi wahaniami przypadkowymi, prowadzi do wysokich błędów.

Podsumowanie

Przeprowadzone badania empiryczne nad wpływem częstości danych na wyniki prognozowania szeregów czasowych za pomocą metod adaptacyjnych przyniosły kilka spostrzeżeń.

Częstość danych z badanego szeregu czasowego wpływa w większym stopniu na zmianę wartości parametrów wygładzania modeli o ich większej liczbie niż w przypadku występowania tylko jednej stałej wygładzania.

Zmiana kierunku przebiegu zjawiska w końcowym okresie, na podstawie którego szacowano model, oraz tuż po wykonaniu prognozy powoduje otrzymanie błędnych wyników prognoz krótkookresowych bez względu na to, z jaką częstością danych statystycznych mamy do czynienia.

Najlepsze prognozy w badanych okresach uzyskiwano generalnie dla danych o większej częstości. Możemy tłumaczyć to faktem, że zbyt duże uśrednianie (wygładzanie) danych statystycznych prowadzi do zaniku istotnych informacji o kształtowaniu się zjawiska w czasie oraz wydłużenia realnego okresu, na podstawie którego był budowany model. Powoduje to, że prognoza budowana jest na podstawie danych zdezaktualizowanych. Inną sprawą jest to, że przedstawienie szeregu czasowego charakteryzującego się nieregularnymi zmianami trendu w postaci mniejszej częstości danych sprawia, że w miejscach dokonywania prognoz powstają zmiany kierunku „trendu” ostatniego segmentu, co ma wpływ jakoś osiągniętych prognoz.

Najniższe wartości błędów prognoz uzyskano dla modelu Holta, co potwierdza tezę, że dzięki większej liczbie parametrów osiąga się większą elastyczność odtwarzania mechanizmów generujących zjawisko. Jest on, w świetle

otrzymanych wyników, najlepszym modelem mogącym mieć zastosowanie w prognozowaniu zjawisk charakteryzujących się częstymi zmianami kierunku przebiegu zjawiska.

Literatura

CIEŚLAK M. (red.), 2001: Prognozowanie gospodarcze. Metody i zastosowania. PWN, Warszawa.

FARNUM N.R., STANTON L.W., 1989: Quantitative Forecasting Methods. PWS-KENT Publishing Company, Boston.

PAWŁOWSKI Z., 1982: Zasady predykcji ekonometrycznej. PWN, Warszawa.

STAŃKO S., 1999: Prognozowanie w rolnictwie. Wydawnictwo SGGW, Warszawa.

ZELIAŚ A., 1997: Teoria prognozy. PWE, Warszawa.

Forecast Results in Face of Data Frequency

Abstract

The purpose of this article is to find relations between forecast results and data frequency used as an example. The example concerns monthly, quarterly and half-yearly sugar prices from London Stock, 1994–2001. The Brown's linear and quadratic as well as Holt's exponential smoothing methods were employed to made the prediction. Obtained results indicate the influence of frequency data on forecasting results.