Łódź, 14.05.2018

Zastosowanie metody iterowanej filtracji do badania zjawisk ekonomicznych na przykładzie idiosynkratycznego „efektu dźwigni”

mgr Piotr Szczepocki
Katedra Metod Statystycznych UŁ
Zebranie Katedry Ekonometrii UŁ

WYDZIAŁ
EKONOMICZNO-SOCJOLOGICZNY
Uniwersytet Łódzki

Używana w prezentacji notacja (por. Brzowoska-Rup i Dawidowicz 2009):

- $x_{i:j} = (x_i,\ldots, x_j)$, to realizacja procesu stochastycznego $\{X_t\}_{t\in N}$ od chwili i do j,
- proces stochastyczny $\{X_t\}_{t\in N}$ ma wartości w przestrzeni $\mathbb{R}^{n_x}$
- $p(x)$ funkcja gęstość rozkładu prawdopodobieństwa zmiennej losowej X
- $p(dx)$ rozkład prawdopodobieństwa zmiennej losowej X
- $\delta: \mathbb{R} \rightarrow \mathbb{R}$ dana wzorem $\delta(x) = \begin{cases} 1 \; dla \; x = 0 \\ 0 \; dla \; x \neq 0 \end{cases},$
- $\chi(A)$ indykator zbioru A

Ukryte łańcuchy Markowa – definicja (Kantas i in. 2015)

WYDZIAŁ EKONOMICZNO-SOCJOLOGICZNY
Uniwersytet Łódzki

Rozważamy parę procesów stochastycznych $\{X_t, Y_t\}_{t\in\mathbb{T}}$, gdzie proces $\{X_t\}_{t\in\mathbb{T}}$ (zwany procesem stanów lub procesem ukrytym) o znanym rozkładzie początkowym $p(dx_0)$ jest procesem o funkcji przejścia (gęstości rozkładu warunkowego)

$$p(x_t \mid x_{t-1})$$

Proces $\{X_t\}_{t\in\mathbb{T}}$ nie jest bezpośrednio obserwowalny, tylko poprzez proces pomiaru $\{Y_t\}_{t\in\mathbb{T}}$. Zakłada się, że proces pomiaru jest warunkowo niezależny (przy danym procesie $\{X_t\}_{t\in\mathbb{T}}$) oraz posiada gęstość rozkładu warunkowego

$$p(y_t \mid x_t)$$

Przykładem takiej przestrzeni stanów może być para procesów $\{X_t, Y_t\}_{t\in\mathbb{T}}$ postaci

$$\begin{cases} Y_t = \Pi X_t + u_t \\ X_t = \Xi X_{t-1} + \eta_t \end{cases} \qquad (1)$$

gdzie u_t i η_t są parami niezależnymi zmiennymi losowymi o rozkładzie normalnym oraz Π i Ξ są ustalonymi macierzami.

Precyzyjną formalną definicję ukrytych łańcuchów Markowa można znaleźć np. w Brzozowska-Rup i Dawidowicz (2009).

Zagadnienie filtracji

WYDZIAŁ EKONOMICZNO-SOCJOLOGICZNY
Uniwersytet Łódzki

Zagadnienie filtracji polega na wyznaczeniu $p(dx_{0:t}|y_{1:t})$ rozkładu warunkowego $X_{0:t}$ pod warunkiem obserwacji $Y_{1:t} = y_{1:t}$ i $p(dx_t|y_{1:t})$ rozkładu X_t pod warunkiem obserwacji $Y_{1:t} = y_{1:t}$ oraz momentów tych rozkładów (np. wartości średniej). Ze wzoru Bayesa otrzymujemy:

$$p(x_t \mid y_{1:t}) = \underbrace{\frac{p(y_t \mid x_t)}{p(y_t \mid y_{1:t-1})} p(x_t \mid y_{1:t-1})}_{korekta} = \frac{p(y_t \mid x_t)}{p(y_t \mid y_{1:t-1})} \underbrace{\int_{R^{n_x}} p(x_t \mid x_{t-1}) p(x_{t-1} \mid y_{1:t-1}) dx_{t-1}}_{predykcja} \tag{2}$$

oraz:

$$p(x_{0:t} \mid y_{1:t}) = \frac{p(x_t \mid x_{t-1}) p(y_t \mid x_t)}{p(y_t \mid y_{1:t-1})} p(x_{0:t-1} \mid y_{1:t-1}) \tag{3}$$

gdzie:

$$p(y_t \mid y_{1:t-1}) = \int_{R^{n_x}} p(y_t \mid x_t) \left\{ \int_{R^{n_x}} p(x_{t-1} \mid y_{1:t-1}) p(x_t \mid x_{t-1}) dx_{t-1} \right\} dx_t$$

Przestrzeń stanów – przykład zastosowania

Klasyczna postać

Podstawowy model stochastycznej zmienności (Taylor 1986):

$$\begin{cases} y_t = \varepsilon_t \exp(h_t / 2) \\ h_t = \mu(1-\phi) + \phi h_{t-1} + \sigma_\eta \eta_t \end{cases} \qquad (4)$$

gdzie $\{y_t\}$ - zaobserwowane zwroty, $\{h_t\}$ - proces logarytmicznej zmienności, $\varepsilon_t \sim N(0,1)(iid), \eta_t \sim N(0,1)(iid)$.
Dodatkowo zakłada się, że $h_0 \sim N\left(\mu, \frac{\sigma_\eta}{\sqrt{1-\phi^2}}\right), |\phi| < 1$.

Parą procesów stochastycznych stanowią $\{h_t, y_t\}_{t\in\mathbb{N}}$, gdzie $\{h_t\}$ jest procesem stanów (procesem ukrytym), $\{y_t\}$ - procesem obserwowanym.
Proces ukryty jest procesem Markowa, o gęstości przejścia:

$$h_t \mid h_{t-1} = h \sim N(\mu(1-\phi) + \theta h, \sigma_\eta) \qquad (5)$$

Proces pomiaru ma gęstość rozkładu warunkowego:

$$y_t \mid h_t = h \sim N(0, \exp(h/2)) \qquad (6)$$

Dodatkowo zakłada się, że $h_0 \sim N\left(\mu, \frac{\sigma_\eta}{\sqrt{1-\phi^2}}\right), |\phi| < 1$.

Przestrzeń stanów

Podejście klasyczne

Uogólniona metoda momentów: Melino, Turnbull (1990)

Metoda *quasi-największej wiarygodności*, w której do oszacowania funkcji wiarygodności używa się filtru Kalmana: Harvey, Ruiz, Shephard (1994), Harvey, Shephard (1996)

Filtry cząsteczkowe (Sekwencyjne metody Monte Carlo, *Sequential Monte Carlo*):
- uczenie parametrów (*parametr learning*): Liu, West (2001), Carvalho, Johannes, Lopes, Polson (2008), Ionides, Bhadra, Atchadé, King (2011, 2015)
- Algorytm EM (*Expectation–maximization):* Kim (2005), Brzozowska-Rup i Dawidowicz (2011)

Połączenie filtru cząsteczkowego z podejściem bayesowskim:
- MCMC w ramach SMC: Gilks, Berzuini (2001)
- SMC w ramach MCMC (*Particle Markov Chain Monte Carlo):* Andrieu, Doucet, Holenstein (2010)

Podejście bayesowskie z wykorzystaniem metody MCMC (*Markov Chain Monte Carlo*): Jacquier, Polson i Rossi (1994), Kim, Shephard i Chib (1998), Pajor (2003)

Podejście bayesowskie

Filtr cząsteczkowy (sekwencyjna metoda Monte Carlo)

- Filtry cząsteczkowe zostały opracowane niezależnie przez różnych autorów (Pitt i Shepard 1999). W szczególności Gordon i inni (1993) wykorzystali filtr cząsteczkowy do niegaussowskich modeli przestrzeni stanów, natomiast Kitagawa (1996) do modelowania szeregów czasowych. Po raz pierwszy termin filtr cząsteczkowy został użyty w pracy Del Moral (1996).

- Filtry cząsteczkowe różnią się między sobą sposobem generowania cząsteczek. Obszerne omówienie poszczególnych rodzajów filtrów cząsteczkowych można znaleźć w monografii Doucet i inni (2001), najbardziej znane to: losowanie z funkcji ważności (*Sequential Importance Sampling*, SIS), losowanie z funkcji ważności ze zwracaniem (*Sequential Importance Resampling*, SIR), filtr pomocniczy (*Auxiliary Particle Filter*)

- Filtry cząsteczkowe znalazły zastosowanie w wielu dziedzinach nauki m.in. w teorii przetwarzaniach sygnałów, chemii molekularnej, fizyce, automatyce czy ekonomii. Przegląd zastosowań można znaleźć w monografii Ristic i inni (2004).

