

多摩大学大学院
経営情報学研究所
博士論文

データ加増型モデルによる
市場ダイナミズムの
統計的予測

峰村 英二

2014年3月

内 容

1. 序論	1
1.1. はじめに	1
1.2. 基礎的考察	1
1.3. 母数推定に関する問題	5
1.4. 新たな統計的方法の導入	8
1.5. 本論の構成	9
2. 理論的背景	10
2.1. 本論で用いる DLM の概要	10
2.2. モデルの評価	14
2.3. データ加増アルゴリズムによる MCMC の概要	17
2.4. EM アルゴリズムによる母数初期期待値の推定	19
2.5. 下位モデル母数候補の推定	22
3. 実証分析	24
3.1. 実際の金利データに基づく実証	24
3.2. 前向き推論(Forward Induction)	33
3.3. 後ろ向き推論(Backward Induction)	34
4. 新たなデータに基づく研究	36
5. おわりに	48
参考文献	53
付 録	56
A.1. 「序論」の補論	56
a.1.1. (参考 1) シグマ集合体と可測性について	56
a.1.2. (参考 2) ブラウン運動	57
a.1.3. (参考 3) 離散時間マルチンゲール	58

a.1.4.	(参考 4) ギルサノフの定理	59
a.1.5.	フィルターの問題について	62
a.1.6.	補論のまとめ	69
A.2.	「2章」の補論	75
a.2.1.	ノーマル・ガンマ分布	75
a.2.2.	(参考) ガンマ分布と逆ガンマ分布	76
a.2.3.	情報割引率について	78
a.2.4.	本論の DLM 体系について	78
a.2.5.	尺度の調整	94
a.2.6.	未知母数 α の提案分布	96
a.2.7.	未知母数 β の提案分布	97

1. 序 論

1.1. はじめに

最近では、市場価格変動などの経済・市場データの評価を巡る理論的かつ実践的な議論が活発に行われ、わが国の金融市場においても住宅や不動産のローン債権を中心として本格的な証券化等が進展している。

この様な動きの背景には、市場で取り引きされる原資産価格や収益率の期待値と変動リスクを計測して予測、制御する理論および技術の発展がある。

しかしながら、現状では、モデルの母数推定問題等を含め、その実用面において、いくつかの問題点が残されているものと考えられる。

ここでは、まず、モデルの背景にある基礎的な観点を浮き彫りにし、その本質を踏まえた上で、いくつかの課題を浮き彫りにする。

1.2. 基礎的考察

ここでは、市場価格変動などの経済・市場データに関する評価理論の1つの例として、「派生証券」の評価に係る理論的考察を紹介する。なお、「派生証券」とは、様々な性格を持つ多様な原資産について、その確率的な変動を把握して理論的に評価可能な形に仕立てたものである¹。

派生証券の価格決定理論はすでに1970年代に確立され、近年に至るまで継続的に発展を続けているが、ここでは、一つの例示として、ある金融資産（株式、等）から派生される証券に関する、きわめて基礎的な価格決定理論を示す。

まず、ある金融資産の収益率 $\{r(t); t \in [0, T]\}$ 自体が以下の確率微分方程式に従って変動しているものと仮定する。

$$dr = \mu(t) dt + \sigma(t) dB, \quad t \in [0, T]; r(0) = r \quad (1.1)$$

1 現在、派生証券として取扱い可能な原資産は多様化している。例えば、ある定められた一定期間内に台風による被害があるかどうか、などの将来的な「天候」を取り引きする「天候デリバティブ」と呼ばれるものも出現している。

ここに、
 dr : 金融資産収益率の増分、 μ : ドリフト期待値、
 σ : 拡散係数 (標準偏差)、
 dB : 標準ブラウン運動増分 である。

次に、金融資産収益率 $\{r(t); t \in [0, T]\}$ から派生される証券価格を $F(r(t), t)$, $t \in [0, T]$ とすると、当該派生証券の評価を行う場合には、 $F(r(t), t)$ の割引現在価値である $e^{-\rho t} F(r(t), t)$ がマルチンゲールとなる確率測度 (分布) Q の下で、現時点 t までの情報に基づく、満期 T における証券価格の割引現在価値 $e^{-\rho T} F(r(T), T)$ の期待値が

$$E^Q [e^{-\rho T} F(r(T), T) | \mathcal{F}_t] = e^{-\rho t} F(r(t), t), \quad t < T \quad (1.2)$$

であること、すなわち、増加系列となる短期金利の観測情報集合 $\mathcal{F}_t := \sigma\{r(s); s \in [0, t]\}$ が与えられた下で、

$$d[e^{-\rho t} F(r(t), t)], \quad 0 \leq t \quad (1.3)$$

のドリフトがゼロとなることが求められる²。

すなわち、ここに、マルチンゲールとは、「時点 t ($t < T$) での証券価格の割引現在価値が、そのまま満期 T における証券価格の割引現在価値の条件付き期待値と同じとなる」ことである³。

(1.3) 式は、以下のとおり表現でき (ただし右辺の F の変数 $r(t), t$ を省略)、

$$d[e^{-r\rho t} F(r(t), t)] = d[e^{-r\rho t}] F + e^{-\rho t} dF \quad (1.4)$$

右辺第2項の dF については「伊藤の公式」⁴ を用いて

2 ドリフトのないブラウン運動は「マルチンゲール」となる。また、ドリフトが未知の場合でも、 \mathcal{F}_t -可測のドリフト期待値を推定できれば、当該期待値を原過程から減じることでマルチンゲール過程に変換できる。詳しくは、付録：A.1.「序論」の補論：「a1.5. フィルターの問題」を参照していただきたい。

3 詳しくは、付録：A.1.「序論」の補論：「a1.3. (参考3) 離散時間マルチンゲール」を参照していただきたい。

4 「伊藤の公式」の概要については、付録：A1.「序論」の補論：「a.1.4.(参考4) ギルサノフの定理」所収の(A.4.5)を参照していただきたい。

$$d[e^{-\rho t}F(r(t),t)] = e^{-\rho t}[-rFdt] \\ + e^{-\rho t}[F_t dt + F_r dr + \frac{1}{2}F_{rr}\sigma^2(t) dt] \quad (1.5)$$

とおくことができる。なお、(1.5)式で、2変数の関数 $F(r,t)$ について、

$$\begin{cases} F_t(r,t) = \frac{\partial}{\partial t} F(r,t) \\ F_r(r,t) = \frac{\partial}{\partial r} F(r,t) \\ F_{rr}(r,t) = \frac{\partial^2}{\partial r^2} F(r,t) \end{cases} \quad \text{とする。}$$

ここでのモデルは、

$$dr = \mu(t) dt + \sigma(t) dB, \quad t \in [0, T]; r(0) = r$$

であり、(1.5)式の右辺第2項に、このモデル式を代入して整理することにより以下を得る。

$$d[e^{-\rho t}F(r(t),t)] = e^{-\rho t}[-rF + F_t + F_r\mu(t) + \frac{1}{2}F_{rr}\sigma^2(t)] dt \\ + e^{-\rho t}\sigma(t)F_r dB \quad (1.6)$$

上式で $e^{-\rho t}F(r(t),t)$ が、 Q, dB の組み合わせのもとでマルチンゲールになるためには、(1.6)式のドリフト項（時間変化 dt に係る項）がゼロとなり、

$$d[e^{-\rho t}F(r(t),t)] = e^{-\rho t}\sigma(t)F_r dB \quad (1.7)$$

となる確率微分方程式にしたがう必要がある。

よって、偏微分方程式

$$-rF + F_t + F_r\mu(t) + \frac{1}{2}F_{rr}\sigma^2(t) = 0, \quad t \in [0, T] \quad (1.8)$$

を $F(r(t),t)$ について解くことにより価格が求まる。

なお、 \mathcal{F}_t -可測ドリフトの期待値が存在し、原系列から抽出できるということを経済学的にとらえれば、「完全市場 (complete market)」の前提から、いわゆる完全情報

の下で、一つの財に一つの価格期待値が成立する「一物一価」の法則が成立していることを表している、とみなすこともできる。

ここで、短期金利を考察の対象となる金融資産収益率とみなし、 $r(t) = r, t \in [0, T]$ の場合を考える。短期金利過程から派生する割引債券の価格関数について

$$\begin{aligned} F(r, t) &\stackrel{def}{=} \exp\{a(t) + b(t)r\}, \quad t \in [0, T]; \\ a(T) &= b(T) = 0. \end{aligned} \tag{1.9}$$

と定義すれば、

$$F_t = \left(\frac{d}{dt}a(t) + \frac{d}{dt}b(t)r\right)F(r, t), \quad F_r = b(t)F(r, t), \quad F_{rr} = b^2(t)F(r, t)$$

であるため、上式を(1.8)式に代入することにより、

$$\left(\frac{d}{dt}b(t) - 1\right)r + \left(\frac{d}{dt}a(t) + \mu(t)b(t) + \frac{\sigma^2(t)b^2(t)}{2}\right) = 0$$

が得られる。よって偏微分方程式(1.8)式が成立するためには、以下が満たされる必要があり、

$$\begin{cases} da(t) = -\mu(t)b(t)dt - \frac{\sigma^2(t)b^2(t)}{2}dt, & t \in [0, T], a(T) = 0, \\ db(t) = dt, & t \in [0, T], b(T) = 0, \end{cases} \tag{1.10}$$

最終的に、ここで示した単純なモデル((1.1)式)に基づいて以下の解が得られる。

$$\begin{cases} a(t) = \frac{1}{2} \int_t^T \sigma^2(u)(T-u)^2 du - \int_t^T \mu(u)(T-u) du, \\ b(t) = -(T-t). \end{cases} \tag{1.11}$$

この様に、本節で紹介した基本的なモデルでは、評価が金融資産の確率的な変動をマルチンゲールに変換する確率測度 Q に基づいて行われる。

1.3. 母数推定に関する問題

従来、確率微分方程式を離散近似したモデルに関する時系列分析において、モデルの母数推定を行うための最も一般的な方法の一つは、「一般化モーメント法 (the generalized method of moments ; GMM, 以下、GMM)」や最尤法(the maximum likelihood estimation ; MLE, 以下、MLE) などである。

しかし、Chan, Karolyi, Longstaff, and Sanders (1992) などの実証結果では、GMM による母数推定では、金利モデルの理論的な母数空間に適合した推定結果が得られない可能性があることが指摘されている。

Chan, Karolyi, Longstaff and Sanders (1992) では、GMM により、

$$dr = (\alpha + \beta r(t))dt + \sigma r(t)^\lambda dB \quad (1.12)$$

ここに、 $r(t)$: 金融資産収益率 $\stackrel{def}{=} 短期金利$ (財務省証券 TB 1ヶ月物金利) ,
 σ : ボラティリティ (拡散係数) , $dB(t)$: 標準ブラウン運動増分

とする短期金利の確率微分方程式を離散近似した時系列モデルを設定し、モデルの未知母数 $\theta = (\alpha, \beta, \sigma, \lambda)$ について推定している。

そして、その推定結果から以下の3点を重要な経験的事実として指摘している。

- (i) $\lambda \neq 0$ ならば、短期金利のボラティリティは金利水準とともに変化するが、実証分析の結果についても $\lambda \neq 0$ となっている。とくに $\lambda > 1$ とするモデルの方が $\lambda \leq 1$ との制約を課したモデルよりも説明力が高い。なお、無制約で推定した場合には、 $\lambda = 1.5$ となる。
- (ii) ここでのモデル母数 β は平均回帰過程を持つモデルにおける調整係数とみなされるが、 β の推定値は有意にゼロと異ならず、短期金利に長期平均水準への回帰メカニズムが機能しているとはいえない。したがって、本推定結果による限り、平均への回

帰を仮定してモデルを複雑にする根拠が薄れる⁵。

(iii) $\lambda = 1.5$ が得られたということは、このモデル上の短期金利は非定常過程であり、確率微分方程式の解が存在しないことを示している⁶。

これら(i)~(iii)の結果を踏まえ、彼らは、短期金利の変動に対してさらに説明力の高いモデルの開発が必要であることを示唆している。しかし、同時に、これらの結果は、離散近似した確率微分方程式モデルに基づく時系列モデルの母数推定の難しさも示していると思われる⁷。

一般的に、GMM については、分析の対象となる確率過程が強定常で必要次数のモーメントが存在すれば当該確率過程の増分に関する確率分布についての情報を必要としないことを利点とする見方がある。しかしながら、その一方で、確率分布に関する情報を一切必要としないために、却って的確な母数推定を行う上での妨げとなっている可能性があるものと考えられる。

今までみてきたとおり、確率微分方程式モデルのドリフト期待値は、増加系列となる観測情報集合 $\mathcal{F}_t := \sigma\{y(s) : s \in [0, t]\}$ が与えられた下で、当該情報集合について、

\mathcal{F}_t - 可測となる必要がある。このため、観測データが追加された場合には、直ちに新たな情報の変化を反映させた形で確率密度の更新ができ、逐次、当該密度布に基づいて観測される確率変数の期待値が計測できることが求められるものと考えられる。したがって、この手続きが担保されなければ、実際の金融資産収益率などの評価に関する有効性が損なわれてしまうおそれがある。

5 短期金利モデルでは、観測される金利変化が長期的な平均水準の回りで変動していると仮定して「平均回帰モデル」が用いられることが多い。

6 未知母数 λ について理論上要請される母数空間は $0 \leq \lambda \leq 1$ の範囲である。1 を超える場合には確率積分が不可能となり確率微分方程式の解が存在しない。一般的には、金利モデルでは $\lambda = 0.5$ (CIR モデルや Hull-White モデル : Hull and White (1993) など)、株式のモデルでは $\lambda = 1$ と仮定されることが多い。有名な Black-Scholes モデルでは $\lambda = 1$ と仮定されている。(1.1) 式で示したモデルでは $\lambda = 0$ である。

7 オプション市場では、多くの市場参加者が (コンピュータ上でプログラムされた) Black-Scholes モデルなどに基づいてオプションの評価を行っているため、モデル母数の推定に関する問題はとくに重要であると考えられる。

また、市場価格変動などの経済・市場データに関する価格変動率などの確率過程に係る経験分布が正規分布よりも裾の厚い分布に従うとする実証的な考察が、問題を一層複雑化させることになる。経済・市場データの評価等に関する基礎的な理論が、観測データの確率的な変動が正規過程に従うという定理から成り立っている場合に、実際の観測確率過程の経験分布が正規分布よりも裾の厚い分布に従うならば、正規過程に基づく予測や制御の範囲では到底起こりえない観測事象が事後的に生起することにより、当初想定した予測や制御の範囲を超える損失等が発生する可能性があり、モデル母数の推定に関する問題をさらに困難にすることが考えられるためである⁸。

この様に、金利や金融資産収益率に関連する理論モデルの母数推定に関する問題の背景には、観測される情報の変化に適合した未知母数の分布の推移（すなわち事後分布）をどのように捉えるか、という課題が存在している⁹。

8 フラクタル理論で著名なベノワ・B・マンデルブロ博士は、1963年に発表した綿花価格の分布の研究を始めとして、商品市場、通貨市場、株式市場のいずれにおいても、市場価格の変動に関する経験分布は正規分布の条件は満たされず、小幅な変動に加えてきわめて大幅な変動が加わった分布の形状（ファット・テール）をとると論じている(Mandelbrot (1963a, 1963b), Mandelbrot and Hudson (2004), Borland *et al* (2005))。

その一つの例として、2002年の通貨市場でドルが円に対して1日に7.92%も下がったことを踏まえ、正規分布で評価した場合は、150億年前のビックバン以降、今日までに1度も生起し得ない価格変動が実際に起こったことになる、などとしている(Mandelbrot and Hudson (2004))。

さらに、一般的な観測データでは、観測する時間間隔が長くなるにつれて「非正規性」が薄れていく傾向を想定するが、殊に市場価格の変動データに関しては、1日、1週間、1ヶ月、1年と、どのような時間単位で相場の変動幅をグラフにしても正規分布に収まらないことについても指摘している(Mandelbrot (2001a, 2001b), Mandelbrot *et al* (1997a, 1997b), Mandelbrot and Hudson (2004), Borland *et al* (2005))。

また、同時に、現在の価格変動が正規分布にならないことは最近の傾向ではなく、1609年に遡って、英国ポンドとオランダ・ギルダーの為替相場の変動幅（超長期時系列データ）を観測しても正規分布には従わないことを実証している。(Mandelbrot and Hudson (2004))。

9 とりわけ、広く用いられているBlack-Scholesモデルは、正規密度にしたがう解を解析的に求めるもので、ドリフトと標準偏差（ボラティリティ）がいずれも固定母数となっている。このため、実務家は、当初設定したモデル母数を市場の証券価格に適合させるように修正する「母数調整(キャリブレーション;calibration)」を行うことがある。

1.4. 新たな統計的方法の導入

金融資産収益率などに関する評価理論の根幹は、「任意の (μ, σ^2) -ブラウン運動に従う確率過程を、いかにして標準ブラウン運動 $((0,1)$ -ブラウン運動)過程に変換させるか」という問題である。

この問題に関しては、前述のとおり、観測過程原系列から、 \mathcal{F}_t -可測ドリフト期待値 ($\mathcal{F}_t := \sigma\{y(s) : s \in [0, t]\}$ は、増加系列となる観測情報集合) を除却してマルチンゲール過程を実現し、適切に推定されたボラティリティで、当該マルチンゲール過程を標準化することにより解決できる。

この方法は極めて単純でありながら、理論的方法¹⁰と同様に、実際的な確率測度変換が可能になるのではないかと考えられる。しかし、従来、この方法が検討されなかったのは、当該ドリフト期待値の逐次推定について、やや困難と想定されたためであろう。

しかしながら、後にフィルターの問題に関して詳しく考察するとおり、Øksendal (1998) の理論的な証明によれば、West and Harrison (1997) の「動的線形モデル (dynamic linear model ; DLM, 以下、単に DLM という。)」などの「状態空間モデル」を適用することによって、 \mathcal{F}_t -可測ドリフト期待値を推定することが可能となる。

また、最近時のボラティリティ推定に関するモデル分析では、Bollerslev (1986) による GARCH モデルを発展的に応用した Nakatsuma (2000) や Vrontos, Dellaportas, and Politis (2000)、統計的シミュレーションを SVM に導入した Jacquier, Polson, and Rossi (1994)、事前分布を設定して考察する Kim, Shephard, and Chib (1995) など、ベイズ統計的な分析方法 (Markov-chain Monte Carlo, Gibbs sampler, 等 ; 以下、MCMC) が数多く採用されている。MCMC は、本来、統計物理学の分野で相互に作用しあう分子からなる状態の平衡状態をシミュレーションするためのものである。この方法では、モデルの事後確率に従って母数空間と観測情報に適合した母数分布から派生させた母数候補を採択することで、採択された十分な数の候補に基づき、母数の事後期待値と周辺事後分布を把握できる。

一方、DLM では、分散母数が未知のとき、予測分布は、観測情報に条件付けられた t 分布にしたがう。このため、分散母数が未知の金利モデルの時系列分析に DLM を適用

¹⁰ 例えば、ギルサノフの定理など。ギルサノフの定理については、付録 : A.1.「序論」の補論 : 「a1.4. (参考4) ギルサノフの定理」を参照していただきたい。

すると、金利の予測分布は、直近までの観測情報を反映した t 分布となる。この結果は、金融資産の収益性などに関する予測分布では、正規分布よりも t 分布などの裾の厚い分布の方が適合度が高いとする先行的研究 (Bollerslev (1987), Watanabe and Asai (2001)) の結論と一致する。したがって、当該金利過程の予測分布から観測レプリケーションを派生させ、当該観測レプリケーションを新たな情報としたデータ加増法 (data augmentation algorithm) を用いることにより、金融資産の収益性などの予測に関して適合度の高いモデル母数の推計を行うことができるものと考えられる。

本論文の最終的な目的は、この様ないくつかの新しい手法に基づき、理論と実証とを接合する統計モデルの母数推定法を研究し、上記の確率測度変換を実際に行うことである。

1.5. 本論の構成

今までの議論を踏まえて、本論で「序論」以下、

- (1) DLMに基づく短期金利過程の \mathcal{F}_t -可測ドリフト期待値の推定と
マルチンゲール変換の実施
- (2) データ加増アプローチを取り入れたベイズ的MCMC法に基づく拡散係数
(ボラティリティ)の推定とマルチンゲール過程の標準ブラウン運動過程への変換の実施
- (3) 本論での測度変換の応用例

について研究し、序論で述べた問題に取り組むとともに、新しい統計的方法の提案を行う。

2. 理論的背景

2.1. 本論で用いるDLMの概要

ここでは、まず、West and Harrison (1997)、および、Minemura (2006) の表記を参考に、本論で用いる DLM について観測・状態・情報・予測の流れを紹介する。

$$\text{Observation : } Y_t = \mathbf{F}'\mathbf{T}_t + v_t, \quad v_t \sim N(0, \sigma_t^2), \quad (t = 1, \dots, N) \quad (2.1)$$

$$\begin{aligned} \text{State : } \quad \mathbf{T}_t &= \mathbf{G}\mathbf{T}_{t-1} + \mathbf{w}_t & \mathbf{w}_t &\sim t(\mathbf{0}, \mathbf{W}_t, h, v_{t-1}), & (t = 1, \dots, N) \\ & & \mathbf{W}_t &\stackrel{\text{def}}{=} \zeta_t^2 \mathbf{W}_t^\dagger = \delta_t \sigma_{t-1}^2 \mathbf{W}_t^\dagger, \quad \mathbf{W}_t^\dagger = \mathbf{I}_h, \\ & & & \text{ここに、}\delta_t\text{は情報割引率、また、}\mathbf{I}_h\text{は}h\text{次元単位行列。} & (2.2) \end{aligned}$$

$$\begin{aligned} \text{Information : } \quad \mathbf{T}_{t-1} | D_{t-1} &\sim t(\mathbf{m}_{t-1}, \mathbf{C}_{t-1}, h, v_{t-1}), & (2.3) \\ \sigma_{t-1}^{-2} | D_{t-1} &\sim \text{Gamma}\left(\frac{v_{t-1}}{2}, \frac{d_{t-1}}{2}\right), \end{aligned}$$

$$\text{with } E[\sigma_{t-1}^{-2} | D_{t-1}] = \frac{v_{t-1}}{d_{t-1}} \stackrel{\text{def}}{=} \frac{1}{S_{t-1}}. \quad (2.4)^{11}$$

ここに、

Y_t : 観測される確率変数	\mathbf{m}_t : 状態期待値行列 ($h \times 1$; unknown)
\mathbf{F} : 回帰ベクトル ($h \times 1$; known) ¹²	\mathbf{C}_t : 状態分散行列 ($h \times h$; unknown)
\mathbf{T}_t : 時変ドリフト (状態ベクトル: $h \times 1$; unknown)	σ_t^2 : 観測誤差分散 (unknown)
\mathbf{G} : 遷移行列 ($h \times h$; known)	\mathbf{W}_t : 状態誤差分散行列 ($h \times h$; unknown)
D_t : 情報集合 (Information Set)	v_t : 自由度 (unknown)
δ_t : 情報割引率 (unknown, 前掲)	ζ_t^2 : 状態誤差分散 (unknown)

11 ここでの S の定義は、West and Harrison (1997) にしたがった。West and Harrison (1997) では、「精度 (precision)」を $x = (\sigma^2)^{-1}$ とおき、 $X | D_t \sim \text{Gamma}(v_t / 2, d_t / 2)$ 、すなわち、

$$p(x | D_t) = \frac{(d_t / 2)^{\frac{v_t}{2}}}{\Gamma(v_t / 2)} x^{v_t / 2 - 1} e^{-x d_t / 2} \propto x^{v_t / 2 - 1} e^{-x d_t / 2}, \quad x > 0.$$

と仮定し、この期待値: $E[X | D_t] = \frac{v_t}{d_t} = \frac{1}{S_t}$ から、 S_t を「精度」の期待値の逆数と定義する。

なお、 $\sigma^{-2} | D_t \sim \text{Gamma}(v_t / 2, d_t / 2)$ のとき、確率変数 $X = \sigma^2$ のしたがう分布は、逆ガンマ分布となるので、その期待値: $E[\sigma^2 | D_t] = \frac{d_t}{v_t - 1}$ を観測誤差分散の推定量とおく。

詳しくは2の補論 A.2. 「ガンマ分布と逆ガンマ分布」を参照していただきたい。

12 本論では、ベクトルも行列も大文字記号で表記する。

ただし、 \mathbf{F} , \mathbf{G} は確定的で任意の有界区間上で有界、 σ_t^2 , ζ_t^2 は未知であるが有界区間上で正と仮定する。

また、 d はガンマ分布の尺度母数の逆数の 2 倍、 h は状態の次元である。さらに、情報集合 $D_t = \{Y_t, D_{t-1}\}$ には、時点 t で得ることが可能な観測に関するすべての情報が含まれる。なお、Minemura (2006) を参照して、予測モデルでは、以下の (2.5) 式、(2.7) 式の関係が示すとおり、さらに階層が仕切られ、下位モデルによって 1 期先観測予測モデルの Q_t^* が推定される構造となっている。

予測:

$$\text{上位モデル} \quad Y_t | D_{t-1} \sim t(f_t, Q_t^*; \nu_{t-1}) \quad (2.5)$$

$$T_t | D_{t-1} \sim t(\mathbf{a}_t, \mathbf{R}_t^*, h, \nu_{t-1}) \quad (2.6)$$

下位モデル (組み込まれた潜在的な「不均一分散モデル」)

$$\begin{aligned} Q_t^* &= \alpha e_{t-1}^2 + \beta Q_{t-1}^* & (t = 2, \dots, N) \\ \text{with } Q_t^* &\equiv Q_t & (t=1) \end{aligned} \quad (2.7)$$

ここに、観測 y_t が与えられたとすると、

$$e_t = y_t - f_t \quad (2.8.1)$$

$$f_t = \mathbf{F}'\mathbf{a}_t \quad (2.8.2)$$

$$\mathbf{a}_t = \mathbf{G}\mathbf{m}_{t-1} \quad (2.8.3)$$

$$\begin{aligned} Q_t &= \mathbf{F}'\mathbf{R}_t\mathbf{F} + S_{t-1} \\ &\stackrel{def}{=} S_{t-1}(\mathbf{F}'\mathbf{R}_t^\dagger\mathbf{F} + 1) \\ &= S_{t-1}Q_t^\dagger \\ (Q_t^\dagger &= \mathbf{F}'\mathbf{R}_t^\dagger\mathbf{F} + 1) \end{aligned} \quad (2.8.4)$$

$$\begin{aligned}
\mathbf{R}_t &= \mathbf{G}\mathbf{C}_{t-1}\mathbf{G}' + \mathbf{W}_t \\
&\stackrel{def}{=} S_{t-1}\mathbf{G}\mathbf{C}_{t-1}^\dagger\mathbf{G}' + \zeta_t^2\mathbf{W}_t^\dagger \\
&= S_{t-1}\mathbf{G}\mathbf{C}_{t-1}^\dagger\mathbf{G}' + \delta_t S_{t-1}\mathbf{W}_t^\dagger \\
&= S_{t-1}(\mathbf{G}\mathbf{C}_{t-1}^\dagger\mathbf{G}' + \delta_t\mathbf{W}_t^\dagger) \\
&= S_{t-1}\mathbf{R}_t^\dagger \\
&\quad (\mathbf{R}_t^\dagger = \mathbf{G}\mathbf{C}_{t-1}^\dagger\mathbf{G}' + \delta_t\mathbf{W}_t^\dagger, \delta_t: \text{情報割引率})
\end{aligned} \tag{2.8.5}$$

さらに、本モデルの観測分散尺度に係る「基本的仮定」として、

$$\begin{aligned}
E[\sigma_t^2 | \sigma_{t-1}^2, D_{t-1}] &\stackrel{def}{=} \sigma_t^{*2} \\
&= \sigma_{t-1}^2 \left(\frac{\sigma_t^{*2}}{\sigma_{t-1}^2} \right) \\
&= \sigma_{t-1}^2 \left(\frac{\sigma_t^{*2} Q_t^\dagger}{\sigma_{t-1}^2 Q_t^\dagger} \right) \\
&\stackrel{def}{=} \sigma_{t-1}^2 \left(\frac{Q_t^*}{Q_t} \right) \quad \text{とおく。}
\end{aligned}$$

なお、「基本的仮説」については、マンデルブロ博士のスケーリング理論により (Mandelbrot *et al* (1997a,1997b))、 σ_t^* に関して以下のとおり表現することが可能である。

$$\begin{aligned}
\sigma_t^* &= \sigma_{t-1}^{2H_{t-1}} \\
H_{t-1} &= \frac{\log_{\sigma_{t-1}^2} \left(\frac{Q_t^*}{Q_t} \right) + 1}{2}
\end{aligned}$$

上式で、常に $Q_t = Q_t^*$ となるならば、 $H_{t-1} = 1/2$ となり、通常のブラウン運動で見られる値に一致する。すなわち、 $2H_{t-1} = 1$ のときに、 $\sigma_{t-1} = \sigma_t^*$ であり、 $2H_{t-1} \neq 1$ のときは、 σ_{t-1} に替わり、 σ_t^* を採用する。

この仮定に基づき、 σ_{t-1}^2 および σ_t^{*2} が未知の場合で、 D_{t-1} が与えられたときの分散母数の推定値について、以下のとおり仮定し再定義する。

$$\begin{aligned}
S_t^* &= S_{t-1} \left(\frac{S_t^*}{S_{t-1}} \right) \\
&= S_{t-1} \left(\frac{S_t^* Q_t^\dagger}{S_{t-1} Q_t^\dagger} \right) \\
&\stackrel{def}{=} S_{t-1} \left(\frac{Q_t^*}{Q_t} \right)
\end{aligned}$$

ここに、 $S_t^* \stackrel{def}{=} E[\sigma_t^2 | D_{t-1}]$
また、 $S_t^* Q_t^\dagger \stackrel{def}{=} Q_t^*$ とおいた。 (2.8.6)

また、これにより、 $\mathbf{R}_t^* \stackrel{def}{=} \left(\frac{Q_t^*}{Q_t} \right) \mathbf{R}_t = S_{t-1} \left(\frac{Q_t^*}{Q_t} \right) \mathbf{R}_t^\dagger$ が得られる。 (2.8.7)

繰返更新:

$$\sigma_t^{-2} | D_t \sim \text{Gamma}(\frac{v_t}{2}, \frac{d_t}{2}) \quad (2.9)$$

$$\mathbf{T}_t | D_t \sim t(\mathbf{m}_t, \mathbf{C}_t, h, v_t) \quad (2.10)$$

ここに、

$$\begin{aligned}
\mathbf{C}_t &= (S_t / S_t^*) [\mathbf{R}_t^* - \mathbf{A}_t \mathbf{A}_t' Q_t^*] \\
&= (S_t / S_{t-1}) [\mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t' Q_t]
\end{aligned} \quad (2.11.1)$$

$$\begin{aligned}
\mathbf{A}_t &= \mathbf{R}_t^* \mathbf{F} / Q_t^* \\
&= \mathbf{R}_t \mathbf{F} / Q_t
\end{aligned} \quad (2.11.2)$$

$$v_t = v_{t-1} + 1 \quad (2.11.3)$$

$$\begin{aligned}
d_t &= d_{t-1} + S_t^* (e_t^2 / Q_t^*) \\
&= d_{t-1} + S_{t-1} (e_t^2 / Q_t)
\end{aligned} \quad (2.11.4)$$

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t e_t \quad (2.11.5)$$

したがって、「基本的仮説」による尺度の変更は、動的線型モデルの体系に何ら影響を及ぼさない。

平滑化:

さらに、観測される確率変数が最大で N 個ある場合、

$$\mathbf{B}_t = \mathbf{C}_t \mathbf{G}' \mathbf{R}_{t+1} \quad (t = 1, \dots, N-1) \quad (2.12.1)$$

$$\begin{cases} \mathbf{a}_N(t) = \mathbf{m}_t + \mathbf{B}_t [\mathbf{a}_N(t+1) - \mathbf{a}_{t-1}] \\ \mathbf{R}_N(t) = \mathbf{C}_t + \mathbf{B}_t [\mathbf{R}_N(t+1) - \mathbf{R}_{t-1}] \mathbf{B}_t' \end{cases} \quad (t = 1, \dots, N-1)$$

$$\text{with } \mathbf{a}_N(N) = \mathbf{m}_N, \mathbf{R}_N(N) = \mathbf{C}_N \quad (t = N) \quad (2.12.2)$$

と、「固定区間 smoother・アルゴリズム」を適用することにより状態の平滑化（スムージング）が可能となる。

本論では、上記の DLM を、以下の時系列モデルの \mathcal{F}_t -可測のドリフト期待値の推定に適用する。

$$Y_t = T_t \Delta + \sigma_t \sqrt{\Delta} \mathcal{Q}(t) ; Y_0 = 0, \quad (t = 1, \dots, N),$$

$$Y_t = \log r_t - \log r_{t-1}$$

$$\mathcal{Q}(t) \ (t=1, \dots, N) \text{ は, 標準正規確率変数の独立な列.} \quad (2.13)$$

なお、DLM を適用した場合、上記の時系列モデルでは、短期金利 $r(t)$ の同時分布は、対数正規分布、予測分布は、対数 t 分布にしたがうことになる。

2.2. モデルの評価

ここでは、未知母数を含むモデルのベイズ的選択理論を詳しく論じている鈴木 (1987;7 章) を参照し、上述の DLM に基づく 1 期先観測予測モデルの評価について、以下で示すとおりモデルの事後確率を用いることとする。

いま、観測可能な確率変数 $Y = \{Y_t\}_{t=1}^N$ のふるまいを表現する確率モデルを要素とする可算集合 \mathfrak{M} があり、実際の観測結果 $y = (y_1, \dots, y_N)$ に基づき、モデル $M \in \mathfrak{M}$ とその未知母数ベクトル $\theta_M \in \Theta_M$ を条件とする密度関数 $p(y|M, \theta_M)$ に Y が確率的に規定され

るとすると、観測結果 y とモデル M に基づく観測 Y の予測分布 $p(y|M)$ は、

$$Y|M \sim p(y|M) = \int_{\Theta_M} p(y|M, \theta_M) p(\theta_M|M) d\theta_M \quad (2.14)$$

と表される。この予測分布とベイズの定理を用いて、モデル M を真とする確率は、

$$P(M|Y=y) = \frac{p(y|M)P(M)}{\sum_{M \in \mathfrak{M}} p(y|M)P(M)} \quad (2.15)$$

によって計算できる。

いま、モデル $M^{(j)}$ と $M^{(k)}$ の比較に関心があるとする。このとき、各々のモデルを真とする事後確率の比は、

$$\frac{P(M^{(j)}|y)}{P(M^{(k)}|y)} = \frac{p(y|M^{(j)})P(M^{(j)})}{p(y|M^{(k)})P(M^{(k)})} \quad (2.16)$$

となる。もし、 $P(M^{(j)}) = P(M^{(k)})$ と仮定するならば、このモデルの事後確率の比は、以下のとおり単純化され、「ベイズ比 (Bayes factor)」となる。

$$B_{jk} = \frac{p(y|M^{(j)})}{p(y|M^{(k)})} = \frac{\int_{\Theta_M^{(j)}} p(y|M^{(j)}, \theta_M^{(j)}) p(\theta_M^{(j)}|M^{(j)}) d\theta_M^{(j)}}{\int_{\Theta_M^{(k)}} p(y|M^{(k)}, \theta_M^{(k)}) p(\theta_M^{(k)}|M^{(k)}) d\theta_M^{(k)}} \quad (2.17)$$

