

# Analiza i prognozowanie szeregów czasowych w R

Adam Zagdański

Instytut Matematyki i Informatyki, Politechnika Wrocławska

QuantUp ([quantup.pl](http://quantup.pl)), Wrocław

Polski Akademicki Zlot Użytkowników R  
16-17 X 2014, Poznań

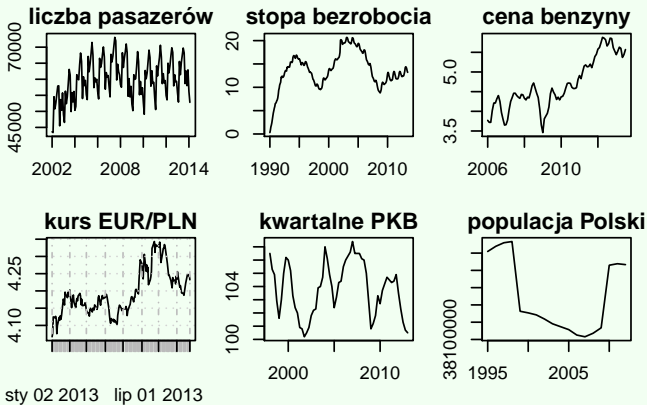
# Plan prezentacji

- 1 Wprowadzenie
- 2 Reprezentacja szeregów w R
- 3 Klasyczne modele i metody
- 4 Wybrane niestandardowe metody
- 5 Referencje

# Czym jest szereg czasowy?

- Szereg czasowy to obserwacje interesującej nas wielkości, zarejestrowane w kolejnych odstępach czasu (np. dniach, miesiącach lub kwartałach).
- Możliwe obszary zastosowań
  - Demografia
  - Socjologia
  - Ekonomia
  - Medycyna
  - Przemysł
- Z szeregami czasowymi często spotykamy się także wtedy, gdy zachodzi konieczność podejmowania ważnych decyzji biznesowych, np. dotyczących kupna/sprzedaży, produkcji, zatrudnienia czy logistyki

# Przykłady szeregów



Rysunek : Przykłady szeregów

# Przykłady analiz

- Prognozowanie cen akcji danej spółki w kolejnym roku
- Przewidywanie zmian cen paliw w kolejnych miesiącach
- Analiza sytuacji na rynku pracy (np. przewidujemy wzrost konkurencji w danej branży, który spowoduje istotną redukcję zatrudnienia)
- Analiza zmian demograficznych, socjologicznych, klimatycznych i ich wpływu na koniunkturę w określonej gałęzi przemysłu

# Typowe cele analizy szeregów czasowych

- Badanie występujących w danych: regularnych cykli, wzorców, trendów
- Prognozowanie wartości szeregów dla przyszłych okresów, na podstawie obserwacji historycznych
- Znalezienie modelu dobrze opisującego przebieg danego zjawiska w czasie

# Dlaczego warto analizować szeregi czasowe w R?

- Podstawowe modele i metody dostępne w wersji bazowej R'a.
- Rosnąca liczba pakietów, w tym najnowsze warianty metod!
- Rosnąca liczba publikacji:
  - Introductory Time Series with R (Use R!) by Paul S.P. Cowpertwait and Andrew V. Metcalfe (Jun 9, 2009).
  - Time Series Analysis: With Applications in R by Jonathan D. Cryer and Kung-Sik Chan (Oct 14, 2009)
  - Forecasting: principles and practice, Rob J Hyndman, George Athanasopoulos, 2013
  - Time Series Analysis and Its Applications: With R Examples by Robert H. Shumway and David S. Stoffer (Nov 25, 2010)
  - Basic Data Analysis for Time Series with R by DeWayne R. Derryberry (Jul 8, 2014)
  - Time Series: Applications to Finance with R and S-Plus by Ngai Hang Chan (Oct 5, 2010)
  - Multivariate Time Series Analysis: With R and Financial Applications by Ruey S. Tsay (Dec 9 2013)