- W ekonomii filtry cząsteczkowego wykorzystuje się m.in w modelach stochastycznej zmienności (podstawowy, z rozkładem t-studenta, z „efektem dźwigni" i „skokami" (por. Lopez i Tsay 2011), z w modelach DSGE (Malik i Pitt 2011, Herbst i Schorfheide 2014) oraz w modelach z parametrami zmiennymi w czasie Da-Silva i in. (2011).

- Zbieżność filtru cząsteczkowego jest przedmiotem wielu prac, m.in. Del Moral (1996) wykazał zgodność, Crisan i Doucet (2002) wykazali zbieżność średniokwadratową, dowód tego twierdzenia przy słabszych założeniach można znaleźć także w Brzozowska-Rup i Dawidowicz (2009).

Filtr cząsteczkowy (sekwencyjna metoda Monte Carlo)

Filtry cząsteczkowe (zwane także sekwencyjnymi metodami Monte Carlo) są rodziną metod aproksymujących rekurencyjnie nieznaną rodzinę rozkładów warunkowych

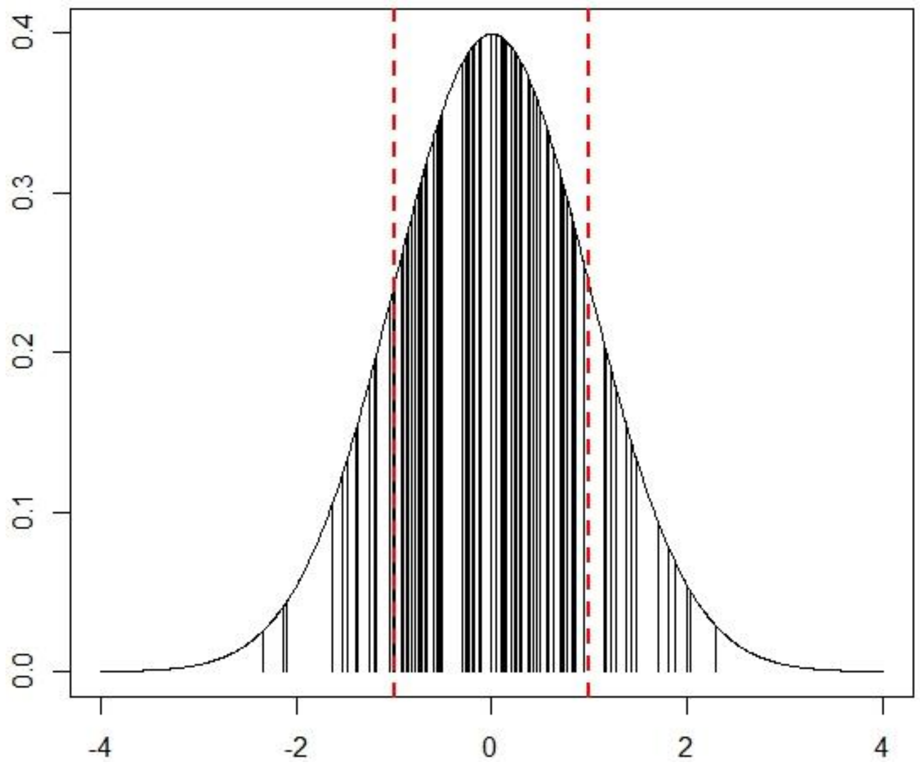
$$p(dx_{0:t} \mid y_{1:t})$$

dla $t = 1, \dots, T$ poprzez kolekcję J par $\left(x_{0:t}^{(j)}, w_i\right)$, zwanych cząsteczkami. Każda cząsteczka składa się z dwóch składowych: pierwsza $x_{0:t}^{(j)}$ odpowiada punktowi z nośnika funkcji $p(x_{0:t}|y_{1:t})$, a drugą składową $w^{(j)}$ jest wagą ważności (*importance weight*).

Rozkład warunkowy $p(dx_{0:t}|y_{1:t})$ (określony na przestrzeni $\mathbb{R}^{(t+1)\,n_X}$) jest aproksymowany przez dyskretny rozkład empiryczny:

$$\hat{p}(dx_{0:t} \mid y_{1:t}) = \sum_{j=1}^{J} w_t^{(j)} \delta\left(x_{0:t} - x_{0:t}^{(j)}\right) \qquad (7)$$

gdzie: $\left\{x_{0:t|t}^{j}\right\}_{j=1}^{J}$ jest ciągiem generowanym z odpowiednio dobranej funkcji ważności $q(\cdot)$.



Rys. 1. Przykład ilustrujący przybliżanie gęstości rozkładu ciągłego poprzez rozkład dyskretny za pomocą losowania Monte Carlo. Na rysunku przedstawiono wykres gęstości rozkładu normalnego standaryzowanego oraz 100 cząsteczek z wylosowanych z rozkładu normalnego standaryzowanego.

$\hat{P}(X \in (-1,1)) = \frac{1}{J}\sum_{j=1}^{J} \delta(x^{(j)} \in (-1,1))$=0,72.

Prawdziwa wartość $P(X \in (-1,1)) = 0{,}683$.

Zauważmy, że :

$$P(X_{0:t} \in B \mid y_{1:t}) = \int_B p(x_{0:t} \mid y_{1:t}) dx_{0:t} = \int_B \frac{p(x_{0:t} \mid y_{1:t})}{q_{0:n}(x_{0:t} \mid y_{1:t})} q_{0:n}(x_{0:t} \mid y_{1:t}) dx_{0:t} \approx \frac{1}{J} \sum_{j=1}^{J} \tilde{w}^{(j)} \delta(x_{0:t}^{(j)} \in B), \quad \tilde{w}^{(j)} = \frac{p(x_{0:t}^{(j)} \mid y_{1:t})}{q_{0:n}(x_{0:t}^{(j)} \mid y_{1:t})} \tag{8}$$

Wagi ważności wyznacza się zatem ze wzoru:

$$\tilde{w}_t^{(j)} = \frac{p(y_t \mid x_t^{(j)}) p(x_t^{(j)} \mid x_{t-1}^{(j)}) p(x_{0:t-1}^{(j)} \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1}) q(x_t^{(j)} \mid x_{0:t-1}^{(j)}, y_{1:t}) q_{0:n-1}(x_{0:t-1}^{(j)} \mid y_{1:t-1})} \propto \tilde{w}_{t-1}^{(j)} \frac{p(y_t \mid x_t^{(j)}) p(x_t^{(j)} \mid x_{t-1}^{(j)})}{q(x_t^{(j)} \mid x_{0:t-1}^{(j)}, y_{1:t})}, \tag{9}$$

$$w_t^{(j)} = \frac{\tilde{w}_t^{(j)}}{\sum_{j=1}^{N} \tilde{w}_t^{(j)}}.$$

Wynika to z równości (3):

$$p(x_{0:t} \mid y_{1:t}) = \frac{p(y_t \mid x_t) p(x_t \mid x_{t-1}) p(x_{0:t-1} \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})}$$

Ponadto: $$q_{0:n}(x_{0:t}^{(j)} \mid y_{1:t}) = q(x_t^{(j)} \mid x_{0:t-1}^{(j)}, y_{1:t}) q_{0:n-1}(x_{0:t-1}^{(j)} \mid y_{1:t-1}) \tag{10}$$

Najczęściej wykorzystuje się jako funkcję ważności funkcję przejścia: $q\left(x_t^{(j)} \middle| x_{0:t-1}^{(j)}, y_{1:t}\right) = \mathrm{p}\left(x_t^{(j)} \middle| x_{t-1}^{(j)}\right)$. Wówczas: (11)

$$\frac{p(y_t \mid x_t^{(j)}) p(x_t^{(j)} \mid x_{t-1}^{(j)})}{q(x_t^{(j)} \mid x_{0:t-1}^{(j)})} = \frac{p(y_t \mid x_t^{(j)}) p(x_t^{(j)} \mid x_{t-1}^{(j)})}{p(x_t^{(j)} \mid x_{t-1}^{(j)}) p(x_{0:t-1}^{(j)} \mid y_{1:t-1})} = p(y_t \mid x_t^{(j)})$$

Uwaga: optymalny (ze względu na minimalizację warunkowej wariancji wag w chwili t) wybór funkcji ważności zaproponowali Liu i Chen (1998).

Filtr cząsteczkowy – algorytm SIR (*Sequential Importance Resampling*)

WYDZIAŁ EKONOMICZNO-SOCJOLOGICZNY
Uniwersytet Łódzki

1. Inicjalizacja algorytmu dla t=0:

Dla $j = 1, \dots, J$:

Losujemy: $x_0^{(j)} \sim p(x_0)$.

