

Research Paper Series

No. 45

マクロ経済変数を用いた住宅ローンの
デフォルト及びプリペイメント分析

松尾 竜悟[†]、竹原 浩太[‡]、室町 幸雄[§]

2023年3月

[†] りそな銀行融資企画部

[‡] 東京都立大学大学院経営学研究科

[§] 東京都立大学大学院経営学研究科

マクロ経済変数を用いた住宅ローンの デフォルト及びプリペイメント分析*

松尾 竜悟[†] 竹原 浩太[‡] 室町 幸雄[§]

2023年3月31日

概要

金融機関において住宅ローンは貸出ポートフォリオの大きな部分を占め、そのリスク管理・収益管理は重要な課題の一つと認識されている。住宅ローンは法人向け融資等、他の貸出と比較して貸出期間は長期に渡り、リスクの水準は時間とともに変化する。本稿では、住宅ローンに内在する「デフォルト」「全額繰上返済」「一部繰上返済」の3つのリスクをマクロ経済変数を共変量とする拡張Coxモデルにより分析し、デフォルトでは「完全失業率（ラグ12か月）」が、全額繰上返済では「金利差（ラグ2か月）」が相応の説明力を有し、かつ、住宅ローンの貸付実行時期に依らず高い説明力を維持することを示した。さらに、残差分析で更なる改善の可能性が示唆されたので、共変量に対する感応度が経過期間により変化する区分Coxモデルによる分析も行ったところ、感応度の期間構造の存在が有意に示唆された。例えば、「全額繰上返済」では影響力の強い金利差に対する感応度が経過期間とともに低下するという結果が得られたが、これはバーンアウト効果として知られる現象と整合的である。また、コーホートごとにダミー変数を追加して分析したところ、2009年頃に貸付実行された住宅ローンはデフォルトが発生しやすく、繰上償還が発生しにくい傾向が見られたが、この時期の住宅ローンは返済比率が高いという傾向と整合的である。

1 はじめに

金融機関において住宅ローンは貸出ポートフォリオの大きな部分を占め、そのリスク管理・収益管理は重要課題の一つと認識されている。昨今の低金利下において金融機関は歴史的超低利の住宅ローンの残高を互いに競いながら積み上げてきたが、ここでもしも市場金利が上昇すると、長期間超低利で固定された貸付（住宅ローン）を長期間保有することになる。住宅ローンのリスク評価で特に重要なのはデフォルト（債務不履行）とプリペイメント（期限前償還）の発生であり、観測データに基づくこれらのリスクの分析・評価が強く求められている。また、住宅ローン債権を裏付として発行され

*本稿の内容は著者の所属組織の公式見解を示すものではなく、全て個人の見解である。

[†]りそな銀行 融資企画部（E-mail: ryugo.matsuo@resonabank.co.jp）

[‡]東京都立大学 大学院経営学研究科（E-mail: ktakehara@tmu.ac.jp）

[§]東京都立大学 大学院経営学研究科（E-mail: muromachi-yukio@tmu.ac.jp）

る RMBS（住宅ローン債権担保証券）は低金利環境下が続く昨今では比較的利回りの高い長期投資商品で、絶対利回りを求める投資家はそのリスク管理に関心を持っているが、その評価・管理においてもデフォルトやプリペイメントによる期限前償還リスクは重視されている¹。住宅ローンのプリペイメントには「全額繰上返済」と「一部繰上返済」があり、「全額繰上返済」（以下では全繰と略）は債務者が住宅売却資金や他金融機関で借り入れた借換資金などにより住宅ローンを完済すること、「一部繰上返済」（以下では一繰と略）は余裕資金などにより住宅ローン残高の一部を返済することをいう。本稿では、「全繰」と「一繰」を総称して「プリペイメント」と呼ぶが、データ分析ではこれらを区別して取り扱う。

RMBS の価格付けで広く用いられるプリペイメント率モデルには、米国の PSA（Public Securities Association）モデルや本邦の PSJ（Prepayment Standard Japan）モデルがある。これらはプリペイメント率を経過期間の確定関数で表現するため簡便で扱いやすい。一方、債務者属性などに依存し、かつ期間構造を持つモデルとしては Cox 比例ハザードモデル（Cox [2]）がよく使われる。また、共変量を時間依存変数に拡張した拡張 Cox モデル²も用いられる。

これらのモデルを用いた本邦の金融分野の先行研究としては、青沼・木島 [13] による定期預金の解約率の分析や、一條・森平 [7] や沓澤 [8] による住宅ローンの分析がある。これらの論文では、共変量として債務者の職業や返済比率などの債務者属性や個々のローンに固有の変数（内的変数と呼ぶ）を用いることが多いが、それらは逐次入手できるとは限らないため、ローン審査時点の情報で評価することになる場合が多い。しかし、その時々々のマクロ経済変数が及ぼす影響も大きいと考えられる。

住宅ローンを分析対象とした先行研究でも共変量にマクロ経済変数は用いられてきたが、その多くは市場金利である。Schwartz and Torous [6] は RMBS 評価のために住宅ローンのプリペイメントリスクの分析を行ったが、そこでは基準となる期間構造はパラメトリックな確率分布に従うと仮定し、さらに「住宅ローンの適用金利と借換金利（長期のトレジャリーレートで代替）の差」「過去の期限前償還額の大きさを表す変数」「季節ダミー」を共変量として使用した。一條・森平 [7] は、本邦の民間金融機関の住宅ローンの「全額繰上返済」の発生強度を拡張 Cox モデルで分析し、特に説明力の高い共変量は「一部繰上返済回数」「職業ダミー」「資金用途ダミー」「住宅ローンの適用金利と市場金利の比」であることを示した。杉村 [11] は、Schwartz and Torous [6] のモデルを改良して本邦の民間金融機関の住宅ローンの「全額繰上返済」「一部繰上返済」「代位弁済（本稿の「デフォルト」に相当する）」の 3 つの競合リスクを分析し、共変量に「住宅ローンの適用金利と市場金利の比」「残存期間」「債務者の年齢」を用いることでモデルの適合度が向上することを示した。Quercia et al. [5] は、米国の住宅ローンを対象に、「失業率」がデフォルトやプリペイメントに与える影響を拡張 Cox