ここで、あらためて、観測データ y_t からドリフトの 1 期先予測値 f_t を差し引いた残差過程である 1 期先予測誤差 e_t を、互いに独立なマルチンゲール過程ととらえ、以下のとおり定義する¹³。

13 f_t は、1 期前の観測情報 D_{t-1} に基づく状態予測期待値であり、当期の観測データ y_t とは関係性がなく、互いに独立している ($y_t \perp f_t$)。したがって、 $e_t (= y_t - f_t) \perp f_t$ となるので、次頁の (2.18) が定義でき、モデルの密度関数 (2.19) の設定が可能となる。

$$e_t = \begin{cases} 0, & (t = 0) \\ y_t - f_t, & (t = 1, \dots, N) \end{cases} \quad (2.18)$$

ここに、

$$e_t | D_{t-1} \sim t(0, Q_t^*; \nu_{t-1}) \quad (t = 1, \dots, N),$$

さらに、本論の上位モデルを M とおき、このモデルの未知母数ベクトルを以下のとおり定めると、

$$\theta_M = (m_0, C_0, \nu_0, d_0, \zeta_t^2, \alpha, \beta)$$

θ_M に基づく、 N 期までの観測結果 $y = (y_1, \dots, y_N)$ を踏まえたモデルの密度関数について、以下のとおり表現できる¹⁴。

$$p(y|M, \theta_M) = \prod_{t=1}^N \left\{ \frac{\Gamma\left(\frac{\nu_{t-1}+1}{2}\right)}{\Gamma\left(\frac{\nu_{t-1}}{2}\right) \sqrt{\pi \nu_{t-1} Q_t^*}} \left(1 + \frac{(y_t - f_t)^2}{\nu_{t-1} Q_t^*}\right)^{-\frac{\nu_{t-1}+1}{2}} \right\}, \quad \nu_0 > 2 \quad (2.19)$$

モデル母数のうち、 $(m_0, C_0, \nu_0, d_0, \zeta_t^2)$ は、DLM の超母数 (hyper parameters) であり、これらが与えられれば、随伴母数 (incidental parameters) である $(m_1, \dots, m_N, C_1, \dots, C_N)$ については、容易に計算することができる。

一方、構造母数 (structural parameters) である (α, β) は、上位モデルに組み込まれた下位モデルの隠れ母数であり、このままでは確率密度の計算は困難である。

さらに、一般的にみて、 (α, β) が隠れ母数であるために、(2.17) 式の積分を行うことは難しい。そこで、以下の 2.3 節で示すシミュレーションを行う必要がある。

¹⁴ 2.1 節でみたとおり、状態誤差分散 ζ_t^2 は、 $\zeta_t^2 = \delta_t \sigma_{t-1}^2$ と、1 期前の観測誤差分散 σ_{t-1}^2 に、情報割引率 δ_t を掛けた値として定義される。ここでは、状態誤差分散 ζ_t^2 について時間を通じて一定と仮定し、これを直接的に推定することにより、併せて情報割引率 δ_t を捕捉する。すなわち、情報割引率 δ_t は、未知ではあるが確定的な時変変数と考え、 δ_t の値は、常に状態誤差分散 ζ_t^2 が一定の値に維持されるように決定されるものとする。

2.3. データ加増アルゴリズムによるMCMCの概要

モデルの母数が未知の場合、当該未知母数について予測分布を用いて効率的な推定を行う場合には、Gelman, Carlin, Stern, and Rubin (1995, 2004) などで紹介されている「Data Augmentation (データ加増) アルゴリズム (以下、DA アルゴリズム)」を導入する必要がある。

DA アルゴリズムは、潜在変数(latent variables)などを人工的に発生させ、母数の事後分布の形状等をすべて評価するシミュレーション方法である。DA アルゴリズムは次の2つのステップからなる。なお、潜在的な観測されない確率変数を Z とおき、その仮定または予想されるデータを z と表記する¹⁵。

- ・抽出ステップ (sampling step) : 観測結果 $Y = y$ に基づく、欠損値、観測されない確率変数など、予想される確率変数 Z の予測分布 $p(z|y)$ から N 個の乱数 z_1, z_2, \dots, z_N を派生させる。
- ・事後ステップ (posterior step) : 加増データ (augmented data) : 実際の観測結果 y と抽出ステップで得られた乱数 z をもとにモンテカルロ法によって、未知母数の条件付き事後分布 $\theta | (Y = y) \sim p(\theta|y)$ を近似する。

なお、抽出ステップにおいては、予想される観測予測値を z とすれば、事後予測分布 $p(z|y)$ は

$$Z|(Y = y) \sim p(z|y) = \int p(z|\theta, y)p(\theta|y)d\theta \quad (2.20)$$

であり、未知母数の条件付き分布 $p(\theta|y)$ について既知であることが前提となる。

¹⁵ 観測されない潜在的な確率変数 Z については、欠損値、母数や超母数の他、場合によっては、人工的に仮定されるデータなど、統計的分析の対象によって様々な設定が可能である。本論の DA アルゴリズムでは、すでに明らかになっているとおり、潜在的な確率変数は、観測予測値を元に得られる1期先予測誤差と、この1期先予測誤差を用いて階層モデル(組み込まれた「不均一分散モデル」)により生成される隠れ母数 (α, β) である。なお、 (α, β) を含む構造母数から随伴母数 $(m_1, \dots, m_N, C_1, \dots, C_N)$ が求められ、さらに、随伴母数が変更されると DLM のモデル構造により、事後的に観測予測値および1期先予測誤差が修正されて推定の手順が循環する。

その具体的な手順は以下のとおりである。

- ① 第 i 回目の $p(\theta|y)$ の近似を $p^{(i)}(\theta|y)$ とする。 $p^{(i)}(\theta|y)$ から、乱数 θ^* を発生させる。
- ② θ^* を所与として、 $p(z|\theta^*, y)$ から、乱数 z^* を発生させる。
- ①, ② を繰り返すことによって $p(z|y)$ から、任意の個数の乱数 z^* を発生させる。

なお、この z^* は、Gelman, Carlin, Stern, and Rubin (1995) の観測される確率変数 Y の複製 y^{rep} と同じよう取り扱われる。

この場合、 $p(\theta|y)$ の積分は次式の近似に基づく。

$$p(\theta|y) = \int p(\theta|y, z)p(z|y)dz \quad (2.21)$$

一般的には、上の $p(\theta|y)$ の積分について z^* を用いた MCMC により近似する。

このような DA アルゴリズムは分布全体を示すことができるので、各未知母数の周辺分布などについて、より確実に現実のデータに即した形で把握することが可能となる。

なお、本論ではモデル M が介在することにより、 $p(z|y)$ を $p(y|M)$ に置き換えて以下の手順を繰り返すことにより、DA アルゴリズムに基づく未知母数の条件付き事後分布を把握することができる。ただし、 N 個の観測があるものとする。

- 1) 各未知母数の初期分布を設定する。
- 2) 所与の観測結果 $y = (y_1, \dots, y_N)$ に基づき、(2.8.2 式) により、 f_t を求める。
- 3) 観測データ y_t から f_t を差し引いた残差過程 (1 期先予測誤差) e_t から階層モデル (組み込まれた「不均一分散モデル」) により、 α, β の候補を推定する。
- 4) 下位モデルにより Q_t^* の候補を求める。
- 5) 上記の 1) ~ 4) までの過程で得られた母数の採択または棄却について、モデルの事後確率に基づいて決定する。

具体的には、モデルの未知母数ベクトルの候補を $\theta_M^{(j)}, \theta_M^{(k)}$ とし、以下の評価関数 $\lambda(\theta_M^{(j)}, \theta_M^{(k)})$ をおき、

$$\lambda(\theta_M^{(j)}, \theta_M^{(k)}) = \min\{\gamma_{jk}, 1\}$$

$$\text{ここに、} \quad \gamma_{jk} = \frac{p(y|M^{(j)}, \theta_M^{(j)})p(\theta_M^{(j)}|M^{(j)})}{p(y|M^{(k)}, \theta_M^{(k)})p(\theta_M^{(k)}|M^{(k)})} \quad (2.22)$$

[0,1] 間の一様乱数 U を発生させ、

$$U \leq \lambda(\theta_M^{(j)}, \theta_M^{(k)})$$

ならば、モデルの未知母数 θ_M を更新し、当該条件を満たさない場合には、 $\theta_M^{(j)} = \theta_M^{(k)}$ とおく。

- 6) 採択された母数推定値による観測 Y の予測分布に基づき、レプリケーション y^{rep} を生成する。
- 7) レプリケーション y^{rep} を新たな観測結果とみなして 2) に戻る。

以下、未知母数 α, β に関するマルコフ過程について収束（過程の既約性と非周期性）が認められる状態になるまで、2) ~ 7) のステップを繰り返す。

以下では、上記のステップのうち、いくつかのポイントについて詳解する。

2.4. EMアルゴリズムによる母数初期期待値の推定

本論のモデルでは状態の次元 h は 1 である。初期に設定する必要のある超母数 (hyper parameters) は、ドリフト分布に係る未知母数 m_0 , C_0 と d_0 , 状態誤差の分布母数 $\zeta_t^{2(0)}$, 観測誤差分散の予測モデルに係る隠れ構造母数の初期値 $a^{(0)}, \beta^{(0)}$ である。また、期首の自由度 ν_0 についても設定する必要がある。

ここでは、EM アルゴリズムに準じた方法で初期時点での母数期待値を決定する。EM アルゴリズムは、欠損値、観測されない確率変数などに関するモデルの未知母数を推計する方法として広く用いられているものである。

本モデルでは、隠れ母数を含む構造母数 (structural parameters) と随伴母数 (incidental parameters) とが互いに階層的に組み込まれたモデル構成をとっていることが、EM アルゴリズムを用いる根拠となっている。

その具体的な手順は以下のとおりである。なお、前述のとおり、潜在的な観測されな

い確率変数を Z とおき、その仮定または予想されるデータを z と表記する。

まず、E ステップで $E\{\log p(\theta_M|z, y)\}$ を計算する。具体的には、 i 時点で最良のモデル未知母数行列 $\hat{\theta}_M^i$ の下で以下を求める。

$$E\{\log p(\theta_M|z, y)\} = \int_Z \log p(\theta_M|z, y) p(z|y, \theta_M^i) dz \quad (2.23)$$

次に、M ステップで、事後周辺密度の期待値に関して、 $E\{\log p(\theta_M|z, y)\} = H(\theta_M|\theta_M^i)$ とおき、 $H(\theta_M|\theta_M^i)$ を最大化する θ_M を $\hat{\theta}_M^{i+1}$ とする。その後、 θ_M が収束するまで、E ステップと M ステップを繰り返す。

本論でのモデル分析においては、モデル未知母数 θ_M とモデル M に基づいて得られる各モーメント（状態期待値および状態分散等）について、初期時点では欠損値 Z とみなすことができるので、直接的には、モデル未知母数 θ_M の暫定的推定値 $\hat{\theta}_M^i$ を条件とする各モーメント¹⁶ の実現値を z と表記すれば、モデル母数の対数周辺事後密度の期待値について、以下のとおり示すことが出来る¹⁷。

$$E\{\log p(\theta_M^i|M, y)\} \propto \int_Z \log p(y|M, z) p(z|M, \theta_M^i) p(\theta_M^i|M) dz \quad (2.24)$$

具体的なモデル母数推定は、事後周辺密度の期待値 $E\{\log p(\theta_M^i|M, y)\}$ を最大化する θ_M を

16 ここでは、各時点（過去・現在・将来）における観測、状態、および、分散の確率密度から求める各々の母数期待値について各モーメントと総称している。

なお、松本、石黒、乾、田邊(2004)で詳しく紹介されている様に、階層ベイズ法は、柔軟なモデリングのための道具として、時系列の解析と予測、確率密度の推定、画像再構成、視覚の認知科学などの様々な分野で注目を浴びている。彼らの記述によれば、従来の統計モデルでは、データ $Y=y$ について母数 $X=x$ で説明する確率モデル $P(y|x)$ を考えるとき、条件となる母数 x の独立な要素をなるべく少なくするのが基本であった。母数が多いほうが、空間的・時間的な変動や統計的分析対象の属性を表現するには有利であるが、単に母数を増やしたのでは、観測データの背後にある法則性を表現する能力が低下し、未知のことが予測困難となるためである。これに対して、階層ベイズ法では、例えば、 x として、各時刻でのシステムの状態などをそのまま使う。その一方で、 x についての事前の知識を事前分布 $P(x|\theta)$ で表現し、ベイズの公式（ベイズの定理）によって求めた事後分布 $P(x|y) \propto P(y|x)P(x|\theta)$ によって推論を行う。ここに、 θ は事前知識の内容やその信頼度などをあらわす母数であり、超母数と呼ばれる。本論の DLM も、各時刻でのシステムの状態などを取り扱う階層ベイズ法に基づくモデル分析である。

17 欠損値等とおく Z については、統計的分析の対象として任意に設定可能である。EM アルゴリズムでは、各モーメントについて欠損値とおいて推定を行っている。

求めて $\hat{\theta}_M^{i+1}$ と更新し、その後、 θ_M について望みの精度が得られるまで同じ手順を繰り返すことにより行われる¹⁸。

本論では、統計的推論の端緒となるモデル母数 θ_M^* を推定するために、この様な EM アルゴリズムを適用し、状態および精度（観測誤差分散母数の逆数）の初期密度

$$\begin{aligned} T_0 | D_0 &\sim t(\mathbf{m}_0, \mathbf{C}_0, h, \nu_0) \\ \sigma_0^{-2} | D_0 &\sim \text{Gamma}(\frac{\nu_0}{2}, \frac{d_0}{2}) \end{aligned}$$

を求めている。

なお、本論の隠れ母数 (α, β) は、その母数空間（制約）に適合した範囲内で設定される限りにおいて、DLM の理論的な体系に変更をもたらすことはなく、再帰的に各モーメントを推定する過程に明示的には影響しない。このため、状態および分散母数の初期密度を求める過程で、隠れ母数 (α, β) の点推定を行うことは可能である¹⁹。

18 θ_M の最適な候補を θ_M^* と表記して $\theta_M = \theta_M^*$ とおき、 θ_M が与えられた後の各モーメントを x で表せば、 θ_M について既知であるとの仮定に基づき、 (x, θ_M) と空間を各モーメントまで拡張した事後同時密度について、以下のとおり示すことができる。

$$p(x, \theta_M | M, y) \propto p(y | M, x) p(x | M, \theta_M^*) p(\theta_M^* | M)$$

これを踏まえると、改めて、

$$p(\theta_M | M, y) \propto \int_X p(y | M, x) p(x | M, \theta_M^*) p(\theta_M^* | M) dx$$

と表現できるが、 θ_M が与えられた下では、DLM モデル M の理論的体系を媒介して各モーメントを計算できるため、密度関数 $p(y | M, \theta_M^*)$ について、次のとおり表記可能である。

$$p(y | M, \theta_M^*) = \int_X p(y | M, x) p(x | M, \theta_M^*) dx$$

ここに、 $p(y | M, \theta_M)$ を自然対数変換した $\log p(y | M, \theta_M)$ をモデル M の対数尤度関数とみなせば、モデル母数 θ_M の対数事後密度 $\log p(\theta_M | M, y)$ について、結局、

$$\log p(\theta_M | M, y) \propto \log p(y | M, \theta_M^*) + \log p(\theta_M^* | M)$$

と示され、具体的なモデル母数の推定は、対数事後密度 $\log p(\theta_M | M, y)$ を θ_M^* を求めることに帰着する。これは、Carlin and Louis (2000), Gelman, Carlin, Stern and Rubin (2004) などで論じられている「階層モデル」による推定方法と基本的に同じである。

19 この様な特徴を持っていることが、 (α, β) が隠れ変数と呼ばれる理由の一つである。

ただし、EM アルゴリズムで得られた隠れ母数 (α, β) の点推定値は、そのままでは、統計的にみて十分に意義のある結果を得ることは難しい。その一方で、隠れ母数 (α, β) は、本論の最終的な目的である任意の確率過程の確率測度変換で重要な役割を果たす不均一分散モデルの根幹をなす母数であり、より詳細な統計的分析が求められる。

隠れ母数 (α, β) の推定値は妥当なものか、その信用区間はどのような範囲か、母数空間を前提とすれば隠れ母数 (α, β) の分布は切断分布となるが、それはどのような形状をとるか、といった、より詳しい推論のためには、隠れ母数 (α, β) の事後確率密度を把握する必要がある。この目的のために、本論では、EM アルゴリズムによる推定結果を端緒とし、あらためて、DA アルゴリズムによるシミュレーションを導入する。

2.5. 下位モデル母数候補の推定

ここでは、すでに (7) 式で示した下位モデルについて、その母数候補の推計を行う。

$$\begin{aligned} Q_t^* &= \alpha e_{t-1}^2 + \beta Q_{t-1}^*, & (t = 2, \dots, N) \\ \text{with } Q_t^* &\equiv Q_1 & (t = 1). \end{aligned}$$

なお、 α, β は、前にみたとおり 1 期先予測モデル（上位モデル）の Q_t^* を規定する隠れ母数である。

まず、母数候補派生のモデルとして以下の Bollerslev 近似式を提示する。

$$e_t^2 = \begin{cases} Q_1^* + \varepsilon_1, & (t = 1) \\ (\alpha + \beta)e_{t-1}^2 + \varepsilon_t - \beta\varepsilon_{t-1}, & (t = 2, \dots, N) \end{cases} \quad (2.25)$$

ここに、

$$\begin{aligned} e_t &= y_t - f_t, \\ e_t &= 0, \quad \varepsilon_t = 0 \quad \text{for } t < 1, \\ \alpha &= 0, \quad \beta = 0, \quad \text{for } t \leq 1, \\ \alpha &\geq 0, \quad \beta \geq 0, \quad \alpha + \beta < 1, \quad \text{for } t > 1, \\ \varepsilon_t &\stackrel{\text{def}}{=} e_t^2 - Q_t^* \quad (t = 1, \dots, N) \\ \varepsilon_t &= \left(\left(\frac{e_t^2}{Q_t^*} \right) - 1 \right) Q_t^* = (z_t - 1) Q_t^*. \end{aligned} \quad (2.26)$$

$$p(z_t|D_{t-1}) \approx \chi^2(1)$$

$$\text{and } E[\varepsilon_t|D_{t-1}] \approx 0, \text{Var}[\varepsilon_t|D_{t-1}] \approx 2Q_t^{*2} \quad (t = 1, \dots, N).$$

ただし、ここでは、効率的に母数を派生させ、円滑に MCMC を行うことを目的とし、 ε_t の密度について Nakatsuma (2000) と同様に、

$$\varepsilon_t | D_{t-1} \sim N(0, 2Q_t^{*2}) \quad (t = 1, \dots, N)$$

と仮定する。

このように仮定することで、回帰分析による母数派生が可能となる。実際に、Nakatsuma (2000) では、未知母数の暫定的な期待値を派生させる段階で回帰分析を適用している。そして、派生された母数候補を最終的に採択するさいの評価関数として、GARCH (1,1) 型モデルのモデル分布 (model distribution) に基づく事後密度比を用いている。

具体的には、付録：A.2. 「2章」の補論：a.2.6. および a.2.7. で詳しく示されているとおり、この仮定に基づき、未知母数 α の提案分布の密度関数について、次のとおり設定でき、

$$f(\alpha | Y_\alpha, X_\alpha, \Sigma, \beta) \propto \varphi(\alpha | \hat{\phi}_\alpha, \hat{\Sigma}_\alpha) I(\alpha | [0, 1 - \beta])$$

ここに、 $\varphi(\cdot | \hat{\phi}_\alpha, \hat{\Sigma}_\alpha)$ は、平均 $\hat{\phi}_\alpha$ 分散行列 $\hat{\Sigma}_\alpha$ の正規密度、 $I(\cdot | [0, 1 - \beta])$ は、区間 $(0 \leq \alpha < (1 - \beta))$ の指標関数

同様にして、未知母数 β の提案分布についても、以下のとおり、おくことができる。

$$f(\beta | Y_\beta, X_\beta, \Sigma, \alpha) \propto \varphi(\beta | \hat{\phi}_\beta, \hat{\Sigma}_\beta) I(\beta | [0, 1 - \alpha])$$

ただし、本論のモデルでは、Nakatsuma (2000) とは異なり、逐次、観測のレプリケーション y^{rep} を派生し、 y^{rep} を基にして推定された母数候補を最終的に採択するさいの評価関数に、モデルの事後確率比の概念に基づく準ベイズ比 ((2.22) 式) を当てはめることにより、理論的な整合性を保持する。

3. 実証分析

3.1. 実際の金利データに基づく実証

ここでは、実際に、自然対数値に変換した「米国財務省証券6ヶ月物金利月次データ (U.S.Treasury bills 6-month rate (monthly))」の change data (1階階差) を用いた金利モデル ((2.13) 式) の考察を行う。なお、6ヶ月は、固定金利債券 (fixed income security) の 利金支払い等やこれに付随する金利スワップ取引の決済のタイミングに合致しており、一定に金利等の金融指標の動向を予測する意義のある期間である。

データの観測期間は1984年8月～2001年11月であり、データの推移は図3-1、図3-2 のとおりである。



図3-1 米国財務省証券6ヶ月物金利(対数変換データ)原系列の推移

[データは月次(1984.8～2001.11)]

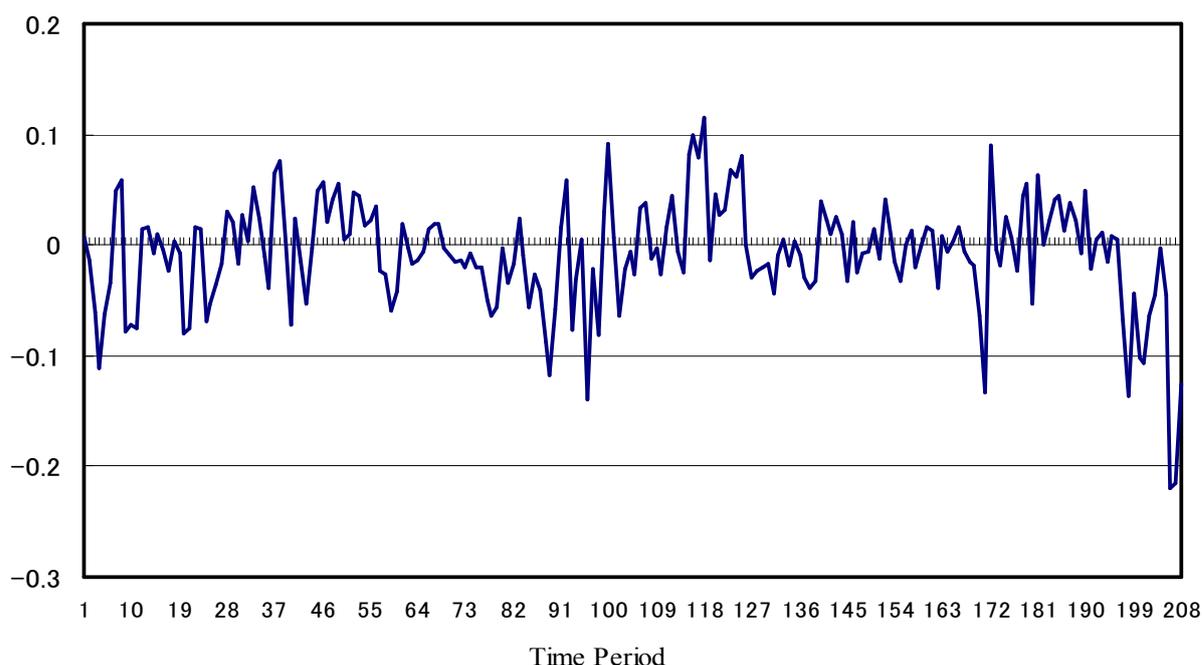


図3-2 米国財務省証券6ヶ月物金利(対数変換後の階差データ)原系列の推移
[データは月次(1984.8~2001.11, $N=208$)]

対数変換を施した原データについて、記述統計的考察を行うと以下のとおりである。対数変換された金利水準、金利変化データともに、尖度、歪度、および、Kolmogorov-Smirnov test の検定結果から総合的に判断して、分布について正規性を認めることが困難であり、とくに、本論の分析対象となる対数金利変化データでは尖度が大きく、「裾の広い分布」であることが想定される。

表3-1 Summary Statistics

T.B. 6-m rate (8/1984 ~ 11/2001 : monthly)	金利水準 (Level) $\log r_t$	金利変化 (Change) $\log r_t - \log r_{t-1}$
観測データ数 (N)	209 (7/1984 データを含む)	208
最大値	2.362739	0.115069
最小値	0.631272	-0.220109
平均値	1.686422	-0.008244
標準偏差	0.303618	0.049003
尖 度	3.346463	5.626126
歪 度	-0.520243	-0.903074
Kolmogorov -Smirnov test	$D = 0.106195$ ($0.01 < p\text{-value} < 0.05$)	$D = 0.097157$ ($0.01 < p\text{-value} < 0.05$)

さて、ここでは、DLMにおける状態の次元 h は 1、さらに、 $F = G = 1$ とおく。
 これらを踏まえて、まず超母数について、

$$m_0 = 0, C_0 = 0.0001, \nu_0 = 3, d_0 = 3 (S_0 = 1), \\ \zeta^{2(0)} = 0, \alpha^{(0)} = 0.15, \beta^{(0)} = 0.84$$

と設定する。期首自由度については、当該分布形が t 分布であることから、1 次および 2 次のモーメントが存在し、かつ、モデルの取り扱いが容易となる最小の自由度として 3 と設定する。ただし、母数 α, β については、前述のとおり、 $\alpha \geq 0, \beta \geq 0, \alpha + \beta < 1$ との制約が適用される。これらの母数の組み合わせによるモデルの周辺対数尤度は 305.125 である。

なお、前述のとおり、状態誤差分散 ζ_t^2 については、時間を通じて定数とし、これを直接推定することにより、併せて情報割引率 δ_t を捕捉する。すなわち、ここでは、情報割引率 δ_t は、未知ではあるが確定的な時変変数と考え、情報割引率の値は、常に状態誤差分散 ζ_t^2 が一定の値に維持されるように決定されるものとする。

次に、EM アルゴリズムに準じた推計を行い、構造母数 (structural parameters) の推定値として、

$$\hat{\zeta}^2 = 0.00013434, \hat{\alpha} = 0.147283, \hat{\beta} = 0.850207$$

を得る。

また、同時に随伴母数 (incidental parameters) についても更新がなされ、所与の観測に基づく平滑化ドリフト期待値について図 3-3 のとおり示すことができる。この結果、モデルの周辺対数尤度は 356.824 に上昇する。

なお、当該ドリフト期待値については、前述のとおり Øksendal (1998, chap.6) が論じている可算加法族 \mathcal{F}_t -可測の期待値: $E(\mu_t | \mathcal{F}_t)$ とみなすことができる。

さらに、本モデルの 1 期先観測予測誤差 $e_t = y_t - f_t$ ($t = 1, \dots, N$) についても図 3-4 のとおり表示される。この過程は互いに独立で、初期値、期待値がともに概ね確実にゼロとなるマルチンゲール過程となっている。

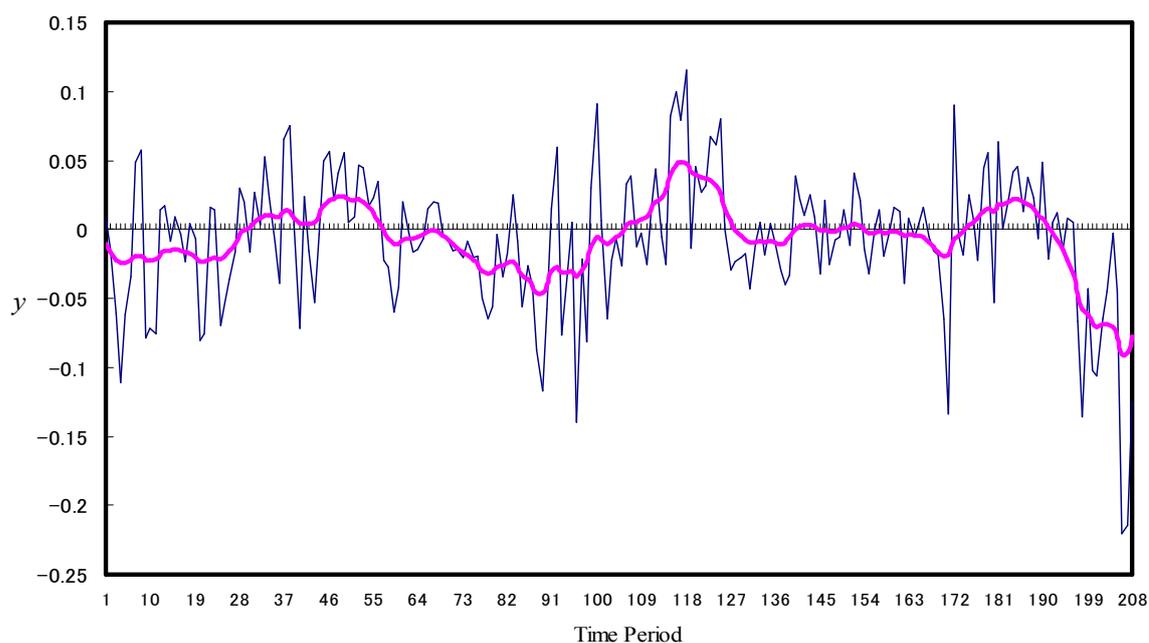


図3-3 EMアルゴリズムに基づく平滑化ドリフト推定結果

(細線:対数金利変化データ、太線:ドリフト推定値)

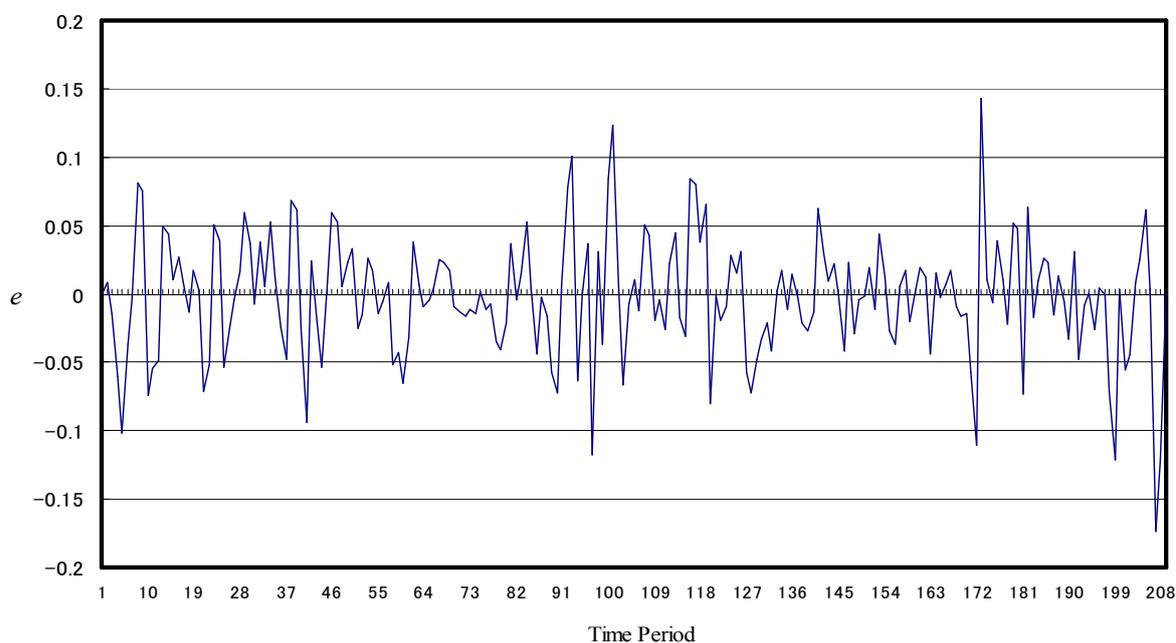


図3-4 1期先予測誤差推移

ここで、上の母数期待値を固定して、2.1節で示したDLMによるベイズ予測で求まる1期先観測予測分布にしたがい、対数金利一階階差データのレプリケーションを50本例

示すと以下の図のとおりである。

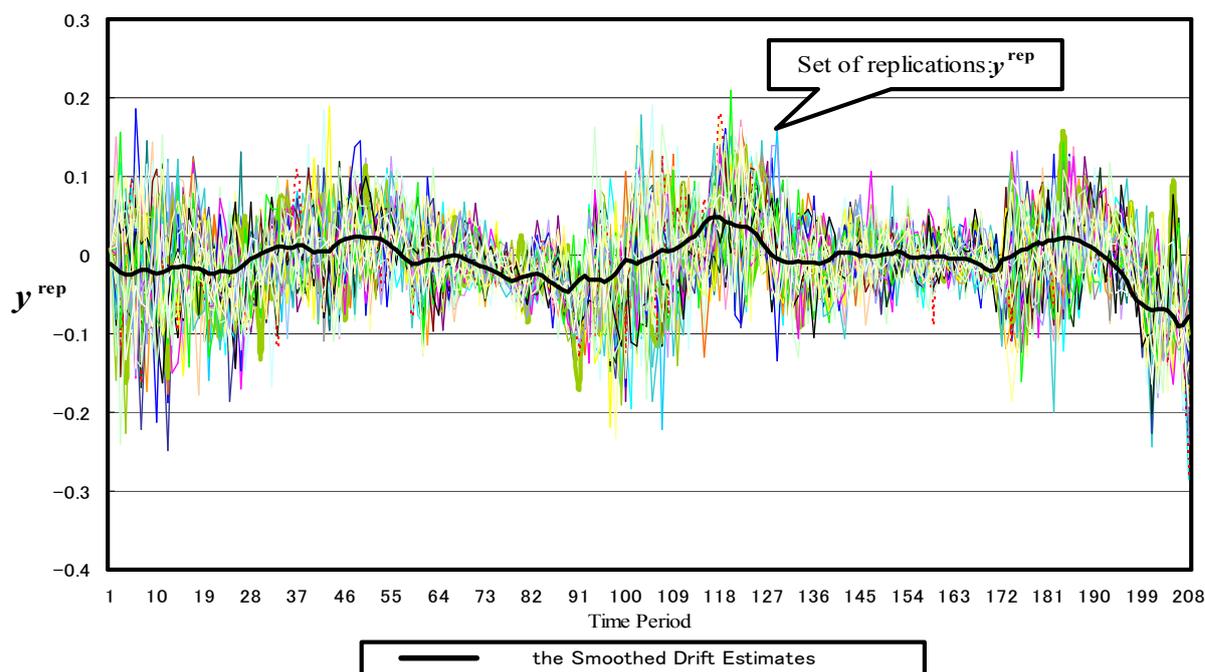


図3-5 観測レプリケーションの例示

これらのレプリケーションは、モデルの 1 期先観測予測分布にしたがい発生させた観測予測値である。これらを新たな観測とみなしてドリフトおよび拡散係数の推定を繰り返し行うことが本論におけるデータ加増型シミュレーションの骨子となる。