Wyznaczamy wagi: $w_0^{(j)} = \frac{1}{J}$.

2. Iteracyjnie dla $t = 1, \dots, T$:

Dla $j = 1, \dots, J$:

Losujemy: $\hat{x}_t^{(j)} \sim q\left(x_t \middle| x_{0:t-1}^{(j)}\right)$

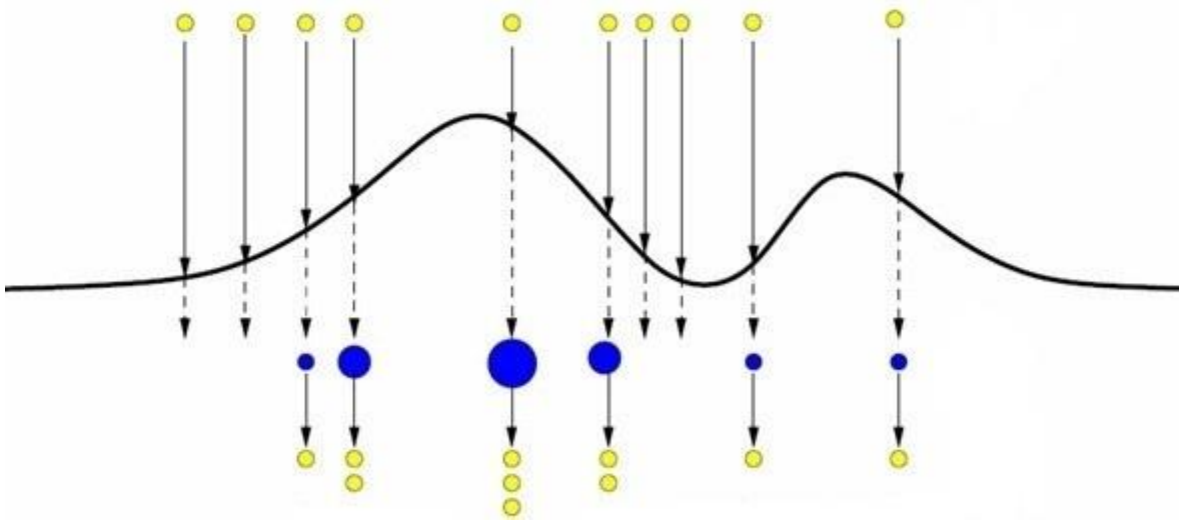
Przyjmujemy: $\hat{x}_{0:t}^{(j)} = (x_{0:t-1}^{(j)}, \hat{x}_t^{(j)})$.

Wyznaczamy wagi: $\widetilde{w}_t^{(j)} = \widetilde{w}_{t-1}^{(j)} p(y_t \mid x_t^{(j)})$.

Normalizujemy wagi: $w_t^{(j)} = \frac{\widetilde{w}_t^{(j)}}{\sum_{j=1}^{J} \widetilde{w}_t^{(j)}}$.

Losujemy (ze zwracaniem) J cząsteczek $x_t^{(J)}$ ze zbioru $\left\{\hat{x}_t^{(j)}, w_t^{(j)}\right\}$.

Przyjmujemy: $x_{0:t}^{(j)} = (x_{0:t-1}^{(j)}, x_t^{(j)})$ oraz $w_t^{(j)} = \frac{1}{\mathrm{J}}$.



Rys. 2. Graficzna prezentacja filtru cząsteczkowego.
Źródło :http://lia.deis.unibo.it/Research/SOMA/MobilityPrediction/filters.shtml

Losowanie ze zwracaniem ma na celu zapobieganiu degeneracji próby, ale jednocześnie prowadzi do degeneracji różnorodności. Miarą różnorodności jest:

$$\hat{N}_{eff} = \frac{1}{\sum_{j=1}^{J} (w_t^{(j)})^2} \; (1 \leq \hat{N}_{eff} \leq J).$$

Filtr cząsteczkowy – wyniki algorytmu

Dla każdego $t = 1, \ldots, T$ można otrzymać następujące oszacowania:

1. Rozkładu warunkowego $X_{0:t}$ pod warunkiem obserwacji $Y_{1:t} = y_{1:t}$:

$$\hat{p}(dx_{0:t} \mid y_{1:t}) = \sum_{j=1}^{J} w_t^{(j)} \delta\left(x_{0:t} - x_{0:t}^{(j)}\right) \qquad (12)$$

2. Momentów rozkładu $X_{0:t}$pod warunkiem obserwacji $Y_{1:t} = y_{1:t}$:

$$E(\varphi(x_{0:t}) \mid y_{1:t}) \approx \sum_{j=1}^{J} w_t^{(j)} \varphi\left(x_{0:t}^{(j)}\right) \qquad (13)$$

3. Rozkładu Y_t pod warunkiem obserwacji $Y_{1:t-1} = y_{1:t-1}$:

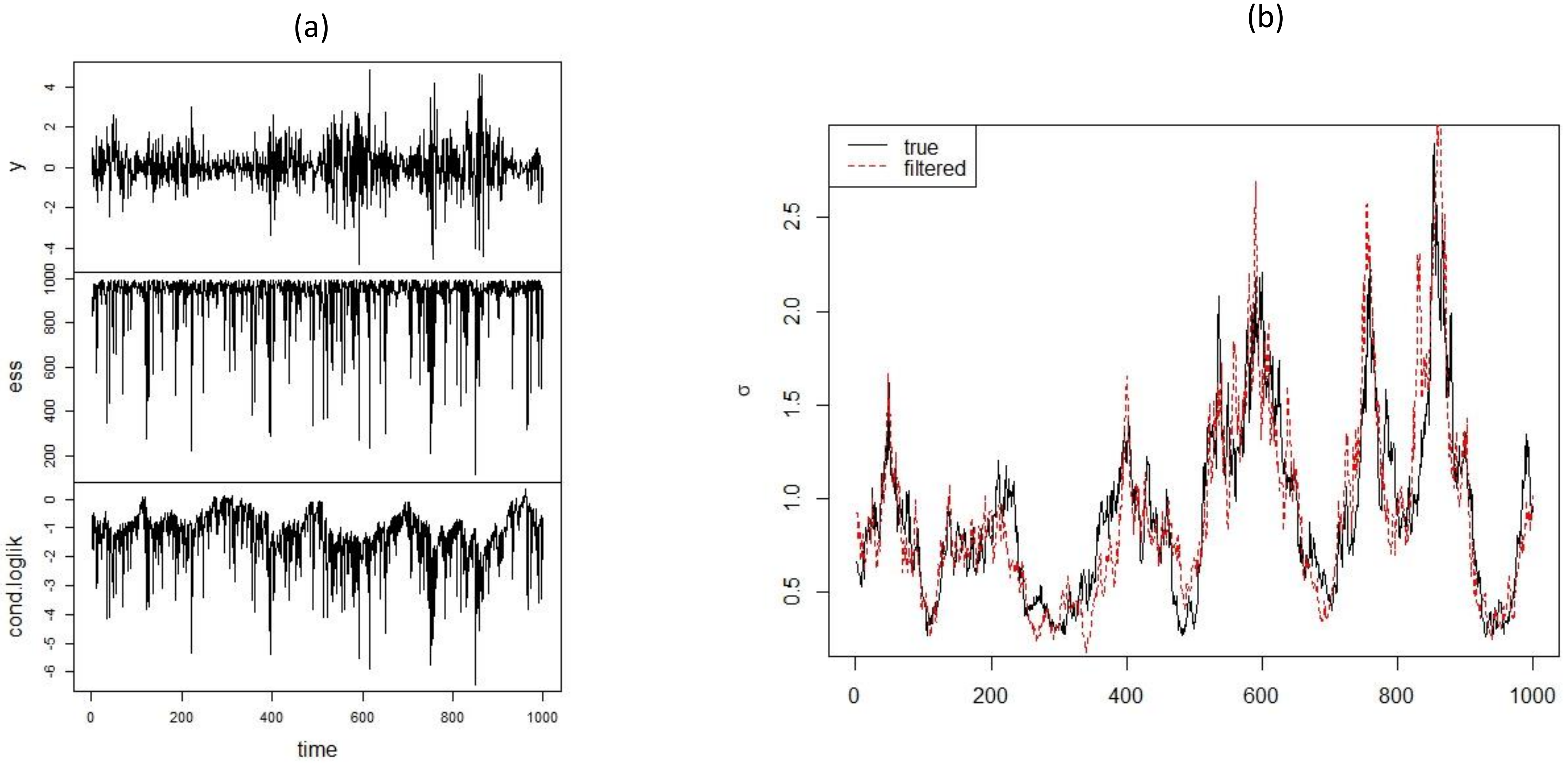
$$\hat{p}(y_t \mid y_{1:t-1}) = \frac{1}{J} \sum_{j=1}^{J} \tilde{w}_t^{(j)} \qquad (14)$$