¹RMBS では、住宅ローン債務者のデフォルトもプリペイメントも元本の削減として扱われるため、キャッシュフローの評価ではどちらも期限前償還とみなすことができる。

²拡張 Cox モデルという呼称は Kleinbaum and Klein [4] の翻訳にならった。

モデルを用いて分析し、特に失業率の「構造的要素」の悪化がデフォルトの発生強度を強め、プリペイメントの発生強度を弱めることを示した。Deng and Haghani [3] は米国における差し押さえ後の住宅ローンのデフォルトとプリペイメントの相互依存性を拡張 Cox モデルで分析した。しかし筆者らの知る限り、本邦の住宅ローンを用いた先行研究では市場金利以外のマクロ経済変数を共変量とする分析は行われていないようである。

そこで本稿では、本邦の住宅金融支援機構の長期固定金利住宅ローンを対象に、金利差以外のマクロ経済変数も共変量として分析を行った。分析に使用するモデルは Cox の比例ハザードモデルをもとに共変量を時間依存変数に拡張したもので、拡張 Cox モデルと呼ばれている。拡張 Cox モデルによる分析結果は既存の結果とおおよそ整合的であったが、残差分析の結果をみると更なる改善の余地が見られた。そこで、共変量に対する感応度が時間区間により変化する区分 Cox モデルによる分析も行い、拡張 Cox モデルと比較検討したところ、共変量に対する感応度の期間構造が有意に存在することがわかり、例えば、住宅ローンの全額繰上返済においてバーンアウト効果と呼ばれる現象と整合的な結果が得られた。さらに残差分析から、推定モデルの適合性がコーホート³に強く依存することが示唆されたので、共変量にコーホートダミー変数を加えた分析も行った。

本稿の構成は以下である。2 節では本研究で使用する拡張 Cox モデルと競合リスクモデルについて述べ、3 節では分析で使用するデータと拡張 Cox モデルによる分析結果を示す。4 節以降では、それぞれのイベント（4 節ではデフォルト、5 節では全額繰上返済、6 節では一部繰上返済）に関して区分 Cox モデルを用いて共変量に対する感応度の期間構造を分析した結果を示し、7 節でまとめる。

2 分析に使用するモデル

2.1 競合リスクモデルと拡張 Cox モデル、区分 Cox モデル

生存時間解析では、基準時点からある事象が発生するまでの時間（以下では「イベント発生時間」という）を分析対象とする。本稿では、住宅ローンの貸付実行時点から契約終了時点までの経過期間に焦点を当てて生存時間解析のフレームワークを適用する⁴。

時間を $t \geq 0$ で表し、 $t = 0$ を基準時点とする。イベント発生時間を表す連続的な非負の確率変数を $T > 0$ とすると⁵、生存関数は $S(t) := P\{T > t\}$ であり、ハザード関数は

$$h(t) := \lim_{\Delta t \rightarrow 0} \frac{P\{t < T \leq t + \Delta t | T > t\}}{\Delta t} \quad (2.1)$$

³貸付実行年が同じ住宅ローンの集合。

⁴以下の記述は Collett [1] の第 2・3 章にならった。

⁵確率変数 T をイベント発生「時刻」（例：死亡時刻）とする場合もある。本稿では観測開始時点（基準時点）が異なる分析対象個体間での比較を論じやすいように、観測開始時点からイベント発生までの「時間」（例：生存時間）とする。

で表現される。(2.1) より,

$$S(t) = \exp\{-H(t)\}, \quad H(t) := \int_0^t h(s)ds \quad (2.2)$$

であり, $S(t)$ と $h(t)$ の期間構造はどちらか一方が決まれば他方も決まる.

本稿では, 「デフォルト」「全部繰上返済」「一部繰上返済」の3種類のイベントを扱う. 複数のイベントを扱う場合, 着目するイベントより先に他のイベントが発生すると, その後の着目イベントの発生が妨げられる. このような場合の分析には競合リスクモデルが使われる. Q 個の異なるタイプのイベントがあり, タイプ q , $q = 1, \dots, Q$ のイベント発生時間を表す確率変数を $T_q > 0$ とし, $T := \min\{T_1, \dots, T_Q\}$ とすると, タイプ q の要因によるハザード関数は

$$h_q(t) := \lim_{\Delta t \rightarrow 0} \frac{P\{t < T_q \leq t + \Delta t | T > t\}}{\Delta t}, \quad q = 1, 2, \dots, Q \quad (2.3)$$

で定義され, 原因別ハザード関数と呼ばれる. $h_q(t)$ は時間 t まで全てのタイプのイベントが発生していないという条件の下で $(t, t + dt]$ 間の一瞬にタイプ q のイベントが発生する強度であり, イベントが一種類しかない場合のハザード関数 $h(t)$ とは異なることに注意されたい. さらに, 本稿の競合リスクモデルでは, 分析対象イベントと打ち切りは独立と仮定して分析する⁶.

観測期間内でのイベント発生時間を $0 < t_{(1)} < t_{(2)} < \dots < t_{(j)}$ とし, 時間 $t_{(j)}$ でイベントが発生した個体数を d_j , 時間 $t_{(j)}$ でイベント発生リスクに晒されている(時間 $t_{(j)}$ の直前 $t_{(j)} -$ までイベントが発生せず, 打ち切りも受けていない) 個体数を r_j とする. このとき, Kaplan–Meier 推定量 $\hat{S}^{KM}(t)$ は,

$$\hat{S}^{KM}(t) := \prod_{j: t_{(j)} \leq t} \left(1 - \frac{d_j}{r_j}\right) \quad (2.4)$$

で定義される. ただし, $0 \leq t < t_{(1)}$ では $\hat{S}^{KM}(t) = 1$ とする. r_j には時間 $t_{(j)}$ でイベントが発生した個体と打ち切りされた個体が含まれていることに注意されたい. また, $t \in [t_{(j)}, t_{(j+1)})$ において発生強度を一定と仮定すると, ハザード関数は

$$\hat{h}^{KM}(t) := \frac{d_j}{r_j(t_{(j+1)} - t_{(j)})}, \quad j = 1, \dots, J-1 \quad (2.5)$$

で定義できる. ただし, $0 \leq t < t_{(1)}$ では $\hat{h}^{KM}(t) = 0$ とする.