ただし、上の図では例示として、EM アルゴリズムで求められた 1 期先観測予測モデルの未知母数推定値（初期期待値）を固定して同一の予測モデルから 50 本のレプリケーションを発生させたが、本論における分析は、将来のある一定期間にわたって固定された同一のモデル分布にしたがうデータが繰り返し観測できることを前提とした未知母数の推論方法を考察するものではない。

本論のレプリケーションは同一モデル分布にしたがう観測の複製ではなく、あくまでもベイズ予測に基づいて生成される、実際には観測することが不可能な予測値である。そして、本論の立場では、レプリケーションのふるまいがドリフト状態などの未知母数ベクトル θ_M の推定に影響し、結果として、データ発生分布 (data generation distribution) となる 1 期先観測予測分布密度自体も逐次変化すると考えるのである²⁰。

20 本論のシミュレーションでは、観測予測とモデル母数推定とが再帰的に行われる。

次に、実際にレプリケーションを新たな観測とみなして、準ベイズ比（(2.22)式）を評価関数とおくMCMCを行い、以下の図3-6および図3-7のとおり、下位モデル母数 α, β について採択された母数候補に関するサンプル・パスを得た。

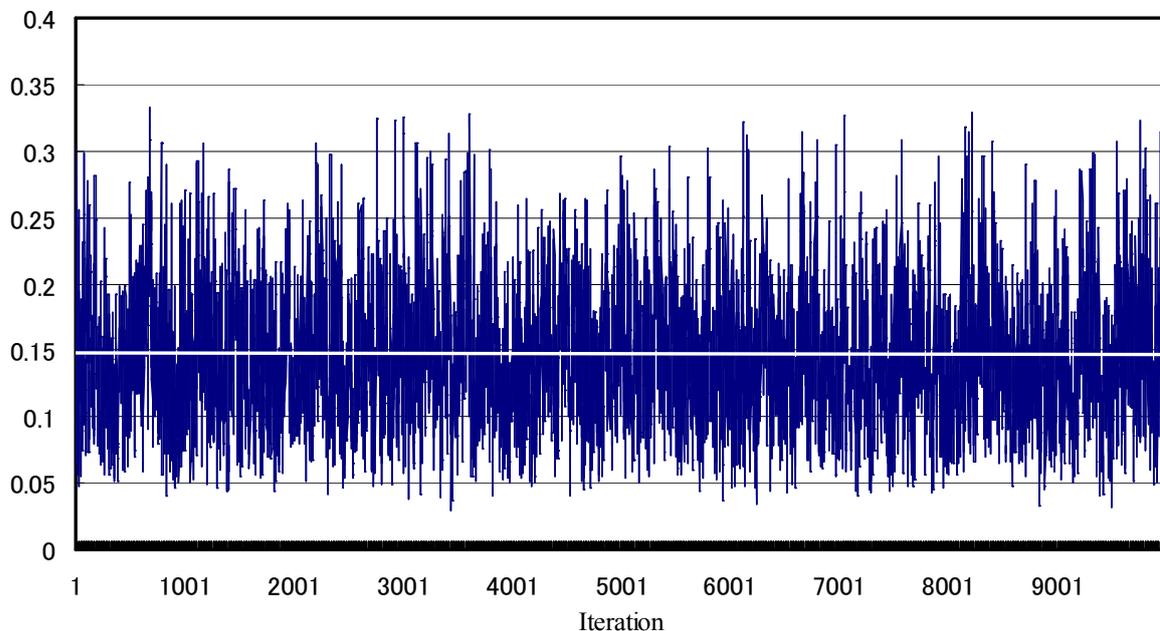


図3-6 未知母数 α に係るサンプル・パス

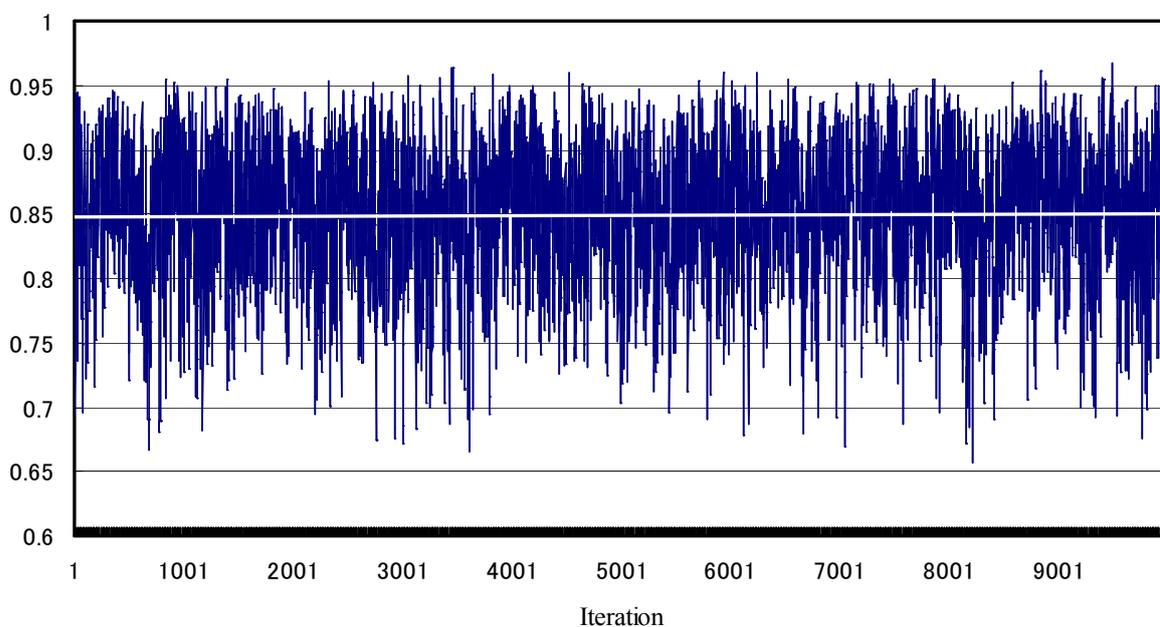


図3-7 未知母数 β に係るサンプル・パス

なお、ここでのシミュレーション回数は 20,000 回である。図 3-6 および図 3-7 は、いずれも初期からの 10,000 回の結果を除いた残りの 10,000 回のサンプル・パスであり、表 3-2 で示されているとおり Geweke の CD 検定により収束を確認している。実際に、計算回数毎に求まる平均値の逐次的な推移をみると、以下の図 3-8 および図 3-9 のとおり、10,000 回以降でほぼ一定値となり、収束していることが分かる。

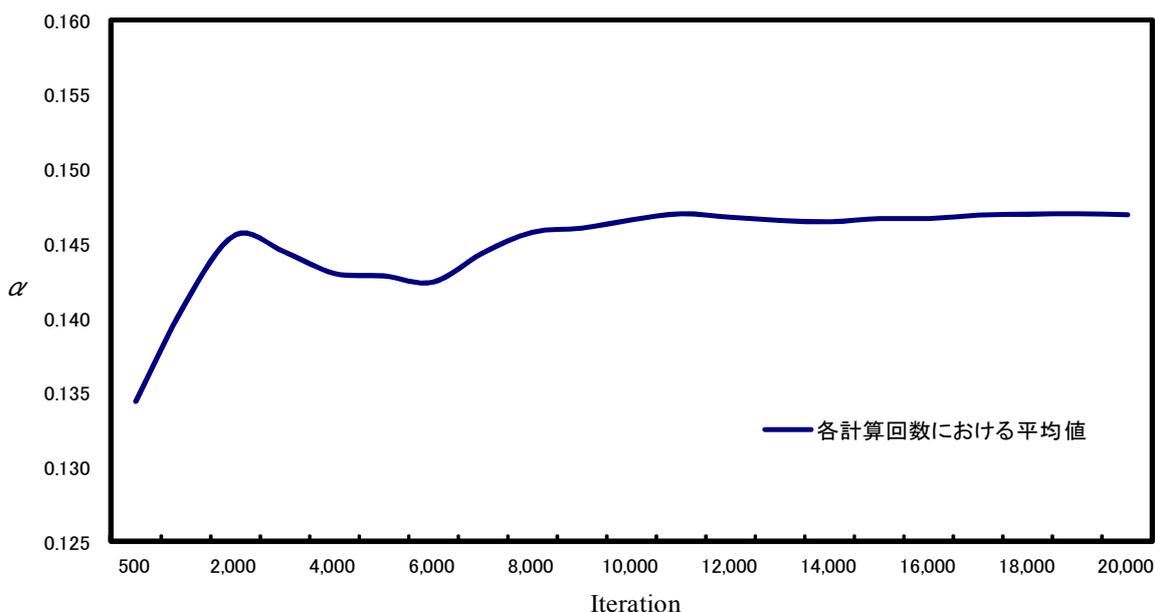


図3-8 未知母数 α に係る計算回数毎の平均値推移

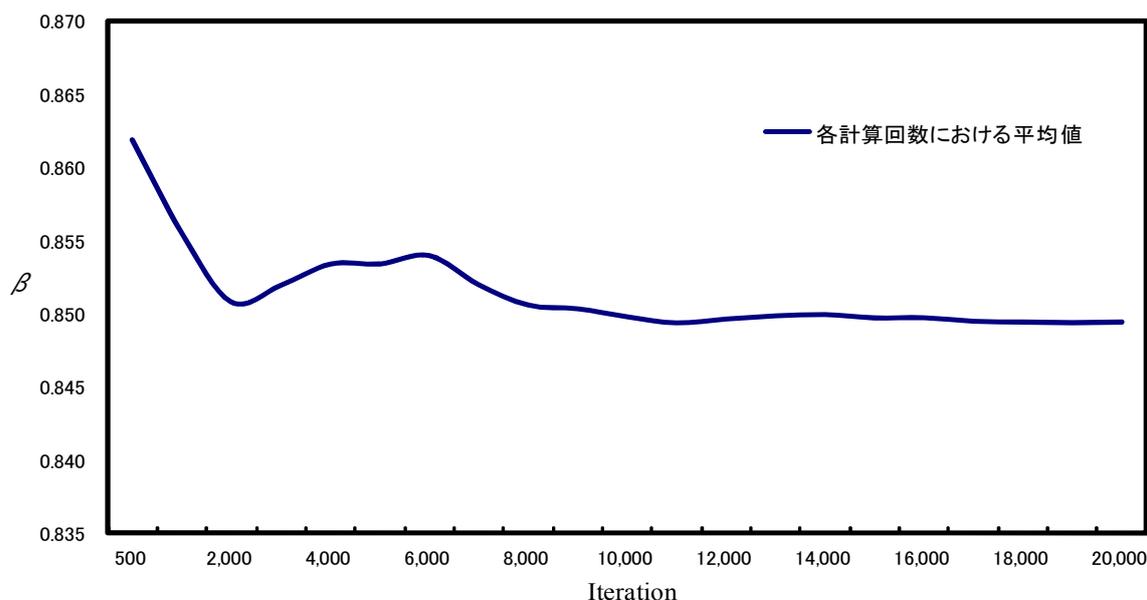


図3-9 未知母数 β に係る計算回数毎の平均値推移

さらに、サンプル・パスに基づき、ヒストグラムを描くと以下の図 3-10、図 3-11 のとおりである。

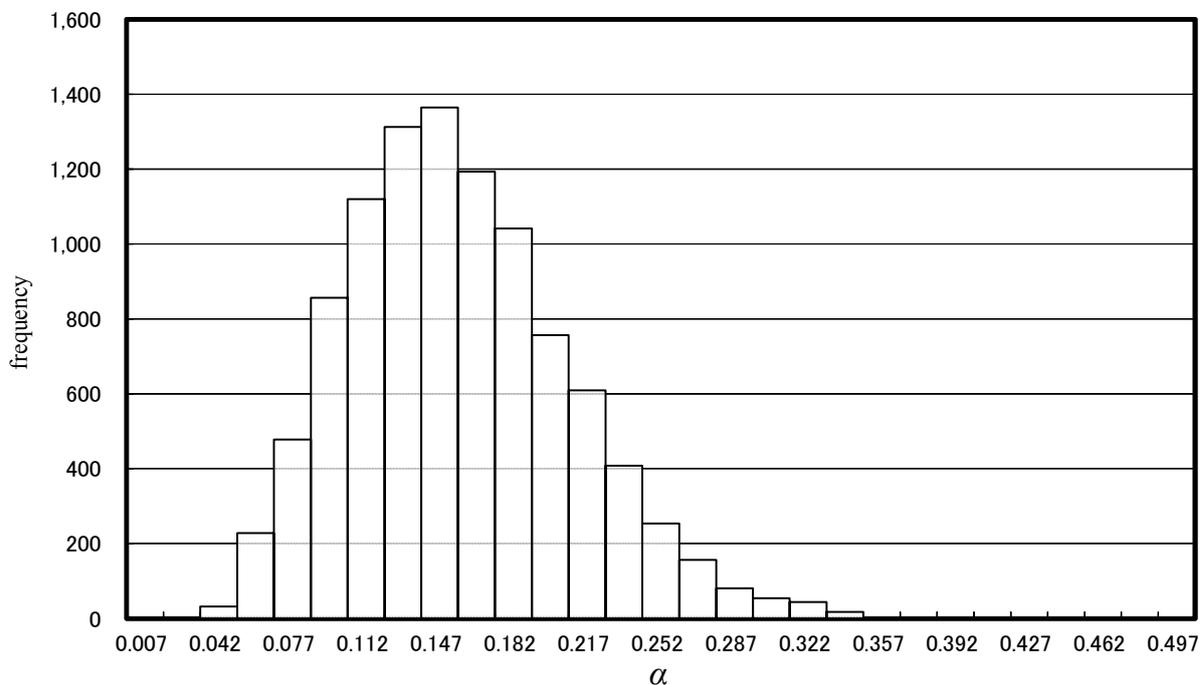


図3-10 未知母数 α に係るサンプル・パスのヒストグラム

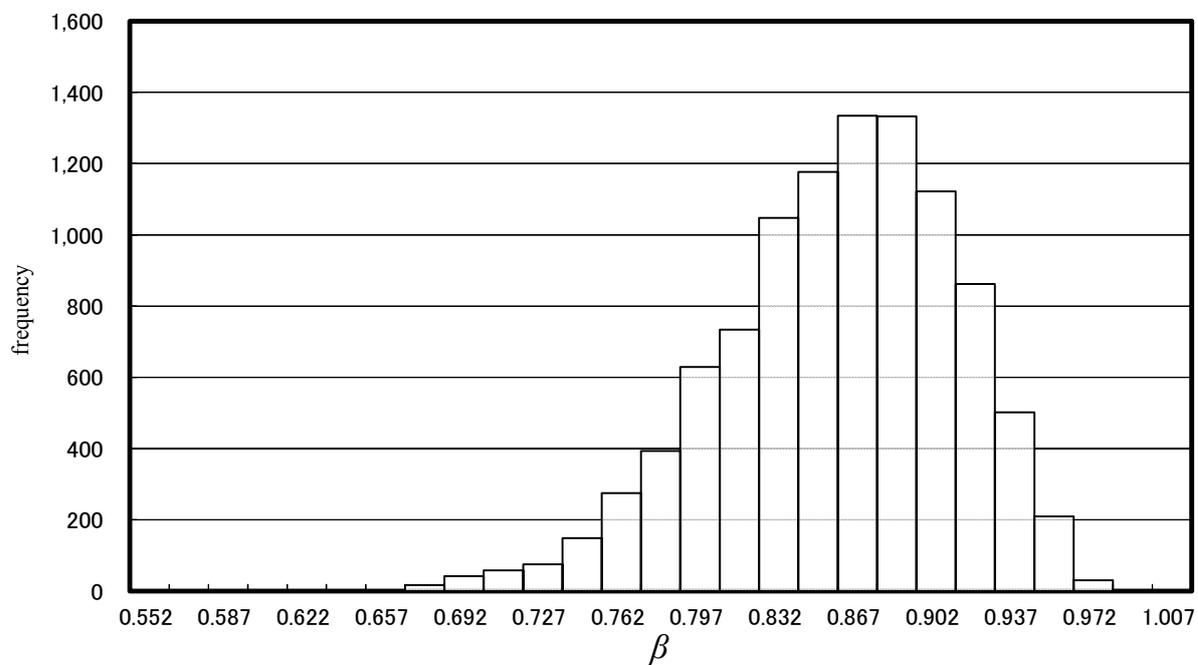


図3-11 未知母数 β に係るサンプル・パスのヒストグラム

図は、いずれも、DLM に階層的に組み込まれたモデルの隠れた未知母数 α 、 β に関する推定事後周辺分布を離散的に表示したグラフとみなすことができる。

あらためて、サンプル・パスの平均、標準偏差を求めると以下の表のとおり。

表3-2 下位モデル未知母数推定結果
(Posterior statistics)

母数	係数	標準誤差	C D	$\hat{\rho}$
α	0.147294	0.053409	-0.111906	0.657958
β	0.849097	0.052418	0.052687	0.665062

Note (1) CD は、Geweke の「convergence diagnostic 統計量」、ここでは、10,000 個のサンプルを前半の 1,000 個、後半の 5,000 個に分割してそれぞれ平均、分散を求め、これに基づき CD を計算している。
(2) $\hat{\rho}$ は、サンプルパスの 1 階自己相関係数推定値である。

シミュレーションで事後的に得られた上記の係数期待値、標準誤差に基づき、以下のとおり、隠れ未知母数 α, β の事後同時密度を示すことができる。

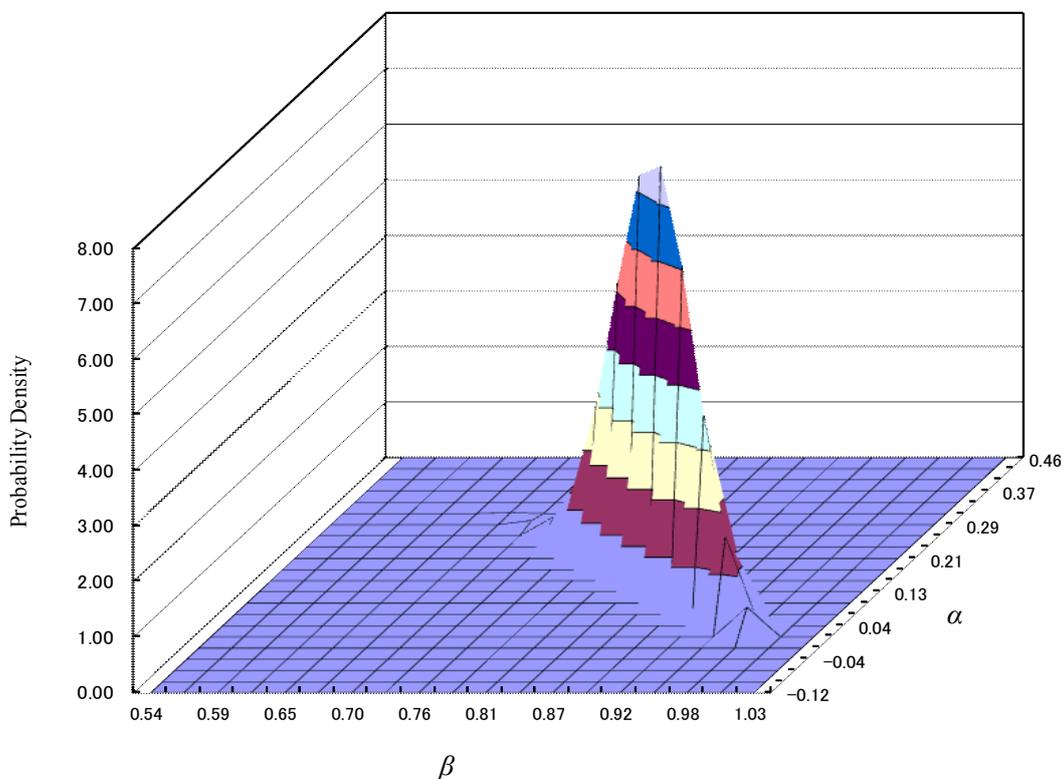


図3-12 未知母数 α, β の事後同時密度

3.2. 前向き推論(Forward Induction)

ここでは、 $N-13$ 期のデータ集合に基づく母数推定値を固定（所与）とし、 $N-12$ 期から N 期までの金利データについて予測分布の推移と実績値とを比較して、実際に多期間にわたる予測の適合性を検証する。

以下の図は、2001 年 11 月期を N 期とおき、2000 年 11 期 ($N-13$ 期) のデータ集合に基づいて推定されたモデル未知母数推定値を固定（所与）とし、2000 年 12 月期 ($N-12$ 期) から 2001 年 11 月期 (N 期) までの 1 年間にわたる 1 期先観測予測分布（対数 t 分布）の推移についてモンテカルロ・シミュレーションを行い、将来的な金利過程を 100 本のサンプル・パスによって表示して実際の金利推移と比較したものである。

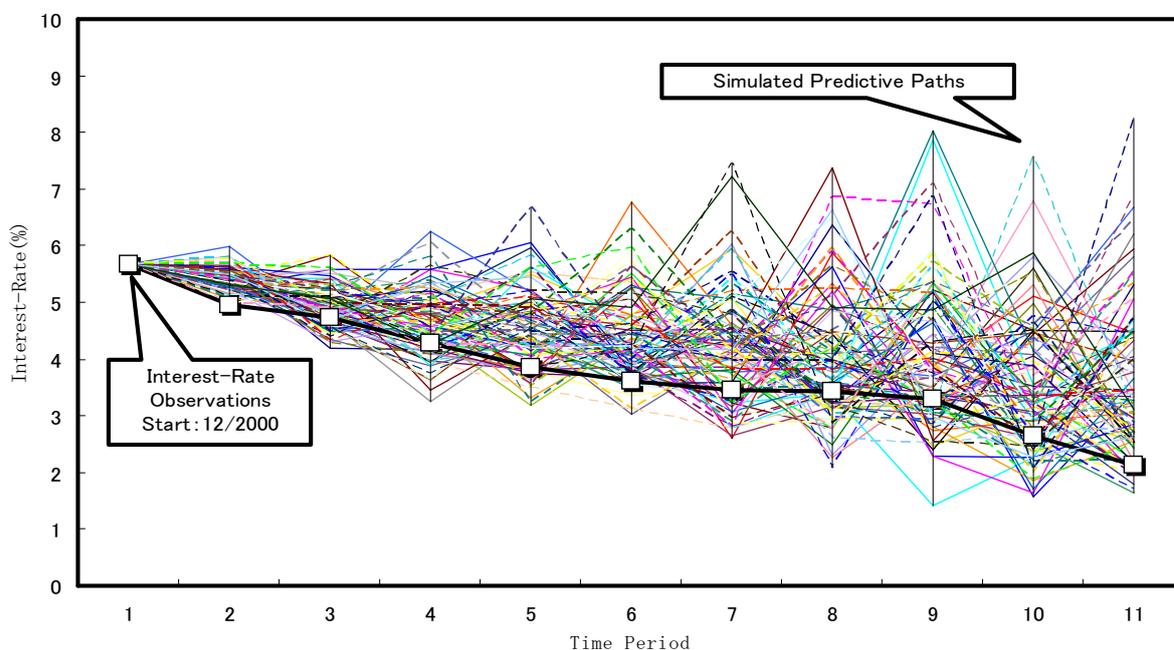


図3-13 金利の観測予測値と実績値の比較 [対数 t 分布に基づくシミュレーション]

この結果によれば、足下の金利データでは低下傾向が見受けられるが、予測分布はこのような傾向に的確に対応している。

このように、本論で提示した予測モデルでは、情報の追加的な更新により母数推定値が修正されることと、金利（観測）の予測分布が裾の広い t 分布にしたがうことにより、固定的なドリフトや拡散係数と標準ブラウン運動にしたがう確率過程を仮定する従来の金利モデルよりも、場合によっては、より正確で情報の変化に対応した金利分布の将来予測が可能となるものと考えられる。

3.3. 後ろ向き推論(Backward Induction)

ここで、所与の金利データの分析に立ち戻り、「後ろ向き推論 (backward induction) 」を行う。

まず、2章2.2節で論じたとおり、互いに独立なマルチンゲール過程である1期先予測誤差 e_t について、観測データ y_t の実績値からドリフトの1期先予測値 f_t を差し引いた残差過程としてとらえ直し、以下のとおり定義する。

$$e_t = \begin{cases} 0 & (t = 0) \\ y_t - f_t & (t = 1, \dots, N) \end{cases} \quad (3.1)$$

さらに、

$$e_t | D_{t-1} \sim t(0, Q_t^*; \nu_{t-1}) \quad (t = 1, \dots, N)$$

ここで、 Q_t^* が適切に推定されているならば、200 を超える自由度からみて、その平方根 $\sqrt{Q_t^*}$ により標準化された直接的な残差過程については、近似的に標準正規過程にしたがうものとみなすことができる。

実際に、 $\sqrt{Q_t^*}$ により標準化された直接的な残差過程のヒストグラムを正規分布の p.d.f. (probability density function) と比較表示すると図 3-14 のとおりである。このヒストグラムをみると、当該標準化過程はかなりの程度正規性を持っていることが分かる。実際に当該標準化過程について Kolmogorov-Smirnov test を行うと、検定統計量 $D = 0.0529$ が得られる。この検定統計量は有意水準 5 % における棄却限界値 (0.094167 ; 観測数 $N = 208$) を大きく下回っており、 p 値は 60 % ~ 70 % の範囲に収まる。したがって、この検定の結果からは「かなりの確からしさをもって当該過程の分布が正規性を持っていないとはいえない」ことになる。

また、標準化された直接的な残差過程の「尖度」は 4.09237、「歪度」は -0.49141 となっている。さらに、当該標準化過程の「標準偏差」は 1.12057 となっており、ここから、当該標準化過程を正規分布で近似する場合には、ほとんど標準正規過程であるとみなすことが可能である。これらの検定結果は、いずれも、概ね Q_t^* が的確に推定されていることを示している。

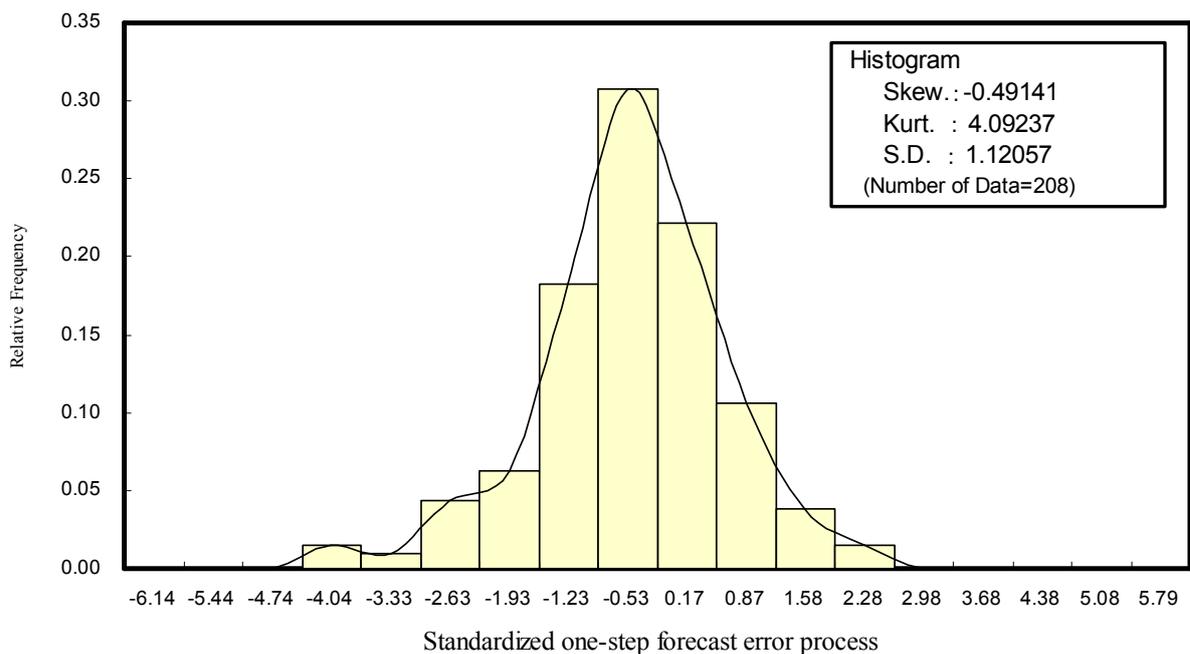


図3-14 $\sqrt{Q_t^*}$ で標準化された e_t のヒストグラム [棒:ヒストグラム,線:平滑化密度]

なお、従来、金利モデルでは、ギルサノフ定理を用いた確率測度変換が一般的であるが、本論のモデル分析による「後ろ向き推論」を適用すれば、上でみたように、金利の過去の履歴について近似的に正規分布で評価することが可能となり、結果として、新たな変換の方法を提言することにつながることになる。

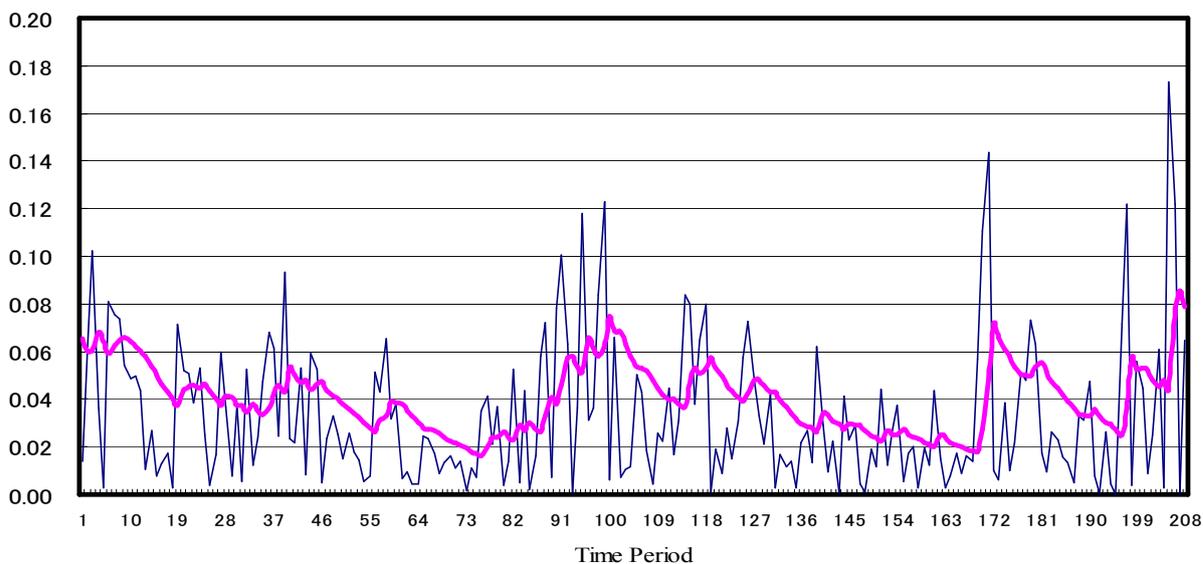


図3-15 $\sqrt{Q_t^*}$ の推定値と e_t の絶対値比較 [太線: $\sqrt{Q_t^*}$ 推定値, 細線: e_t の絶対値]

4. 新たなデータに基づく研究²¹

既に考察を行ったとおり、経済学や経営学の分野における時系列データを取り扱う際に、多くの場合問題とされることは、実際に観測されるデータの分布を求めると、正規分布と比較して、かなりの程度、裾の厚い分布となることが多いことである。

一方で、経済時系列分析の基礎理論は、観測データの確率的な変動が、原則として、標準正規過程に従うという定理を用いている。このため、もし観測データの確率過程の経験分布が、正規分布よりも裾の厚い分布に従う場合は、正規分布に基づく予測や制御の範囲では到底起こりえない観測事象が事後的に生起することにより、予測や制御の有効性が損なわれるおそれがある。

あらためて、この問題について、以下の図 4-1 に即して説明する。図 4-1 は正規密度と金利、株価、外国為替などの変動データで一般的に観測される経験密度を比較して表示したものである。

一般に、正規密度に基づく評価の背景にある考え方は、観測において大きな外れ値が発生したとしても、その値が全体の平均に影響することはほとんど無い、ということである。したがって、一定の回数を平均すれば平均値がほぼ定まり、次の試行や観測の結果がどのようなものであったとしても平均値には影響を与えないとする。また、このことは、正規分布には大きな変位の出現する確率が急速に小さくなる特徴があることを示唆している。

これに対して、正規分布と比較して、かなりの程度、裾の厚い分布では、突然の暴落などの極端な値動きが正規分布から予測される以上の確率で起こり、かつ、その変位が、従前の観測データ全体の変位と同程度の影響度を持つ可能性がある。このため、市場価格のリスク評価などがきわめて難しくなることが考えられる。

なお、これに類する問題は、単に市場価格の変動のみならず、様々な社会的な現象に関わるデータを観測する際にも多く見受けられるのではないかと考えられる。

さらに、問題を複雑化させる要素として、観測時間の経過とともに変化する観測過程

21 本章の取りまとめにあたって、多摩大学経営情報学部・同大学院経営情報学研究所の今泉 忠教授からご指導をいただいた。

の「ずれ」、すなわち、確率的に変化するドリフトの存在がある。図 4-1 で示した裾の厚い分布になんらかのドリフトが加わった場合、観測過程のサンプル・パスは、例えば、図 4-2 の様になり、ヒストグラムを描くとやや歪みのある形状となる（図 4-3）。