Na podstawie całej próby można wyznaczyć oszacowanie (logarytmu) funkcji wiarygodności:

$$\hat{L}(\theta, y_1, \ldots, y_T) = \prod_{t=1}^{T} \hat{p}(y_t \mid y_{1:t-1}) = \prod_{t=1}^{T} \left(\frac{1}{J} \sum_{j=1}^{J} \tilde{w}_t^{(j)} \right) \qquad (15)$$

$$\ln \hat{L}(\theta, y_1, \ldots, y_T) = \ln \left(\prod_{t=1}^{T} \hat{p}(y_t \mid y_{1:t-1}) \right) = \sum_{t=1}^{T} \ln \hat{p}(y_t \mid y_{1:t-1}) = \sum_{t=1}^{T} \ln \left(\frac{1}{J} \sum_{j=1}^{N} \tilde{w}_t^{(j)} \right) \qquad (16)$$

Estymator funkcji wiarygodności dla dowolnego $T \geq 2$ jest nieobciążony. Estymator logarytmu funkcji wiarygodności jest obciążony.

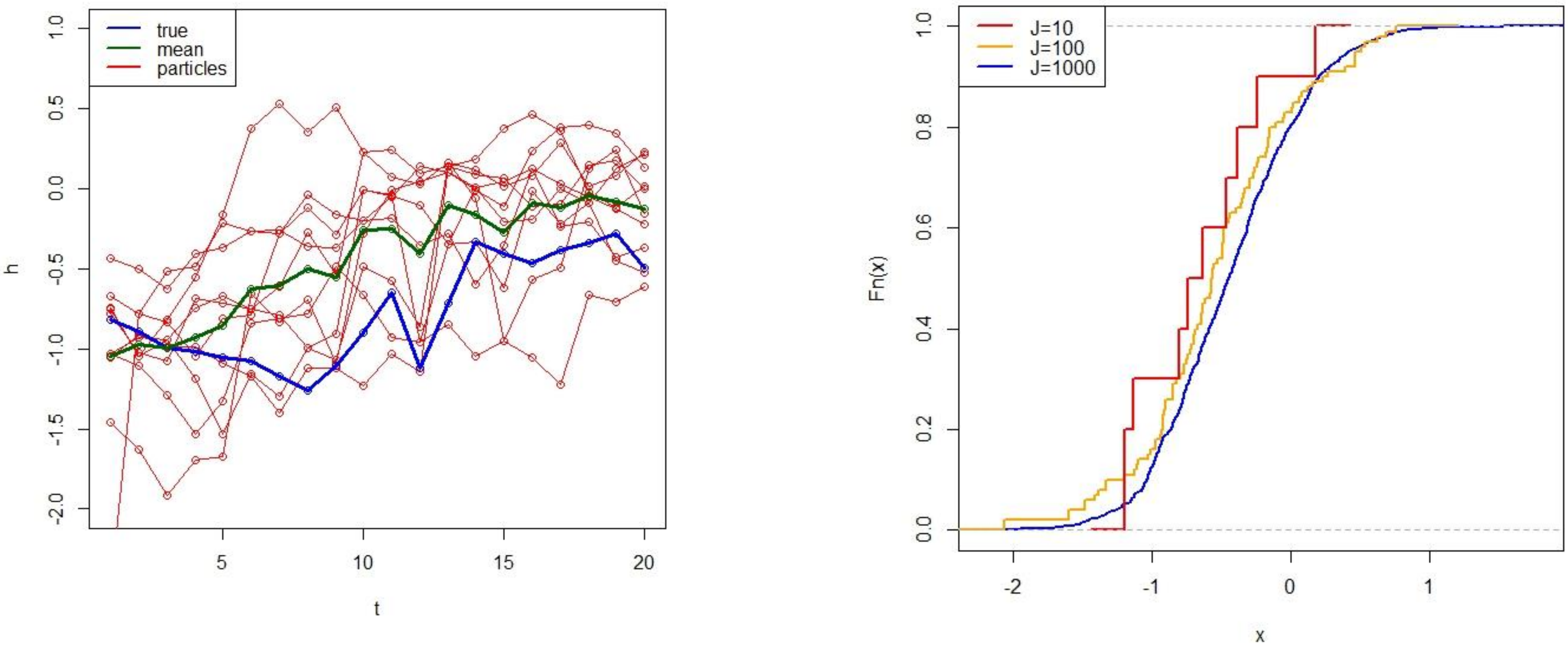


Rys. 3. Przykład zastosowania filtru cząsteczkowego dla klasycznego modelu stochastycznej zmienności: (a góra) symulacja trajektorii logarytmicznych zwrotów, (a środek) wykres efektywnej liczby cząsteczek, (a dół) $\ln \hat{p}(y_t|y_{1:t-1})$, (b) wykres trajektorii procesu zmienności (kolor czarny) oraz średniej arytmetycznej $\hat{p}(x_t|y_{1:t})$ dla t=1,…,1000 (kolor czerwony). Symulacja dla parametrów: μ=-0,5, ϕ=0,98, σ_η=0,2; T=1000.

Źródło: opracowanie własne przy użyciu pakietu pomp programu R Cran.

(a)

(b)



Rys. 4. Przykład zastosowania filtru cząsteczkowego dla klasycznego modelu stochastycznej zmienności: (a) trajektorie 10 cząsteczek (kolor czerwony), średniej arytmetycznej $\hat{p}(x_t|y_{1:t})$ (kolor zielony), logarytmicznego procesu zmienności (kolor niebieski) dla t=1,…,20 oraz (b) empiryczne dystrybuanty rozkładu prawdopodobieństwa $p(x_{20}|y_{1:20})$ dla 10,100,1000 cząsteczek. Symulacja dla parametrów: μ=-0,5, ϕ=0,98, σ_η=0,2.

Źródło: opracowanie własne przy użyciu programu R Cran.

Filtr cząsteczkowy – estymacja parametrów

Przyjmując, że gęstości $p(x_t|x_{t-1},\theta)$, $p(x_0,\theta)$ i $p(y_t|x_t,\theta)$ zależą od pewnego wektora parametrów θ. Wówczas:

$$L(\theta, y_1,\ldots, y_T) = \prod_{t=1}^{T} p(y_t \mid y_{1:t-1},\theta) \qquad (17)$$

Trudności związane z wykorzystaniem filtru cząsteczkowego do estymacji parametrów:
- estymator (logarytmu) funkcji wiarygodności nie jest funkcją ciągła parametrów w przypadku najczęściej stosowanych wariantów filtru cząsteczkowego (m. in. SIR, pomocniczy filtr cząsteczkowy*)*
- estymator logarytmu funkcji wiarygodności funkcji wiarygodności jest estymatorem obciążonym

Proponowane w literaturze metody estymacji (Kantas i in. 2015):

- Algorytm EM (*Expectation–maximization)* to iteracyjna metoda wyznaczania maksimum funkcji wiarygodności składająca się w każdej iteracji z dwóch kroków: Krok E (*Expectation*) i krok M (*Maximization*), w kroku E wykorzystuje się filtr cząsteczkowy do otrzymania symulacyjnie warunkowych wartości oczekiwanych (dla modeli SV algorytm EM został przedstawiony w Kim (2005), Brzozowska-Rup i Dawidowicz (2011))

- ciągła aproksymacja dystrybuanty empirycznej losowania ze zwracaniem zaproponowana przez Malik i Pitt (2011)

- uczenie parametrów (*parametr learning*): Liu, West (2001), Storvik (2002), Carvalho, Johannes, Lopes, Polson (2008), Ionides, Bhadra, Atchadé, King (2011, 2015)

- *Particle Markov Chain Monte Carlo:* Andrieu, Doucet, Holenstein (2010)

Filtr cząsteczkowy – estymacja parametrów

Uczenie parametrów (*parametr learning*):

- Liu, West (2001): metoda ta łączy (i) pomocniczy filtr cząsteczkowy, (ii) aproksymację $p(y_t|y_{1:t-1},\theta)$ przez mieszaninę rozkładów normalnych, (iii) sztucznie wprowadzoną ewolucję wektora parametrów $\theta_t = \theta_{t-1} + \varsigma_t$, gdzie $\varsigma_t \sim N(0, W_t)$

$$\hat{p}(y_t \mid y_{1:t-1}, \theta) = \frac{1}{J}\sum_{j=1}^{J} N(m^{(j)}, V; \theta_t), \quad (18)$$

$$m^{(j)} = a\theta_t^{(j)} + (1-a)\bar{\theta}_t, \ \ \bar{\theta}_t = \frac{1}{J}\sum_{j=1}^{J}\theta_t^{(j)}, V = \frac{1-a^2}{J}\sum_{j=1}^{J}(\theta_t^{(j)} - \bar{\theta}_t)(\theta_t^{(j)} - \bar{\theta}_t)^T$$

- Storvik (2002), Carvalho, Johannes, Lopes, Polson (2008) –modyfikacje metody Liu i Westa z wykorzystaniem statystyk dostatecznych

- Ionides i in. (2011, 2015) – iteracyjne powtarzanie filtru cząsteczkowego z parametrami wylosowanymi z wielowymiarowego rozkładu zaburzeń z malejącą z kroku na krok wariancją

Iterowana filtracja (Ionides i in. 2015)

Algorytm:

Niech: M – liczba iteracji, J – liczba cząsteczek, T – liczba okresów.