イベント発生時間の分布がさまざまな変数に依存する際によく利用されるのが Cox [2] の比例ハザードモデルである. イベント発生に影響を与える変数群を p 次元共変量ベクトル $\mathbf{x} = (x_1, \dots, x_p)^\top$ とし, 個体 i の p 次元共変量ベクトルを $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ として, 個体 i のハザード関数を

$$h_i(t; \mathbf{x}_i) = h_0(t) \exp(\boldsymbol{\beta}^\top \mathbf{x}_i) \quad (2.6)$$

⁶分析対象イベントと打ち切りの間に依存関係がある場合, 推定結果には偏りが生じうる. また, 一方のイベントは打ち切りと同じ効果を持つが, もう一方のイベントには打ち切りの効果がないという半競合リスクという概念も提案されており, 本稿の研究に応用可能と思われる. 詳細は西川 [12] や江村・道前 [9] などを参照.

と表現する．ただし， $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$ は共変量に対する感応度を示す p 次元定係数ベクトル， $h_0(t)$ は時間 t のみの関数で，ベースラインハザード関数と呼ばれる．このとき，

$$H_i(t; \mathbf{x}_i) := \int_0^t h_i(s; \mathbf{x}_i) ds = H_0(t) \exp(\boldsymbol{\beta}^\top \mathbf{x}_i), \quad H_0(t) := \int_0^t h_0(s) ds$$

であり，個体 i の条件付生存関数は

$$S_i(t; \mathbf{x}_i) := S_0(t)^{\exp(\boldsymbol{\beta}^\top \mathbf{x}_i)}, \quad S_0(t) := \exp(-H_0(t))$$

と表現され， $S_0(t)$ はベースライン生存関数と呼ばれる．

比例ハザードモデルの共変量を時間依存変数に拡張したものが拡張 Cox モデルで，ハザード関数は， p 次元共変量ベクトルを $\mathbf{x}_i(t) = (x_{i1}(t), \dots, x_{ip}(t))^\top$ として，

$$h_i(t; \mathbf{x}_i(t)) = h_0(t) \exp(\boldsymbol{\beta}^\top \mathbf{x}_i(t)) \quad (2.7)$$

で表現される．通常の Cox モデルでは異なる個体 a, b に対するハザード比 $\frac{h_a(t; \mathbf{x}_a)}{h_b(t; \mathbf{x}_b)}$ が時間 t に依存せず一定となるが，拡張 Cox モデルではハザード比が

$$\frac{h_a(t; \mathbf{x}_a)}{h_b(t; \mathbf{x}_b)} = \frac{\exp(\boldsymbol{\beta}^\top \mathbf{x}_a(t))}{\exp(\boldsymbol{\beta}^\top \mathbf{x}_b(t))} = \exp(\boldsymbol{\beta}^\top (\mathbf{x}_a(t) - \mathbf{x}_b(t)))$$

のように時間 t に依存するため，比例ハザード性が成立しないことに注意されたい．

本稿では共変量にマクロ経済変数などを採用するため，拡張 Cox モデルを用いて分析を行うが，

$$h_i(t; \mathbf{x}_i(t)) = h_0(t) \exp(\boldsymbol{\beta}^\top \mathbf{x}_i(t)), \quad \mathbf{x}_i(t) = \mathbf{y}(t_0^i + t)$$

と定式化する．ただし， t_0^i は i 番目の住宅ローンの貸付実行時点， t は t_0^i からの経過期間， $\mathbf{y}(t_0^i + t)$ は暦上の時刻 $(t_0^i + t)$ における時間依存変数ベクトルである．したがって，住宅ローンの貸付実行時点が異なる個体 a, b に対する共変量は，同じ時間依存変数であっても $\mathbf{x}_a(t) \neq \mathbf{x}_b(t)$ となる．

さらに本研究では，時間 t を分割した各区間内で共変量の感応度が一定値をとる区分 Cox モデルを用いて感応度の期間構造について検討する．区分 Cox モデルのハザード関数は， $V + 1$ 個のノード $t_v, v = 0, 1, \dots, V, 0 = t_0 < t_1 < \dots < t_V$ を使い，時間 t を V 個の区間 $s_v = [t_{v-1}, t_v), v = 1, \dots, V$ に分割して，

$$h_i(t; \mathbf{x}_i(t)) := h_0(t) \exp \left\{ \sum_{v=1}^V \boldsymbol{\beta}_v^\top \mathbf{x}_i(t) 1_{s_v}(t) \right\} \quad (2.8)$$

で表現される．ここで， $1_{s_v}(t)$ は $t \in s_v$ のとき 1， $t \notin s_v$ のとき 0 をとる定義関数， $\boldsymbol{\beta}_v = (\beta_{v1}, \dots, \beta_{vp})^\top, v = 1, \dots, V$ は p 次元定数ベクトルである．

2.2 共変量に対する感応度とベースラインハザード関数の推定

本項では Collett [1] の 3.3 節をもとにモデルの感応度パラメータ β の推定について拡張 Cox モデルを使って記述するが、どのモデルでもロジックは同じである。

部分尤度によるパラメータ推定ではタイデータの取り扱いに注意が必要である。タイデータとは、複数個体で分析対象イベントや打ち切りが同じ時間に発生しているデータのことである。実際の分析データでは、一定の時間間隔でイベントの発生を計測することが多く、タイデータは頻繁に現れる。本稿で分析する住宅ローンのデータは月次データで、1ヶ月間のイベント発生件数が当月末時点で記録されるため、タイデータの取り扱いが必要となる。