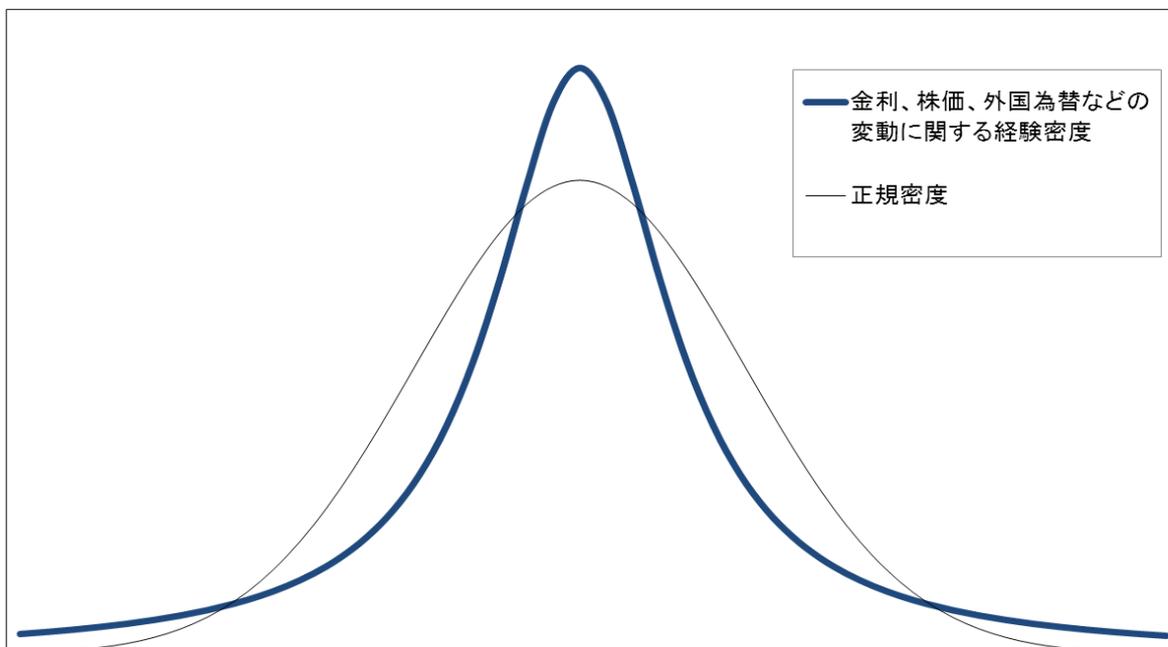


図4-1 正規分布と裾の厚い分布の密度関数

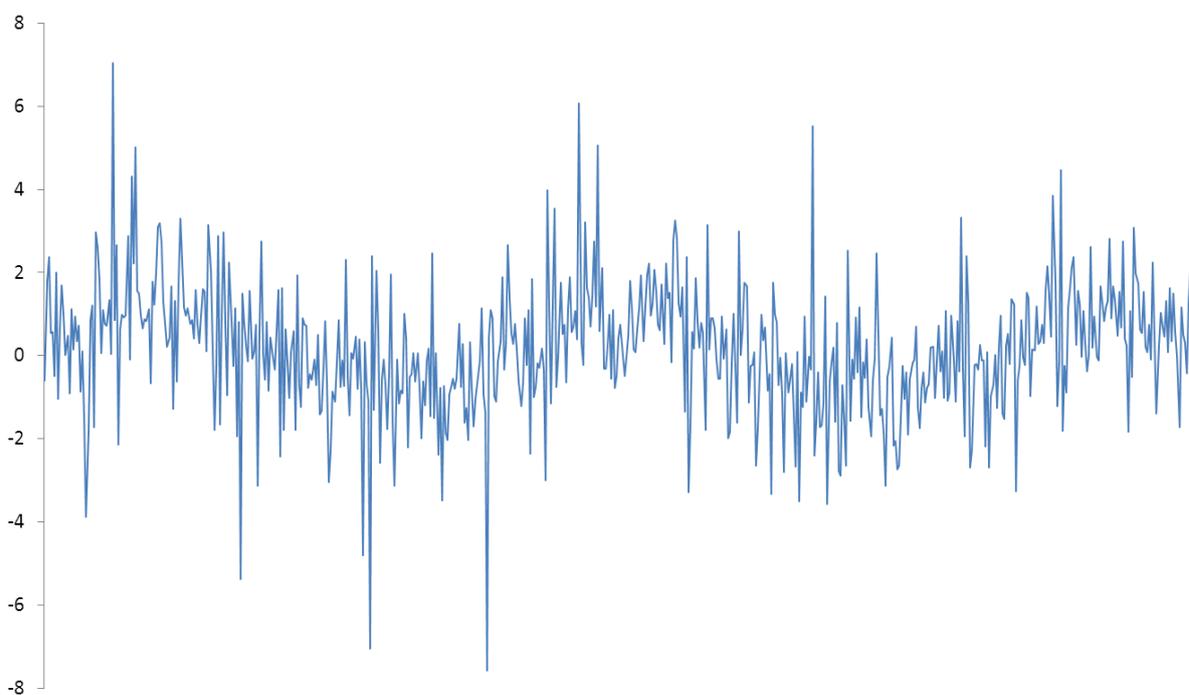


図4-2 ドリフトのある裾の厚い分布に従う確率過程のサンプル・パス

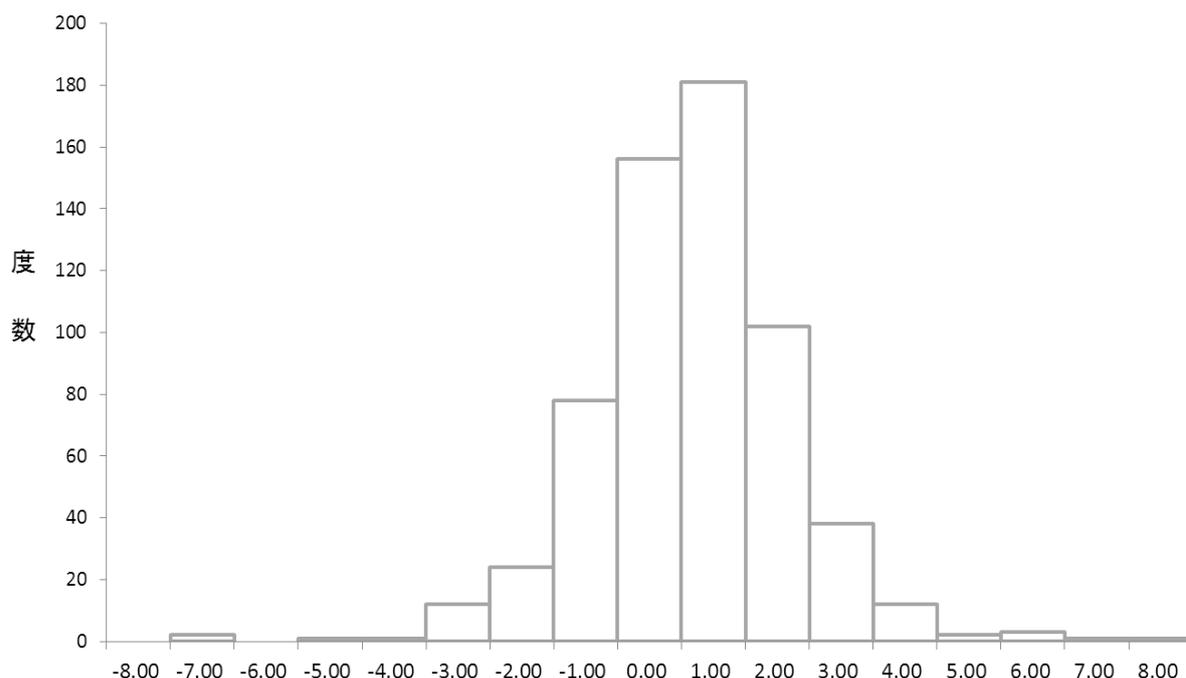


図4-3 ドリフトのある裾の厚い分布に従う確率過程のサンプル・パスのヒストグラム

この場合は、先にみた裾の厚い分布の問題点に加え、新たに観測時点によって得られる平均値が不規則に変動するリスクが加わり、確率過程の評価がさらに困難となるため、何らかの方法により観測過程のずれを除却して分布の歪みを補正することが必要である。なお、その際、ずれを除却した観測過程を推定モーメントで基準化することにより正規過程に変換することができれば、複雑な動きを示す確率過程に対して頑健性のある評価が可能となるものと考えられる。ただし、この方法は単なる確率変数の変数変換とは異なるものである。なぜならば、裾の厚い分布を固定的なモーメント等を用いて変数変換しても、平均の位置や相対的な密度が変わるだけで、その基本的特性（裾の長さや歪みなど）は、依然として従前の分布と同じだからである。これに対して、ここで求められているものは、確率過程の背後にある法則性を解明し、確率変数を規定する分布そのものを、従前の分布の特性いかににかかわらず、正規分布へ変換する方法である。

マンデルブロ博士は、上でみた様な複雑な確率過程の評価の問題について、スケーリング理論に基づくフラクタル的な接近法を用いることにより、現実に即した形で分析可能となることを示している。その根幹をなす観点は、まず第一に、商品市場、通貨市場、株式市場など観測される市場価格の観測データのほとんどが、図 4-1 で示した様な裾の

厚い分布に従うと考えることである。

その際に参考となる重要な研究例が、Mandelbrot and Hudson (2004) で紹介されている、水理学者ハーストの「ナイル川の水位の記録における長期記憶解析」である。このナイル川の研究に関しては、Mandelbrot (1963a) で示された綿花価格に潜むスケーリング性が水理問題に関する研究から発見された事例として、とくにマンデルブロ博士が重要視しているものである。

ハーストは、今から 100 年程前、ナイル川にダムを建設するさいに、洪水時にナイル川がどのくらい氾濫するかを予測している。そのとき、特に課題となった問題は、「ときには水位がきわめて高く、また別の年には極端に低くなる」というナイル川の流量変動に関する観測情報を、どのようにして統計的推論の俎上に載せるか、ということであった。

一般的には、年ごとの数量の変動は統計的に独立と考え、100 年以上の観測データがあれば、漸近的に正規分布に基づく推論が行われることが多く、例えば、「過去 25 年機能していたダムを参考にして、今後 100 年間使えるダムを設計するには、時間が 4 倍になるので、ダムの高さは 2 倍にしておけばよい」と推論される。

しかし、ハーストは、以下に示す独自の解析方法の結果により、ダムの高さはもっと高く見積もっておかねばならないと結論付けた。

※ ハーストのナイル川の水位の記録における長期記憶解析

(Rescaled range (R/S) analysis)

・基準統計量

$$\begin{aligned} Q(\tau) &= R(\tau) / S(\tau) \\ Q(\tau) &\propto \tau^H \end{aligned} \tag{4.1}$$

ここに、

$$\begin{aligned} R(\tau) &= \max_{1 \leq k \leq \tau} \sum_{j=1}^k (x_j - E[\tilde{x}_\tau | D_\tau]) - \min_{1 \leq k \leq \tau} \sum_{j=1}^k (x_j - E[\tilde{x}_\tau | D_\tau]) \\ S(\tau) &= \sqrt{\frac{1}{\tau} \sum_{j=1}^{\tau} (x_j - E[\tilde{x}_\tau | D_\tau])^2} \end{aligned}$$

(確率変数にティルド[~]を付す。)

$E[\cdot | D_\tau]$ は、期間 τ における観測情報： $D_\tau = \{\tilde{x}_1, \dots, \tilde{x}_\tau\}$ に基づく期待値

ハーストが、上記 H (本稿では、これをハースト指数と呼ぶ。) を推定したところ、 H は有意に $1/2$ よりも大きく、したがって、「ナイル川の水量は、互いに独立に同じ正規分布にしたがう過程ではない」と結論付けられた。実際にハーストが推定した結果では、 H は $3/4$ であり、「ナイル川の水量は $3/4$ 乗で拡大していく (正規分布よりも裾が長く厚い分布にしたがう)」ということであった。

この結果にしたがえば、「過去 25 年機能していたダムを参考にして、今後 100 年間使えるダムを設計するには、時間が 4 倍になるので、ダムの高さは従来の 2.8 倍以上にしておかなければならない」ことになる。

マンデルブロ博士は、このハーストの考えについて、ナイル川の流量変動パターンと株価の変動が共通していること、その他、国際的な原油価格の変動、ロンドン市場の金価格、規制が撤廃された米国の電力市場などに関する研究においても同様の結果が得られたことなどから、ナイル川の流量変動パターンをフラクタル幾何学の典型的な例としてとらえ、自らの金融市場に関する研究手法の柱の一つに位置づけている。

なお、ハーストの考え方を一般化すると、以下のとおりとなる。

時系列観測データの時点 t における q 次モーメントを $\sigma^q(c\tau)$ と表し、次のとおり定義する。

$$\begin{aligned}\sigma^q(c\tau) & \stackrel{def}{=} E[|\tilde{x}(t+c\tau) - \tilde{x}(t)|^q] \\ & = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^T |\tilde{x}(t+c\tau) - \tilde{x}(t)|^q\end{aligned}\tag{4.2}$$

ただし、 $0 < c < 1$

次に、ハースト指数の導出にあたり、正規密度に基づいて連続化すれば、以下の関係式が得られる。

$$\sigma^q(c\tau) \propto c^q \tau^{qH_q}\tag{4.3}$$

(ここに、 H_q を「一般化ハースト指数」と呼ぶ。)

さらに、単純化のため $\sigma^q(c\tau) = c^q \tau^{qH_q}$ とおけば、「一般化ハースト指数」： H_q について、以下のとおり求まる。

$$\begin{aligned}
H_q &= \left(\frac{\log \sigma^q(c\tau)}{\log c^q \tau^q} \right) \\
&= \frac{1}{q} \left(\frac{\log \sigma^q(c\tau)}{\log c\tau} \right) \\
&= \frac{1}{q} \left\{ \log_{c\tau} \left(\frac{\sigma^q(c\tau)}{c\tau} \right) + 1 \right\}
\end{aligned} \tag{4.4}$$

本稿では、峰村・今泉(2013)を参照し、上記を「基本的仮説」とする。

次に重要な観点は、観測過程のドリフトを逐次推定し捕捉することである。その際、確率論等を参照し、推定されたドリフト期待値を観測過程から除却することで確実にマルチンゲール過程を実現することが求められる。

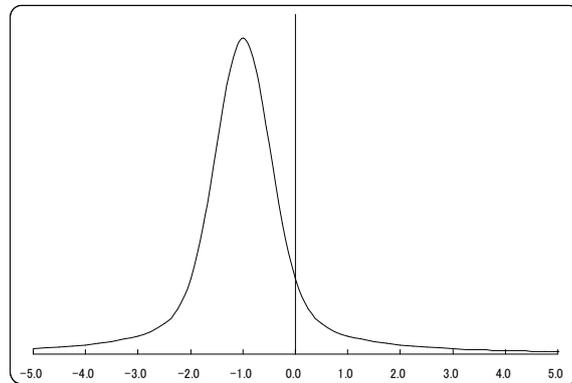
これらの観点を統合することにより、異時点間の観測結果において以前の観測時点ではみられなかった構造的な変化や大きな変動がデータの生成過程で生じたとしても、逐次的確に、当該データの背後にある「ドリフト」と「ボラティリティ」を解析し、観測過程の経験密度を正規密度へ変換できれば、正確で頑健的な確率過程の評価が可能となることが期待できる。

本稿では、上述の課題及び視点を踏まえ、確率論の理論と統計的推論とを接合した、新たな確率測度変換の手法を提案する。その具体的な手順は以下のとおりである(図5-4)。

- (1) 観測過程のドリフト期待値の逐次推定
- (2) 観測過程のモーメント推定
- (3) 観測過程からドリフト期待値を除却することにより、期待値ゼロのマルチンゲール過程を実現
- (4) 推定モーメントに基づくマルチンゲール過程の標準化を実施
- (5) 観測過程の正規密度への確率測度変換

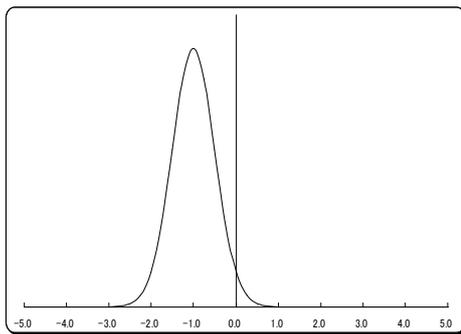
以下では、本稿の提案手法を実際のデータに適用する。

<典型的な金利、株価、外国為替などの変動に関する分布(ドリフト付き「べき分布」)>

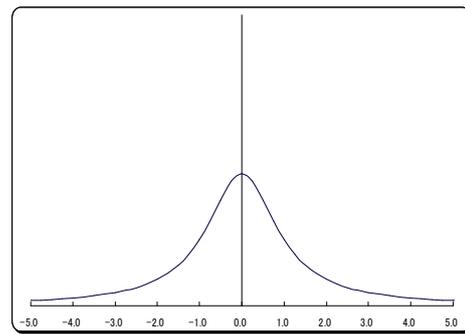


本稿の方法

(1) 観測過程のドリフト期待値の
逐次推定

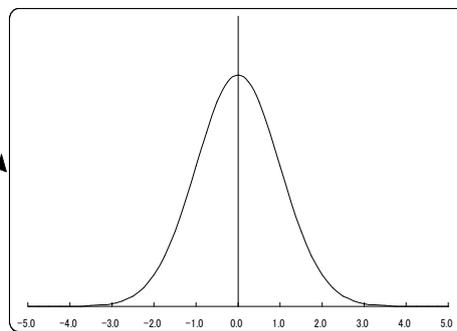


(2) 観測過程のモーメント推定



(5) 観測過程の(標準)正規過程への
確率測度変換

(3) 観測過程
からドリフト期
待値を除却す
ることにより、
期待値ゼロの
マルチンゲー
ル過程を
実現



(4) 推定モー
メントに基づく
マルチンゲー
ル過程の
標準化を
実施

図4-4 提案する確率測度変換の方法

ここでは、Minemura(2006)と同様に、「米国財務省証券 6 ヶ月物金利（月次データ）」に関する統計的分析を行う。6 ヶ月物金利は、固定金利債券の利金支払い金利スワップ取引に関する決済等のタイミングに合致しており、重要な金利指標の一つとなっている。このため、当該データを月次の長期時系列で眺めると、金融市場の変動を如実に表す指標ともなりうると思われる。

本稿におけるデータ観測期間は、1994 年 9 月から 2011 年 12 月までとする。この間、2000 年代半ばから、米国における金融緩和による住宅市場の活況とその反動、サブプライム・ローン問題の表面化と GSE の破綻、リーマン・ブラザーズの破綻、ギリシャ危機の表面化による欧州経済の混乱等が立て続けに発生し、米国の金融市場等において、かつて無い程の大きな衝撃的変化が生起している。

具体的に原系列データをみると図 4-5 のとおり示される。



図4-5 米国財務省証券 6 ヶ月物金利原データ（期間：1994 年 9 月～ 2011 年 12 月； $N=208$ ）

この図によれば、2007 年 7 月に GSE 問題が発生し、また、2008 年 9 月にリーマン・ブラザーズが破綻した影響を受けて金融政策が大幅緩和に転じたことを受けて、2007 年の年央から金利が急激に低下する傾向にあった中、さらに、2009 年年末から 2010 年上

半期にギリシャ危機が表面化した後、金利水準はさらに一段低下し、歴史的にみても、きわめて低い水準で推移している。

また、この金利推移を対数変化率でみると、2009年以降に金利水準がきわめて低水準となってから金利の変化率の変動幅が大きくなる傾向が生じ、足下において振幅の激しい乱高下状態が続いている（図4-6）。

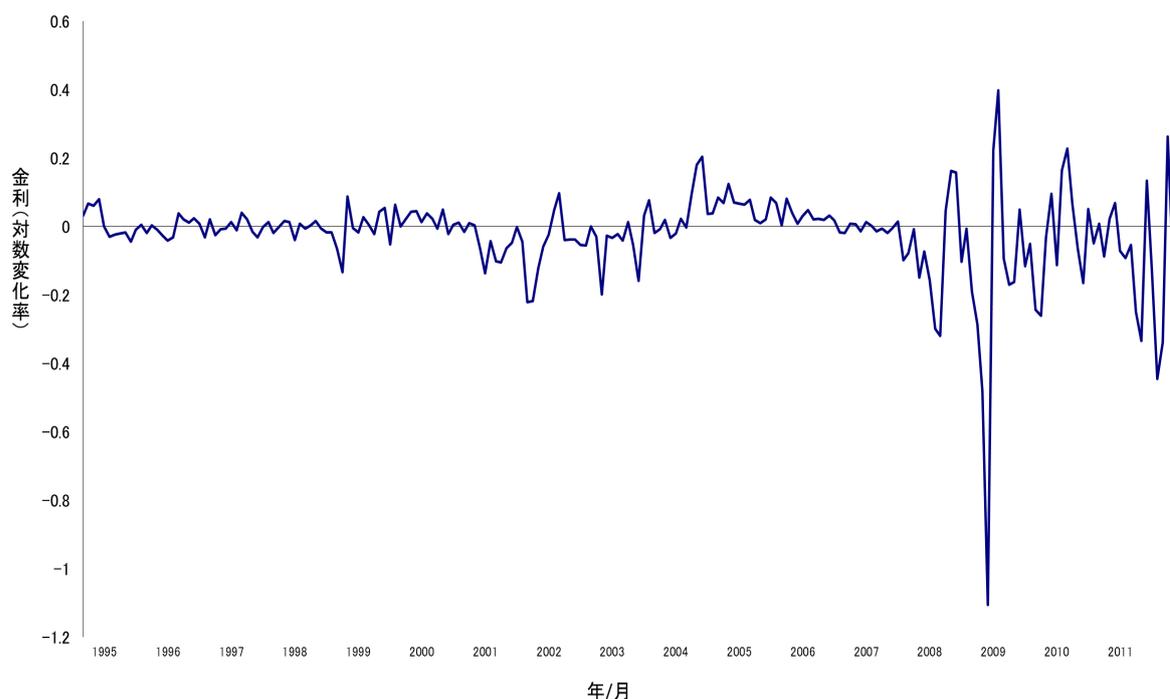


図4-6 米国財務省証券6ヶ月物金利対数変化率（期間：1994年9月～2011年12月； $N=208$ ）

当該変化率の確率密度の裾は、正規分布と比較してかなり厚くなるものと想定される。実際、標準化された当該過程のヒストグラムを示すと図4-7のとおりである。これによれば、標準化後標本過程の歪度は -3.22674 、尖度は 27.06034 となっており、この標本過程は、明らかに正規分布に従っていない。この結果から、金融市場における指標金利の一つとして採用されることの多い米国財務省証券6ヶ月物金利の対数変化率は正規性を担保しておらず、正規密度に準拠する評価法では、正確なリスク評価が難しいことが示された。

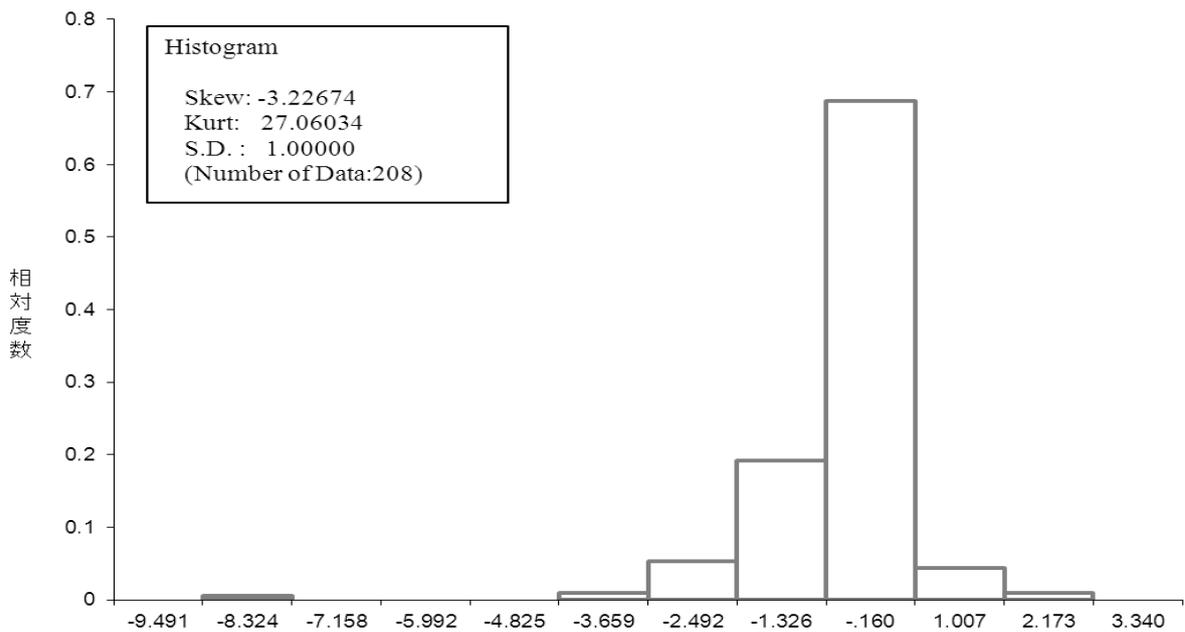


図4-7 標準化後米国財務省証券6ヶ月物金利対数変化率のヒストグラム
(期間: 1994年9月~2011年12月; N=208)

この米国財務省証券6ヶ月物金利の対数変化率について、DLMを適用して推定されたドリフト期待値(平滑化後)を表示すると図4-8のとおりである。図では、ドリフト期待値が対数変化率の変動に応じてシフトしていることが分かる。

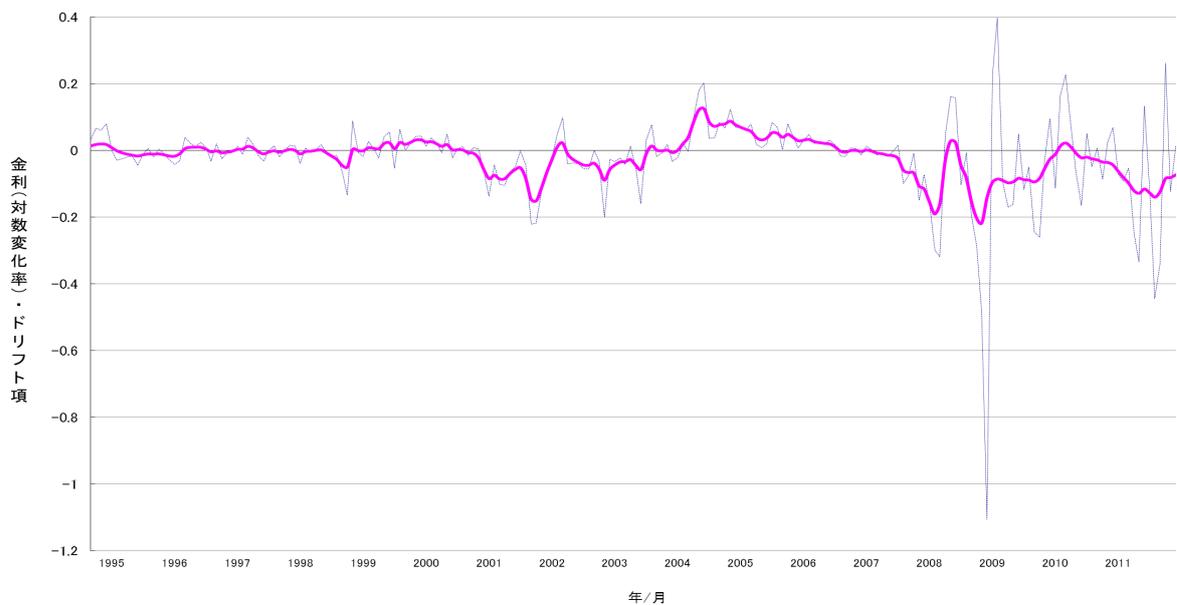


図4-8 米国財務省証券6ヶ月物金利対数変化率のドリフト推定値
(細線: 標本過程原系列, 太線: ドリフト推定値)

次に、構造モデルにより観測予測誤差モーメント推定し、当該モーメントに基づいて観測予測誤差原系列を標準化して、図 4-9 および図 4-10 の結果を得る。

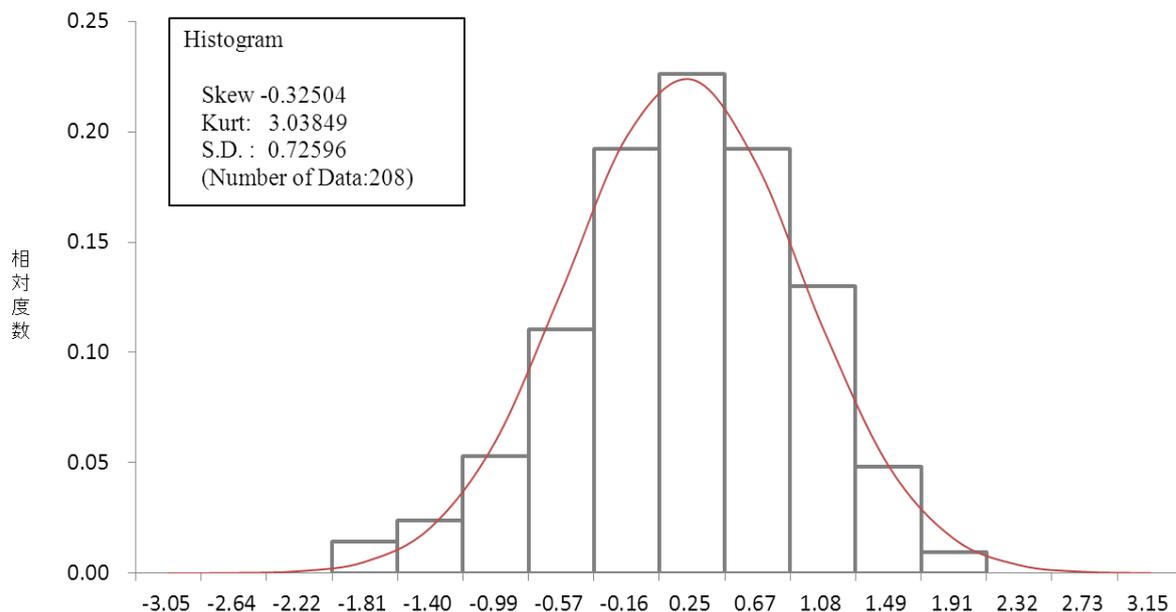


図4-9 $\sqrt{Q^*(t)}$ により標準化された $e(t)$ のヒストグラム
(棒:ヒストグラム, 線:正規密度による近似線)

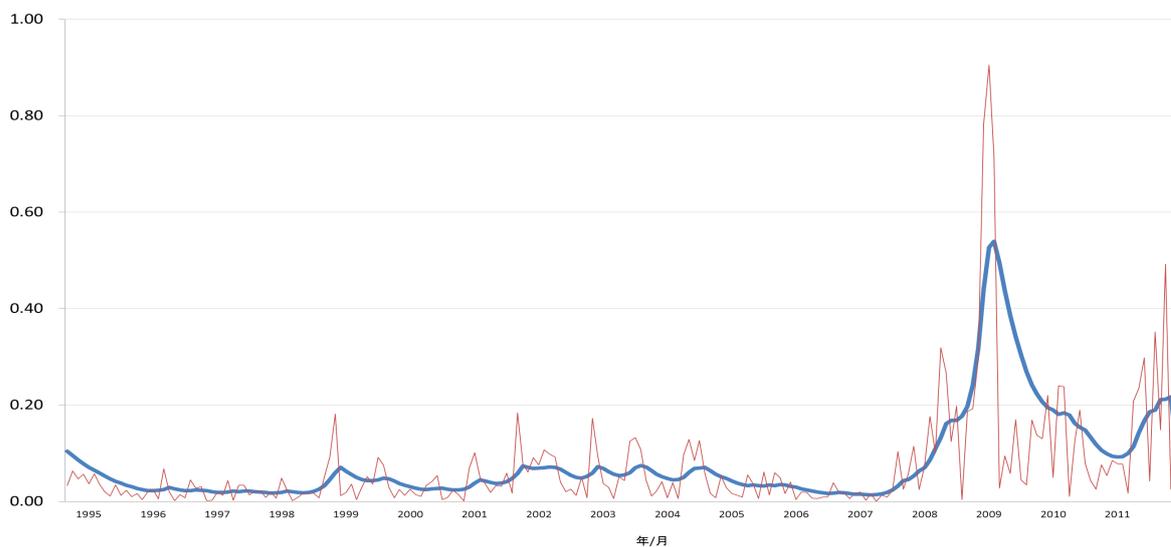


図4-10 $\sqrt{Q^*(t)}$ の推定値と $e(t)$ の絶対値比較
(太線: $\sqrt{Q^*(t)}$ の推定値, 細線: $e(t)$ の絶対値)

図 4-9 のヒストグラムから示される標準化された観測予測誤差の「歪度」は -0.32505、「尖度」は 3.03849、「標準偏差」は 0.72596 となっており、ほとんど標準正規過程に変換されたものとみなすことが可能である。

これらの結果、観測される確率過程が従う分布について、裾の厚い分布が想定される場合や、確率過程そのものがドリフト変動を伴う過程の場合では、本稿の方法²²を用いることで、原系列過程の基本的な特性を保持したまま扱いやすい正規過程に確率測度を変換でき、確率過程の予測等に資することができる。

22 ここでは、今まで論じたモデルを踏まえ、観測予測誤差： $e(t)$ 、および、モーメント： $Q^*(t)$ に基づく考察を行う。なお、 $e(t) \stackrel{def}{=} y_t - \mathbf{F}'\mathbf{a}_N(t)$ ($1 \leq t < N$) と定義する。

$$e(t) \sim t(0, Q^*(t); \nu_N)$$

$$Q^*(t) \stackrel{def}{=} \left(\frac{S^*(t)}{S_N}\right)(\mathbf{F}'\mathbf{R}_N(t)\mathbf{F} + S_N)$$

ここに、 $S^*(t)$ については、以下のとおり定義する。

$$S^*(t) \stackrel{def}{=} \begin{cases} \mu_t S_t^* + w^2 [S^*(t+1) - \mu_t S_t^*] & (1 \leq t < N) \\ S_N & (t = N) \end{cases}$$

ただし、ここでは、 $w^2 = 0.5$ 、 $\mu_t = \left(\frac{S_t}{S_{t-1}}\right)$ とおいた。

5. おわりに

本研究の意義については、以下の①～⑨のとおり、総括することができる²³。

- ① 確率過程の評価に関する論点整理
- ② 確率微分方程式と確率論（「フィルターの問題」およびマルチンゲール理論）の融合
- ③ 観測データに対する確実な確率測度（確率密度）変換の方法の確立
- ④ 不均一分散モデルを組み込んだ階層的 DLM（動的線型モデル）による新たなモデルの提案
- ⑤ 上記モデルの母数推定に関する新たな試み：「データ加増型ベイズ予測 (DABF)」（EM アルゴリズムによる初期値の推定とデータ加増型ベイズ予測による MCMC の組み合わせ）の提案
- ⑥ MCMC による階層モデルの隠れ変数の事後密度捕捉
- ⑦ 本論のモデルおよび母数推定方法を用いた、裾の厚い分布を持つ観測データ（米国財務省証券金利データ）の正規密度への確率測度変換の実証的分析
- ⑧ 前向き推論に基づく将来予測
- ⑨ 後ろ向き推論（正規密度による回顧的予測）に基づく観測データ評価

まず、①について、市場価格変動などの経済・市場データの評価を行う上で、当該観測過程を標準正規密度へ変換することが重要なファクターの一つに含まれておりながら、従来の固定母数によるモデル (Black-Scholes モデル) に基づく母数推計方法 (GMM や MLE) を用いたときに実現が困難な場合があることを指摘した。

次に、②および③について、①の考察を踏まえ、確率論に基づく「フィルターの問題」

23 本論の取りまとめにあたっては、全章の数式表現を含む内容について東京大学および多摩大学・大学院名誉教授 鈴木 雪夫博士の全面的な指導を受けており、とくに、①～③の考察、研究および方法に関連する序論 (p.1 ~ p.9) および、付録：A1. 「序論」の補論 (p.55 ~ p.68) で示した主要な数式表現については、そのすべてに関して鈴木博士の指導による校閲を経ている。

の厳密な検討に基づき、状態空間表現に準拠した確率微分方程式による最良の可測推定（「 \mathcal{F}_t - 可測ドリフト期待値」の推定）を行い、当該期待値を原系列から除却することによりマルチンゲール過程を実現し、さらに、適切に推定された拡散係数（ボラティリティ）で当該過程を標準化することで、リスク中立的確率を実現する方法を提案した。

また、④のとおり、上述の金融資産収益率の拡散係数を推定するために、不均一分散モデルを組み込んだ階層的 DLM を提案した。その際、不均一分散モデルを組み込むことにより、DLM の理論的体系と矛盾を生じることが無いように、不均一分散モデルのモデル母数（未知母数）について「隠れ変数」として取り扱う工夫を行った。