Dla m=1,…,M:

1. $\Theta_{0,j}^{F,m} \sim h_0(\theta \mid \Theta_j^{m-1}; \sigma_m)$ dla j=1,…,J

   $X_{0,j}^{F,m} \sim f_0(x_0; \Theta_j^m)$ dla j=1,…,J

2. Dla t=1,…,T:

   $\Theta_{t,j}^{P,m} \sim h_t(\theta \mid \Theta_{t-1,j}^{F,m-1}; \sigma_m)$

   $X_{t,j}^{P,m} \sim f(x_t \mid X_{t,j}^{F,m}; \Theta_{t,j}^{P,m})$ dla j=1,…,J

   $w_{t,j}^m = g(y_t \mid X_{t,j}^{P,m}; \Theta_{t,j}^{P,m})$ dla j=1,…,J

   Losujemy $k_{1:J}$ wg prawdopodobieństwa: $\mathbb{P}(k_j = i) = w_{t,i}^m / \sum_{u=1}^J w_{t,u}^m$

3. Przyjmujemy: $\Theta_j^m = \Theta_{T,j}^{F,m}$ dla j=1,…,J

4. Zwiększamy m=m+1 i wracamy do 2.

Wkład algorytmu:

Symulator: $f_0(x_0; \theta)$

Symulator: $f(x_t \mid x_{t-1}; \theta)$

Ewaluator: $g(y_t \mid x_t; \theta)$

Dane: $y_{1:T}$

Początkowa wartość wektora parametrów: Θ_j^0 dla j=1,…,J

Rozkład zaburzeń: $h_t(\theta \mid \varphi; \sigma_m)$ t=0,1,…,T

Ciąg zaburzeń: σ_m dla m=1,…,M

Wynik algorytmu:

Końcowy wektor parametrów:

Θ_j^M dla j=1,…,J

Iterowana filtracja

Zalety:
1) Podejście typu „plug-and-play"
2) Możliwość wykorzystania metody w bardzo wielu zagadnieniach
3) Zaimplementowane w pakiecie pomp programu R Cran
4) Teoretyczne własności estymatora
5) Dokładność estymacji porównywalna z konkurencyjnymi metodami (m.in. *PMCM*)
6) Klasyczne podejście „częstościowe"

Wady:
1) Podejście typu „plug-and-play"
2) Metoda aproksymuje gradient funkcji wiarygodności
3) „Przekleństwo wielowymiarowości"
4) Czas obliczeń

Twierdzenie 5 z (Ionides i in. 2011)

Niech $\{a_m\}_{m\in\mathbb{N}}, \{\tau_m\}_{m\in\mathbb{N}}, \{\sigma_m\}_{m\in\mathbb{N}}, \{J_m\}_{m\in\mathbb{N}}$ niech będą nieujemnymi ciągami takimi, że $\tau_m \to 0, \sigma_m/\tau_m \to 0$, $\sum_{m=1}^{\infty} a_m = 0, \sum_{m=1}^{\infty} a_m J_m \tau_m^{-2} < \infty$. Niech:

$$\hat{\theta}_{m+1} = \hat{\theta}_m + a_m \sum_{t=1}^{T} \left(V_{t,m}^{P}\right)^{-1} \left(\overline{\theta}_{t,m}^{F} - \overline{\theta}_{t-1,m}^{F}\right) \quad (19)$$

gdzie:

$$\overline{\theta}_{t,m}^{F} = \frac{1}{J_m} \sum_{j=1}^{J_m} \Theta_{t,j}^{F,m} \qquad V_{t,m}^{P} = \frac{1}{J_m - 1} \sum_{j=1}^{J_m} \left(\Theta_{t,j}^{P,m} - \overline{\theta}_{t,m}^{F}\right)\left(\Theta_{t,j}^{P,m} - \overline{\theta}_{t,m}^{F}\right)^{T}$$

Wówczas $\lim_{m\to\infty} \hat{\theta}_m = \hat{\theta}_{MLE}$ z prawdopodobieństwem 1. Warunki te spełnione są m.in. dla $a_m = m^{-1}, \tau_m^2 = m^{-1}, \sigma_m^2 = m^{-(1+\delta)}, J_m = m^{(1/2+\delta)}$ (Ionides i in. 2011).

Rys historyczny o „efekcie dźwigni”

- „Efekt dźwigni” można formalne zdefiniować jako ujemny związek pomiędzy $\mathbb{E}(h_{t+1}|y_t)$ (przyszłą zmiennością w okresie $t+1$) oraz y_t (obecny zwrot w okresie t) (Yu 2004)
- Black (1976) zauważył, że zmienność rośnie, gdy cena aktywa spada. Zaproponował następujące wytłumaczenie zwane „efektem dźwigni”: spadek cen akcji danej spółki powoduje identyczny spadek kapitału własnego spółki; jednocześnie rynkowa cena długu pozostaje niezmieniona, co powoduje wzrost współczynnika zadłużenia do kapitału własnego (*debt-to-equity ratio*) tzw. „dźwigni finansowej”
- Christie (1982) przedstawił teoretyczne uzasadnienie „efektu dźwigni” w obrębie twierdzeń Modigiliani-Millera (fundamentalnych twierdzeń dotyczących finansów przedsiębiorstw)
- Hipoteza „efektu dźwigni” spotkała się liczną krytyką (Suska 2015):
  - wzrost zmienności wydaje się zazwyczaj zbyt duży, aby mógł być tłumaczony zmianą współczynnika zadłużenia do kapitału własnego
  - wzrostowi cen nie zawsze towarzyszy spadek zmienności, czasem obserwuje się nawet dodatnią zależność
  - założenie, że inwestorzy znają wartość współczynnika zadłużenia do kapitału własnego jest trudne do obronienia
  - „Efekt dźwigni” jest zazwyczaj silniejszy dla indeksów giełdowych niż pojedynczych spółek
- Alternatywnie asymetrię zmienności można tłumaczyć hipotezą sprzężenia zwrotnego zmienności (*volatility feedback effect*): jeżeli zmienność można wycenić, to oczekiwany wzrost zmienności (wzrost ryzyka) powoduje także wzrost wymaganej przez inwestorów stopy zwrotu, czyli spadek bieżącej wartości spółki (spadek ceny)

Modele stochastycznego „efektu dźwigni”:

- Bandi and Renó (2012) rozważali modele stochastycznej zmienności, w których „efekt dźwigni” zależy od procesu zmienności, model został oszacowany nieparametrycznie na podstawie danych wysokiej częstotliwości

- Yu (2012) zaproponował model stochastycznej zmienności, w którym „efekt dźwigni” zależy od znaku i wielkości zaburzeń logarytmicznych stóp zwrotu:

$$h_t = \varphi h_{t-1} + \gamma \sum_{i=1}^{m+1} \left(\rho_i \varepsilon_{t-1} + \sqrt{1-\rho_i^2}\,\eta_t \right) \chi\left(\tau_{i-1} \geq \varepsilon_{t-1} \geq \tau_i \right) \qquad (20)$$

- Veraart i Veraart (2012) zaproponowali uogólnienia modeli stochastycznej zmienności Hestona (1993) oraz Barndorff-Nielsena i Shepharda (2001), w którym parametr ρ_t jest liniową transformacją procesu Jacobiego

$$\begin{cases} \rho_t = 2g_t - 1 \\ dg_t = (\xi - g_t)dt + \theta\sqrt{g_t(1-g_t)}\,dW_t \end{cases} \qquad (21)$$

- Breto (2016) zaproponował model stochastycznej zmienności, w którym parametrρ_t jest błądzeniem losowym przekształconym odwrotną transformacją Fishera

Idiosynkratyczny „efekt dźwigni”