I 個の観察対象個体を考え、各個体は観察期間内にイベントが発生または打ち切られることで観測を終了するものとし、その時間を $t_i, i = 1, \dots, I$ とする。また、観測期間内に J 個のイベントが発生し、 $I - J$ 個が打ち切られたものとする。

まず、タイデータがない場合を考える。イベント発生時間を $t_{(1)} < \dots < t_{(J)}$ とし、 $t_{(0)} = 0$ とする。 $t_{(j)}, j = 1, \dots, J$ でイベントが発生した個体を \mathbf{x} に添字 (j) を付けて表現すると、部分尤度関数は、

$$\tilde{L}(\beta) = \prod_{j=1}^J \frac{\exp(\beta^\top \mathbf{x}_{(j)}(t_{(j)}))}{\sum_{\ell \in \mathcal{R}(t_{(j)})} \exp(\beta^\top \mathbf{x}_\ell(t_{(j)}))} \quad (2.9)$$

で与えられる。ただし、 $\mathcal{R}(t)$ は時間 t でイベント発生リスクのある個体の集合⁷（以下ではリスク集合という）で、 $\mathbf{x}_\ell(t_{(j)})$ は時点 $t_{(j)}$ で打ち切られた個体 ℓ の時点 $t_{(j)}$ における共変量ベクトルである。この部分尤度関数を最大化することで、パラメータ β に対する推定値 $\hat{\beta}$ が得られる。

タイデータがある場合、同時点におけるイベント発生個体の順序を考慮する。最も正確な評価式である Exact 法（省略）は、タイデータが多くなると計算時間が膨大になるので、Breslow 法や Efron 法などの近似式がしばしば使用される。観測期間内のイベント発生時間を $t_{(1)} < \dots < t_{(J)}$ ⁸、時間 t にイベント発生した個体の集合を $\mathcal{D}(t)$ 、 $\mathcal{D}(t_{(j)})$ の要素数を d_j とすると、Breslow 法では部分尤度関数を

$$\tilde{L}(\beta) = \prod_{j=1}^J \frac{\prod_{k \in \mathcal{D}(t_{(j)})} \exp(\beta^\top \mathbf{x}_k(t_{(j)}))}{\left(\sum_{\ell \in \mathcal{R}(t_{(j)})} \exp(\beta^\top \mathbf{x}_\ell(t_{(j)})) \right)^{d_j}} \quad (2.10)$$

⁷時間 t の直前 $t-$ でイベントも打ち切りも発生していない個体からなる集合。

⁸タイデータがない場合と異なり、ここでは J がイベント発生時点の総数であることに注意。

で, Efron 法では

$$\tilde{L}(\boldsymbol{\beta}) \simeq \prod_{j=1}^J \frac{\prod_{k \in \mathcal{D}(t_{(j)})} \exp(\boldsymbol{\beta}^\top \mathbf{x}_k(t_{(j)}))}{\prod_{d=1}^{d_j} \left(\sum_{\ell \in \mathcal{R}(t_{(j)})} \exp(\boldsymbol{\beta}^\top \mathbf{x}_\ell(t_{(j)})) - \frac{d-1}{d_j} \sum_{k \in \mathcal{D}(t_{(j)})} \exp(\boldsymbol{\beta}^\top \mathbf{x}_k(t_{(j)})) \right)} \quad (2.11)$$

で近似する. Efron 法による近似式 (2.11) は Breslow 法による近似式 (2.10) に比べて精度が高く⁹, Breslow 法の方が計算時間は短い.

西川 [12] によると, イベント発生が独立な場合, 競合リスクにおける原因別ハザード関数の感応度 $\boldsymbol{\beta}$ は, イベント別の部分尤度関数をそれぞれ独立に最大化することで同時に, 競合リスクモデル全体としての部分尤度関数も最大化する. ただし, 各イベントに対応する部分尤度関数を考える際, (2.10) や (2.11) において, 集合 $\mathcal{D}(t_{(j)})$ を時間 $t_{(j)}$ で分析対象イベントが発生した個体のみの集合とし, リスク集合 $\mathcal{R}(t_{(j)})$ には時間 $t_{(j)}$ でイベントが発生する全ての個体を含めることになる¹⁰.

本稿では, 部分尤度関数の計算には近似的な Breslow 法または Efron 法を用いる. また, パラメータの推定値 $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)^\top$ のそれぞれには帰無仮説 $H_0^i: \hat{\beta}_i = 0$ に対する漸近正規性を利用した検定を行う. また, モデル適合度の検定にはワルド検定を使用する.

時間区間 $[t_{(j)}, t_{(j+1)})$, $j = 0, \dots, J-1$ におけるベースラインハザード関数の推定値 $\hat{h}_0(t_{(j)})$ は, ξ_j に関する非線形方程式

$$\sum_{k \in \mathcal{D}(t_{(j)})} \frac{\exp(\hat{\boldsymbol{\beta}}^\top \mathbf{x}_k(t_{(j)}))}{1 - \xi_j} = \sum_{\ell \in \mathcal{R}(t_{(j)})} \exp(\hat{\boldsymbol{\beta}}^\top \mathbf{x}_\ell(t_{(j)})) \quad (2.12)$$

の解を $\hat{\xi}_j$ とし,

$$\hat{h}_0(t_{(j)}) = 1 - \hat{\xi}_j \quad (2.13)$$

で与えられる. (2.12) はタイデータが存在する場合には解析的に解けないので, Newton-Raphson 法などの反復法で解くことになるが, 本稿では一條・森平 [7] で提案された,

$$\log \tilde{\xi}_j = \frac{-d_j}{\sum_{\ell \in \mathcal{R}(t_{(j)})} \exp(\hat{\boldsymbol{\beta}}^\top \mathbf{x}_\ell(t_{(j)}))} \quad (2.14)$$

で与えられる $\tilde{\xi}_j$ を反復法の初期値として収束を早める手法を採用する.