## Reprezentacja szeregów w R: klasa *ts*

- Przykład: stopa bezrobocia w Polsce w latach 1990-2013, dane miesięczne.

```
> print(bezrobocie)
```

```
##           Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  
##   Nov  Dec  
## 1990  0.3  0.8  1.5  1.9  2.4  3.1  3.8  4.5  5.0  5.5  5  
##     .9  6.5  
## 1991  6.6  6.8  7.1  7.3  7.7  8.4  9.4  9.8 10.5 10.8 11  
##     .1 12.2  
## 1992 12.1 12.4 12.1 12.2 12.3 12.6 13.1 13.4 13.6 13.5 13  
##     .5 14.3
```

```
> class(bezrobocie)
```

```
## [1] "ts"
```



# Reprezentacja szeregów w R: klasa *ts*

```
> # data pierwszej i ostatniej obserwacji  
> start(bezrobocie)
```

```
## [1] 1990      1
```

```
> end(bezrobocie)
```

```
## [1] 2013      6
```

```
> # liczba obserwacji na jednostkę czasu  
> frequency(bezrobocie)
```

```
## [1] 12
```

## Reprezentacja szeregów w R: pozostałe klasy

- *xts* – extensible time series,
- *mts* – multiple time series,
- *msts* – multi-seasonal time series,
- *its* – irregularly spaced time series,
- ...

# Klasyczne modele i metody analizy szeregów

- Pakiet *stats*
  - **wykresy i podstawowe narzędzia:** `window`, `plot.ts`, `monthplot`, `acf`, `pacf`, `ccf`,...
  - **wygładzanie i dekompozycja:** `filter`, `decompose`, `stl`,...
  - **modele i algorytmy:** `ar`, `arima`, `arima.sim`, `HoltWinters`,...
- Pakiet *forecast*
  - **wykresy i podstawowe narzędzia:** `tsdisplay`, `seasonplot`, `BoxCox`, `nsdiffs`, `monthdays`, `bizdays`, `seasadj`,...
  - **prognozowanie:** `meanf`, `snaive`, `rwf`, `tslm`, `ses`, `holt`, `hw`, `auto.arima`, `arfima`, `ets`, `splinef`, `forecast`, `nnetar`, `tbats`, `accuracy`,...

# Klasyczne modele i metody: wykres szeregu

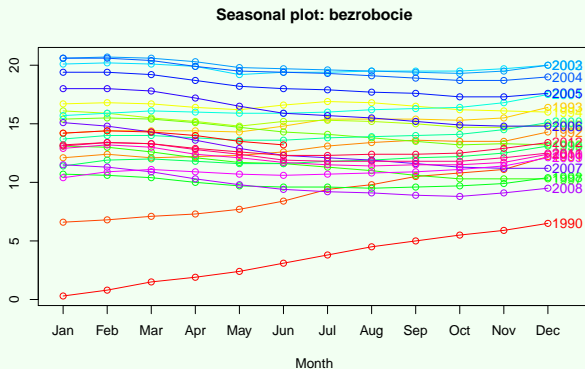
```
> plot(bezrobocie, main="stopa bezrobocia w Polsce")
```



Rysunek : Wykres szeregu

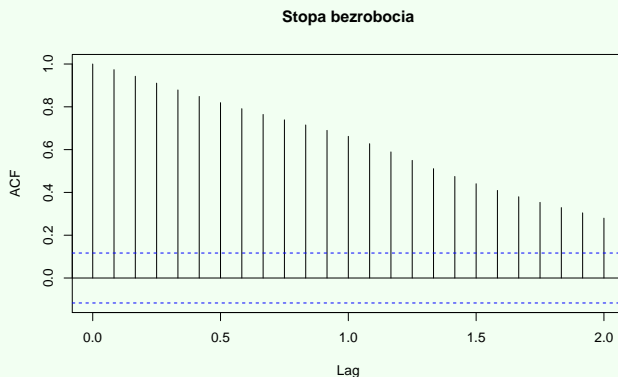
# Klasyczne modele i metody: wykres sezonowy

```
> seasonplot(bezrobocie, col=rainbow(23), year.labels=TRUE)
```



# Klasyczne modele i metody: ACF (funkcja autokorelacji)

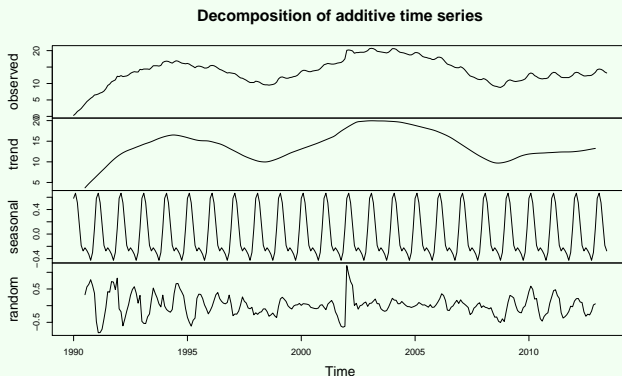
```
> acf(bezrobocie, main="Stopa bezrobocia")
```