Harvey, Ruziz i Shephard (1996) zaproponowali następujący model stochastyczny uwzględniający „efekt dźwigni”:

$$\begin{cases} y_t = \varepsilon_t \exp(h_t/2) \\ h_{t+1} = \mu(1-\phi) + \phi h_t + \sigma_\eta \eta_t \end{cases}, \qquad (22)$$

$$[\varepsilon_t \quad \eta]^T \sim N(0,\Sigma), \quad \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

$\Leftrightarrow$

$$\begin{cases} y_t = \varepsilon_t \exp(h_t/2) \\ h_{t+1} = \mu(1-\phi) + \phi h_t + \sigma_\eta \rho y_t \exp(-h_t/2) + \sigma_\eta \sqrt{1-\rho^2}\,\xi_t \end{cases}, \qquad (23)$$

$$\xi_t = (\eta_t - \rho\varepsilon_t)/\sqrt{1-\rho^2}, \quad [\varepsilon_t \quad \xi_t]^T \sim N(0, I_2),$$

Bretó (2014) zaproponował model idiosynkratycznego „efektu dźwigni”, w którym współczynnik asymetrii ρ_t jest procesem stochastycznym (błądzeniem losowym przekształconym odwrotną transformacją Fishera):

$$\rho_t = \frac{\exp(2g_t)-1}{\exp(2g_t)+1}, \qquad (24)$$

$$g_{t+1} = g_t + \sigma_\nu \nu_t$$

$$\nu_t \sim N(0,1)$$

Transformacja Fishera:

$$f:(-1,1) \to R,$$

$$f(x) = \operatorname{arctanh} x = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right).$$

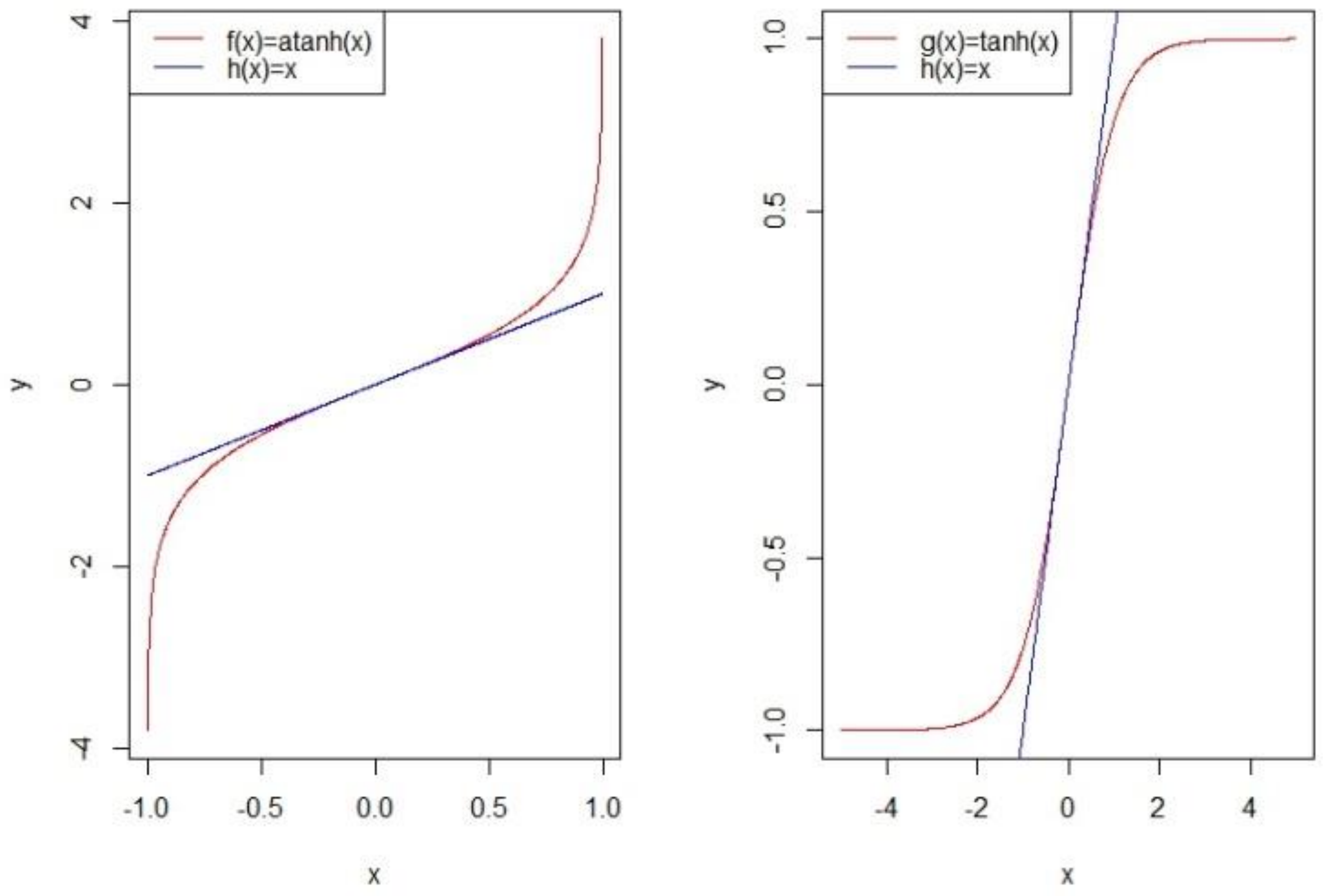
Odwrotna transformacja Fishera:

$$g: R \to (-1,1),$$

$$g(x) = \tanh x = \frac{\exp(2x)-1}{\exp(2x)+1}$$

Tw. Fisher (1915)
Niech (X,Y) ma dwuwymiarowy rozkład normalny o parametrze korelacji ρ. Oznaczmy przez r współczynnik korelacji z próby. Wówczas $f(r)$ ma rozkład normalny o wartości oczekiwanej $f(\rho)$ i odchyleniu standardowym równym $\frac{1}{\sqrt{N-3}}$, gdzie f jest transformacją Fishera.

WYDZIAŁ
EKONOMICZNO-SOCJOLOGICZNY
Uniwersytet Łódzki



Rys. 5. Transformacja Fishera i odwrotna transformacja Fishera.

Źródło: opracowanie własne przy użyciu programu R Cran.

Idiosynkratyczny „efekt dźwigni”

WYDZIAŁ EKONOMICZNO-SOCJOLOGICZNY
Uniwersytet Łódzki

Dane: indeks WIG, dzienne logarytmiczne stopy zwrotu przemnożone przez 100
Okres: 2.01.1996-31.12.2016 (5258 obserwacji)
Źródło: stooq.pl

Początkowa wartość wektora parametrów: 10 losowych wartości z przedziałów $\mu \in (-1; 1), \phi \in (0{,}9; 1), \sigma_\eta \in (0{,}01; 1)$, $\sigma_\nu \in (0{,}01; 1), G_0 \in (-0{,}5; 0{,}5), \rho \in (-1{,}0)$.