(2.12) の解 $\hat{\xi}_j$ を用いると, ベースライン生存関数の推定値は, $t_{(\alpha)} \leq t < t_{(\alpha+1)}$ を満たす自然数を α とし,

$$\hat{S}_0(t) = \prod_{j=1}^{\alpha} \hat{\xi}_j \quad (2.15)$$

⁹Collett [1] は, どちらの近似式でも大差はないと述べている. 本稿の分析でも推定値の差が小さいことを確認した.

¹⁰対象としないイベントが発生した個体は打ち切りを受けたとみなすことに相当する.

で与えられる階段関数となる。ただし、 $0 \leq t < t_{(1)}$ では $\hat{S}_0(t) = 1$ とする。また、時間区間 $[t_{(j)}, t_{(j+1)})$ 内でイベント発生強度は一定と仮定すると、ベースラインハザード関数の推定値は

$$\hat{h}_0(t) = \frac{1 - \hat{\xi}_j}{t_{(j+1)} - t_{(j)}}, \quad t \in [t_{(j)}, t_{(j+1)}) \quad (2.16)$$

で与えられる。ただし、 $0 \leq t < t_{(1)}$ では $\hat{h}_0(t) = 0$ とする¹¹。条件付生存関数 $\hat{S}_i(t)$ の推定値は

$$\hat{S}_i(t) = \exp \left\{ - \int_0^t \hat{h}_i(s; \mathbf{x}_i(s)) ds \right\} = \exp \left\{ - \int_0^t \hat{h}_0(s) \exp \left(\hat{\boldsymbol{\beta}}^\top \mathbf{x}_i(s) \right) ds \right\}$$

で与えられる。

2.3 統計学的残差 (Cox-Snell 残差)

生存解析における残差分析では Cox-Snell 残差がよく用いられるので、本稿でも使用する。観測データ j の Cox-Snell 残差 R_j^{CS} は、 $t_j > 0$ をイベント発生時間または打ち切り時間（観測終了時間）、 $\hat{S}_j(\cdot)$ をモデルによる j の生存関数の推定値、累積ハザード率の推定値を $\hat{H}_j(\cdot)$ として、

$$R_j^{CS} = -\log \hat{S}_j(t_j) = \hat{H}_j(t_j) \quad (2.17)$$

で定義される。具体的な表現はモデルによって異なり、観測データ j の累積ベースラインハザード率とベースラインハザード率の推定値をそれぞれ $\hat{H}_0(\cdot)$ 、 $\hat{h}_0(\cdot)$ とすると、Cox 比例ハザードモデルでは

$$R_j^{CS} = \exp \left(\hat{\boldsymbol{\beta}}^\top \mathbf{x}_j \right) \hat{H}_0(t_j), \quad (2.18)$$

拡張 Cox モデルでは

$$R_j^{CS} = \int_0^{t_j} \hat{h}_0(u) \exp \left(\hat{\boldsymbol{\beta}}^\top \mathbf{x}_j(u) \right) du, \quad (2.19)$$

区分 Cox モデルでは

$$R_j^{CS} = \int_0^{t_j} \hat{h}_0(u) \exp \left(\sum_{v=1}^V \hat{\boldsymbol{\beta}}_v^\top \mathbf{x}_j(u) 1_{s_v}(u) \right) du \quad (2.20)$$

である。適切なモデルで推定された場合、 R_j^{CS} はパラメータ 1 の指数分布（標準指数分布）に従う。

3 拡張 Cox モデルによる分析

3.1 データと分析方法

本研究では、住宅金融支援機構（以下、機構と略）が資産担保証券（以後、機構 MBS と略）の投資家向けに開示している「償還履歴データ（コアデータ）」¹²を使用する。これは 1997 年以降の毎年

¹¹ここでは Collett [1] の記述に倣った。(2.13) は時間区間に対するハザード確率、(2.16) は瞬間的な強度である。

¹²https://www.jhf.go.jp/investor/shisan_tanpo/kihatsu.html から取得した。データ等の詳細は住宅金融支援機構市場資金部 [15] を参照。

3月に貸付実行された機構の長期固定金利住宅ローンで、機構と提携した民間金融機関が販売・貸付を実行し、機構が買い取りを行ったローン債権の、月次の返済履歴データである。このうち1997年から2019年までの毎年3月に貸付実行された住宅ローンの2020年3月までの返済履歴データを分析する。

分析するイベントは「デフォルト」「全額繰上返済（全繰）」「一部繰上返済（一繰）」の3種類で、一條・森平 [7] にならい、月々のデフォルト金額やプリペイメント金額を一定値で割った値をイベント発生数とする¹³。本来ならば「一繰」を実行した個体はデータから除外されずに残り、再度イベントを繰り返すこともあるが、本稿の分析では「デフォルト」「全繰」と同様にイベント観測時点以降のデータから除外される点に注意されたい。

共変量には「完全失業率（季節調整値）」「金利差」「日経平均株価」「鉱工業生産指数（季節調整済指数）」を使用する。ここで、「金利差」は各時点における住宅ローンの残存債権金額に対する適用金利の加重平均値（以下では加重平均金利という）と円金利スワップレート10年物の差とする。円金利スワップレート10年物と「日経平均株価」は日次データから月次平均値を求めて使用する。また、共変量では最長12ヶ月（1ヶ月、2ヶ月、3ヶ月、6ヶ月、9ヶ月、12ヶ月）間のタイムラグも考慮する。