Rysunek : Funkcja autokorelacji (ACF)

# Klasyczne modele i metody: Dekompozycja i wygładzanie

```
> bezrobocie.dekomp <- decompose(bezrobocie)
> plot(bezrobocie.dekomp)
```



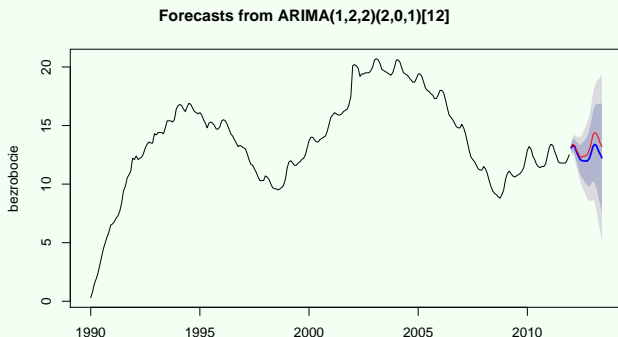
## Prognozowanie: Narzędzia dostępne w pakiecie *forecast*

- Pakiet *forecast* pozwala na automatyczne prognozowanie z wykorzystaniem dwóch najbardziej popularnych metod:
  - **modele ARIMA**
  - **modele ETS** (ExponenTialSmoothing lub ErrorTrendSeason)
- Optymalne modele dla poszczególnych szeregów wybierane są całkowicie automatycznie, z wykorzystaniem kryteriów statystycznych oceniających dokładność dopasowania.
- Oprócz prognoz punktowych otrzymujemy także przedziały predykcyjne (domyślnie, dla dwóch poziomów ufności: 80% i 95%).
- **Case study:** Adam Zagdański, *Jak zautomatyzować prognozowanie szeregów wykorzystując R?*,  
<http://quantup.pl/wp-content/uploads/2014/03/AutomatyzacjaKonstrukcjiPrognoz.pdf>



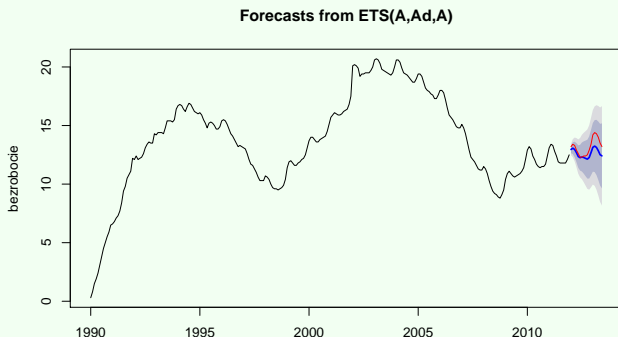
# Prognozowanie: ARIMA

```
> library(forecast)
> bezrobocie.arma <- auto.arma(bezrobocie.learn)
> bezrobocie.arma.forecast <- forecast(bezrobocie.arma, h=
  length(bezrobocie.test))
```



# Prognozowanie: ETS

```
> library(forecast)
> bezrobocie.ets <- ets(bezrobocie.learn)
> bezrobocie.ets.forecast <- forecast(bezrobocie.ets, h=
  length(bezrobocie.test))
```



# Prognozowanie: ARIMA vs ETS

```
> accuracy(bezrobocie.arima.forecast, bezrobocie.test)[,2:6]
```

```
##           RMSE      MAE      MPE  MAPE      MASE
## Training set 0.2267 0.1416 0.07223 1.123 0.06979
## Test set    0.6721 0.5569 4.14115 4.141 0.27447
```

```
> accuracy(bezrobocie.ets.forecast, bezrobocie.test)[,2:6]
```

```
##           RMSE      MAE      MPE  MAPE      MASE
## Training set 0.2450 0.1606 -0.1228 1.949 0.07915
## Test set    0.6769 0.5444  3.9909 4.002 0.26829
```