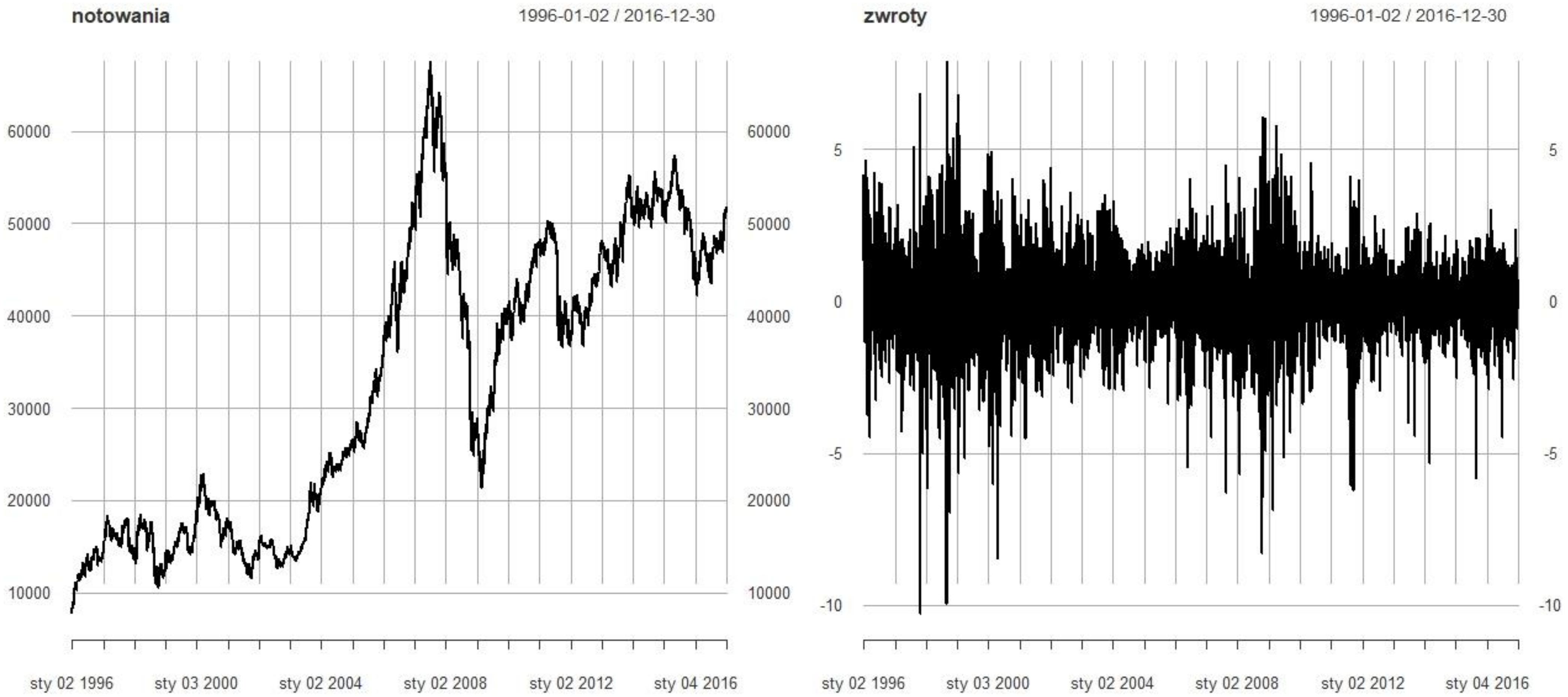
Rozkład zaburzeń: $h_t(\theta \mid \varphi; \sigma\ )$ to wielowymiarowy rozkład normalny o wektorze wartości oczekiwanych φ i macierzy kowariancji $\sigma\ \Sigma$, gdzie Σ jest macierzą diagonalną.

Ciąg zaburzeń σ_m: ciąg geometryczny o ilorazie ½. Liczba cząsteczek: 1000. Liczba iteracji: 150.

Tablica 1. Otrzymane oszacowania parametrów.

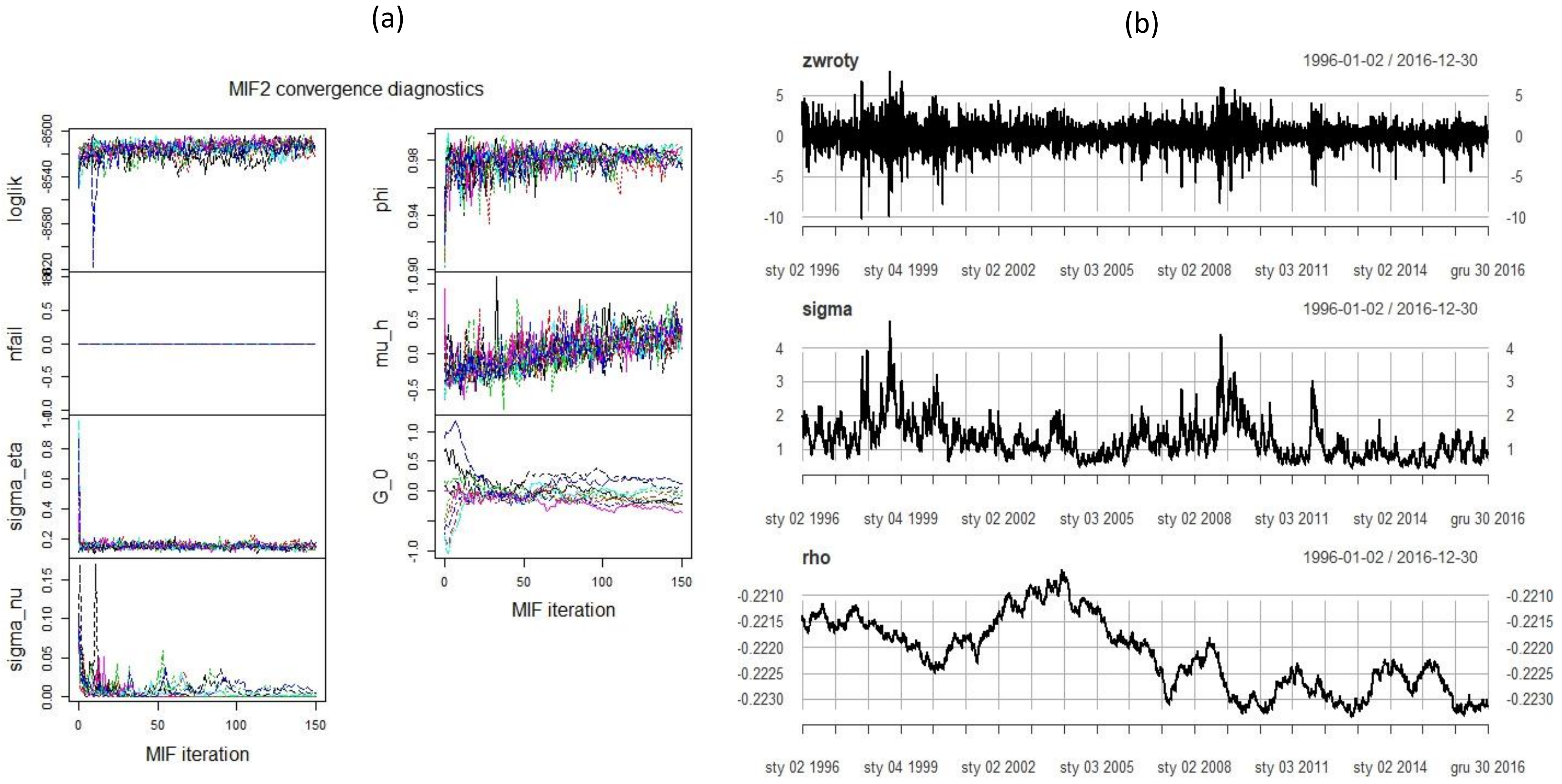
Model	μ	ϕ	σ_η	ρ	σ_ν	lnL
Harvey i Shephard (1996)	0,4752	0,9841	0,1477	-0,247	-	-8511,579
Bretó (2014)	0,3641	0,9816	0,1591	-	0,0261	-8509,219

Źródło: opracowanie własne przy użyciu pakietów : pomp programu R Cran.



Rys. 6. Dzienne notowania indeksu WIG w badanym okresie (a) oraz dzienne logarytmiczne stopy zmian (b).

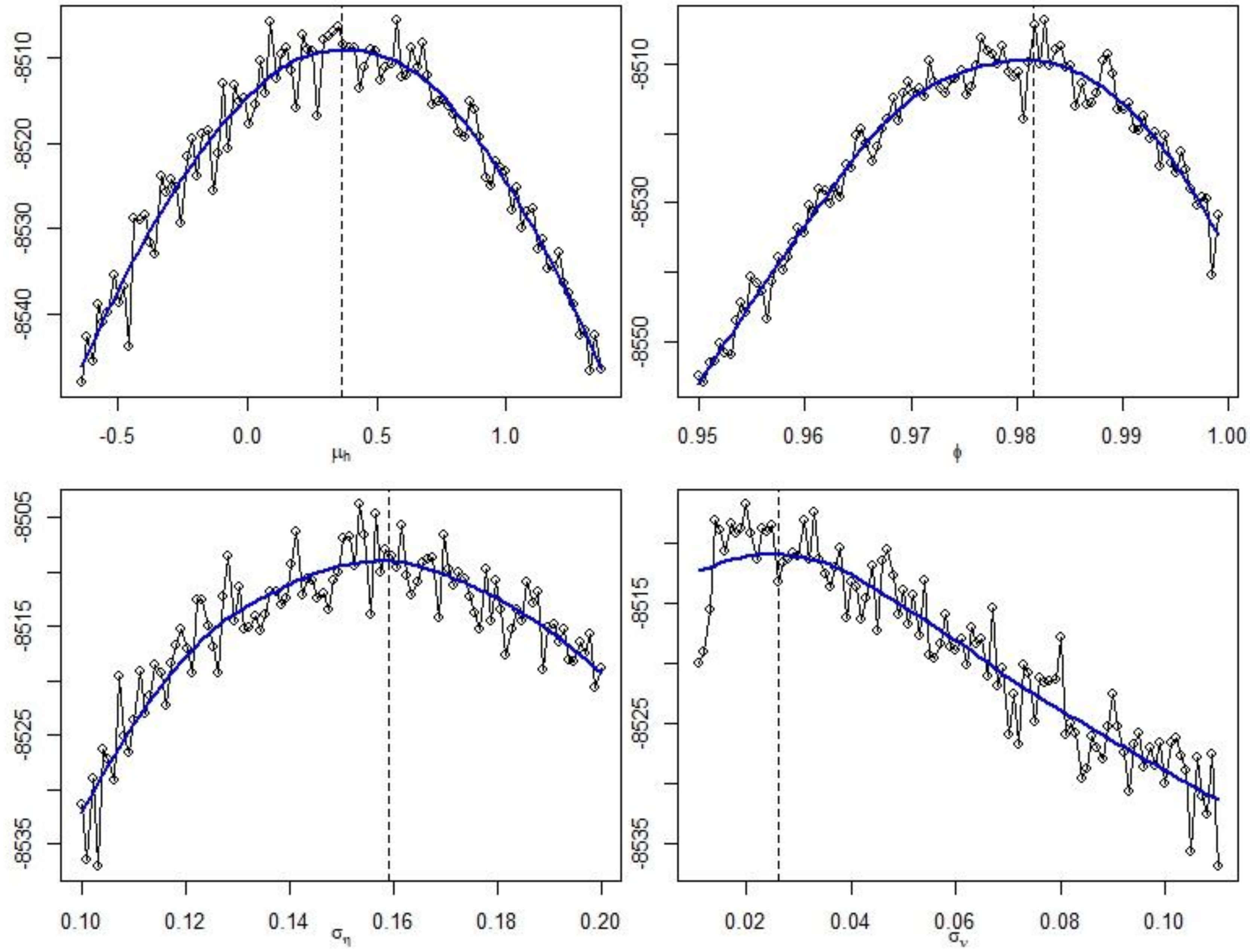
Źródło: opracowanie własne przy użyciu pakietu pomp programu R Cran.