「デフォルト」「全繰」「一繰」の3つのイベントを競合リスクとし、それぞれの原因別ハザード関数を拡張Coxモデルで表現して、共変量に対する感応度やベースラインハザード関数を推定する。共変量候補のマクロ経済変数は分析対象期間の平均と標準偏差で標準化して使用する。このため、推定結果 $\hat{\beta}$ の各成分の絶対値は該当する共変量への依存性の強さを示すと解釈できる。また、部分尤度を \tilde{L} 、パラメータの推定値を $\hat{\beta}$ 、パラメータ数（=共変量ベクトルの次数）を p として、

$$AIC(\hat{\beta}, p) = -2\log\tilde{L}(\hat{\beta}) + 2p \quad (3.1)$$

で与えられるAICを評価規準とし、変数のタイムラグも考慮した総当たり法でAICが最小となる共変量の組み合わせを求める。なお、事前分析によりBreslow法とEfron法の推定結果の差は小さいことが確認できたため、総当たり法による共変量の最適な組み合わせと推定値の算出には、まず計算負荷の軽いBreslow法を適用し、得られた最適な組み合わせにEfron法を適用して最終的な推定値を算出する。

3.2 共変量に対する感応度

表1a, 1b, 1cに、それぞれのモデルにおける推定値を示す。各パラメータの p 値とWald検定の p 値によると、これらのモデルはどれも統計的に有意である。なお、3.1節で挙げた4つのマクロ経済変数のうち「日経平均株価」はどのモデル・イベントの場合でも選ばれなかった。

¹³本研究では一定値を1,000,000とした。この一定値を変更しても、以下で示す結果はあまり変わらなかった。

表 1: 各種イベントに対する拡張 Cox モデル
(a) デフォルト

モデル名	Wald 検定統計量 (p 値) AIC	共変量 (L : ラグの大きさ)		感応度	標準誤差	p 値
			β			
拡張 Cox モデル	5,015.0(0.0) 1,296,254	失業率 (L12)	β	0.12157	0.00502	0.0
		金利差 (L3)	β	0.36808	0.00744	0.0
		鉱工業 生産指数 (L0)	β	-0.05852	0.00512	0.0

(b) 全額繰上返済

モデル名	Wald 検定統計量 (p 値) AIC	共変量 (L : ラグの大きさ)		感応度	標準誤差	p 値
			β			
拡張 Cox モデル	75,330.3(0.0) 31,618,306	失業率 (L0)	β	-0.06919	0.00107	0.0
		金利差 (L2)	β	0.36163	0.00154	0.0
		鉱工業 生産指数 (L6)	β	-0.11292	0.00117	0.0

(c) 一部繰上返済

モデル名	Wald 検定統計量 (p 値) AIC	共変量 (L : ラグの大きさ)		感応度	標準誤差	p 値
			β			
拡張 Cox モデル	43,159.1(0.0) 5,546,755	失業率 (L2)	β	0.42998	0.00253	0.0
		金利差 (L12)	β	0.14796	0.00382	0.0
		鉱工業 生産指数 (L9)	β	0.18350	0.00198	0.0

表 1a によると、デフォルトに関しては「完全失業率」や「金利差」に比べて「鉱工業生産指数」の影響力（ここではパラメータの絶対値）が低い。「完全失業率」に関しては、 $\beta > 0$ であることとラグ 12ヶ月¹⁴が最適モデルに選ばれたことは、単変量による分析の結果（詳細は略、以下でも単変量による分析結果は略）とも整合的である。表 1b によると、全線に関してはラグ 2ヶ月の「金利差」の影響力が高く、事前の単変量による分析結果と整合的である。また、「金利差」が重要という結果は一條・森平 [7] や杉村 [11] の結果とも整合的である。一方、「完全失業率」に関しては、推定値は統計的に有意であるが、影響力は相対的に低く、感応度の符号は負になるなど、単変量による事前分析と異なる。表 1c によると、一線の感応度はどれも統計的に有意に正である。このうち「完全失業率」の影響力が最も大きいだが、この結果は、収入の低下を示唆する失業率の上昇が一線の発生増加を暗示するため、直感にやや反している。一方、「鉱工業生産指数」が高いのは景気が良い状態であり、債務者の資金繰りが良好で余裕資金を一線に充てやすくなると考えられる。また、「金利差」が高いほど債務者には一線による残高低減のインセンティブが働くと考えられるので、「鉱工業生産指数」と「金利差」に関する結果は経済学的に尤もらしい。なお、「金利差」と「鉱工業生産指数」のラグはそれぞれ 12ヶ月と 9ヶ月であり、デフォルトの「完全失業率」の 12ヶ月とともに推定結果の中では長めであった。

さらに、分析対象を前半の融資（貸付実行年 1997 年から 2008 年まで、経過期間は最長 276ヶ月）と後半の融資（貸付実行年 2009 年から 2019 年まで、経過期間は最長 132ヶ月）に分け、それぞれに対して最適モデルと同じ共変量を用いてパラメータ推定を行うことで感応度の安定性を調べたところ¹⁵、デフォルトでは「完全失業率」、全線では「金利差」が融資の実行時期に依らず高い影響力を有することがわかったが、一線では「完全失業率」以外の共変量の符号は前半と後半で異なり、安定して高い影響力を持つ共変量は見当たらなかった。

3.3 ベースラインハザード関数

図 1a-1c は、それぞれ「デフォルト」「全線」「一線」の最適モデルにおけるベースラインハザード関数¹⁶（青線）と、全てのデータを使用して Kaplan–Meier 推定量から (2.5) より算出した各経過期間のハザード確率の実測値（黒線）で、どちらも年率換算値である。

図 1a によると、デフォルトでは 3, 8, 12 月に対応する経過期間でベースラインハザード関数も実測値も高くなるという季節性が見られる。また、ベースラインハザード関数には経過期間に対して緩やかな上昇傾向も見られる。図 1b によると、全線のベースラインハザード関数は経過期間 3 年頃から季節性が顕著になり、10 年を過ぎる頃まで続くが、その後は低減する。また、経過期間に関して

¹⁴表 1 の中の $L12$ はラグ 12ヶ月のマクロ経済変数であることを示す。他も同様である。

¹⁵結果は松尾 [10] を参照。

¹⁶本稿の分析では各共変量を標準化しているため、ここで示すベースラインハザード関数は、共変量が平均値をとるときハザード関数、いわば平均的なハザード関数とみなすことができる。