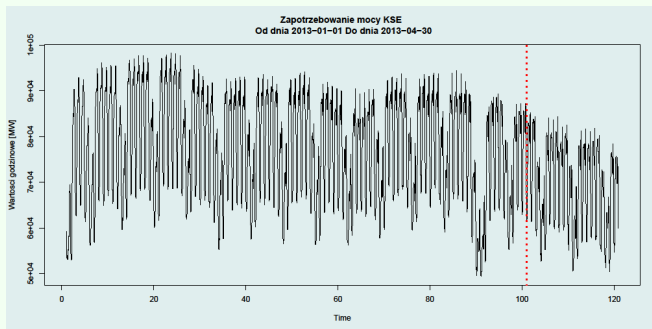
# Wybrane zaawansowane i niestandardowe metody

- Modelowanie szeregów o złożonej sezonowości
- Narzędzia dla traderów
- Analiza eksploracyjna szeregów
- Pozostałe zagadnienia

## Modelowanie szeregów o złożonej sezonowości

- Wiele szeregów cechuje występowanie niestandardowych zachowań sezonowych.
- Przykłady: szeregi o ułamkowym (niecałkowitym) okresie, szeregi o dużej częstotliwości i wielookresowej sezonowości.
- Większość standardowych metod opracowano z myślą o prostych wzorcach wahań sezonowych (np.:  $s=12$ ,  $s=4$ ).
- Istnieją również specjalistyczne modele i metody dedykowane szeregom o skomplikowanych wzorcach sezonowości.
- Kilka najnowszych propozycji: De Livera, A.M., Hyndman, R.J., Snyder, R. D. (2011), *Forecasting time series with complex seasonal patterns using exponential smoothing*, JASA, 106(496), 1513-1527.
- Implementacja w R: algorytmy *bats* i *tbats* (pakiet *forecast*)

# Modelowanie szeregów o złożonej sezonowości

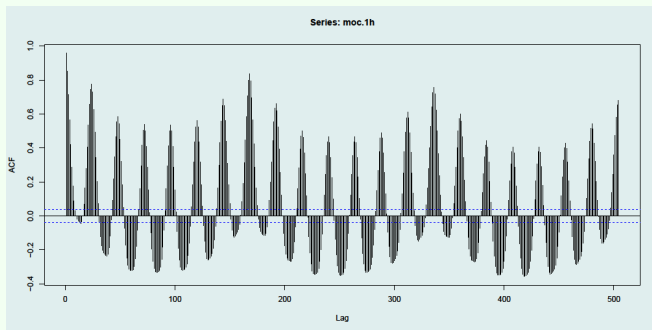


Rysunek : Zapotrzebowanie mocy KSE

- **Case study:** Adam Zagdański, *Prognozowanie zapotrzebowania mocy Krajowego Systemu Energetycznego (KSE)*. <http://quantup.pl/wp-content/uploads/2014/03/CaseStudyZapotrzebowanieMocyKSE.pdf>

# Modelowanie szeregów o złożonej sezonowości

- Funkcja autokorelacji (ACF)



**Rysunek** : ACF dla szeregu KSE. Widoczna sezonowość dobową ( $s=24$ ) oraz tygodniową ( $s=168$ ).

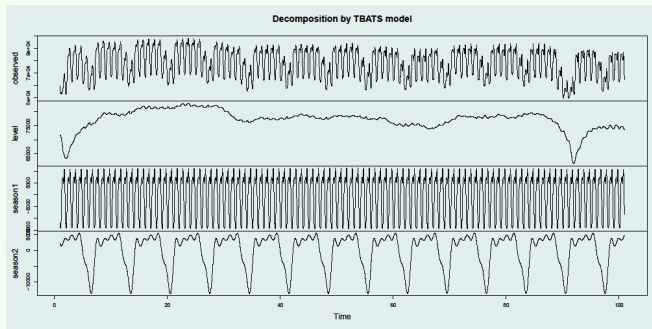
# Modelowanie szeregów o złożonej sezonowości

```
> library(forecast)
> # konwersja na obiekt msts (Multi-Seasonal Time Series)
> moc.1h <- msts( moc.1h, seasonal.periods=c(24, 168), ts.
  frequency=24 )
>
> # podzial danych na czesc uczaca i testowa
> # .....
>
> # dopasowanie modelu TBATS
> moc.1h.tbats.fit <- tbats(moc.1h.train, seasonal.periods=c
  (24,168))

> # parametry algorytmu TBATS (Trigonometric Exponential
  smoothing state space model with Box-Cox transformation,
  ARMA errors, Trend and Seasonal components)
> #
> # tbats(y, use.box.cox=NULL, use.trend=NULL,
> #       use.damped.trend=NULL, seasonal.periods=NULL,
> #       use.arma.errors=TRUE, use.parallel=TRUE,
> #       num.cores=NULL, bc.lower=0, bc.upper=1,...)
```