Rys. 7. (a) wykres diagnostycznych dla metody iterowanej filtracji oraz wykresy: stóp zmian (b góra), procesu zmienności (b środek), idiosynkratyczny „efekt dźwigni” (c dół).

Źródło: opracowanie własne przy użyciu pakietu pomp programu R Cran.

WYDZIAŁ EKONOMICZNO-SOCJOLOGICZNY
Uniwersytet Łódzki



Rys. 8. Jednowymiarowe profile funkcji wiarygodności wyznaczone za pomocą filtru cząsteczkowego (pozostałe parametry jak w optimum uzyskanym metodą iterowanego algorytmu) - linia czarna, lokalnie wygładzone wielomianami stopnia drugiego – linia niebieska.

Źródło: opracowanie własne przy użyciu pakietu pomp programu R Cran.

Łódź, 14.05.2018

Dziękuję za uwagę

WYDZIAŁ EKONOMICZNO-SOCJOLOGICZNY
Uniwersytet Łódzki

Literatura:

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle markov chain monte carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72(3), 269-342.

Bandi, F. M., & Renò, R. (2012). Time-varying leverage effects. *Journal of Econometrics*, *169*(1), 94-113.

Black, F. (1976). Studies of Stock Price Volatility Changes. In: Proceedings of the 1976 Meeting of the Business and Economic Statistics Section, American Statistical Association, Washington DC, 177-181.

Bretó, C. (2014). On idiosyncratic stochasticity of financial leverage effects. Statistics & Probability Letters, 91, 20-26.

Brzozowska-Rup, K., & Dawidowicz, A. L. (2009). Metoda filtru cząsteczkowego. *Matematyka Stosowana: matematyka dla społeczeństwa*, *10*(51), 69-107.

Brzozowska-Rup, K., & Dawidowicz, A. L. (2011). Parameter estimation for nonlinear state-space models using particle methods combined with EM algorithm, FindEcon Monograph Series Advance in financial Market Analysis, Łódź University Press, 9.

Carvalho, C., Johannes, M., Lopes, H., and Polson, N. (2008) Particle Learning and Smoothing. Discussion Paper 2008-32, Duke University Dept. of Statistical Science.

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of financial Economics*, *10*(4), 407-432.

Crisan, D., & Doucet, A. (2002). A survey of convergence results on particle filtering methods for practitioners. IEEE Transactions on signal processing, 50(3), 736-746.

Da-Silva, C. Q., Migon, H. S., & Correia, L. T. (2011). Dynamic Bayesian beta models. Computational Statistics & Data Analysis, 55(6), 2074-2089.

Del Moral, P. (1996). Nonlinear filtering using random particles. Theory of Probability & Its Applications, 40(4), 690-701.

Doucet, A., de Freitas, N., Gordon, N.J., (2001). Sequential Monte Carlo Methods in Practice. Springer-Verlag, New York.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika, 10(4), 507-521.
Gilks, W. R., & Berzuini, C. (2001). Following a moving target—Monte Carlo inference for dynamic Bayesian models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(1), 127-146.
Gordon, N., Salmond, D. and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proceedings-F, 140, 107-13.
Harvey, A., Ruiz, E., & Shephard, N. (1994). Multivariate stochastic variance models. The Review of Economic Studies, 61(2), 247-264.
Harvey, A. C., & Shephard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business & Economic Statistics*, *14*(4), 429-434.
Herbst, E., & Schorfheide, F. (2014). Sequential Monte Carlo sampling for DSGE models. Journal of Applied Econometrics, 29(7), 1073-1098.
Ionides, E. L., Bhadra, A., Atchadé, Y., & King, A. (2011). Iterated filtering. The Annals of Statistics, 39(3), 1776-1802.
Ionides, E. L., Nguyen, D., Atchadé, Y., Stoev, S., & King, A. A. (2015). Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. Proceedings of the National Academy of Sciences, 112(3), 719-724.
Jacquier, E., N. G. Polson and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371-389.
Kantas, N., Doucet, A., Singh, S. S., Maciejowski, J., & Chopin, N. (2015). On particle methods for parameter estimation in state-space models. Statistical science, 30(3), 328-351.

Kastner, G., Frühwirth-Schnatter, S., & Lopes, H. F. (2014). Analysis of exchange rates via multivariate Bayesian factor stochastic volatility models. In *The Contribution of Young Researchers to Bayesian Statistics* (pp. 181-185). Springer, Cham.

Kim, S., Shephard, N. and Chib, S. (1998). Stochastic Volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies, 65, 361-93.

Kim, J. (2005). *PARAMETER ESTIMATION IN STOCHASTIC VOLATILITY MODELS WITH MISSING DATA USING PARTICLE METHODS AND THE EMALGORITHM* (Doctoral dissertation, University of Pittsburgh). Dostęp online (27.04.2018): http://d-scholarship.pitt.edu/8265/1/JeongeunKim_2005.pdf

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. Journal of Computational and Graphical Statistics, 5, 1-25.

Liu, J. S., & Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. Journal of the American statistical association, 93(443), 1032-1044.

Liu, J. and West, M. (2001). Combined parameters and state estimation in simulation-based filtering. In: Sequential Monte Carlo Methods in Practice (by Doucet, A., J. F. G. De Freitas and N. Gordon) 97-233. Springer-Verlag, New York.

Lopes, H. F., & Tsay, R. S. (2011). Particle filters and Bayesian inference in financial econometrics. Journal of Forecasting, 30(1), 168-209.

Malik, S., & Pitt, M. K. (2011). Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics*, *165*(2), 190-209.

Melino, A., & Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility. Journal of econometrics, 45(1-2), 239-265.

Pajor, A. (2003). Procesy zmienności stochastycznej SV w bayesowskiej analizie finansowych szeregów czasowych. *Monografie: prace doktorskie/Akademia Ekonomiczna w Krakowie*, (2).

Pitt, M.K., Shephard, N., (1999). Filtering via simulation: auxiliary particle filter. Journal of the American Statistical Association 94, 590–599.

Ristic, B., Arulampalam, S., & Gordon, N. (2003). Beyond the Kalman filter: Particle filters for tracking applications. Artech house.

Storvik, G. (2002). Particle filters for state-space models with the presence of unknown static parameters. IEEE Transactions on signal Processing, 50(2), 281-289.

Suska, J. (2015). Modelling Leverage Effect in a Financial Time Series. *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Finanse. Rynki finansowe. Ubezpieczenia*, (73 Ryzyko, zarządzanie, wartość), 843-852.

Taylor, S. (1986). Modelling financial time series. Chichester: Wiley.

Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics*, *127*(2), 165-178.

Veraart, A. E., & Veraart, L. A. (2012). Stochastic volatility and stochastic leverage. *Annals of Finance*, *8*(2-3), 205-233.