は、期間5年頃までは上昇し、その後減少に転じるという明確な期間構造が見える。金利による影響を取り除いてもこのような期間構造が見られることは特筆すべき点である。図1cによると、一線のベースラインハザード関数にも実測値にも暦月による季節性として、およそ1年ごとの明確なピークが全期間を通して顕著に見られる。また、経過期間に対しては緩やかな減少傾向も見られる。

なお、どの事象に関しても長期部分ではベースラインハザード関数が大きく変動するが、これは過去のイベントの発生により母数である債務者総数が次第に減少していくためと考えられる。

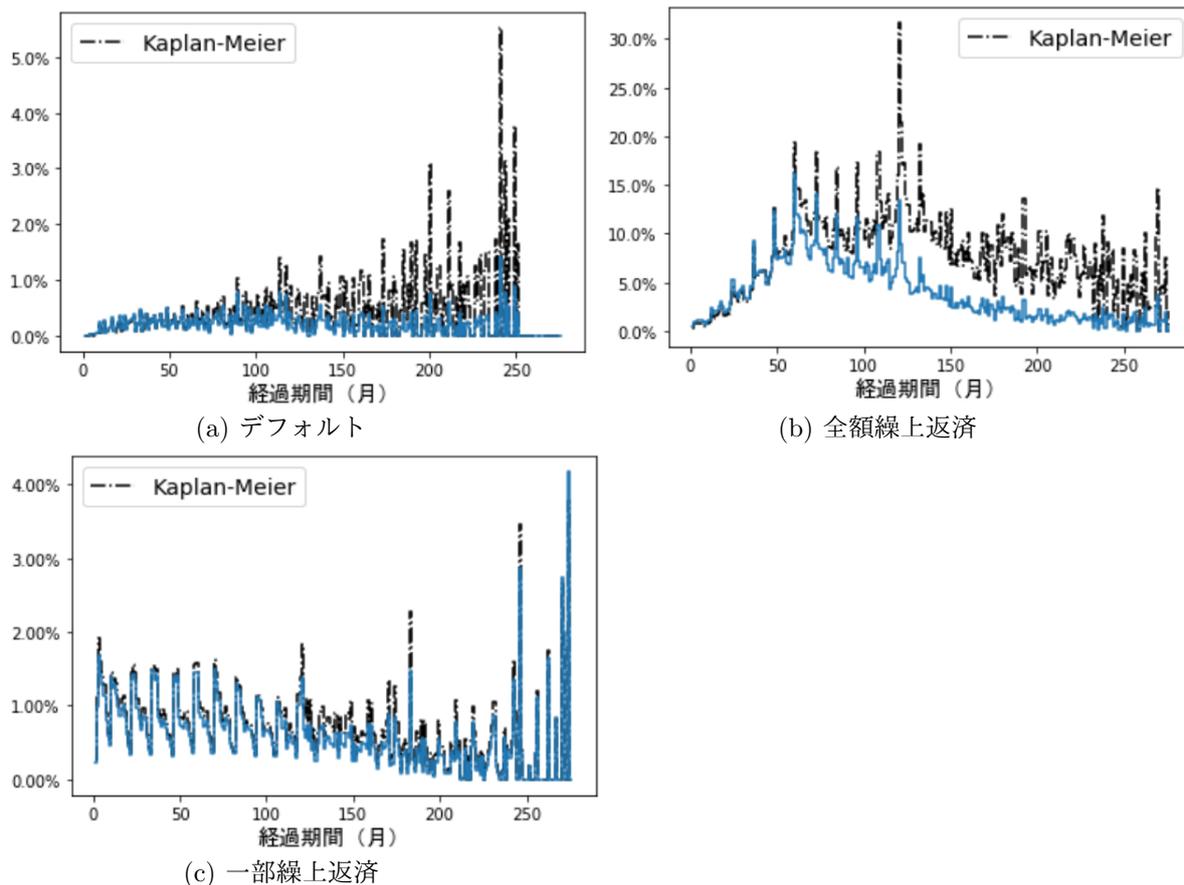


図 1: 拡張 Cox モデルにおけるベースラインハザード関数と実測平均値（年率換算値）

3.4 Cox-Snell 残差を用いた分析

3.4.1 分析方法

本節では、(2.17) で定義した Cox-Snell 残差 R_j^{CS} を用いたモデルの評価を行う。具体的には、Cox-Snell 残差 R_j^{CS} を横軸に、また (2.4) で導入した生存関数に対する Kaplan-Meier 推定量 $\hat{S}^{KM}(t)$ を用いて定義される

$$R_j^{KM} = -\log \hat{S}^{KM}(t_j) \quad (3.2)$$

を縦軸にした散布図 $\{(R_j^{CS}, R_j^{KM})\}_{j=0,1,\dots,J}$ を用いる。これは、仮に $\hat{S}_j(t_j) = \hat{S}^{KM}(t_j)$, すなわちイベント発生時間（または打ち切り時間） t_j に対する生存関数のモデルによる推定値 $\hat{S}_j(t_j)$ が Kaplan-Meier 推定量 $\hat{S}^{KM}(t_j)$ と一致する場合、散布図の点が 45 度線上に並び視覚的に当てはまりを確認しやすいためである¹⁷。

これに加えて、経過時間方向に関するデータへのフィッティングの違いを明確に示すため、A 節で定義される $(-\log \hat{S}^{KM})(\cdot)$ の逆関数 $(-\log \hat{S}^{KM})^{-1}(\cdot)$ を用いて、散布図 $\{(R_j^{CS}, R_j^{KM})\}_{j=0,1,\dots,J}$ に対して

$$\begin{cases} \hat{t}_j^{CS} &= (-\log \hat{S}^{KM})^{-1}(R_j^{CS}) \\ \hat{t}_j^{KM} &= (-\log \hat{S}^{KM})^{-1}(R_j^{KM}) = t_j \end{cases} \quad (3.3)$$