さらに、隠れ変数となるモデル母数を推定する方法として「データ加増型ベイズ予測 (DABF)」を提案し、その具体的な手順を取りまとめた (⑤, ⑥)²⁴。

実際に、⑦のとおり、不均一分散モデルを組み込んだ階層的 DLM と「データ加増型ベイズ予測 (DABF)」を、裾の厚い分布を持つ観測データ（米国財務省証券金利データ）の分析に当てはめて実証を行うと、マルチンゲール過程を実現可能な \mathcal{F}_t - 可測ドリフト期待値が推定でき、同時に、残差過程を正規密度に変換可能な拡散係数を捕捉できた。また、不均一分散モデルのモデル母数についても、隠れ変数でありながら事後密度を把握可能となった。

この様に、①～⑦の考察、研究および方法により、従来では困難であった確率測度変換プロセスの実現が可能となる方法の開発について、一定の条件の下で成功したものと考えることができる。また、①～⑦には、首尾一貫した理論的な背景があり、基礎的な

24 「データ加増型ベイズ予測 (⑤)」では、まず、分散母数が未知の DLM の枠組みにしたがって時系列モデルを組み立てる。本論の内容に即してみれば、時系列モデルとして、ブラウン運動にしたがう金利モデルを離散化して近似的にあてはめており、モデル上で拡散係数として取り扱われる観測誤差分散については階層的に組み込まれた下位モデル（不均一分散モデル）により推定されるように設定している。

次に、観測の 1 期先予測分布に基づいて観測レプリケーションを発生させ、これらを新たな観測とみなして母数の修正を繰り返すシミュレーションを行う。その結果、観測および母数の「事前」、「予測」、「事後」分布についても、逐次、一斉に更新がなされる。そのさい、モデル母数の初期値の設定については EM アルゴリズムを適用し、推定されたモデル母数の採択するさいの評価関数については準ベイズ比を用いる。このデータ加増法に基づくシミュレーションは、階層的に組み込まれた構造母数（未知母数 α, β ）のサンプルパスの収束が確認された時点で終了する。このように、本論におけるモデル分析では、基本的なベイズ統計的手法がいくつも組み合わせられた構造となっている (⑦)。

最後に、最終的に得られたモデル母数に基づき、考察の対象となる確率過程について、現時点までの情報を十分に反映した「前向き推論 (⑧)」と「後ろ向き推論 (⑨)」を行う。

考察、研究と応用的な側面（実証）とが適切に結び付いた成果が得られたものと想定される²⁵。

さらに、本論の研究は、マンデルブロ博士の提唱したスケーリング理論に基づくマルチフラクタル・モデルに則り、金融資産収益率などの変動の背後にあるスケーリング性を分析するモデルをも踏まえており、金融資産収益率などの変動過程を標準正規過程に変換するプロセスにおいて、確率的ドリフト項と時変観測誤差分散を抽出し評価できる利点を備えている。したがって、分析対象となる確率過程の分布の裾が厚くなる傾向を持つとき、その背景に関してドリフト変動と観測誤差分散変化の両面から要因分析できる。

なお、本稿のモデル分析に基づいて正確な確率測度変換が可能になれば、金融実務面などで、今まで以上に金融資産価格収益率などの効率的な評価が可能となることに加え、金融市場の高リスク性（市場においては、既存の金融理論では決して起こりえないリスクが現実には頻繁に起こること）に対する対応手段等についても選択肢が広がることも想定される。

この様に、本論の方法に基づけば、

- ・「前向き推論」では、常に最新の情報を反映した予測が可能となる。
- ・観測レプリケーションを用いたシミュレーションにより、将来的にありうべき観測の動向をある程度評価したモデル母数推定が可能となる。
- ・「後ろ向き推論」では、過去の観測を正規過程として評価できる。
- ・「前向き推論」と「後ろ向き推論」とを組み合わせることにより、従来困難であった様々な応用が可能となる余地が広がる。

などが期待できる。

25 本論の巻末に添付しているとおり、①～⑨の考察、研究および方法について別途作成した論文が、査読を経て、*Journal of Applied Statistics* 誌に収録された。

本論で示した「データ加増型ベイズ予測法」は、任意の確率過程の「将来」、「現在」、「過去」の推移および履歴について、追加的な情報を取り入れて逐次修正された確率測度に準拠して適切な評価を行う方法であり、「将来」に対する予測の結果を「現在」の状態を把握するモデルの修正に用いたり、「過去」の状態の再評価に用いたりすることが可能となる一方で、「現在」や「過去」の修正を通じて、再び「将来」に対する予測の精度を高めることができる。

このような特徴を持つ予測法は、経済学等の分野のみならず、確率過程の予測や制御を行うさいにも応用可能な方法であり、マンデルブロ博士が、商品市場、通貨市場、株式市場における市場価格の変動と乱流状態にある流体のエネルギー散逸量の変動や大型河川の流量変動などには、いくつもの相似点があることを指摘している（Mandelbrot (1972), Mandelbrot and Hudson (2004)) ことを踏まえれば、さらに広範な範囲で応用される可能性を秘めているとみることもできよう。

今後は、より複雑な確率過程を分析するモデルにも対応できるように、継続的な研究を続けて行きたいと考えている²⁶。

26 昨今では、「ティックデータ（市場における約定と気配に関する価格や量等の時系列データ）」を含む巨大で多様な観測データに基づく金融市場の分析や研究も数多く行われている。「ティックデータ」の変動は、比較的短期の記憶に大きな影響を受けることが実証されているものの、ドリフトの影響度は相対的に少なく、またその一方で、変位の幅が極めて大きくなることも想定される。今後、この様な観測データの予測や評価に関して、さらなる研究を進めたいと考えている。また、金融市場における取引方法の変更等の制度的な構造変化（例えば、HFT導入）によっても、観測される金融市場のデータに新たな歪みを生じさせる可能性がある。本論のドリフト推定方法に関し、この様な構造変化によって生じうる歪みや偏りの程度を計測可能となるように発展させることも重要であると考えられる。

謝 辞

本論文をまとめるにあたり、長きにわたり研究の進捗を温かくお見守りいただき、終始、懇切なご指導、ご鞭撻とともに、格別のご高配をいただきました多摩大学経営情報学部 今泉 忠教授に深く感謝申し上げます。

学位論文審査において、貴重なご指導、ご助言をいただきました多摩大学大学院経営情報学研究科 岡太 彬訓教授、温かい励ましとともに貴重なご示唆をいただきました多摩大学大学院経営情報学研究科 宇佐美 洋教授に深く感謝申し上げます。

金融市場分析の新展開、データ解析法の進展等を踏まえ、貴重なご教示とともに、温かく、懇切なご指導をいただきました、進化経済学の泰斗でいらっしゃる中央大学商学部 有賀 裕二教授、時系列分析に関するモデリングや統計的予測法などに関しまして、数多くの貴重なご教示をいただきました中央大学理工学部 渡邊 則生教授に深く感謝申し上げます。

本研究を進める際に、熱心なご指導をいただきました多摩大学 松谷 泰行名誉教授に深く感謝申し上げます。

その他、本研究を進める過程で、様々なご教示、ご支援をいただきました数多くの先生方、大学関係者の皆様方に心より感謝申し上げます。

最後になりますが、大学院在学時より、確率論、統計学の基礎的概念から懇切なご指導、ご鞭撻をいただき、確率論と統計理論との関連性、数学的に完結した統計理論の美しさ、奥深さと、不確実性の下での推論、判断、意思決定や動的な逐次推論法などを含む幅広い有用性とにつつまして、興味深い問題、論点等を踏まえ、数多くご教示いただきました東京大学、多摩大学名誉教授 鈴木 雪夫博士に衷心より深甚の謝意を表します。

本論文を鈴木 雪夫博士に捧げます。

参 考 文 献

- Bollerslev, T. (1986), Generalised Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31, 307-327.
- Bollerslev, T. (1987), A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rate of Return, *Review of Economics and Statistics*, 64, 542-547.
- Borland, L., Bouchaud, J. P., Muzy, J. F., and Zumbach, G. (2005), The Dynamics of Financial Markets - Mandelbrot's Multifractal Cascades, and Beyond, *Wilmott magazine*, Issue 16, 86-96.
- Carlin, B.P. and Louis, T.A. (2000), *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd edn, Boca Raton: Chapman & Hall/CRC.
- Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders. (1992), An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance*, 47, 1209-1227.
- Chib, S., and E. Greenberg.(1994), Bayes inference in regression models with ARMA (p,q) errors, *Journal of Econometrics*, 64, 183-206.
- Chib, S., and E. Greenberg.(1995), Understanding the Metropolis - Hastings Algorithm, *American Statistician*, 49, 327-335.
- Engle, R. F. (1982), Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 987-1008.
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995), *Bayesian Data Analysis*, Chapman & Hall.
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004), *Bayesian Data Analysis*, 2nd edn, Boca Raton: Chapman & Hall/CRC.

- Hull, J. and White, A.(1993), One-factor interest rate models and the valuation of interest-rate derivative securities, *Journal of Financial and Quantitative Analysis*, 28, 235-254.
- Jacquier, E., N. G. Polson, and P. E. Rossi. (1994), Bayesian Analysis of Stochastic Volatility Model, *Journal of Business & Economic Statistics*, 12, 371-389.
- Kim, S., N. Shephard, and S. Chib. (1998), Stochastic Volatility : Likelihood Inference and Comparison with ARCH Models, *Review of Economic Studies*, 65, 361-393.
- Mandelbrot, B. B (1963a), The Variation of Certain Speculative Prices, *Journal of Business*, 36, 394-419.
- Mandelbrot, B. B (1963b), New Methods in Statistical Economics, *Journal of Political Economy*, 71, 421-440.
- Mandelbrot, B. B (1972), Lognormal hypothesis and distribution of energy dissipation in intermittent turbulence, *Statistical Models & Turbulence*, New York., Springer, 333-351.
- Mandelbrot, B. B, Calvet, L, and Fisher, A. (1997a), A Multifractal Model of Asset Returns, *Discussion Papers of the Cowles Foundation for Economics at Yale University: Paper #1164*
- Mandelbrot, B. B, Calvet, L, and Fisher, A. (1997b), A Multifractality of the Deutschmark/US Dollar Exchange Rate, *Discussion Papers of the Cowles Foundation for Economics at Yale University: Paper #1165*
- Mandelbrot, B. B (2001a), Scaling in Financial Prices: III. Cartoon Brownian Motions in Multifractal Time , *Quantitative Finance*, 1, 427-440.
- Mandelbrot, B. B (2001b), Scaling in Financial Prices: IV. Multifractal Concentration , *Quantitative Finance*, 1, 641-649.

Mandelbrot, B. B and Hudson, R. L . (2004), *THE (MIS) BEHAVIOR OF MARKETS: A Fractal View of Risk, Ruin, and Reward*, Cambridge, MA : Perseus Books.

Minemura, E. (2006), An Interest-Rate Model Analysis Based on Data Augmentation Bayesian Forecasting, *Journal of Applied Statistics*, 33, 1085-1104.

Nakatsuma, T. (2000), Bayesian Analysis of ARMA-GARCH Models: A Markov Chain Sampling Approach, *Journal of Econometrics*, 95, 57-69.

Øksendal, B. (1998), *Stochastic Differential Equations; An Introduction with Applications*, 5th edn, Berlin: Springer-Verlag.

Vrontos, I.D., Dellaportas, P., and Politis, D. N. (2000), Full Bayesian Inference for GARCH and EGARCH Models, *Journal of Business & Economic Statistics*, 18, 187-198.

Watanabe, T. and M. Asai (2001), Stochastic Volatility Models with Heavy-Tailed Distributions: A Bayesian Analysis, *Bank of Japan, IMES Discussion Paper Series*, No.2001-E-17.

West, M. and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, 2nd edn, New York: Springer-Verlag.

森村 英典, 木島 正明 (1991), *ファイナンスのための確率過程*, 日科技連.

鈴木 雪夫 (1987), *統計学*, 朝倉書店.

松本 隆, 石黒 真木夫, 乾 敏郎, 田邊 國士 (2004), *階層ベイズモデルとその周辺 -時系列・画像・認知への応用-*, 岩波書店.

峰村 英二, 今泉 忠 (2013), 「動的線型モデルによるファット・テール分布の確率測度変換」, 『経営・情報研究 多摩大学研究紀要』, 17, 69-96.

山本 拓 (1988), *経済の時系列分析*, 創文社.

付 録

A1 「序論」の補論²⁷

a.1.1. (参考1) シグマ集合体と可測性について

集合 Ω 上のシグマ集合体： \mathcal{F} とは、 Ω の部分集合を要素とする集合で、以下の性質を満たすものであると定義する。

- (1) $\phi \in \mathcal{F}$,
- (2) $F \in \mathcal{F} \Rightarrow F^C \in \mathcal{F}$. ただし、 $F^C = \Omega / F$ は F の Ω での補集合,
- (3) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow A := \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

(Ω, \mathcal{F}) を可測空間と呼ぶ。さらに、写像 $P: \mathcal{F} \rightarrow [0,1]$ で次の性質を持つものを (Ω, \mathcal{F}) 上の確率測度という。

- (a) $P(\phi) = 0, P(\Omega) = 1$,
- (b) もし $A_1, A_2, \dots \in \mathcal{F}$ かつ $\{A_i\}$ が互いに素ならば、

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

(Ω, \mathcal{F}, P) を確率空間といい、 P -外測度がゼロとなる集合 G

$$P^*(G) := \inf\{P(F) ; F \in \mathcal{F}, G \subset F\} = 0$$

がすべて \mathcal{F} に属する時、確率空間は完備であるという。

なお、 \mathcal{F} に属する Ω の部分集合 F を \mathcal{F} -可測集合 (事象) と呼ぶ。

確率空間： (Ω, \mathcal{F}, P) 上の関数 $Y: \Omega \rightarrow \mathbf{R}^n$ が、すべての開集合 $U \subset \mathbf{R}^n$ に対して、

$$Y^{-1}(U) := \{\omega \in \Omega ; Y(\omega) \in U\} \in \mathcal{F}$$

を満たすとき、 Y は、 \mathcal{F} -可測 (measurable) であるという。

²⁷ 補論では、本論の内容に係わる基礎的または基本的な項目について「参考」として記載し、補足すべき重要な内容と区別して表記する。

a.1.2. (参考2) ブラウン運動

$\{Y_t\}$ をブラウン運動過程とすると、その定義は以下のとおり与えられる。

(B1) 増分 $[Y_{t+s} - Y_s]$ は正規分布 $N(0, \sigma^2 t)$ にしたがう。

(B2) すべての排反な区間 $[t_i, t_{i+s}]$ における増分 $[Y(t_i) - Y(t_{i+s})]$ は独立。

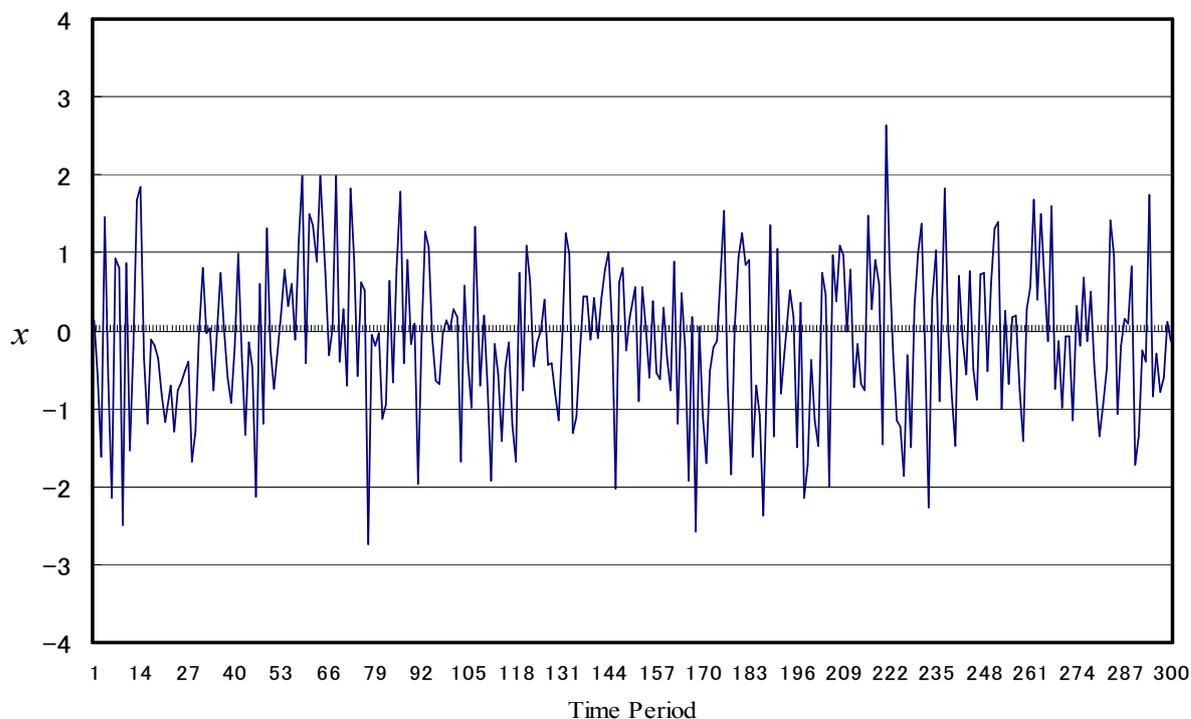
(B3) Y_0 (初期値) = 0. また、 Y_t は、 t の関数として、 $t=0$ において連続。

(Y_t は、すべての t で連続。)

(B1) から、ブラウン運動は正規分布の分散母数である σ^2 によって特徴付けられるが、この σ^2 (場合によっては σ) を拡散係数 (ボラティリティと呼ばれることが多い) と呼び、とくに、 $\sigma^2 = 1$ として、標準正規分布により特徴付けられた過程を標準ブラウン運動 ((0,1)-ブラウン運動) と呼ぶ。

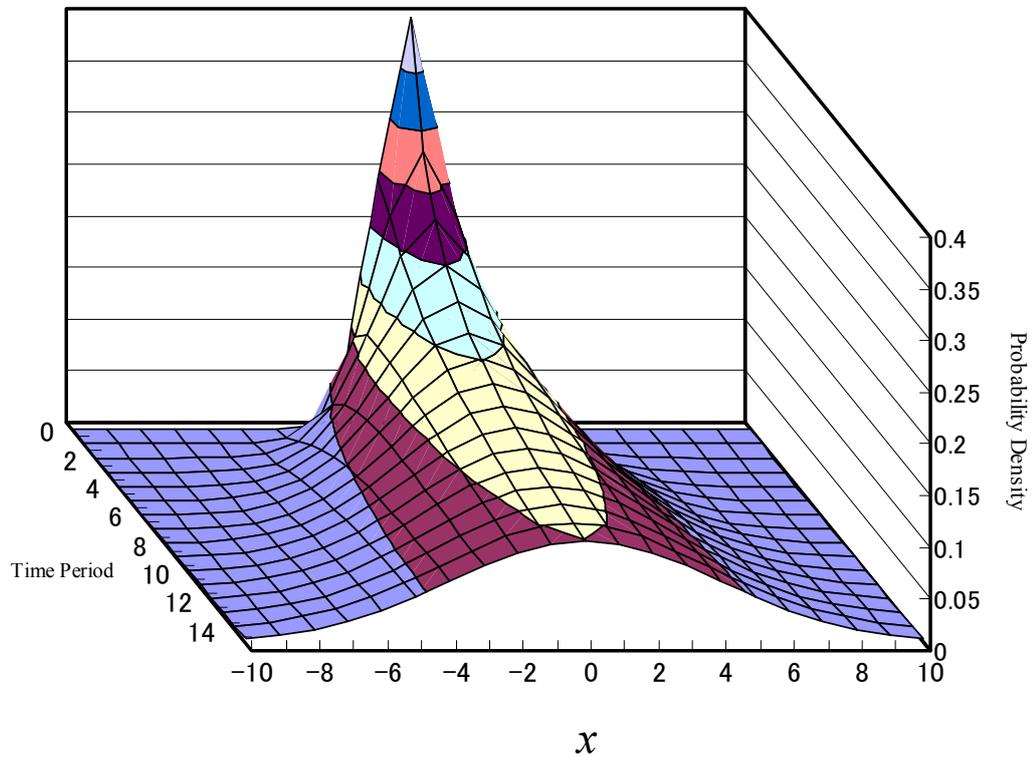
なお、ドリフト係数: μ を持ち、増分: $[Y_{t+s} - Y_s]$ が正規分布 $N(\mu t, \sigma^2 t)$ にしたがうブラウン運動は、 (μ, σ^2) -ブラウン運動と呼ばれる。

ここから、標準ブラウン運動の増分は標準正規過程となり、そのサンプル・パスは以下のとおり、平均ゼロの独立したマルチンゲールとして示される。



図A2-1 標準正規過程のサンプル・パス

さらに、標準ブラウン運動に基づく拡散過程の模式図は、図 A2-2 の様になる（ただし、分布の裾は $-\infty$ から $+\infty$ の範囲）。



図A2-2 標準ブラウン運動による拡散過程の推移グラフ(0期起点15期まで)

a.1.3. (参考3) 離散時間マルチンゲール

$\{X_n\}_{n \in \{0\} \cup \mathbb{N}}$ を完備確率空間 (Ω, \mathcal{F}, P) 上の確率過程とし、部分シグマ集合体の系列 $\{\mathcal{F}_n\}_{n \in \{0\} \cup \mathbb{N}}$ は増大系列であるとする。さらに、すべての n に対して X_n は、 \mathcal{F}_n - 可測（すなわち、確率過程 $\{X_n\}_{n \in \{0\} \cup \mathbb{N}}$ は $\{\mathcal{F}_n\}_{n \in \{0\} \cup \mathbb{N}}$ に適合している (adapted)）とする。このとき、

- (i) すべての n に対して、 $E[|X_n|] < \infty$ (期待値が存在)
- (ii) $m \geq n$ ならば、 $E[X_m | \mathcal{F}_n] = X_n$

が成り立つとき、 $\{X_n\}$ を $\{\mathcal{F}_n\}$ に関するマルチンゲールと呼ぶ。

なお、ドリフトのないブラウン運動はマルチンゲールである。実際、ここで、あらめ

て、 $\{X_n, n \geq 0\}$ をドリフトのない $(0, \sigma^2)$ -ブラウン運動とし、 \mathcal{F}_n を当該ブラウン運動の時点 n までの履歴からなるシグマ集合体とすれば、

$$E[X_{n+m} | \mathcal{F}_n] = E[X_{n+m} - X_n | \mathcal{F}_n] + E[X_n | \mathcal{F}_n]$$

であり、ブラウン運動は独立増分を持つことから、増分： $X_{n+m} - X_n$ は、時点 n までの履歴： \mathcal{F}_n と独立である。また、 X_n は、 \mathcal{F}_n -可測であるから、

$$E[X_{n+m} - X_n | \mathcal{F}_n] = E[X_{n+m} - X_n] = 0;$$

$$E[X_{n+m} | \mathcal{F}_n] = E[X_n | \mathcal{F}_n] = X_n$$

となり、ドリフトのないブラウン運動はマルチンゲールであることが示される。

a.1.4. (参考4) ギルサノフの定理

$T > 0$ を固定し、2つの確率過程 $\{\mu(t, \omega); t \in [0, T]\}$ 、および、 $\{\sigma(t, \omega); t \in [0, T]\}$ を、シグマ集合体の族 $\{\mathcal{F}_t; t \in [0, T]\}$ に関して発展的可測 (progressively measurable) であり、かつ、

$$P\left\{\omega; \int_0^T \mu(t, \omega) dt < \infty\right\} = 1 \tag{A.4.1}$$

および、 $\{\sigma(t, \omega); t \in [0, T]\}$ についても、

$$\int_0^T \left[\int_{\Omega} \sigma(t, \omega)^2 P(d\omega) \right] dt < \infty \tag{A.4.2}$$

を満たすものと仮定する。

なお、一般に、確率過程 $X = \{X_t(\omega); t \in \mathbf{R}^+\}$ が、 $\{\mathcal{F}_t; t \in \mathbf{R}^+\}$ に関して発展的可測であるとは、任意に $t (\geq 0)$ をとめたとき、 $X_s(\omega): [0, t] \times \Omega \rightarrow \mathbf{R}$ が、 $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$ に関して可測であること、すなわち、任意の $A \in \mathcal{B}$ に対して、

$$\{(s, \omega) \in [0, t] \times \Omega ; X_s(\omega) \in A\} \in \mathcal{B}([0, t]) \otimes \mathcal{F}_t \quad \forall A \in \mathcal{B}$$

であることである。なお、ここに、 $\mathcal{B}([0, t])$ は、実数直線上の閉区間 $[0, t]$ に含まれるボレル集合の全体からなるシグマ集合体：ボレル集合体 (Borel field) である。

発展的・可測な確率過程 $\{\mu(t, \omega) ; t \in [0, T]\}$ と $\{\sigma(t, \omega) ; t \in [0, T]\}$ を係数に持つ確率微分方程式

$$dX = \mu(t) dt + \sigma(t) dB, \quad 0 \leq t \leq T; \quad X(0) = 0 \quad (\text{A.4.3})$$

の解は、

$$X(t) = \int_0^t \mu(u) du + \int_0^t \sigma(u) dB(u), \quad 0 \leq t \leq T \quad (\text{A.4.4})$$

で与えられる。ただし、なお、ここでは、表現を簡潔にするために、とくに必要がなければ、 ω を明記しないこととする。

(A.4.4) 式の右辺第一項はサンプルパスごとの通常の積分であるから有界変動を持つ確率過程であり、第二項は標準ブラウン運動に関する確率積分であるから、2乗可積分でサンプルパスが連続なマルチンゲールである。

ここで、新たな関数について、 $g(t, x) \in C^2([0, \infty) \times \mathbf{R})$ (g は、 $[0, \infty) \times \mathbf{R}$ 上で2回連続的・微分可能) と仮定し、

$$Y(t) = g(t, X(t))$$

とおけば、

$$dY = \frac{\partial}{\partial t} g(t, X) dt + \frac{\partial}{\partial x} g(t, X) dX + \frac{1}{2} \frac{\partial^2}{\partial x^2} g(t, X) \cdot (dX)^2 \quad (\text{A.4.5})$$

を得る。ただし、 $(dX)^2 = dX \cdot dX$ は、次の計算規則で求められる。

$$dt \cdot dt = dt \cdot dB = dB \cdot dt = 0, \quad dB \cdot dB = dt$$

上記を「伊藤の公式」という。なお、「伊藤の公式」の詳しい証明や導出については、Øksendal (1998) の4章などを参照願いたい。

さて、以上の結果を踏まえ、以下では、 $\{X(t)\}$ について、 $f(x) = \exp\{x\}$ により、新たに $Y(t) = \exp\{X(t)\}$ と指数変換した確率過程を $\{Y(t)\}$ とおく。なお、あらためて、 $\{\mu(t)\}$ と $\{\sigma(t)\}$ は、一様に有界と仮定する。

$Y(t)$ について伊藤の公式をあてはめると、

$$\frac{dY}{Y} = \left\{ \mu(t) + \frac{\sigma^2(t)}{2} \right\} dt + \sigma(t) dB \quad (\text{A.4.6})$$

が得られる。上式で

$$\mu(t) \equiv -\frac{\sigma^2(t)}{2}$$

とおき、(A.4.6) 式のドリフト項を消去させた上で積分形式で表現すると以下となる。

$$Y(t) - Y(0) = \int_0^t Y(u) \sigma(u) dB(u), \quad 0 \leq t \leq T \quad (\text{A.4.7})$$

ここで、 $Y(t) = \exp\{X(t)\}$ であるので、

$$Y(t) = \exp \left\{ \int_0^t \sigma(u) dB(u) - \frac{1}{2} \int_0^t \sigma(u)^2 du \right\}, \quad t \in [0, T] \quad (\text{A.4.8})$$

となる。(A.4.7) 式と $\{\sigma(t)\}$ の一様有界性より $\{Y(t)\}$ は (A.4.2) 式の条件を満たし、さらに、 $\{Y(t)\sigma(t)\}$ も同様に (A.4.2) 式の条件を満たす。したがって、その確率積分である (A.4.7) 式は、サンプルパスが連続で 2 乗可積分なマルチンゲールである。結局、(A.4.8) 式で与えられる指数的確率過程 $\{Y(t)\}$ についても 2 乗可積分マルチンゲールになることが分かる。

いま、事象 $A \in \mathcal{F}_T$ に対して、(A.4.8) 式で定義される $Y(T)$ に基づいて、

$$\tilde{P}(A) \equiv E[Y(T)\mathbf{I}_A] = \int_A Y(T, \omega) P(d\omega) \quad (\text{A.4.9})$$

により、 \mathcal{F}_T から非負実数の集合への写像 \tilde{P} を定義する。ここで、 \mathbf{I}_A は、事象 A の指標関数である。ここに、 $\{Y(t)\}$ はマルチンゲールであるから、 $E[Y(0)] = E[Y(t)] = 1$ となるので、 $\tilde{P}(\Omega) = \int_\Omega Y(T, \omega) P(d\omega) = 1$ が成り立つ。また、指標関数の定義から排反な事象列 $\{A_i\}$

に対して、 $\mathbf{I}_{\cup_i A_i} = \sum_i \mathbf{I}_{A_i}$ が成立し、

$$\tilde{P}(\cup_i A_i) = E[Y(T)\sum_i \mathbf{I}_{A_i}] = \sum_i E[Y(T)\mathbf{I}_{A_i}] = \sum_i \tilde{P}(A_i)$$

となり、 \tilde{P} が可測空間 (Ω, \mathcal{F}_T) 上の確率測度となること分かる。 E は、もとの確率測度 P に関する期待値であるため、(A.4.9) は P から \tilde{P} への確率測度の変換を与える。このとき、

$$\tilde{B}(t) = B(t) - \int_0^t \sigma(u) du$$

と定義された $\{\tilde{B}(t)\}$ は、確率測度 \tilde{P} に関して、標準ブラウン運動であることが知られている (森村, 木島 (1991))。これをギルサノフの定理と呼ぶ。

a.1.5. フィルターの問題について

Øksendal (1998) によれば、DLM により適切に求められた状態推定量は、最良の可測推定量であり、最良の線形推定量である。

一般に、 \hat{X}_t が観測 $\{Y_s; s \leq t\}$ に基づく状態 X_t の可測推定、あるいは単に推定であるとは、観測される確率過程 $\{Y_s(\omega); s \leq t\}$ の生成するシグマ集合体 \mathcal{S}_t ($\mathcal{S}_t \subset \mathcal{F}_t$) として、

$$\hat{X}_t \text{ は } \mathcal{S}_t \text{ - 可測} \tag{A.5.1}$$

となることをいい、さらに \hat{X}_t が最良の推定量であるときは、

$$\int_{\Omega} |X_t - \hat{X}_t|^2 dP = E[|X_t - \hat{X}_t|^2] = \inf \{ E[|X_t - Z|^2]; Z \in K_t \} \tag{A.5.2}$$

が成り立つことである。ここで、 (Ω, \mathcal{F}, P) を、原点から発する 1 次元ブラウン運動に対応する確率空間とし、 E を P に関する期待値とする。

$$K_t := K(Y, t) := \{Z : \Omega \rightarrow \mathbf{R}^n ; Z \in L^2(P) \text{ かつ } Z \text{ は } \mathcal{G}_t\text{-可測}\} \quad *A1 \quad (A.5.3)$$

(A.5.2) を満たす最良の可測推定量は、ヒルベルト空間 $L^2(P)$ の部分空間 K_t への正射影 P_{K_t} を用いて実現される。したがって、以下が成り立つ。

$$\hat{X}_t = P_{K_t}(X_t) = E[X_t | \mathcal{G}_t]$$

一方、観測 $\{Y_s ; s \leq t\}$ に基づく最良の線形推定量は、以下の形の線形結合

$$c_0 + c_1 Y_{s_1}(\omega) + \dots + c_k Y_{s_k}(\omega), \quad s_k \leq t, c_k \in \mathbf{R}$$

からなる線形空間の $L^2(P)$ に関する閉包 (closure) を $L_t = L(Y, t)$ とすれば、 $L^2(P)$ から L_t への正射影 $P_{L_t}(X_t)$ となる。

$L^2(P)$ に属する確率変数である $X, Y_s ; s \leq t$, 任意の $n \geq 1$ と、 $s_1, \dots, s_n \leq t$ に対して、

$$(X, Y_{s_1}, \dots, Y_{s_n}) \in \mathbf{R}^{n+1}$$

がガウス型確率変数であるときに限り、状態 X の最良線形推定量と最良可測推定量は一致し、

$$P_{K_t}(X_t) = E[X_t | \mathcal{G}_t] = P_{L_t}(X_t)$$

が成り立つ (Øksendal (1998))。

ここで、以下の状態空間表現を提示する。

$$\text{(線形な観測)} \quad dY_t = G(t)X_t dt + \sigma(t)dV_t, \quad G(t), \sigma(t) \in \mathbf{R} \quad (A.5.4)$$

$$\text{(線形な系)} \quad dX_t = F(t)X_t dt + \xi(t)dW_t, \quad F(t), \xi(t) \in \mathbf{R} \quad (A.5.5)$$

ただし、ここでは、 G, F, σ, ξ は、すべて確定的および既知で、任意の有界区間上有界

*A1 可測写像 $f : \Omega \rightarrow \mathbf{R}$, $\{\omega \in \Omega | f(\omega) < c\} \in F$, が $\int_{\Omega} |f(\omega)|^2 P(d\omega) < \infty$, を満たすとき、一般に、「2乗可積分」あるいは「 $L^2(P)$ に属する」という。 $L^2(P)$ は、 $\|H\|_{L^2} := \sqrt{\int_{\Omega} |f(\omega)|^2 P(d\omega)}$ をノルムとするヒルベルト空間で、 $L^2(P) = L^2(\Omega, P)$ であり、 K_t について、(A.5.3) 式のとおり定義できる。

であると仮定する。さらに、 X_0 はガウス型確率変数で、1次元ブラウン運動 $\{V_t\}, \{W_t\}$ と独立、 $\sigma(t), \zeta(t)$ は、有界区間上一様に正であると仮定する。なお、 $Y_0=0$ とおく。

この状態空間表現をもとにして、1次元線形フィルターの問題について考える。

「フィルターの問題」とは、 $0 \leq s \leq t$ で、(A.5.4)を満たす観測が与えられたとき、この観測に基づく系の状態 X_t の「最良の可測推定量」： $\hat{X}_t = E[X_t | \mathcal{F}_t]$ がどのように与えられるか、という問題を考察するものである。直感的には、これは観測から最良の方法でノイズを濾過（フィルター）することを意味しており、その濾過の方法を考察するものである。

この問題を考察するために、まず、あらためて以下の線形結合（ガウス型確率変数）

$$c_0 + c_1 Y_{t_1} + \dots + c_k Y_{t_k} \quad (0 \leq t_k \leq T, c_k \in \mathbf{R})$$

からなる線形空間の $L^2(P)$ における「閉包」について $L_t = L(Y, T)$ と定義して、以下のイノベーション過程 N_t を考える。

$$N_t = Y_t - \int_0^t G(s) \hat{X}_s ds, \quad \text{ただし、} G(s) \hat{X}_s = P_{L(Y, s)}(G(s) X_s). \quad (\text{A.5.6})$$