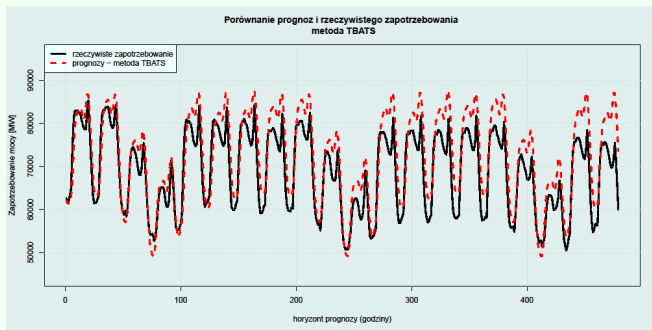
# Modelowanie szeregów o złożonej sezonowości



Rysunek : Dekompozycja szeregu KSE

# Modelowanie szeregów o złożonej sezonowości

```
> # prognozy  
> moc.1h.tbats.forecast <- forecast(moc.1h.tbats.fit, h=n.test  
  )
```



Rysunek : Prognozy dla danych KSE

# Narzędzia dla traderów

- Pakiet *quantmod* (Quantitative Financial Modelling Framework)

```
> library(quantmod)
>
> # pobranie notowań z ustalonego okresu
> getSymbols( Symbols="FB", from="2014-05-01", src="yahoo")
>
> # wykres świecowy
> chartSeries(FB, type = "candlesticks", up.col='green',dn.
  col='red')
```

# Narzędzia dla traderów (pakiet quantmod)

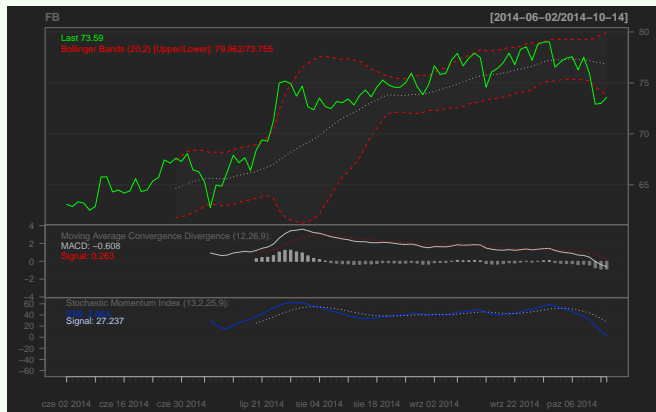


Rysunek : Notowania giełdowe

## Narzędzia dla traderów (pakiet quantmod)

```
> # wykres liniowy
> chartSeries(FB, type="line")
>
> # Bollinger Bands
> addBBands()
>
> # MACD == Moving Average Convergence Divergence
> addMACD()
>
> # STS == Stochastic Oscillator / Stochastic Momentum Index
> addSMI()
```