このイノベーション過程は、以下を満たす。

$$\left\{ \begin{array}{l} (1) N_t \text{ は直交増分をもつ。} \\ (2) E[N_t^2] = \int_0^t \sigma^2(s) ds. \\ (3) L(N, t) = L(Y, t) \quad (\forall t \geq 0). \end{array} \right.$$

なお、 dN_t については、状態空間表現の (A.5.4) 式から、

$$dN_t = G(t)(X_t - \hat{X}_t)dt + \sigma(t)dV_t \quad (\text{A.5.7})$$

が成り立つ。

・(1),(2),(3)の証明

(1) $s < t$, $Z \in L(Y, s)$ とする。 $r \geq s$ ならば、 $X_r - \hat{X}_r \perp L(Y, r)$ となり、1次元ブラウン運動 V が独立増分をもつことから、次の等式を得る。

$$\begin{aligned}
E[(N_t - N_s)Z] &= E\left[\left(\int_s^t G(r)(X_r - \hat{X}_r)dr + \int_s^t \sigma(r)dV_r\right)Z\right] \\
&= \int_s^t G(r)E\left[(X_r - \hat{X}_r)Z\right]dr + E\left[\left(\int_s^t \sigma(r)dV_r\right)Z\right] = 0.
\end{aligned}$$

$N_u \in L(Y, u), u \geq 0$, であるから、これより、 N_t が直交増分をもつことが分かる。

(2) 伊藤の公式を $g(t, x) = x^2$ として適用すると、以下が成り立つ。

$$d(N_t^2) = 2N_t dN_t + \frac{1}{2}2(dN_t)^2 = 2N_t dN_t + \sigma^2 dt$$

$$E(N_t^2) = E\left[\int_0^t 2N_s dN_s\right] + \int_0^t \sigma^2(s) ds$$

$$\int_0^t N_s dN_s = \lim \sum N_{t_j} [N_{t_{j+1}} - N_{t_j}]$$

上記の \lim は、時間区間を限りなく一様に細分化したものの極限である。

さらに N は直交増分をもつから、

$$E\left[\int_0^t 2N_s dN_s\right] = 0.$$

よって、 $E[N_t^2] = \int_0^t \sigma^2(s) ds$ が証明された。

(3) あらゆる線形結合 $c_0 + c_1 Y_{t_1} + \dots + c_k Y_{t_k}$ ($0 \leq t_i \leq T, c_j \in \mathbf{R}$) のなす線形空間の $L^2(P)$ での閉包を $L(Y, T)$ と定義すると、 $L(Y, T)$ に属する関数について、次のとおり表現できる。

$$L(Y, T) = \left\{c_0 + \int_0^T f(t) dY_t; f \in L^2[0, T], c_0 \in \mathbf{R}\right\}.$$

よって、各 $s < r < t$ に対して、

$$G(r)\hat{X}_r = c(r) + \int_0^r g(r, s) dY_s.$$

となる $g(r, \cdot) \in L^2[0, r], c(r) \in \mathbf{R}$ が存在し、関数 $f \in L^2[0, T]$ に対し、

$$\begin{aligned} \int_0^t f(s) dN_s &= \int_0^t f(s) dY_s - \int_0^t f(r) G(r) \hat{X}_r dr = \int_0^t f(s) dY_s - \int_0^t f(r) \left[\int_0^r g(r,s) dY_s \right] dr - \int_0^t f(r) c(r) dr \\ &= \int_0^t \left[f(s) - \int_s^t f(r) g(r,s) dr \right] dY_s - \int_0^t f(r) c(r) dr \end{aligned}$$

となる。ヴォルテラ型積分方程式の一般論から、任意の $h \in L^2[0,t]$ について、

$$f(s) - \int_s^t f(r) g(r,s) dr = h(s)$$

となる $f \in L^2[0,t]$ が存在する (Øksendal(1998))。とくに、 $h = \chi_{[0,t_1]}(0 \leq t_1 \leq t)$ とし、対応する f をとれば、

$$\int_0^t f(s) dN_s + \int_0^t f(r) c(r) dr = \int_0^t \chi_{[0,t_1]}(s) dY_s = Y_{t_1}$$

となる。したがって、 $L(N,t) \supset L(Y,t)$ が成り立つ。

また、 $L(N,t) \subset L(Y,t)$ が成り立つことは自明である。よって、 $L(N,t) = L(Y,t)$ ($\forall t \geq 0$) である。

(「証明」終わり)

さらに、 $\sigma(t)$ が有界区間上一様に正であることにより、新たな確率過程 $R_t(\omega)$ を次の式で定義する。

$$dR_t = \frac{1}{\sigma(t)} dN_t(\omega), \quad t \geq 0, \quad R_0 = 0. \quad (\text{A.5.8})$$

この R_t については、以下が成り立つ。

- $$\left\{ \begin{array}{l} \text{(i) } R_t \text{ は連続的な経路をもつ.} \\ \text{(ii) } R_t \text{ は直交増分をもつ.} \\ \text{(iii) } E[R_t] = 0 \quad \text{かつ} \quad E[R_t R_s] = \min(s, t). \end{array} \right.$$

ここで、

$$L(R,t) = L(N,t) = L(Y,t)$$

が成り立ち、

$$P_{L(R,t)}(X_t) = \hat{X}_t = P_{L(N,t)}(X_t)$$

となる。

なお、(i), (ii) の成立は、 N_t の性質から明らかである。

・(iii)の証明

$E[R_t] = 0$ は、 $E[N_t] = 0$ より明らか。

さらに、伊藤の公式から、

$$d(R_t^2) = 2R_t dR_t + (dR_t)^2 = 2R_t dR_t + dt$$

が成り立つ。

R_t は直交増分をもち、

$$E[R_t^2] = E\left[\int_0^t ds\right] = t$$

となる。

R_t が直交増分を持つので、右辺の第一項の期待値はゼロ ($E[R_t dR_t] = 0$) である。

ゆえに、 $s < t$ ならば、

$$E[R_t R_s] = E[(R_t - R_s)R_s] + E[R_s^2] = E[R_s^2] = s \text{ となる。} \quad (\text{「証明」終わり})$$

なお、最良の線形推定量 \hat{X}_t については、

$$\hat{X}_t = E[X_t] + \int_0^t \frac{\partial}{\partial s} E[X_t R_s] dR_s \quad (\text{A.5.9})$$

と表現できる。

$\frac{\partial}{\partial s} E[X_t R_s] = f(s, t)$ とおけば、

$$L(R, t) = \left\{ c_0 + \int_0^t f(s, t) dR_s; f(\cdot, t) \in L^2[0, t], c_0 \in \mathbf{R} \right\}$$

が成り立つが、 R は、既にみたように標準ブラウン運動（有界マルチンゲール）であり、上式の $f \in L^2[0, t]$, $c_0 \in \mathbf{R}$ が存在し、(A.5.9) を踏まえて、

$$\hat{X}_t = c_0(t) + \int_0^t f(s, t) dR_s, \quad c_0(t) = E[X_t] \quad (\text{A.5.10})$$

と整理できる（マルチンゲールの表現定理）。

したがって、(A.5.10) の期待値をとれば、 $c_0(t) = E[\hat{X}_t] = E[X_t]$ を得る。

さらに、(A.5.10) から、状態推定量 \hat{X}_t の変化について、

$$d\hat{X}_t = c'_0(t)dt + f(t,t)dR_t + \left(\int_0^t \frac{\partial}{\partial t} f(s,t)dR_s \right) dt \quad (\text{A.5.11})$$

という等式が得られる。

なお、 R_s については、(A.5.7) と (A.5.8) により、

$$R_s = \int_0^s \frac{G(u)}{\sigma(u)} (X_u - \hat{X}_u) du + V_s$$

となるため、

$$X_u - \hat{X}_u = e_u$$

とおけば、

$$E[X_t R_s] = \int_0^s \frac{G(u)}{\sigma(u)} E[X_t e_u] du$$

を得る。したがって、

$$f(s,t) = \frac{\partial}{\partial s} E[X_t R_s] = \frac{G(s)}{\sigma(s)} E[X_t e_s]$$

であり、同様に考えて、

$$f(t,t) = \frac{G(t)}{\sigma(t)} E[X_t e_t]$$

となるが、 $E[X_t e_t]$ については、 $\hat{X}_t \perp e_t$ であり、 $E[e_t] = 0$ であることから、

$$E[X_t e_t] = E[(\hat{X}_t + e_t)e_t] = E[\hat{X}_t e_t + e_t^2] = E[\hat{X}_t]E[e_t] + E[e_t^2] = E[e_t^2]$$

ここで、 $E[e_t^2]$ は $t \geq 0$ における平均自乗誤差であり、

$$f(t,t) = \frac{G(t)}{\sigma(t)} E[e_t^2]$$

と表すことができる。

これを踏まえて、(A.5.11) について状態空間表現 (A.5.4)、(A.5.5) の G, F をあてはめ、さらに、 $E[e_t^2] = S(t)$ とおくと、

$$d\hat{X}_t = c'_0(t)dt + \frac{G(t)S(t)}{\sigma(t)} dR_t + \left(\int_0^t f(s,t)dR_s \right) F(t)dt$$

を得る。さらに、

$$\begin{aligned}
c_0'(t) &= \frac{d}{dt} c_0(t) = F(t)c_0(t), \\
dR_t &= \frac{1}{\sigma(t)} [dY_t - G(t)\hat{X}_t dt], \\
\left(\int_0^t f(s,t) dR_s \right) F(t) dt &= F(t)(\hat{X}_t - c_0(t)) dt,
\end{aligned}$$

という関係式を代入して、最終的に、

$$d\hat{X}_t = \left(F(t) - \frac{G^2(t)S(t)}{\sigma^2(t)} \right) \hat{X}_t dt + \frac{G(t)S(t)}{\sigma^2(t)} dY_t,$$

$$\text{初期条件 : } \hat{X}_0 = E[X_0]. \quad (\text{A.5.12})$$

なる確率微分方程式を得る。そして、この確率微分方程式から状態の最良の可測推定となる解： $\hat{X}_t = E[X_t | \mathcal{F}_t]$ が得られる。

この状態空間表現から推定された母数は最良の線形推定であり、同時に最良の可測推定となるため、ここでの方法に基づいて金利の確率微分方程式モデルの時変ドリフト項について適切に状態推定すれば最良可測推定量が得られ、原系列過程から当該ドリフトを減じた過程はマルチンゲール残差となる。

さらに、(A.5.8) の関係式のとおり、マルチンゲール残差を拡散係数で除すことにより、標準ブラウン運動に変換することも可能である。

本論では、上記の概念を参考にしつつ、さらにベイズ統計的方法を導入して、まったく新しい金利過程の測度（分布）変換の方法を提言する。

a.1.6. 補論のまとめ

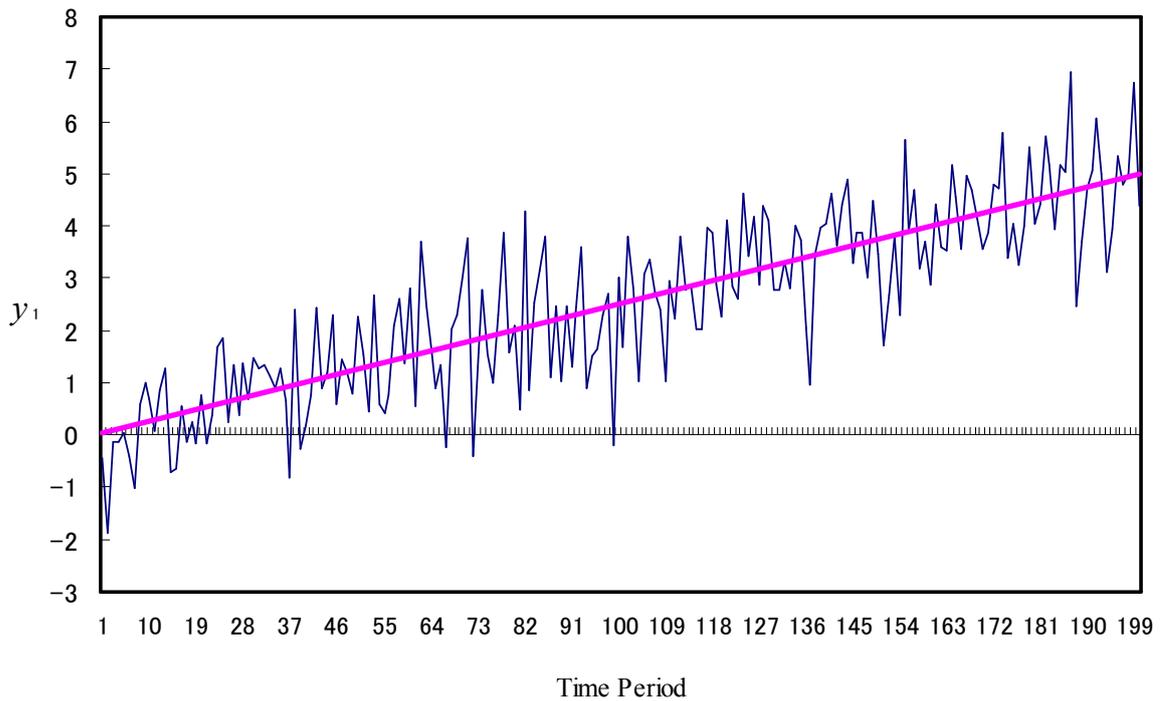
ここで、追加的な観測情報に基づく確率密度の逐次修正が不可能な場合、正規密度による評価が困難になることをごく単純で簡単な例を用いて提示する。

例えば、金融資産収益率の変化を定義する確率微分方程式について、あらためて以下の単純なモデルで表現できるものとする。

$$dr = \alpha dt + \sigma dB$$

ここに、 r : 金融資産収益率, α : ドリフト項,
 σ : ボラティリティ (拡散係数), dB : 標準ブラウン運動増分

一部の教科書や入門書で例示されているように、 α が確定的で定数、さらに $\sigma=1$ とおいた場合、金融資産収益率の変化 $dr(=y_1)$ について以下のようなサンプル・パスを得る。

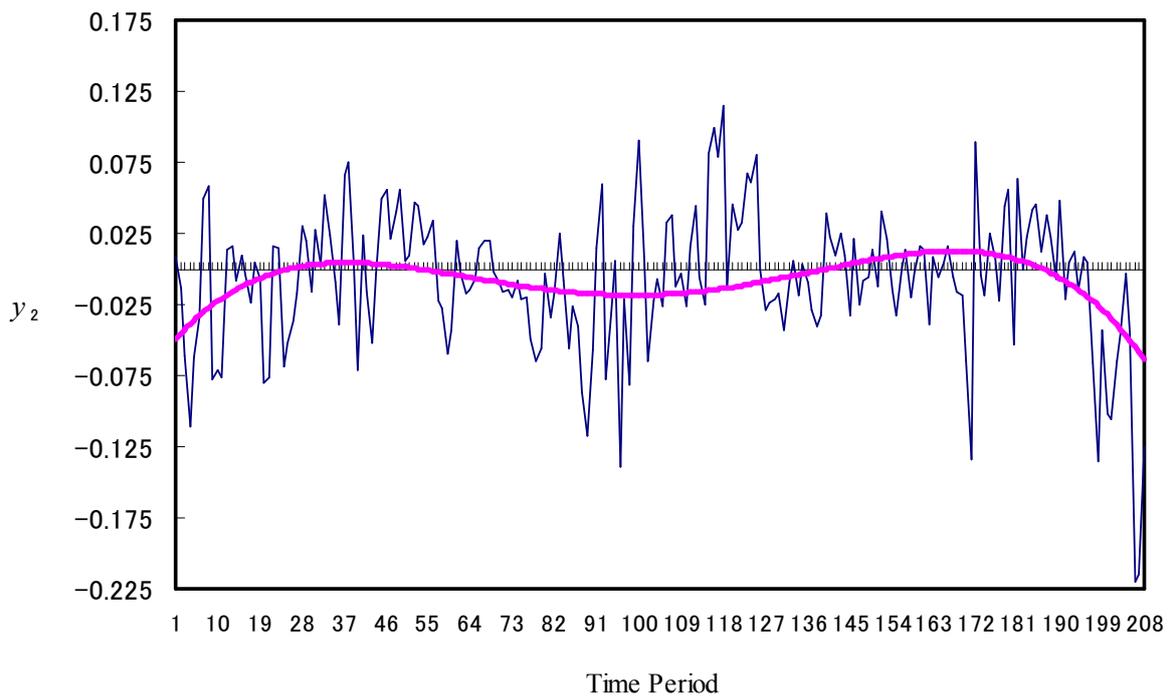


図A6-1 $\alpha = 0.025, \sigma = 1$ の場合のサンプル・パス ($N=200$)

したがって、このサンプル・パスからドリフトを減じれば、マルチンゲール・イノベーション過程 (標準正規過程) のサンプル・パスが得られ、標準的な正規密度に従って金融資産収益率の変化について評価できる。

しかし、問題は以下のようなサンプル・パスに直面した場合である^{*A2}。

*A2 このサンプル・パスは、既に3章で考察した「米国財務省証券6ヶ月物金利 (月次; 1984年8月~2001年11月, $N = 208$)」の自然対数変換値の一階階差データである。

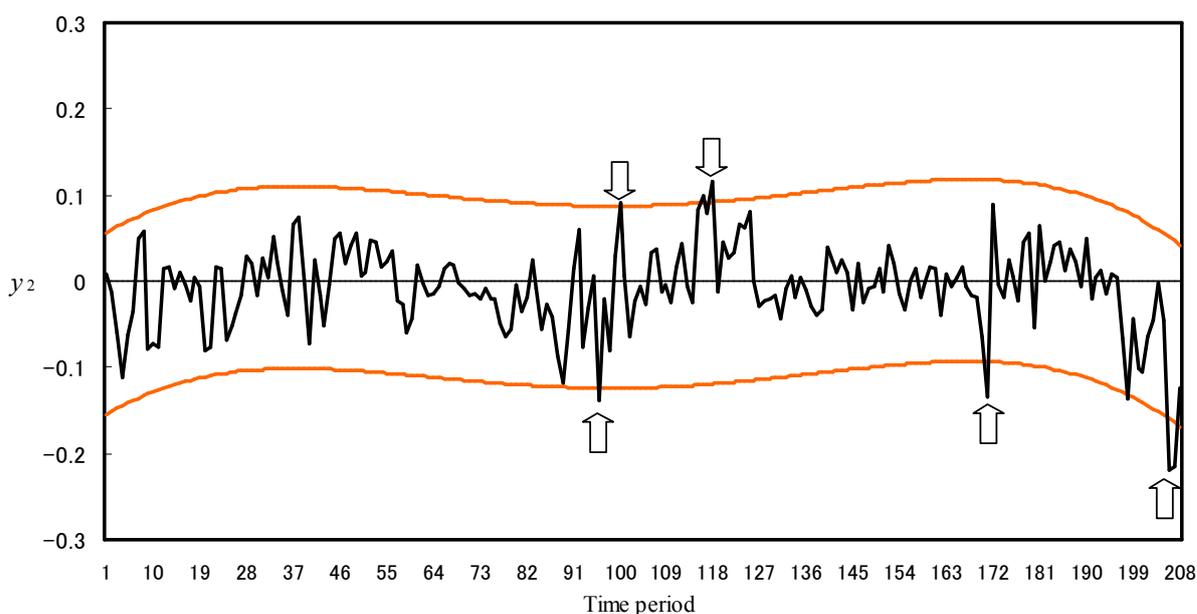


図A6-2 α と σ がともに未知のサンプル・パス ($N=208$)

上の例では、ドリフトが、緩やかながらやや湾曲的な形状で推移していることが想定されるため^{*A3}、ひとまず、単純な多項式推定を用いて非線形ドリフトを推定している。

この観測過程の確率密度について、あえて正規密度を仮定し、ドリフトと拡散係数（ボラティリティ）推定値を固定した上で当該観測データの確率密度（分布）を求めると次の図 A6-3 のとおりとなる。

*A3 このような傾向は一般に金利データ特有の特徴と考えられている。なお、ここでの推定ドリフトを観測データから減じてもマルチンゲール過程にはならない。また、線形トレンド（Black-Scholesモデルの様な定数としてのドリフト）をあてはめた場合にはドリフト期待値はきわめて緩やかな傾きを持った直線になる。



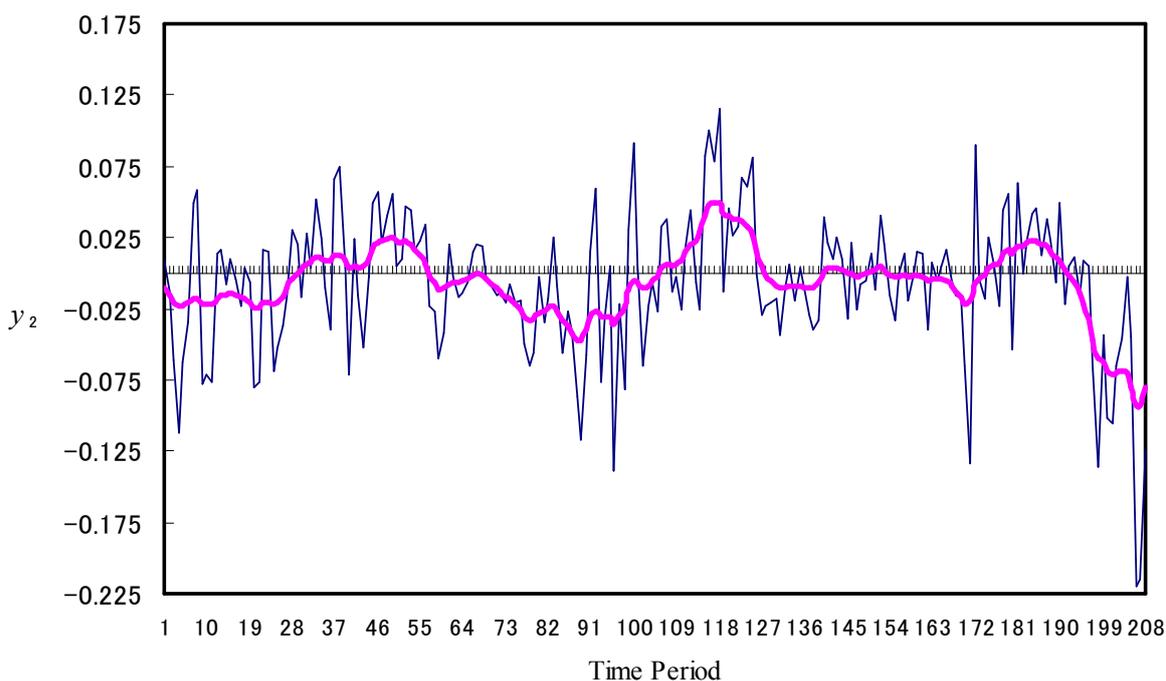
図A6-3 観測の推定確率密度 ($N=208$; 正規密度を仮定)

ここでは、確率 99.0 % の信頼区間（太線）内で計測できない観測データ (outliers) がいくつか（5個）見受けられており、とくに金利低下局面で大きな外れ値がある。このため、当該データを正規密度で十分に評価することはやや困難であると考えられる。

このように単純な多項式推定（＝事後的な観測情報の変化を反映させない時間のみの関数）では、観測や将来予測分布に関し、正規密度に基づく評価が困難となるおそれがある。

一方、同じ観測データについて DLM を用いてドリフト推定を行うと図 A6-4 の結果が得られる。DLM で推定されたドリフト推定値は推定時点までの増大する過去の観測情報をすべて織り込んだ推定量であり、 \mathcal{F}_t 可測のドリフト期待値となる^{*A4}。

*A4 具体的な推定モデルおよび方法は、本論の 2 章、3 章で詳しく論じられている。なお、DLM により推定されたドリフト期待値を観測過程から減じるとマルチンゲール過程が実現する（3 章で提示）。なお、ここでは「固定 smoother アルゴリズム」に基づく平滑化後のドリフト期待値を提示した。



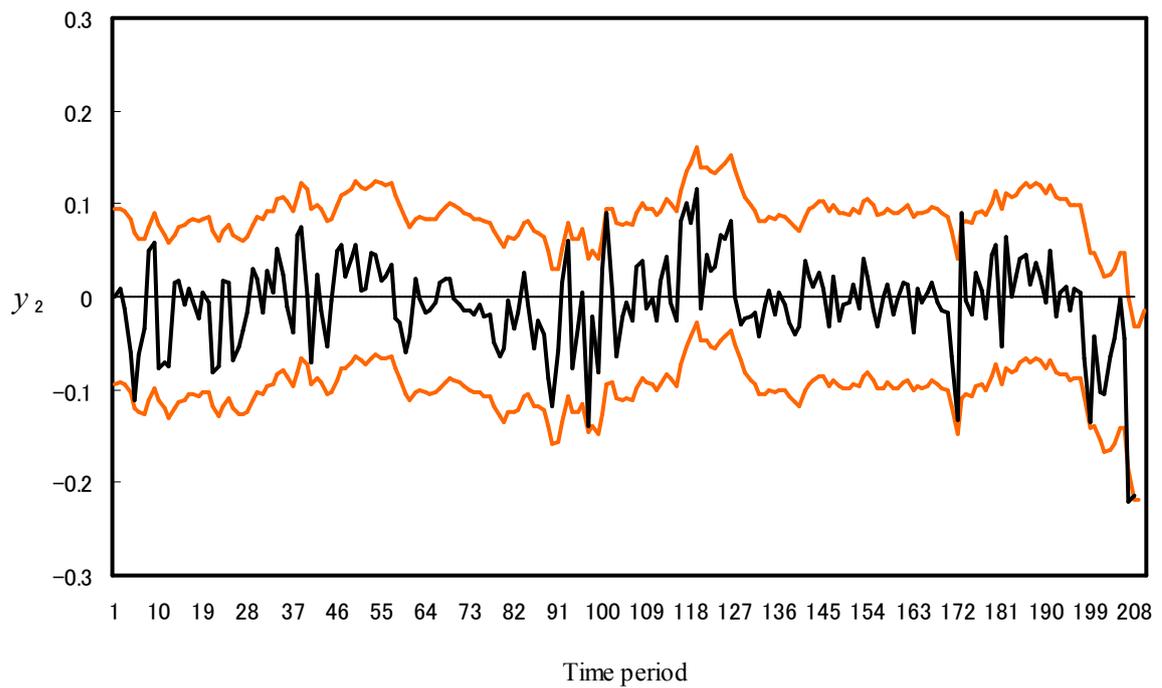
図A6-4 DLMによる \mathcal{F}_t -可測ドリフト期待値の推定結果 ($N=208$)

上の図では、ドリフト推定値がサンプル・パスの変化に即応してシフトしていることが分かる。

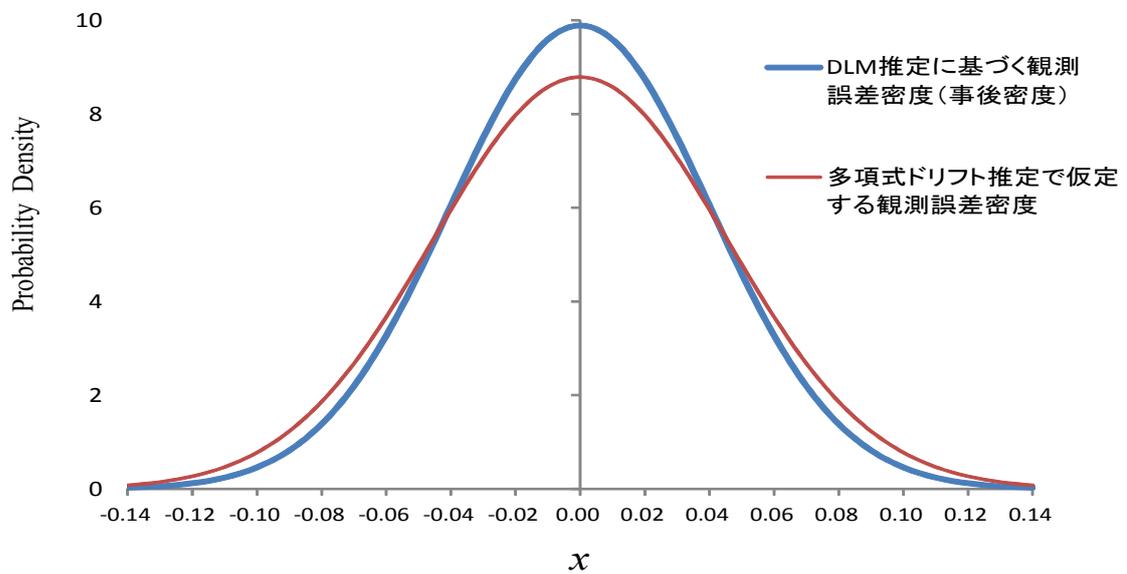
さらに、DLMでは、観測過程の「事後密度」を正規密度として適切に把握することができる。この図では、先ほどの例と異なり、概ねすべての観測データが確率99.0%の確率密度内に適合している。したがって、見かけ上変動幅が大きいデータに対しても、情報変化に即応して適切に母数が逐次修正されて推定されることにより、事後的にみて正規密度で評価することが可能となっている^{*A5}。

なお、DLMによる未知母数推定の観測誤差密度の方が裾が狭くなっており、正規密度に適合するとともに、より効率的な推定結果を併せて得ることができる。

*A5 図A6-5で表示した確率密度は、観測に係る事後密度である。ベイズ統計学では、観測とモデル未知母数の双方について、「事前分布」、「予測分布」、「事後分布」を得ることができる。「事後分布」は「事前分布」設定以降の情報変化を合理的に反映させた分布となる。なお、本論の研究では、「予測分布」を設定するさいに、新たにボラティリティ（拡散係数）に関する予測期待値を更新するモデルを採用している。当該モデルについては2章で詳しく論じられている。



図A6-5 DLM推定に基づく観測の事後確率密度 (N=208)



図A6-6 観測誤差密度

この結果から、DLMによる未知母数推定により、効率的な確率密度（分布）の把握が可能となる事が分かる。

A.2. 「2章」の補論

a.2.1. ノーマル・ガンマ分布

一般に、

$$\begin{cases} X | \sigma^{-2} \sim N(\mathbf{m}, \sigma^2 \Sigma) & \mathbf{m} \in \mathbf{R}^h, \Sigma; h\text{次正定符号行列} \\ \sigma^{-2} \sim \text{Gamma}(v/2, d/2) & v > 0 \text{ and } d > 0, \end{cases}$$

ここに、 $E[X] \equiv E[X | \sigma^{-2}] = \mathbf{m}$, $\text{Var}[X | \sigma^{-2}] = \sigma^2 \Sigma$.