# Narzędzia dla traderów (pakiet quantmod)

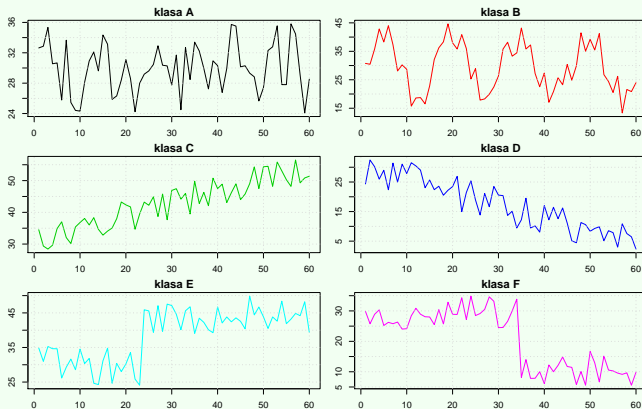


Rysunek : Analiza techniczna

## Analiza eksploracyjna szeregów

- Przykład: grupowanie kart kontrolnych
- Mamy 6 różnych klas (typów) kart kontrolnych, związanych z różnymi sytuacjami, jakie mogą wystąpić w pracy maszyn:
  - **A** - zachowanie normalne,
  - **B** - zmiany cykliczne,
  - **C** - rosnący trend,
  - **D** - malejący trend,
  - **E** - przesunięcie w górę,
  - **F** - przesunięcie w dół.
- Źródło: Alcock R.J. and Manolopoulos Y. Time-Series Similarity Queries Employing a Feature-Based Approach. 7th Hellenic Conference on Informatics. August 27-29. Ioannina, Greece 1999.

# Analiza eksploracyjna szeregów

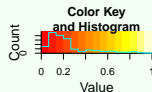


Rysunek : Przykładowe karty kontrolne

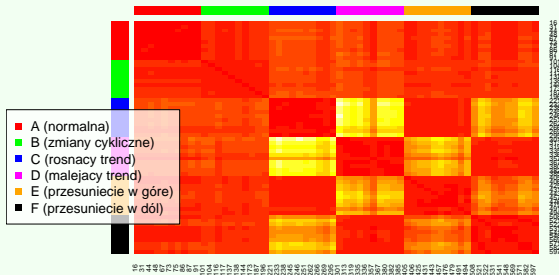


# Odległości na bazie DTW (Dynamic Time Warping)

```
> library(TSclust)
> karty.odl.DTW <- proxy::dist(karty, method="DTW", by_rows=
  TRUE)
```

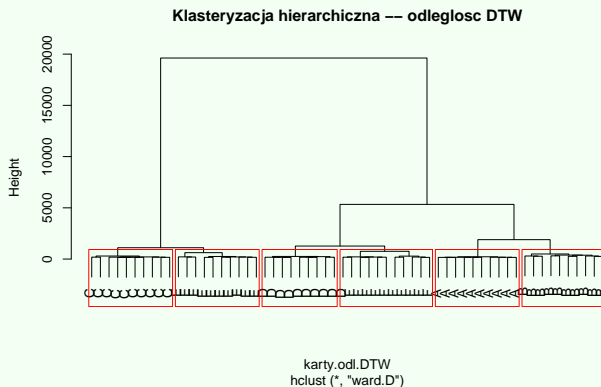


macierz odległości DTW



# Klastrowanie hierarchiczne na bazie odległości DTW

```
> hc.DTW <- hclust( karty.odl.DTW, method=linkage.method)
```



Rysunek : Dendrogram

## Pozostałe zagadnienia

- Modele wektorowe szeregów czasowych
- Dynamiczne modele regresyjne (transfer function models)
- Nieliniowe szeregi czasowe
- Prognozowanie szeregów na bazie sieci neuronowych
- Identyfikacja obserwacji odstających (ang. outliers)
- Wykrywanie punktów zmiany i segmentacja szeregów
- Analiza finansowych szeregów czasowych
- Testy pierwiastków jednostkowych (unit roots) i testy kointegracji
- Analiza w dziedzinie częstotliwości (analiza spektralna)
- ...
- **CRAN Task View: Time Series Analysis**  
<http://cran.r-project.org/web/views/TimeSeries.html>
- McLeod, Yu, Mahdi, *Time Series Analysis with R*, Handbook of Statistics, Volume 30, 2012, 661-712.

## Referencje

- **[1]** Rob J. Hyndman & Yeasmin Khandakar (2008), *Automatic Time Series Forecasting: The forecast Package for R*, Journal of Statistical Software, American Statistical Association, vol.27(i03).
- **[2]** *CRAN Task View: Time Series Analysis* <http://cran.r-project.org/web/views/TimeSeries.html>
- **[3]** Adam Zagdański, *Jak zautomatyzować prognozowanie szeregów wykorzystując R?*  
<http://quantup.pl/wp-content/uploads/2014/03/AutomatyzacjaKonstrukcjiPrognoz.pdf>
- **[4]** Adam Zagdański, *Case study: Prognozowanie zapotrzebowania mocy Krajowego Systemu Energetycznego (KSE)*. <http://quantup.pl/wp-content/uploads/2014/03/CaseStudyZapotrzebowanieMocyKSE.pdf>