のとき、 (X, σ^{-2}) の従う分布をノーマル・ガンマ分布といい、 $N\text{-Gamma}(\mathbf{m}, \Sigma, \frac{v}{2}, \frac{d}{2})$ と記す。

(i) (X, σ^{-2}) の密度関数は、

$$\begin{aligned} p(\mathbf{x}, \sigma^{-2}) &= (2\pi)^{-\frac{h}{2}} |\sigma^2 \Sigma|^{-\frac{h}{2}} \exp\left\{-\frac{\sigma^{-2}}{2} [(\mathbf{x} - \mathbf{m})' \Sigma^{-1} (\mathbf{x} - \mathbf{m})]\right\} \\ &\quad \times \frac{d^{v/2}}{2^{v/2} \Gamma(\frac{v}{2})} (\sigma^{-2})^{\frac{v}{2}-1} \exp\left\{-\frac{d}{2} \sigma^{-2}\right\} \\ &\propto (\sigma^{-2})^{\frac{v+h}{2}-1} \exp\left\{-\frac{\sigma^{-2}}{2} [(\mathbf{x} - \mathbf{m})' \Sigma^{-1} (\mathbf{x} - \mathbf{m}) + d]\right\}, \quad \sigma^{-2} > 0 \end{aligned}$$

で与えられ、

(ii) したがって、 $X = \mathbf{x}$ を得たときの σ^{-2} の条件付き密度は、

$$\begin{aligned} p(\sigma^{-2} | \mathbf{x}) &\propto d^{*\frac{v^*}{2}} (\sigma^{-2})^{\frac{v^*}{2}-1} \exp\left\{-\frac{d^*}{2} \sigma^{-2}\right\}, \quad \sigma^{-2} > 0, \\ \sigma^{-2} | \mathbf{x} &\sim \text{Gamma}\left(\frac{v^*}{2}, \frac{d^*}{2}\right), \quad \text{ここに、} v^* = v + h, \quad d^* = d + (\mathbf{x} - \mathbf{m})' \Sigma^{-1} (\mathbf{x} - \mathbf{m}). \end{aligned}$$

で与えられる。

(iii) また、 X の周辺分布は、

$$\begin{aligned} p(\mathbf{x}) &= p(\mathbf{x}, \sigma^{-2}) / p(\sigma^{-2} | \mathbf{x}) \propto [v + (\mathbf{x} - \mathbf{m})' \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m})]^{-(v+h)/2}, \\ &\quad \propto [1 + \frac{1}{v} (\mathbf{x} - \mathbf{m})' \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m})]^{-(v+h)/2}, \quad \mathbf{x} \in \mathbf{R}^h. \end{aligned}$$

ここに、 $\mathbf{C} = E[\sigma^2] \Sigma$

と、 h 次元の Student- t 分布にしたがう。

(iv) さらに、 $\sigma^{-2} \sim \text{Gamma}(\frac{v}{2}, \frac{d}{2})$ より、 $E[\sigma^{-2}] = \frac{v}{d}$ は、容易に示される。

本論で示される情報に係る分布の取り扱いに関しては、基本的に、(i) ~ (iii) の論理が用いられる。

a.2.2. (参考) ガンマ分布と逆ガンマ分布

(1) ガンマ分布

確率密度関数が、正数 α, β を母数として、

$$f(x | \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) \cdot \frac{1}{\beta} I_{(0, \infty)}(x)$$

である確率分布を母数 α, β のガンマ分布といい、これを $\text{Gamma}(\alpha, \beta)$ で表す。

α を形の母数、 β を尺度母数という。

なお、 $\Gamma(\alpha)$ はガンマ関数であり、

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad (\alpha > 0)$$

で定義される。

$X \sim \text{Gamma}(\alpha, \beta)$ のとき、 $Z = X / \beta \sim \text{Gamma}(\alpha, 1)$ であるから、

$$\begin{aligned} E[X^n] &= E[(\beta Z)^n] = \beta^n E[Z^n] \\ &= (\alpha + n - 1)(\alpha + n - 2) \cdots \alpha \beta^n \quad (n = 1, 2, \dots) \end{aligned}$$

$$E[X] = \alpha\beta, \quad \text{Var}[X] = \alpha\beta^2$$

となる。

本論では、「精度 (precision : 観測誤差分散の逆数)」の分布について、

$$\sigma_t^{-2} | D_t \sim \text{Gamma}\left(\frac{\nu_t}{2}, \frac{d_t}{2}\right)$$

とおいているが、 $\nu_t = 2\alpha$ 、 $d_t = 2\beta^{-1}$ であり、ここから、

$$E[\sigma_t^{-2} | D_t] = \frac{\nu_t}{d_t}$$

が得られる。

(2) 逆ガンマ分布

$Z \sim \text{Gamma}(\alpha, \beta)$ のとき、確率変数 $X = Z^{-1}$ のしたがう分布を逆ガンマ分布といい、 $\text{Gamma}^{-1}(\alpha, \beta)$ と表す。その密度関数は、

$$f(x | \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \left(\frac{1}{\beta x}\right)^{\alpha-1} \exp\left(-\frac{1}{\beta x}\right) \cdot \frac{1}{\beta x^2} I_{(0, \infty)}(x)$$

である。

$X \sim \text{Gamma}^{-1}(\alpha, \beta)$ のとき、

$$E[X^n] = E[Z^{-n}] = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} z^{-n} \left(\frac{z}{\beta}\right)^{\alpha-1} \exp\left(-\frac{z}{\beta}\right) \frac{dz}{\beta} = \frac{\Gamma(\alpha-n)}{\Gamma(\alpha)} \beta^{-n} \quad (n < \alpha)$$

これから、

$$E[X] = \frac{1}{(\alpha-1)\beta} \quad (\alpha > 1), \quad \text{Var}[X] = \frac{1}{(\alpha-1)^2(\alpha-2)\beta^2} \quad (\alpha > 2)$$

が得られる。

先にみたとおり、 $\sigma_t^{-2} | D_t \sim \text{Gamma}\left(\frac{v_t}{2}, \frac{d_t}{2}\right)$ であるとき、 $\sigma_t^2 | D_t \sim \text{Gamma}^{-1}\left(\frac{v_t}{2}, \frac{d_t}{2}\right)$ と考えれば、 $v_t = 2\alpha$, $d_t = 2\beta^{-1}$ に基づき、

$$E[\sigma_t^2 | D_t] = \frac{d_t}{v_t - 1}$$

が得られる^{*A6}。

*A6 逆ガンマ分布の期待値は、推測統計学の「不偏分散」の定義に類するものになる。

なお、West and Harrison (1997) は、「精度 (観測誤差分散母数の逆数)」の期待値についてひとまず、 $\frac{v_t}{d_t} \stackrel{\text{def}}{=} \frac{1}{s_t}$ と定義し、後に逆ガンマ分布の期待値と一致する様に観測誤差散を再定義している。本論も West and Harrison (1997) に準拠している。

A.2.3. 情報割引率について

情報割引率 δ_t は、時変ドリフトの形状を決定する重要な要素である。本論の情報割引率は、次のとおり、状態誤差分散行列を規定するものとして定義している。

$$\mathbf{W}_t \stackrel{def}{=} \zeta_t^2 \mathbf{W}_t^\dagger = \delta_t \sigma_{t-1}^2 \mathbf{W}_t^\dagger, \quad \mathbf{W}_t^\dagger = \mathbf{I}_h, \quad \mathbf{I}_h \text{ は } h \text{ 次元単位行列}$$

この定義により、状態誤差分散 ζ_t^2 については、 $\zeta_t^2 = \delta_t \sigma_{t-1}^2$ と、1 期前の観測誤差分散 σ_{t-1}^2 に、情報割引率 δ_t を掛けた値としている。

a.2.4. 本論のDLM体系について

ここでは、本論の DLM の体系におけるモデルの情報更新等に関して、理論的な考察を行う。

(1) 統計モデルの基本的な組み合わせ

まず、分散母数について未知とする前提の下で、以下の①～⑤に示す統計モデルの組み合わせを考える。なお、各変数および母数の定義は、本論と同じである。この統計モデルの組み合わせを元にして、(2)～(10)のバリエーションが得られる。なお、単純化のために、情報割引率 δ_t については常に 1 と仮定し、 $\sigma_{t-1}^2 = \zeta_t^2$ とおく。

$$\textcircled{1} Y_t | (t, \sigma_t^{-2}) \sim N(\mathbf{F}'t_t, \sigma_t^2) \quad \text{[本論(2.1)式]}$$

$$\begin{aligned} \textcircled{2} T_t | (D_{t-1}, t_{t-1}, \sigma_{t-1}^{-2}) &\sim N(\mathbf{G}t_{t-1}, \zeta_t^2 \mathbf{W}_t^\dagger) \\ &= N(\mathbf{G}t_{t-1}, \sigma_{t-1}^2 \mathbf{W}_t^\dagger) \quad \text{[単純化のため、}\delta_t\text{を1と仮定]} \end{aligned}$$

$$\textcircled{3} T_{t-1} | (D_{t-1}, \sigma_{t-1}^{-2}) \sim N(\mathbf{m}_{t-1}, \sigma_{t-1}^2 \mathbf{C}_{t-1}^\dagger)$$

$$\textcircled{4} \sigma_t^{-2} | D_t \sim \text{Gamma}(v_t/2, d_t/2) \text{ for any } v > 0 \text{ and } d > 0$$

[本論(2.4)式]

$$\text{ここから、} E[\sigma_t^{-2} | D_t] = \frac{v_t}{d_t} \stackrel{def}{=} \frac{1}{S_t} \quad \text{[West and Harrison (1997) の「精度」の定義に準拠]}$$

$$\textcircled{5} E[\sigma_t^2 | D_{t-1}, \sigma_{t-1}^{-2}] \stackrel{def}{=} \sigma_t^{*2} = \sigma_{t-1}^2 \left(\frac{Q_t^*}{Q_t} \right)$$

$$\left(\begin{array}{l} \text{上記から、} \\ E[\sigma_t^{-2} | D_{t-1}, \sigma_{t-1}^{-2}] \stackrel{def}{=} \sigma_t^{*-2} = \left[\sigma_{t-1}^2 \left(\frac{Q_t^*}{Q_t} \right) \right]^{-1} \text{ と考える。} \end{array} \right) \quad \text{[モデルの基本的な仮定]}$$

(2) 状態更新式(分散母数の条件がない場合)

$$\mathbf{T}_t | (D_{t-1}, \mathbf{t}_{t-1}, \sigma_{t-1}^{-2}) \sim N(\mathbf{G}\mathbf{t}_{t-1}, \sigma_{t-1}^{-2} \mathbf{W}_t^\dagger)$$

$$\sigma_{t-1}^{-2} | D_{t-1} \sim \text{Gamma}(v_{t-1}/2, d_{t-1}/2)$$

より、

$$p(\mathbf{t}_t, \sigma_{t-1}^{-2} | D_{t-1}, \mathbf{t}_{t-1})$$

$$\stackrel{\text{def}}{=} n(\mathbf{t}_t | \mathbf{G}\mathbf{t}_{t-1}, \sigma_{t-1}^{-2} \mathbf{W}_t^\dagger) \cdot \text{gamma}(\sigma_{t-1}^{-2} | v_{t-1}/2, d_{t-1}/2) \quad [\text{ノーマル・ガンマ密度関数}]$$

$$= (2\pi)^{-\frac{h}{2}} |\sigma_{t-1}^{-2} \mathbf{W}_t^\dagger|^{-\frac{h}{2}} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} [(\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})' \mathbf{W}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})]\right\} \cdot \frac{d_{t-1}^{\frac{v_{t-1}}{2}}}{2^{\frac{v_{t-1}}{2}} \Gamma\left(\frac{v_{t-1}}{2}\right)} (\sigma_{t-1}^{-2})^{\frac{v_{t-1}-1}{2}} \exp\left\{-\frac{d_{t-1}}{2} \sigma_{t-1}^{-2}\right\}$$

次に、 $\mathbf{T}_t = \mathbf{t}_t$ を得た後、 $\sigma_{t-1}^{-2} | (D_{t-1}, \mathbf{t}_t, \mathbf{t}_{t-1})$ の条件付き密度について、

$$p(\sigma_{t-1}^{-2} | D_{t-1}, \mathbf{t}_t, \mathbf{t}_{t-1})$$

$$\stackrel{\text{def}}{=} \text{gamma}\left(\sigma_{t-1}^{-2} \left| \frac{v_{t-1} + h}{2}, \frac{d_{t-1} + [(\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})' \mathbf{W}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})]}{2} \right.\right)$$

$$= \frac{[d_{t-1} + (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})' \mathbf{W}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})]^{\frac{v_{t-1}+h}{2}}}{2^{\frac{v_{t-1}+h}{2}} \Gamma\left(\frac{v_{t-1}+h}{2}\right)} (\sigma_{t-1}^{-2})^{\frac{v_{t-1}+h-1}{2}} \exp\left\{-\frac{d_{t-1} + [(\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})' \mathbf{W}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})]}{2} \sigma_{t-1}^{-2}\right\}$$

とおけば、

$$\begin{aligned} p(\mathbf{t}_t | D_{t-1}, \mathbf{t}_{t-1}) &= \int_0^{+\infty} p(\mathbf{t}_t, \sigma_{t-1}^{-2} | D_{t-1}, \mathbf{t}_{t-1}) d\sigma_{t-1}^{-2} \\ &= p(\mathbf{t}_t, \sigma_{t-1}^{-2} | D_{t-1}, \mathbf{t}_{t-1}) / p(\sigma_{t-1}^{-2} | D_{t-1}, \mathbf{t}_t, \mathbf{t}_{t-1}) \end{aligned}$$

$$\begin{aligned}
& (2\pi)^{-\frac{h}{2}} |\sigma_{t-1}^2 \mathbf{W}_t^\dagger|^{-\frac{h}{2}} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} [(\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})' \mathbf{W}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})]\right\} \cdot \frac{d_{t-1}^{\frac{\nu_{t-1}}{2}}}{2^{\frac{\nu_{t-1}}{2}} \Gamma\left(\frac{\nu_{t-1}}{2}\right)} (\sigma_{t-1}^{-2})^{\frac{\nu_{t-1}-1}{2}} \exp\left\{-\frac{d_{t-1}}{2} \sigma_{t-1}^{-2}\right\} \\
& = \frac{[d_{t-1} + (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})' \mathbf{W}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})]^{\frac{\nu_{t-1}+h}{2}} (\sigma_{t-1}^{-2})^{\frac{\nu_{t-1}+h}{2}-1} \exp\left\{-\frac{d_{t-1} + [(\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})' \mathbf{W}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})]}{2} \sigma_{t-1}^{-2}\right\}}{2^{\frac{\nu_{t-1}+h}{2}} \Gamma\left(\frac{\nu_{t-1}+h}{2}\right)} \\
& \propto \frac{(\sigma_{t-1}^{-2})^{\frac{\nu_{t-1}+h}{2}-1} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} [d_{t-1} + (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})' \mathbf{W}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})]\right\}}{[d_{t-1} + (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})' \mathbf{W}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})]^{\frac{\nu_{t-1}+h}{2}} (\sigma_{t-1}^{-2})^{\frac{\nu_{t-1}+h}{2}-1} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} [d_{t-1} + (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})' \mathbf{W}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})]\right\}} \\
& \propto [d_{t-1} + (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})' \mathbf{W}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})]^{-\frac{\nu_{t-1}+h}{2}} \\
& \propto \left[1 + \frac{1}{\nu_{t-1}} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})' \left(\frac{\nu_{t-1}}{d_{t-1}}\right) \mathbf{W}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})\right]^{-\frac{\nu_{t-1}+h}{2}} \\
& \propto \left[1 + \frac{1}{\nu_{t-1}} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})' (S_{t-1} \mathbf{W}_t^\dagger)^{-1} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})\right]^{-\frac{\nu_{t-1}+h}{2}} \\
& \propto \left[1 + \frac{1}{\nu_{t-1}} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})' \mathbf{W}_t^{-1} (\mathbf{t}_t - \mathbf{G}\mathbf{t}_{t-1})\right]^{-\frac{\nu_{t-1}+h}{2}}
\end{aligned}$$

よって、

$$\mathbf{T}_t | (D_{t-1}, \mathbf{t}_{t-1}) \sim t(\mathbf{G}\mathbf{t}_{t-1}, \mathbf{W}_t, h, \nu_{t-1})$$

ここに、 $\mathbf{W}_t = \left(\frac{d_{t-1}}{\nu_{t-1}}\right) \mathbf{W}_t^\dagger \stackrel{\text{def}}{=} \left(\frac{d_{t-1}}{\nu_{t-1}}\right) \mathbf{I}_h = S_{t-1} \mathbf{I}_h$ 、 \mathbf{I}_h は、 h 次元単位行列
(d_t の定義については、後述)

$$E[\mathbf{T}_t | D_{t-1}, \mathbf{t}_{t-1}] = \mathbf{G}\mathbf{t}_{t-1}, \quad \text{Var}[\mathbf{T}_t | D_{t-1}, \mathbf{t}_{t-1}] = \left(\frac{\nu_{t-1}}{\nu_{t-1}-1}\right) \mathbf{W}_t \quad (\nu_0 > 2).$$

これは、本論の (2.2) 式に該当する。

(3) 状態予測式 (分散母数の条件がある場合)

$$\mathbf{T}_t | (D_{t-1}, \sigma_{t-1}^{-2}, \mathbf{t}_{t-1}) \sim N(\mathbf{G}\mathbf{t}_{t-1}, \sigma_{t-1}^2 \mathbf{W}_t^\dagger)$$

$$\mathbf{T}_{t-1} | (D_{t-1}, \sigma_{t-1}^{-2}) \sim N(\mathbf{m}_{t-1}, \sigma_{t-1}^2 \mathbf{C}_{t-1}^\dagger)$$

より、

$$\begin{aligned}
p(\mathbf{t}_t | D_{t-1}, \sigma_{t-1}^2) &= \int_{\mathbf{R}^h} p(\mathbf{t}_t | D_{t-1}, \sigma_{t-1}^2, \mathbf{t}_{t-1}) p(\mathbf{t}_{t-1} | D_{t-1}, \sigma_{t-1}^2) d\mathbf{t}_{t-1} \\
&\stackrel{def}{=} \int_{\mathbf{R}^h} n(\mathbf{t}_t | \mathbf{G}\mathbf{t}_{t-1}, \sigma_{t-1}^2 \mathbf{W}_t^\dagger) n(\mathbf{t}_{t-1} | \mathbf{m}_{t-1}, \sigma_{t-1}^2 \mathbf{C}_{t-1}^\dagger) d\mathbf{t}_{t-1} \\
&\propto \exp \left\{ -\frac{\sigma_{t-1}^2}{2} [(\mathbf{t}_t - \mathbf{a}_t)' \mathbf{R}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{a}_t)] \right\}
\end{aligned}$$

ここに、

$$\begin{aligned}
\mathbf{a}_t &= \mathbf{G}\mathbf{m}_{t-1} \\
\mathbf{R}_t^\dagger &= (\mathbf{G}\mathbf{C}_{t-1}^\dagger \mathbf{G}' + \mathbf{W}_t^\dagger) \quad *A7 \\
\mathbf{T}_t | (D_{t-1}, \sigma_{t-1}^2) &\sim N(\mathbf{a}_t, \sigma_{t-1}^2 \mathbf{R}_t^\dagger)
\end{aligned}$$

$$\begin{aligned}
*A7 \quad E(\tilde{\mathbf{t}}_t | D_{t-1}, \sigma_{t-1}^2) &= E(E(\tilde{\mathbf{t}}_t | \mathbf{G}, D_{t-1}, \sigma_{t-1}^2, \tilde{\mathbf{t}}_{t-1}) | D_{t-1}, \sigma_{t-1}^2) = E(\mathbf{G}\tilde{\mathbf{t}}_{t-1} | D_{t-1}, \sigma_{t-1}^2) \\
&= \mathbf{G}E(\tilde{\mathbf{t}}_{t-1} | D_{t-1}, \sigma_{t-1}^2) = \mathbf{G}\mathbf{m}_{t-1} \stackrel{def}{=} \mathbf{a}_t \quad \text{とおいた。}
\end{aligned}$$

また、

$$\begin{aligned}
E(\tilde{\mathbf{t}}_t \tilde{\mathbf{t}}_t' | D_{t-1}, \sigma_{t-1}^2, \mathbf{t}_{t-1}) &= E((\mathbf{G}\mathbf{t}_{t-1} + \tilde{\mathbf{w}}_t)(\mathbf{G}\mathbf{t}_{t-1} + \tilde{\mathbf{w}}_t)' | D_{t-1}, \sigma_{t-1}^2, \mathbf{t}_{t-1}) \\
&= E(\mathbf{G}\mathbf{t}_{t-1} \mathbf{t}_{t-1}' \mathbf{G}' + 2\mathbf{G}\mathbf{t}_{t-1} \tilde{\mathbf{w}}_t' + \tilde{\mathbf{w}}_t \tilde{\mathbf{w}}_t' | D_{t-1}, \sigma_{t-1}^2, \mathbf{t}_{t-1}) \\
&= \mathbf{G}\mathbf{t}_{t-1} \mathbf{t}_{t-1}' \mathbf{G}' + \sigma_{t-1}^2 \mathbf{W}_t^\dagger \\
E(\tilde{\mathbf{t}}_{t-1} \tilde{\mathbf{t}}_{t-1}' | D_{t-1}, \sigma_{t-1}^2) &= \sigma_{t-1}^2 \mathbf{C}_{t-1}^\dagger + \mathbf{m}_{t-1} \mathbf{m}_{t-1}', \quad (\sigma_{t-1}^2 \mathbf{C}_{t-1}^\dagger \stackrel{def}{=} \text{Var}(\tilde{\mathbf{t}}_{t-1} | D_{t-1}, \sigma_{t-1}^2)) \\
E(\tilde{\mathbf{t}}_t \tilde{\mathbf{t}}_t' | D_{t-1}, \sigma_{t-1}^2) &= E(\mathbf{G}\tilde{\mathbf{t}}_{t-1} \tilde{\mathbf{t}}_{t-1}' \mathbf{G}' + \sigma_{t-1}^2 \mathbf{W}_t^\dagger | D_{t-1}, \sigma_{t-1}^2) \\
&= \mathbf{G}E(\tilde{\mathbf{t}}_{t-1} \tilde{\mathbf{t}}_{t-1}' | D_{t-1}, \sigma_{t-1}^2) \mathbf{G}' + \sigma_{t-1}^2 \mathbf{W}_t^\dagger \\
&= \mathbf{G}(\sigma_{t-1}^2 \mathbf{C}_{t-1}^\dagger + \mathbf{m}_{t-1} \mathbf{m}_{t-1}') \mathbf{G}' + \sigma_{t-1}^2 \mathbf{W}_t^\dagger \\
\text{Var}(\tilde{\mathbf{t}}_t | D_{t-1}, \sigma_{t-1}^2) &= E(\tilde{\mathbf{t}}_t \tilde{\mathbf{t}}_t' | D_{t-1}, \sigma_{t-1}^2) - (E(\tilde{\mathbf{t}}_t | D_{t-1}, \sigma_{t-1}^2))^2 \\
&= \mathbf{G}(\sigma_{t-1}^2 \mathbf{C}_{t-1}^\dagger + \mathbf{m}_{t-1} \mathbf{m}_{t-1}') \mathbf{G}' + \sigma_{t-1}^2 \mathbf{W}_t^\dagger - \mathbf{G}\mathbf{m}_{t-1} \mathbf{m}_{t-1}' \mathbf{G}' \\
&= \sigma_{t-1}^2 (\mathbf{G}\mathbf{C}_{t-1}^\dagger \mathbf{G}' + \mathbf{W}_t^\dagger) \\
&= \sigma_{t-1}^2 \mathbf{R}_t^\dagger \quad (\mathbf{R}_t^\dagger = \mathbf{G}\mathbf{C}_{t-1}^\dagger \mathbf{G}' + \mathbf{W}_t^\dagger)
\end{aligned}$$

(上の表現では、とくに確率変数にティルド $[\sim]$ を付して区分している。
また、単純化のために、情報割引率 δ については 1 とおいた。)

(4) 状態予測式(分散母数の条件がない場合)

$$\begin{aligned} \mathbf{T}_t | (D_{t-1}, \sigma_{t-1}^{-2}) &\sim N(\mathbf{a}_t, \sigma_{t-1}^2 \mathbf{R}_t^\dagger) \\ \sigma_{t-1}^{-2} | D_{t-1} &\sim \text{Gamma}(v_{t-1}/2, d_{t-1}/2) \end{aligned}$$

より、

$$\begin{aligned} p(\mathbf{t}_t, \sigma_{t-1}^{-2} | D_{t-1}) & \stackrel{def}{=} n(\mathbf{t}_t | \mathbf{a}_t, \sigma_{t-1}^{-2} \mathbf{R}_t^\dagger) \cdot \text{gamma}(\sigma_{t-1}^{-2} | v_{t-1}/2, d_{t-1}/2) \quad [\text{ノーマル・ガンマ密度関数}] \\ &= (2\pi)^{-\frac{h}{2}} |\sigma_{t-1}^2 \mathbf{R}_t^\dagger|^{-\frac{h}{2}} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} [(\mathbf{t}_t - \mathbf{a}_t)' \mathbf{R}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{a}_t)]\right\} \cdot \frac{d_{t-1}^{\frac{v_{t-1}}{2}}}{2^{\frac{v_{t-1}}{2}} \Gamma\left(\frac{v_{t-1}}{2}\right)} (\sigma_{t-1}^{-2})^{\frac{v_{t-1}-1}{2}} \exp\left\{-\frac{d_{t-1}}{2} \sigma_{t-1}^{-2}\right\} \end{aligned}$$

次に、 $\mathbf{T}_t = \mathbf{t}_t$ を得た後、 $\sigma_{t-1}^{-2} | (D_{t-1}, \mathbf{t}_t)$ の条件付き密度について、

$$\begin{aligned} p(\sigma_{t-1}^{-2} | D_{t-1}, \mathbf{t}_t) & \stackrel{def}{=} \text{gamma}\left(\sigma_{t-1}^{-2} \left| \frac{v_{t-1} + h}{2}, \frac{d_{t-1} + [(\mathbf{t}_t - \mathbf{a}_t)' \mathbf{R}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{a}_t)]}{2}\right.\right) \\ &= \frac{[d_{t-1} + (\mathbf{t}_t - \mathbf{a}_t)' \mathbf{R}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{a}_t)]^{\frac{v_{t-1} + h}{2}} (\sigma_{t-1}^{-2})^{\frac{v_{t-1} + h}{2} - 1} \exp\left\{-\frac{d_{t-1} + [(\mathbf{t}_t - \mathbf{a}_t)' \mathbf{R}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{a}_t)]}{2} \sigma_{t-1}^{-2}\right\}}{2^{\frac{v_{t-1} + h}{2}} \Gamma\left(\frac{v_{t-1} + h}{2}\right)} \end{aligned}$$

とおけば、

$$\begin{aligned} p(\mathbf{t}_t | D_{t-1}) &= \int_0^{+\infty} p(\mathbf{t}_t, \sigma_{t-1}^{-2} | D_{t-1}) d\sigma_{t-1}^{-2} \\ &= p(\mathbf{t}_t, \sigma_{t-1}^{-2} | D_{t-1}) / p(\sigma_{t-1}^{-2} | D_{t-1}, \mathbf{t}_t) \end{aligned}$$

$$\begin{aligned}
& (2\pi)^{-\frac{h}{2}} |\sigma_{t-1}^2 \mathbf{R}_t^\dagger|^{-\frac{h}{2}} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} [(\mathbf{t}_t - \mathbf{a}_t)' \mathbf{R}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{a}_t)]\right\} \cdot \frac{d_{t-1}^{\frac{v_{t-1}}{2}}}{2^{\frac{v_{t-1}}{2}} \Gamma\left(\frac{v_{t-1}}{2}\right)} (\sigma_{t-1}^{-2})^{\frac{v_{t-1}-1}{2}} \exp\left\{-\frac{d_{t-1}}{2} \sigma_{t-1}^{-2}\right\} \\
= & \frac{\frac{[d_{t-1} + (\mathbf{t}_t - \mathbf{a}_t)' \mathbf{R}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{a}_t)]^{\frac{v_{t-1}+h}{2}}}{2^{\frac{v_{t-1}+h}{2}} \Gamma\left(\frac{v_{t-1}+h}{2}\right)} (\sigma_{t-1}^{-2})^{\frac{v_{t-1}+h}{2}-1} \exp\left\{-\frac{d_{t-1} + [(\mathbf{t}_t - \mathbf{a}_t)' \mathbf{R}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{a}_t)]}{2} \sigma_{t-1}^{-2}\right\}}{(\sigma_{t-1}^{-2})^{\frac{v_{t-1}+h}{2}-1} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} [d_{t-1} + (\mathbf{t}_t - \mathbf{a}_t)' \mathbf{R}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{a}_t)]\right\}} \\
\propto & \frac{[d_{t-1} + (\mathbf{t}_t - \mathbf{a}_t)' \mathbf{R}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{a}_t)]^{\frac{v_{t-1}+h}{2}} (\sigma_{t-1}^{-2})^{\frac{v_{t-1}+h}{2}-1} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} [d_{t-1} + (\mathbf{t}_t - \mathbf{a}_t)' \mathbf{R}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{a}_t)]\right\}}{[d_{t-1} + (\mathbf{t}_t - \mathbf{a}_t)' \mathbf{R}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{a}_t)]^{\frac{v_{t-1}+h}{2}}} \\
& \propto [d_{t-1} + (\mathbf{t}_t - \mathbf{a}_t)' \mathbf{R}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{a}_t)]^{-\frac{v_{t-1}+h}{2}} \\
& \propto \left[1 + \frac{1}{v_{t-1}} (\mathbf{t}_t - \mathbf{a}_t)' \left(\frac{v_{t-1}}{d_{t-1}} \mathbf{R}_t^{\dagger-1}\right) (\mathbf{t}_t - \mathbf{a}_t)\right]^{-\frac{v_{t-1}+h}{2}} \\
& \propto \left[1 + \frac{1}{v_{t-1}} (\mathbf{t}_t - \mathbf{a}_t)' (S_{t-1} \mathbf{R}_t^\dagger)^{-1} (\mathbf{t}_t - \mathbf{a}_t)\right]^{-\frac{v_{t-1}+h}{2}} \\
& \propto \left[1 + \frac{1}{v_{t-1}} (\mathbf{t}_t - \mathbf{a}_t)' \mathbf{R}_t^{\dagger-1} (\mathbf{t}_t - \mathbf{a}_t)\right]^{-\frac{v_{t-1}+h}{2}}
\end{aligned}$$

よって、

$$\mathbf{T}_t | D_{t-1} \sim t(\mathbf{a}_t, \mathbf{R}_t, h, v_{t-1})$$

ここに、

$$\mathbf{R}_t \stackrel{def}{=} \left(\frac{d_{t-1}}{v_{t-1}}\right) \mathbf{R}_t^\dagger = S_{t-1} \mathbf{R}_t^\dagger,$$

$$E[\mathbf{T}_t | D_{t-1}] = \mathbf{a}_t, \quad Var[\mathbf{T}_t | D_{t-1}] = \left(\frac{v_{t-1}}{v_{t-1}-1}\right) \mathbf{R}_t$$

(d_t の定義については、後述)

(5) 観測予測式(分散母数の条件がある場合)

$$Y_t | (D_{t-1}, \mathbf{t}_t, \sigma_{t-1}^{-2}) \sim N(\mathbf{F}' \mathbf{t}_t, \sigma_{t-1}^2)$$

$$\mathbf{T}_t | (D_{t-1}, \sigma_{t-1}^{-2}) \sim N(\mathbf{a}_t, \sigma_{t-1}^2 \mathbf{R}_t^\dagger)$$

より、

$$\begin{aligned}
p(y_t | D_{t-1}, \sigma_{t-1}^{-2}) &= \int_{\mathbf{R}^h} p(y_t | D_{t-1}, \mathbf{t}_t, \sigma_{t-1}^{-2}) p(\mathbf{t}_t | D_{t-1}, \sigma_{t-1}^{-2}) d\mathbf{t}_t \\
&\stackrel{\text{def}}{=} \int_{\mathbf{R}^h} n(y_t | \mathbf{F}'\mathbf{t}_t, \sigma_{t-1}^{-2}) n(\mathbf{t}_t | \mathbf{a}_t, \sigma_{t-1}^{-2} \mathbf{R}_t^\dagger) d\mathbf{t}_t \\
&\propto \exp \left\{ -\frac{\sigma_{t-1}^{-2}}{2} \left[\frac{(y_t - \mathbf{F}'\mathbf{a}_t)^2}{Q_t^\dagger} \right] \right\} \\
&\propto \exp \left\{ -\frac{\sigma_{t-1}^{-2}}{2} \left[\frac{(y_t - f_t)^2}{Q_t^\dagger} \right] \right\}
\end{aligned}$$

よって、

$$\begin{aligned}
Y_t | (D_{t-1}, \sigma_{t-1}^{-2}) &\sim N(f_t, \sigma_{t-1}^2 Q_t^\dagger) \\
&= N(f_t, Q_t) \quad *A8
\end{aligned}$$

ここで、 $f_t = \mathbf{F}'\mathbf{a}_t$ とおいた。

したがって、

$$\begin{aligned}
E[Y_t | D_{t-1}, \sigma_{t-1}^{-2}] &= f_t, \\
\text{Var}[Y_t | D_{t-1}, \sigma_{t-1}^{-2}] &= \sigma_{t-1}^2 Q_t^\dagger = \sigma_{t-1}^2 (\mathbf{F}'\mathbf{R}_t^\dagger \mathbf{F} + 1) = Q_t
\end{aligned}$$

$$\begin{aligned}
*A8 \quad E(\tilde{y}_t | D_{t-1}, \sigma_{t-1}^{-2}) &= E(E(\tilde{y}_t | \mathbf{F}, D_{t-1}, \sigma_{t-1}^{-2}, \tilde{\mathbf{t}}_t) | D_{t-1}, \sigma_{t-1}^{-2}) = E(\mathbf{F}'\tilde{\mathbf{t}}_t | D_{t-1}, \sigma_{t-1}^{-2}) \\
&= \mathbf{F}'E(\tilde{\mathbf{t}}_t | D_{t-1}, \sigma_{t-1}^{-2}) = \mathbf{F}'\mathbf{a}_t = f_t \quad \text{とおいた。}
\end{aligned}$$

また、

$$\begin{aligned}
E(\tilde{y}_t^2 | D_{t-1}, \sigma_{t-1}^{-2}, \mathbf{t}_t) &= E((\mathbf{F}'\mathbf{t}_t + \tilde{v}_t)^2 | D_{t-1}, \sigma_{t-1}^{-2}, \mathbf{t}_t) \\
&= E(\mathbf{F}'\mathbf{t}_t \mathbf{t}_t' \mathbf{F} + 2\mathbf{F}'\mathbf{t}_t \tilde{v}_t + \tilde{v}_t^2 | D_{t-1}, \sigma_{t-1}^{-2}, \mathbf{t}_t) = \mathbf{F}'\mathbf{t}_t \mathbf{t}_t' \mathbf{F} + \sigma_{t-1}^2
\end{aligned}$$

$$E(\tilde{\mathbf{t}}_t \tilde{\mathbf{t}}_t' | D_{t-1}, \sigma_{t-1}^{-2}) = \sigma_{t-1}^2 \mathbf{R}_t^\dagger + \mathbf{a}_t \mathbf{a}_t', \quad (\sigma_{t-1}^2 \mathbf{R}_t^\dagger \stackrel{\text{def}}{=} \text{Var}(\tilde{\mathbf{t}}_t | D_{t-1}, \sigma_{t-1}^{-2}))$$

$$\begin{aligned}
E(\tilde{y}_t^2 | D_{t-1}, \sigma_{t-1}^{-2}) &= E(\mathbf{F}'\tilde{\mathbf{t}}_t \tilde{\mathbf{t}}_t' \mathbf{F} + \sigma_{t-1}^2 | D_{t-1}, \sigma_{t-1}^{-2}) = \mathbf{F}'E(\tilde{\mathbf{t}}_t \tilde{\mathbf{t}}_t' | D_{t-1}, \sigma_{t-1}^{-2})\mathbf{F} + \sigma_{t-1}^2 \\
&= \mathbf{F}'(\sigma_{t-1}^2 \mathbf{R}_t^\dagger + \mathbf{a}_t \mathbf{a}_t')\mathbf{F} + \sigma_{t-1}^2
\end{aligned}$$

$$\begin{aligned}
\text{Var}(\tilde{y}_t | D_{t-1}, \sigma_{t-1}^{-2}) &= E(\tilde{y}_t^2 | D_{t-1}, \sigma_{t-1}^{-2}) - (E(\tilde{y}_t | D_{t-1}, \sigma_{t-1}^{-2}))^2 \\
&= \mathbf{F}'(\sigma_{t-1}^2 \mathbf{R}_t^\dagger + \mathbf{a}_t \mathbf{a}_t')\mathbf{F} + \sigma_{t-1}^2 - \mathbf{F}'\mathbf{a}_t \mathbf{a}_t' \mathbf{F} \\
&= \sigma_{t-1}^2 (\mathbf{F}'\mathbf{R}_t^\dagger \mathbf{F} + 1) \\
&= \sigma_{t-1}^2 Q_t^\dagger = Q_t \quad (Q_t^\dagger = \mathbf{F}'\mathbf{R}_t^\dagger \mathbf{F} + 1)
\end{aligned}$$

(上の表現では、とくに確率変数にティルド[~]を付している。)

(6) 観測予測式(分散母数の条件がない場合)

$$Y_t | (D_{t-1}, \sigma_{t-1}^{-2}) \sim N(f_t, \sigma_{t-1}^2 Q_t^\dagger)$$

$$\sigma_{t-1}^{-2} | D_{t-1} \sim \text{Gamma}(v_{t-1}/2, d_{t-1}/2)$$

より、

$$p(y_t, \sigma_{t-1}^{-2} | D_{t-1})$$

$$\stackrel{\text{def}}{=} n(y_t | f_t, \sigma_{t-1}^{-2} Q_t^\dagger) \cdot \text{gamma}(\sigma_{t-1}^{-2} | v_{t-1}/2, d_{t-1}/2) \quad [\text{ノーマル・ガンマ密度関数}]$$

$$= (2\pi\sigma_{t-1}^2)^{-\frac{1}{2}} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} \left[\frac{(y_t - f_t)^2}{Q_t^\dagger}\right]\right\} \cdot \frac{d_{t-1}^{\frac{v_{t-1}}{2}}}{2^{\frac{v_{t-1}}{2}} \Gamma\left(\frac{v_{t-1}}{2}\right)} (\sigma_{t-1}^{-2})^{\frac{v_{t-1}-1}{2}} \exp\left\{-\frac{d_{t-1}}{2} \sigma_{t-1}^{-2}\right\}$$

次に、 $Y_t = y_t$ を得た後、 $\sigma_{t-1}^{-2} | (D_{t-1}, y_t)$ の条件付き密度について、

$$p(\sigma_{t-1}^{-2} | D_{t-1}, y_t)$$

$$\stackrel{\text{def}}{=} \text{gamma}\left(\sigma_{t-1}^{-2} \left| \frac{v_{t-1}+1}{2}, \frac{d_{t-1} + \frac{(y_t - f_t)^2}{Q_t^\dagger}}{2}\right.\right)$$

$$= \frac{\left[d_{t-1} + \frac{(y_t - f_t)^2}{Q_t^\dagger} \right]^{\frac{v_{t-1}+1}{2}}}{2^{\frac{v_{t-1}+1}{2}} \Gamma\left(\frac{v_{t-1}+1}{2}\right)} (\sigma_{t-1}^{-2})^{\frac{v_{t-1}+1}{2}-1} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} \left[d_{t-1} + \frac{(y_t - f_t)^2}{Q_t^\dagger} \right]\right\}$$

とおけば、

$$\begin{aligned} p(y_t | D_{t-1}) &= \int_0^{+\infty} p(y_t, \sigma_{t-1}^{-2} | D_{t-1}) d\sigma_{t-1}^{-2} \\ &= p(y_t, \sigma_{t-1}^{-2} | D_{t-1}) / p(\sigma_{t-1}^{-2} | D_{t-1}, y_t) \end{aligned}$$

$$\begin{aligned}
& (2\pi\sigma_{t-1}^2)^{-\frac{1}{2}} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} \left[\frac{(y_t-f_t)^2}{Q_t^\dagger}\right]\right\} \cdot \frac{d_{t-1}^{\frac{v_{t-1}}{2}}}{2^{\frac{v_{t-1}}{2}} \Gamma\left(\frac{v_{t-1}}{2}\right)} (\sigma_{t-1}^{-2})^{\frac{v_{t-1}-1}{2}} \exp\left\{-\frac{d_{t-1}}{2} \sigma_{t-1}^{-2}\right\} \\
= & \frac{\left[d_{t-1} + \frac{(y_t-f_t)^2}{Q_t^\dagger}\right]^{\frac{v_{t-1}+1}{2}}}{2^{\frac{v_{t-1}+1}{2}} \Gamma\left(\frac{v_{t-1}+1}{2}\right)} (\sigma_{t-1}^{-2})^{\frac{v_{t-1}+1}{2}-1} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} \left[d_{t-1} + \frac{(y_t-f_t)^2}{Q_t^\dagger}\right]\right\} \\
& (\sigma_{t-1}^{-2})^{\frac{v_{t-1}+1}{2}-1} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} \left[d_{t-1} + \frac{(y_t-f_t)^2}{Q_t^\dagger}\right]\right\} \\
\propto & \frac{\left[d_{t-1} + \frac{(y_t-f_t)^2}{Q_t^\dagger}\right]^{\frac{v_{t-1}+1}{2}} (\sigma_{t-1}^{-2})^{\frac{v_{t-1}+1}{2}-1} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} \left[d_{t-1} + \frac{(y_t-f_t)^2}{Q_t^\dagger}\right]\right\}}{\left[d_{t-1} + \frac{(y_t-f_t)^2}{Q_t^\dagger}\right]^{\frac{v_{t-1}+1}{2}}} \\
& \propto \left\{1 + \frac{1}{v_{t-1}} \left[\frac{(y_t-f_t)^2}{\left(\frac{d_{t-1}}{v_{t-1}}\right) Q_t^\dagger}\right]\right\}^{-\frac{v_{t-1}+1}{2}} \\
& \propto \left\{1 + \frac{1}{v_{t-1}} \left[\frac{(y_t-f_t)^2}{S_{t-1} Q_t^\dagger}\right]\right\}^{-\frac{v_{t-1}+1}{2}} \\
& \propto \left\{1 + \frac{1}{v_{t-1}} \left[\frac{(y_t-f_t)^2}{Q_t}\right]\right\}^{-\frac{v_{t-1}+1}{2}}
\end{aligned}$$

ここで、 Q_t について、

$$S_{t-1} Q_t^\dagger = Q_t$$

とおいた。

よって、

$$Y_t | D_{t-1} \sim t(f_t, Q_t; v_{t-1})$$

また、

$$E[Y_t | D_{t-1}] = f_t, \quad \text{Var}[Y_t | D_{t-1}] = \left(\frac{v_{t-1}}{v_{t-1}-1}\right) Q_t$$

である。

(7) 「基本的仮説」に基づく各分布

① 状態予測分布(分散母数の条件がある場合)

本論のモデルの基本的な仮定において、

$$\begin{aligned} E[\sigma_t^2 | D_{t-1}, \sigma_{t-1}^2] &\stackrel{def}{=} \sigma_t^{*2} \\ &= \sigma_{t-1}^2 \left(\frac{Q_t^*}{Q_t} \right) \end{aligned}$$

であることを考慮し、尺度： $\sigma_{t-1}^2 \mathbf{R}_t^\dagger$ について、 $\sigma_t^{*2} \mathbf{R}_t^\dagger$ と再定義し、
本論の状態予測分布(分散母数の条件のある場合)について、

$$\mathbf{T}_t | (D_{t-1}, \sigma_{t-1}^{-2}) \sim N(\mathbf{a}_t, \sigma_t^{*2} \mathbf{R}_t^\dagger)$$

と示す。なお、ここに、

$$E[\mathbf{T}_t | D_{t-1}, \sigma_{t-1}^{-2}] = \mathbf{a}_t, \quad \text{Var}[\mathbf{T}_t | D_{t-1}, \sigma_{t-1}^{-2}] = \sigma_t^{*2} \mathbf{R}_t^\dagger$$

である。

② 状態予測分布(分散母数の条件がない場合)

本論のモデルの基本的な仮定に基づき、

$$\begin{aligned} E[\sigma_t^2 | D_{t-1}] &\stackrel{def}{=} S_t^* \\ &= S_{t-1} \left(\frac{Q_t^*}{Q_t} \right) \end{aligned}$$

とおき、尺度： $S_{t-1} \mathbf{R}_t^\dagger \stackrel{def}{=} \mathbf{R}_t$ について、 $S_t^* \mathbf{R}_t^\dagger \stackrel{def}{=} \mathbf{R}_t^*$ と再定義し、

本論の状態予測分布(分散母数の条件のない場合)について、

$$\mathbf{T}_t | D_{t-1} \sim t(\mathbf{a}_t, \mathbf{R}_t^*, h, \nu_{t-1})$$

と示す。なお、ここに、

$$E[\mathbf{T}_t | D_{t-1}] = \mathbf{a}_t, \quad \text{Var}[\mathbf{T}_t | D_{t-1}] = \left(\frac{V_{t-1}}{V_{t-1}-1}\right)\mathbf{R}_t^*$$

である。 これは、本論の(2.6)式に該当する。

③ 観測予測分布(分散母数の条件がある場合)

①と同様に、本論のモデルの基本的な仮定において、

$$\begin{aligned} E[\sigma_t^2 | D_{t-1}, \sigma_{t-1}^2] &\stackrel{def}{=} \sigma_t^{*2} \\ &= \sigma_{t-1}^2 \left(\frac{Q_t^*}{Q_t} \right) \end{aligned}$$

であることを考慮し、尺度： $\sigma_{t-1}^2 Q_t^\dagger \stackrel{def}{=} Q_t$ について、 $\sigma_t^{*2} Q_t^\dagger \stackrel{def}{=} Q_t^*$ と再定義し、
本論の観測予測分布(分散母数の条件のある場合)について、

$$Y_t | (D_{t-1}, \sigma_{t-1}^{-2}) \sim N(f_t, Q_t^*)$$

と示す。なお、ここに、

$$E[Y_t | D_{t-1}, \sigma_{t-1}^{-2}] = f_t, \quad \text{Var}[Y_t | D_{t-1}, \sigma_{t-1}^{-2}] = Q_t^*$$

である。

④ 観測予測分布(分散母数の条件がない場合)

②と同様に、本論のモデルの基本的な仮定に基づいて、

$$\begin{aligned} E[\sigma_t^2 | D_{t-1}] &\stackrel{def}{=} S_t^* \\ &= S_{t-1} \left(\frac{Q_t^*}{Q_t} \right) \end{aligned}$$

とおき、尺度： $S_{t-1} Q_t^\dagger \stackrel{def}{=} Q_t$ について、 $S_t^* Q_t^\dagger \stackrel{def}{=} Q_t^*$ と再定義し、
本論の状態予測分布(分散母数の条件のない場合)について、

$$Y_t | D_{t-1} \sim t(f_t, Q_t^*; \nu_{t-1})$$

と示す。なお、ここに、

$$E[Y_t | D_{t-1}] = f_t, \quad \text{Var}[Y_t | D_{t-1}] = \left(\frac{v_{t-1}}{v_{t-1}-1}\right)Q_t^*$$

である。これは、本論の(2.5)式に該当する。

(8) 観測予測分布に係る d_t について

既にみたとおり、

$$f(y_t | D_{t-1}, \sigma_{t-1}^{-2}) \propto (\sigma_{t-1}^{-2})^{\frac{1}{2}} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} \left[\frac{(y_t - f_t)^2}{Q_t^\dagger}\right]\right\}$$

$$\begin{aligned} f(\sigma_t^{-2} | D_t) &\propto f(\sigma_{t-1}^{-2} | D_{t-1}) f(y_t | D_{t-1}, \sigma_{t-1}^{-2}) \\ &\propto (\sigma_{t-1}^{-2})^{\frac{(v_{t-1}+1)}{2}-1} \exp\left\{-\frac{\sigma_{t-1}^{-2}}{2} \left[d_{t-1} + \frac{(y_t - f_t)^2}{Q_t^\dagger}\right]\right\} \end{aligned}$$

となることから、観測予測分布に係る d_t について、

$$d_t = \left[d_{t-1} + \left\{ \frac{(y_t - f_t)^2}{Q_t^\dagger} \right\} \right] = \left[d_{t-1} + S_{t-1} \left\{ \frac{(y_t - f_t)^2}{Q_t} \right\} \right]$$

とおくことができる。

ここでは、前述のとおり、 $S_{t-1}Q_t^\dagger = Q_t$ とおく。

$E[\sigma_t^2 | D_{t-1}] \stackrel{def}{=} S_t^* = S_{t-1} \left(\frac{Q_t^*}{Q_t} \right)$ と仮定する場合は、

$$d_t = \left[d_{t-1} + S_t^* \left\{ \frac{(y_t - f_t)^2}{Q_t^*} \right\} \right]$$

$$S_t^* \stackrel{def}{=} S_{t-1} \left(\frac{Q_t^*}{Q_t} \right), \quad Q_t^* \stackrel{def}{=} S_t^* Q_t^\dagger$$

と定義する。

(9) 状態方程式と1期先予測分布

上でみた観測と状態それぞれの1期先予測分布(分散母数の条件なし)を以下のとおり整理する。

$$\begin{cases} T_t | D_{t-1} \sim t(\mathbf{a}_t, \mathbf{R}_t^*, h, \nu_{t-1}) \\ Y_t | D_{t-1} \sim t(f_t, Q_t^*; \nu_{t-1}) \end{cases}$$

一般に、多変量 t 分布

$$X \sim t(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p, \nu), \quad X' = (X_1, \dots, X_p)$$

について

$$X' = (X_1, X_2), \quad X'_1 = (X_1, \dots, X_{p_1}) \quad (p - p_1 = p_2 > 0)$$

と区分けした場合、

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

と分割すると、

$$X_1 | (X_2 = \mathbf{x}_2) \sim t(\boldsymbol{\mu}_{1,2}, \frac{\nu}{\nu + p_2} \tau_2 \boldsymbol{\Sigma}_{11,2}, p_1, \nu + p_2)$$

$$\boldsymbol{\mu}_{1,2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2), \quad \boldsymbol{\Sigma}_{11,2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21},$$

$$\tau_2 = 1 + \frac{1}{\nu} (\mathbf{x}_2 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2),$$

$$E[X_1 | X_2 = \mathbf{x}_2] = \boldsymbol{\mu}_{1,2},$$

$$Var[X_1 | X_2 = \mathbf{x}_2] = \frac{\nu + p_2}{\nu + p_2 - 2} \cdot \frac{\nu}{\nu + p_2} \tau_2 \boldsymbol{\Sigma}_{11,2} = \frac{\nu}{\nu + p_2 - 2} \tau_2 \boldsymbol{\Sigma}_{11,2} \quad (\nu + p_2 > 2)$$

が成り立つ。

本論のモデルでは、

$$\boldsymbol{\mu} = \begin{pmatrix} \mathbf{a}_t \\ f_t \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{R}_t^* & \mathbf{R}_t^* \mathbf{F} \\ \mathbf{F}' \mathbf{R}_t^* & Q_t^* \end{pmatrix}$$

となるが、 Y_t はスカラーであるので、条件付き分布として以下が得られる。

$$\mathbf{T}_t | (D_{t-1}, y_t) = \mathbf{T}_t | D_t \sim t(\boldsymbol{\mu}_{1:2t}, \frac{\nu_{t-1}}{\nu_{t-1}+1} \boldsymbol{\tau}_{2t} \boldsymbol{\Sigma}_{1:2t}, h, \nu_t)$$

$$\boldsymbol{\mu}_{1:2t} \stackrel{def}{=} \mathbf{m}_t = \mathbf{a}_t + \mathbf{F}' \mathbf{R}_t^* (y_t - f_t) / Q_t^*,$$

$$\boldsymbol{\Sigma}_{1:2t} = \mathbf{R}_t^* - \mathbf{R}_t^* \mathbf{F} \mathbf{F}' \mathbf{R}_t^* / Q_t^*$$

なお、 τ_{2t} については、DLMの理論的な体系を考慮して、

$$\tau_{2t} \stackrel{def}{=} 1 + \frac{1}{\nu_{t-1}} \left\{ \frac{(y_t - f_t)^2}{Q_t} \right\} = 1 + \frac{1}{\nu_{t-1}} \left\{ \frac{(y_t - f_t)^2}{Q_t^*} \right\} \left(\frac{Q_t^*}{Q_t} \right) \quad \text{とおく。}$$

ここで、

$$d_t = \left[d_{t-1} + S_{t-1} \left\{ \frac{(y_t - f_t)^2}{Q_t} \right\} \right] = \left[d_{t-1} + S_t^* \left\{ \frac{(y_t - f_t)^2}{Q_t^*} \right\} \right]$$

であることに留意し、 $\nu_{t-1} + 1 = \nu_t$ であることを踏まえると、

$$\begin{aligned} \tau_{2t} &= \left(\frac{d_t}{d_{t-1}} \right) = 1 + \frac{1}{\nu_{t-1}} \left\{ \frac{(y_t - f_t)^2}{Q_t} \right\}, \\ \left(\frac{\nu_{t-1}}{\nu_{t-1}+1} \right) \tau_{2t} &= \left(\frac{\nu_{t-1}}{\nu_t} \right) \tau_{2t} = \left(\frac{\nu_{t-1}}{\nu_t} \right) \left(\frac{d_t}{d_{t-1}} \right) = \left(\frac{S_t}{S_{t-1}} \right), \\ \left(\frac{S_t}{S_{t-1}} \right) &= \left(\frac{\nu_{t-1}}{\nu_t} \right) \left[1 + \frac{1}{\nu_{t-1}} \left\{ \frac{(y_t - f_t)^2}{Q_t} \right\} \right], \end{aligned}$$

$$\begin{aligned} S_t &= \left(\frac{\nu_{t-1}}{\nu_t} \right) S_{t-1} + \left(\frac{S_{t-1}}{\nu_t} \right) \left\{ \frac{(y_t - f_t)^2}{Q_t} \right\} \\ &= \frac{d_{t-1} + S_{t-1} \left\{ \frac{(y_t - f_t)^2}{Q_t} \right\}}{\nu_t} \\ &= \frac{d_{t-1} + S_t^* \left\{ \frac{(y_t - f_t)^2}{Q_t^*} \right\}}{\nu_t}, \end{aligned}$$

$$\begin{aligned} \left(\frac{S_t}{S_t^*} \right) &= \left(\frac{\nu_{t-1}}{\nu_t} \right) \left[1 + \frac{1}{\nu_{t-1}} \left\{ \frac{(y_t - f_t)^2}{Q_t^*} \right\} \left(\frac{Q_t^*}{Q_t} \right) \right] \\ &= \left(\frac{S_t}{S_{t-1}} \right). \end{aligned}$$

が得られる。

これらの結果を踏まえ、前掲の条件付き分布の公式を考慮すると、新たに以下が定義でき、

$$\mathbf{C}_t \stackrel{def}{=} \left(\frac{v_{t-1}}{v_t}\right) \tau_{2t} \boldsymbol{\Sigma}_{1^*2t} = \left(\frac{S_t}{S_t^*}\right) \boldsymbol{\Sigma}_{1^*2t} = \left(\frac{S_t}{S_t^*}\right) [\mathbf{R}_t^* - \mathbf{R}_t^* \mathbf{F} \mathbf{F}' \mathbf{R}_t^* / Q_t^*]$$

$$S_t = \left[d_{t-1} + S_t^* \left\{ \frac{(y-f)^2}{Q_t^*} \right\} \right] / v_t$$

さらに、

$$(y_t - f_t) = e_t,$$

$$\mathbf{R}_t^* \mathbf{F} / Q_t^* = \mathbf{R}_t \mathbf{F} / Q_t = \mathbf{A}_t \quad \text{とおくと、}$$

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t^* e_t$$

$$\mathbf{C}_t = \left(\frac{S_t}{S_t^*}\right) [\mathbf{R}_t^* - \mathbf{A}_t \mathbf{A}_t' Q_t^*]$$

$$= \left(\frac{S_t}{S_t^*}\right) [\mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t' Q_t]$$

よって、 \mathbf{C}_t が通常の DLM の体系に基づく定義に戻り、

$$T_t | (D_{t-1}, y_t) \sim P(t_t | D_t) = t(\mathbf{m}_t, \mathbf{C}_t, h, v_t)$$

$$E[T_t | D_t] = \mathbf{m}_t, \quad Var[T_t | D_t] = \left(\frac{v_{t-1}+1}{v_{t-1}+1-2}\right) \mathbf{C}_t = \left(\frac{v_t}{v_t-2}\right) \mathbf{C}_t \quad (v_t > 2)$$

同様に考えて、

$$T_{t-1} | D_{t-1} \sim P(t_{t-1} | D_{t-1}) = t(\mathbf{m}_{t-1}, \mathbf{C}_{t-1}, h, v_{t-1})$$

これは、本論の(2.3)式に該当する。

(10) 平滑化方程式

N 期までの情報集合 D_N に基づく回顧的な状態 T_{N-k} の周辺分布は、次のとおり求められる。

なお、 $1 \leq k < N$ である。

$$p(\mathbf{t}_{N-k} | D_N) = \int_{\mathbf{R}^h} p(\mathbf{t}_{N-k} | \mathbf{t}_{N-k+1}, D_N) p(\mathbf{t}_{N-k+1} | D_N) d\mathbf{t}_{N-k+1}$$

ここで、 $f(\mathbf{t}_{N-k} | \mathbf{t}_{N-k+1}, D_N)$ については、以下のとおり定義される。

$$p(\mathbf{t}_{N-k} | \mathbf{t}_{N-k+1}, D_N) = \frac{p(\mathbf{t}_{N-k} | \mathbf{t}_{N-k+1}, D_{N-k})p(\mathbf{y} | \mathbf{t}_{N-k}, \mathbf{t}_{N-k+1}, D_{N-k})}{p(\mathbf{y} | \mathbf{t}_{N-k+1}, D_{N-k})}$$

$$p(\mathbf{t}_{N-k} | \mathbf{t}_{N-k+1}, D_{N-k}) \propto p(\mathbf{t}_{N-k} | D_{N-k})p(\mathbf{t}_{N-k+1} | \mathbf{t}_{N-k}, D_{N-k})$$

ここに、 $\mathbf{y} = \{y_{N-k+1}, \dots, y_N\}$

また、状態モデル式に基づき、ここでは、

$$\mathbf{T}_{N-k+1} | (D_{N-k}, \sigma_{N-k}^{-2}, \mathbf{t}_{N-k}) \sim N(\mathbf{G}\mathbf{t}_{N-k}, \sigma_{N-k}^2 \mathbf{W}_{N-k+1}^\dagger)$$

と定義し、さらに、状態方程式と状態予測式から、

$$\begin{cases} \mathbf{T}_{N-k} | (D_{N-k}, \sigma_{N-k}^{-2}) \sim N(\mathbf{m}_{N-k}, \sigma_{N-k}^2 \mathbf{C}_{N-k}^\dagger) \\ \mathbf{T}_{N-k+1} | (D_{N-k}, \sigma_{N-k}^{-2}) \sim N(\mathbf{a}_{N-k+1}, \sigma_{N-k}^2 \mathbf{R}_{N-k+1}^\dagger) \end{cases} \quad \text{とおく。}$$

さらに、上記を踏まえて得られるモーメントを $\mathbf{h}_N(N-k)$, $\sigma_N^2 \mathbf{H}_N^\dagger(N-k)$ とおき、次のとおり定義する^{*A9}。

$$\begin{aligned} \mathbf{h}_N(N-k) &= \mathbf{m}_{N-k} + \mathbf{C}_{N-k}^\dagger \mathbf{G}' \mathbf{R}_{N-k+1}^{\dagger-1} [\mathbf{t}_{N-k+1} - \mathbf{a}_{N-k+1}] \\ &= \mathbf{m}_{N-k} + \mathbf{B}_{N-k} [\mathbf{t}_{N-k+1} - \mathbf{a}_{N-k+1}] \end{aligned}$$

$$\begin{aligned} \mathbf{H}_N^\dagger(N-k) &= \mathbf{C}_{N-k}^\dagger - \mathbf{C}_{N-k}^\dagger \mathbf{G}' \mathbf{R}_{N-k+1}^{\dagger-1} \mathbf{G} \mathbf{C}_{N-k}^\dagger \\ &= \mathbf{C}_{N-k}^\dagger - \mathbf{B}_{N-k} \mathbf{R}_{N-k+1}^\dagger \mathbf{B}'_{N-k} \end{aligned}$$

$$\mathbf{B}_N = \mathbf{C}_N^\dagger \mathbf{G}' \mathbf{R}_{N+1}^{\dagger-1}$$

これらの結果に基づき、あらためて、以下を定義する。

$$\mathbf{T}_{N-k} | (D_{N-k}, \mathbf{t}_{N-k+1}, \sigma_{N-k}^{-2}) \sim N(\mathbf{h}_N(N-k), \sigma_{N-k}^2 \mathbf{H}_N^\dagger(N-k))$$

*A9 1期先予測分布を導出した際の議論を踏まえれば、多変量正規分布 $N(\boldsymbol{\mu}, \sigma^2 \boldsymbol{\Sigma})$ において、

$$\boldsymbol{\mu} = \begin{pmatrix} \mathbf{m}_{N-k} \\ \mathbf{a}_{N-k+1} \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{C}_{N-k}^\dagger & \mathbf{C}_{N-k}^\dagger \mathbf{G}' \\ \mathbf{G} \mathbf{C}_{N-k}^\dagger & \mathbf{R}_{N-k+1}^\dagger \end{pmatrix}$$

となるので、これより、 $\mathbf{h}_N(N-k)$, $\mathbf{H}_N^\dagger(N-k)$ が得られる。

次に、

$$\mathbf{T}_{N-k+1} | (D_N, \sigma_N^{-2}) \sim N(\mathbf{a}_N(N-k+1), \sigma_N^2 \mathbf{R}_N^\dagger(N-k+1))$$

とおくと、

$$p(\mathbf{t}_{N-k} | D_N, \sigma_N^{-2}) = \int_{\mathbf{R}^h} p(\mathbf{t}_{N-k} | D_N, \sigma_N^{-2}, \mathbf{t}_{N-k+1}) p(\mathbf{t}_{N-k+1} | D_N, \sigma_N^{-2}) d\mathbf{t}_{N-k+1}$$

から、

$$\mathbf{T}_{N-k} | (D_N, \sigma_N^{-2}) \sim N(\mathbf{a}_N(N-k), \sigma_N^2 \mathbf{R}_N^\dagger(N-k))$$

$$\text{ここに、} \begin{cases} \mathbf{a}_N(N-k) = E(\mathbf{h}_N(N-k) | D_N) \\ \sigma_N^2 \mathbf{R}_N^\dagger(N-k) = \sigma_N^2 \mathbf{H}_N^\dagger(N-k) + \text{Var}(\mathbf{h}_N(N-k) | D_N) \end{cases}$$

したがって、

$$\begin{cases} \mathbf{a}_N(N-k) = \mathbf{m}_{N-k} + \mathbf{B}_{N-k}[\mathbf{a}_N(N-k+1) - \mathbf{a}_{N-k+1}] \\ \mathbf{R}_N^\dagger(N-k) = \mathbf{C}_{N-k}^\dagger - \mathbf{B}_{N-k}[\mathbf{R}_{N-k+1}^\dagger - \mathbf{R}_N^\dagger(N-k+1)]\mathbf{B}_{N-k}' \end{cases}$$

$$\mathbf{a}_N(N) = \mathbf{m}_N, \mathbf{R}_N^\dagger(N) = \mathbf{C}_N^\dagger$$

が得られる。

さらに、上の式で、 $(N-k)$ を $t (t=1, \dots, N)$ とおくと、

$$\mathbf{T}_t | (D_N, \sigma_N^{-2}) \sim N(\mathbf{a}_N(t), \sigma_N^2 \mathbf{R}_N^\dagger(t))$$

$$\mathbf{a}_N(t) = \mathbf{m}_t + \mathbf{B}_t[\mathbf{a}_N(t+1) - \mathbf{a}_{t+1}]$$

$$\mathbf{R}_N^\dagger(t) = \mathbf{C}_t^\dagger + \mathbf{B}_t[\mathbf{R}_{t+1}^\dagger - \mathbf{R}_N^\dagger(t+1)]\mathbf{B}_t'$$

$$\mathbf{a}_N(N) = \mathbf{m}_N, \mathbf{R}_N^\dagger(N) = \mathbf{C}_N^\dagger$$

上記から、観測誤差分散 σ_N^2 が未知の場合は、

$$\begin{aligned} \mathbf{T}_t | D_N &\sim t(\mathbf{a}_N(t), S_N \mathbf{R}_N^\dagger(t), h, \nu_N) \quad [S_N = E(\sigma_N^2 | D_N)] \\ &= t(\mathbf{a}_N(t), \mathbf{R}_N(t), h, \nu_N) \end{aligned}$$

とおく。

これは、本論の(2.12.1)式、(2.12.2)式に関連する。

A.2.5. 尺度の調整

(2.5), (2.6) の尺度については、(2.8.6) ~ (2.8.7) で示すとおりの事後的な調整を行っており、以下のとおり、いずれも尺度を $(S_t^* / S_{t-1}) = (Q_t^* / Q_t)$ 倍になるように設定している。これは、精度の高い予測を行うために、ガンマ分布の期待値では求めることができ

ない尺度母数の予測期待値について階層モデルの母数推定量を用いて把握して、観測の1期先予測分布に反映させるための試みである^{*A10}。

$$E[\sigma_t^{-2} | D_{t-1}] \stackrel{def}{=} S_t^{*-1} = \left[S_{t-1} \left(\frac{Q_t^*}{Q_t} \right) \right]^{-1}$$

$$\mathbf{R}_t^* \stackrel{def}{=} \left(\frac{Q_t^*}{Q_t} \right) \mathbf{R}_t = S_t^* \mathbf{R}_t^\dagger, \quad (\mathbf{R}_t = S_{t-1} \mathbf{R}_t^\dagger, \quad Q_t = S_{t-1} (\mathbf{F}' \mathbf{R}_t^\dagger \mathbf{F} + 1))$$

$$Q_t^* \stackrel{def}{=} \mathbf{F}' \mathbf{R}_t^* \mathbf{F} + S_t^* = S_t^* (\mathbf{F}' \mathbf{R}_t^\dagger \mathbf{F} + 1)$$

*A10 このような尺度変換は、すでに指摘したとおり、マンデルブロ博士の「スケーリング理論」に基づくフラクタル的分析手法に類するものである。マンデルブロ博士のスケーリング理論に関する基本形の一つは、水理学者ハーストの研究（ナイル川の水位の記録における長期記憶解析：Rescaled range analysis）で提案されたハースト指数 H で規定される。

（参考）「ハーストのナイル川の水位の記録における長期記憶解析（Rescaled range (R/S) analysis）」

・基準統計量

$$Q(\tau) = R(\tau) / S(\tau)$$

$$Q(\tau) \propto \tau^H$$

ここに、

$$R(\tau) = \max_{1 \leq k \leq \tau} \sum_{j=1}^k (x_j - E[\tilde{x}_\tau | D_\tau]) - \min_{1 \leq k \leq \tau} \sum_{j=1}^k (x_j - E[\tilde{x}_\tau | D_\tau])$$

$$S(\tau) = \sqrt{\frac{1}{\tau} \sum_{j=1}^{\tau} (x_j - E[\tilde{x}_\tau | D_\tau])^2}$$

（確率変数にティルド[~]を付す。）

$E[\cdot | D_\tau]$ は、期間 τ における観測情報： $D_\tau = \{\tilde{x}_1, \dots, \tilde{x}_\tau\}$ に基づく期待値

この R/S 解析は、時系列データに長期相関があるかどうかを調べる場合に広く用いられており、とくに通常の正規分布で近似できる現象とは大きく異なる性質を持つデータを解析する際に特に重要性を持っている。R/S 解析の公式が示すことは「想定した期間における観測データの最大値と最小値の差が、毎回ランダムであると想定した場合の広がりとは比べて大きい小さいか」ということである。もし、それがきわめて大きなものであれば、時系列データの変化そのものの推移において増加や減少が持続することにより、最大値や最小値をさらに大きくする性質があるかどうかを検討しなければならない。変動が互いに独立（ランダム）ならば分子と分母の比は 1:2 となり、 $H=1/2$ となる。解析の結果、この値から外れた値を取る場合は長期相関の存在を疑うべきと考える。実際、変動の幅が大きい場合は $H > 1/2$ となり、長時間持続する変動が起こっていることになる。逆に、 $H < 1/2$ の時は反相関となり、変動しても元の値の水準に急に引き戻され、振幅が大きくなる傾向があることを示している (Mandelbrot *et al.* (1997a), Mandelbrot and Hudson (2004))。

本論では、観測誤差分散母数の尺度（ただし、観測誤差分散母数が未知の場合）について、

$$S_t^{*-1} \stackrel{def}{=} E[\sigma_t^{-2} | D_{t-1}] = \left[S_{t-1} \left(\frac{Q_t^*}{Q_t} \right) \right]^{-1}$$

と変換することを提案するが、本変換は上でみたハースト指数 H によるスケーリング性の把握方法と基本的に同じ構造を持っている。

なお、(2.7) のとおり、 Q_t^* ($t=1, \dots, N$) については、

$$Q_t^* = \alpha e_{t-1}^2 + \beta Q_{t-1}^* \quad (t=2, \dots, N)$$

$$\text{with } Q_1^* \equiv Q_1 \quad (t=1)$$

と求められる。 $((\alpha, \beta)$ は、DLM の体系からは直接的に明示されない隠れ母数である。)

ただし、このように修正された尺度を用いても、(2.11.1) ~ (2.11.5) で示されているとおり、その母数空間（制約）に適合した範囲内で設定される限りにおいて、再帰的に各モーメントを計算する過程には一切影響を与えないので、DLM の体系に変更をもたらすことはない。

A.2.6. 未知母数 α の提案分布

未知母数 α の提案分布の密度関数

$$f(\alpha | Y_\alpha, X_\alpha, \Sigma, \beta) \propto \varphi(\alpha | \hat{\phi}_\alpha, \hat{\Sigma}_\alpha) I(\alpha | [0, 1-\beta]) \quad (\text{A.6.1})$$

ここに、 $\varphi(\cdot | \hat{\phi}_\alpha, \hat{\Sigma}_\alpha)$ は、平均 $\hat{\phi}_\alpha$ 分散行列 $\hat{\Sigma}_\alpha$ の正規密度とし、 $I(\cdot | [0, 1-\beta])$ は、区間 $(0 \leq \alpha < (1-\beta))$ の指標関数となる。

$$\hat{\phi}_\alpha = \hat{\Sigma}_\alpha (X_\alpha' \Lambda^{-1} Y_\alpha + \Sigma_{\alpha_0}^{-1} \phi_{\alpha_0})$$

$$\hat{\Sigma}_\alpha = (X_\alpha' \Lambda^{-1} X_\alpha + \Sigma_{\alpha_0}^{-1})^{-1}$$

$$\Lambda = \text{diag} \{2Q_1^{*2}, \dots, 2Q_N^{*2}\}$$

$$Y_\alpha = (e_1^2, \dots, e_N^2)'$$

$$X_\alpha = (\bar{e}_0^2, \dots, \bar{e}_{N-1}^2)'$$

ϕ_{α_0} : α の事前期待値, Σ_{α_0} : α の事前分散

なお、

$$\bar{e}_t^2 = \begin{cases} 0 & (t=0) \\ e_t^2 + \beta \bar{e}_{t-1}^2 & (t=1, \dots, N) \end{cases}$$

$$e_t^2 = \bar{e}_t^2 - \beta \bar{e}_{t-1}^2 \quad (t=1, \dots, N)$$

明らかに、上式から、

$$\varepsilon_t \approx e_t^2 - \alpha \bar{e}_{t-1}^2 \quad (t=1, \dots, N)$$

A.2.7. 未知母数 β の提案分布

未知母数 β の提案分布は以下のとおりである。

$$f(\beta | Y_\beta, X_\beta, \Sigma, \alpha) \propto \varphi(\beta | \hat{\phi}_\beta, \hat{\Sigma}_\beta) I(\beta | [0, 1 - \alpha]) \quad (\text{A.7.1})$$

ここに、

$$\hat{\phi}_\beta = \hat{\Sigma}_\beta (X'_\beta \Lambda^{-1} Y_\beta + \Sigma_{\beta_0}^{-1} \phi_{\beta_0})$$

$$\hat{\Sigma}_\beta = (X'_\beta \Lambda^{-1} X_\beta + \Sigma_{\beta_0}^{-1})^{-1}$$

$$\Lambda = \text{diag}\{2Q_t^{*2}, \dots, 2Q_N^{*2}\}$$

$$Y_\beta = (\varepsilon_1(\beta^\dagger) + \xi_1(\beta^\dagger)\beta^\dagger, \dots, \varepsilon_N(\beta^\dagger) + \xi_N(\beta^\dagger)\beta^\dagger)'$$

$$X_\beta = (\xi_1(\beta^\dagger), \dots, \xi_N(\beta^\dagger))'$$

ϕ_{β_0} : β の事前期待値, Σ_{β_0} : β の事前分散

ここで、 β^\dagger は、MCMC によって逐次採択される未知母数 β に係る推定量である。(ただし、未知母数に係る初期推定量は EM アルゴリズムによって推定される。)

さらに、 β^\dagger で評価された下位モデルの残差を $\varepsilon_t(\beta^\dagger)$ と表すと、モデル残差について、以下のとおり近似することができる、

$$\varepsilon_t(\beta) \approx \varepsilon_t(\beta^\dagger) + \xi_t(\beta - \varepsilon_{t-1}(\beta^\dagger)),$$

$$\xi_t(\beta^\dagger) = \left. \frac{\partial \varepsilon_t(\beta)}{\partial \beta} \right|_{\beta = \beta^\dagger}.$$

$\xi_t(\beta^\dagger)$ については、以下のとおり定義される^{*A11}。

$$\xi_t(\beta^\dagger) = \begin{cases} 0 & (t=1) \\ \varepsilon_{t-1}(\beta^\dagger) + \xi_{t-1}(\beta^\dagger)\beta^\dagger & (t=2, \dots, N) \end{cases}$$

*A11 ここで示した定義は、基本的に、時系列分析における自己回帰移動平均 (ARMA 型) モデル等における微係数の求め方と同じである。詳しくは、山本 (1988) の 3 章および 4 章を参照願いたい。