のように変形する。 $(\hat{t}_j^{CS}, \hat{t}_j^{KM}) = (\hat{t}_j^{CS}, t_j)$ は前述の散布図における各点 (R_j^{CS}, R_j^{KM}) の各要素を同じ単調関数で変換した値なので、 (R_j^{CS}, R_j^{KM}) の性質を維持しつつ経過時間方向に関する当てはまりを観察することが可能となる。本稿ではこれを、「時間変換された Cox-Snell 残差」と呼ぶ。なお、イベント発生までの最終時間 t_J について、ある $0 < k < J$ が存在して

$$\hat{S}^{KM}(t_J) > \hat{S}_j(t_j) \quad \forall j \geq k \quad (3.4)$$

となる、すなわち時点 t_k （あるいはそれ以降）までのイベント発生率（ハザード関数）を、モデルが実際の観測よりも高く見積もっている場合、(3.3) 式における逆関数での変換が全て t_J を返し散布図上で（異なる t_j に対して） \hat{t}_j^{CS} が一直線に並ぶことがあることも併せて指摘しておく。

¹⁷同様の分析は、例えば Collett [1] の 4 章でも見られる。

3.4.2 分析結果

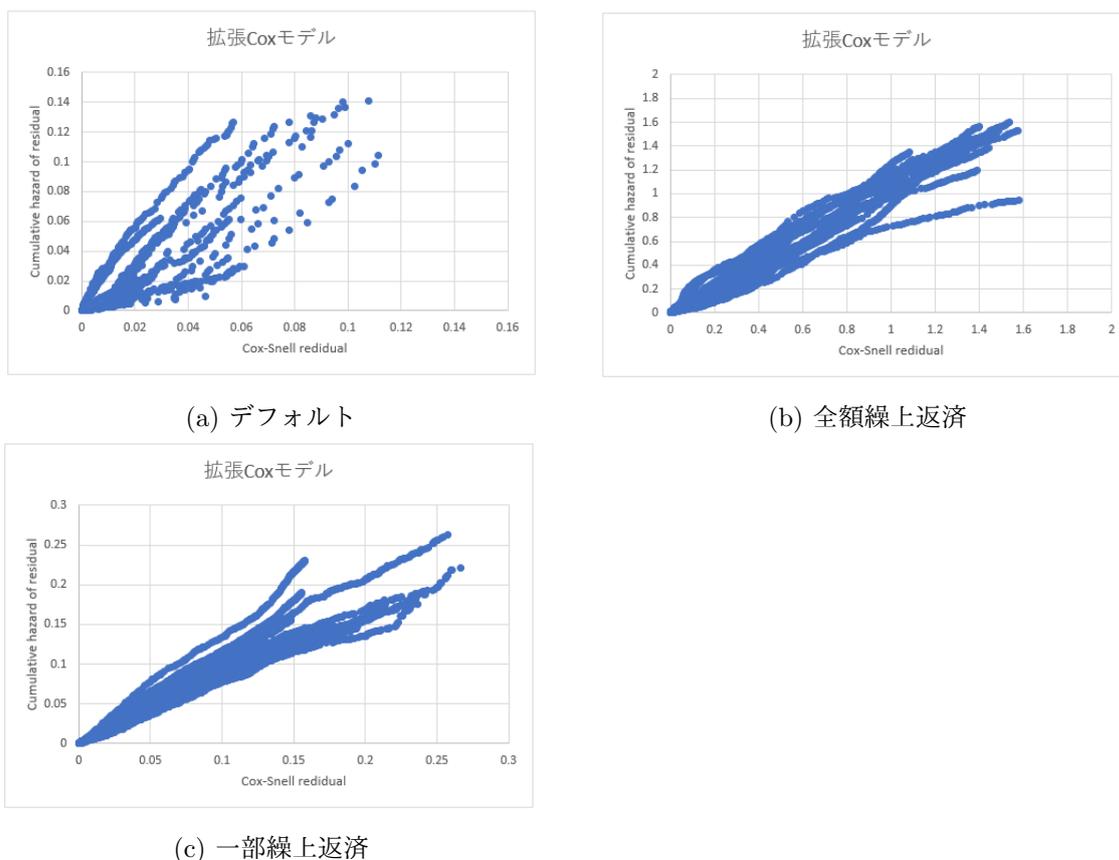
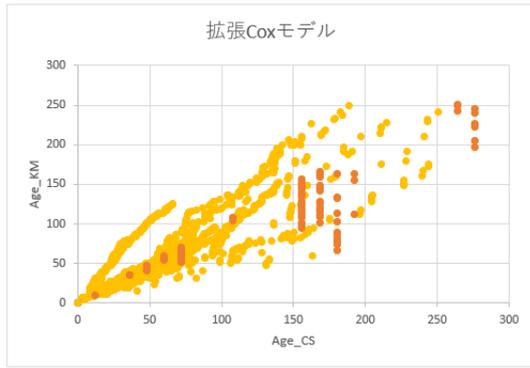


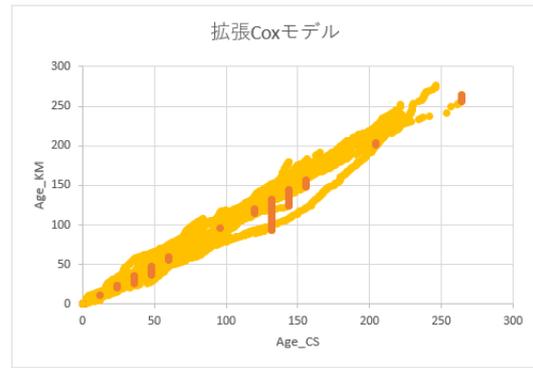
図 2: Cox-Snell 残差プロット (拡張 Cox モデル)

本節では散布図を用いて Cox-Snell 残差の視覚的検証を行う。

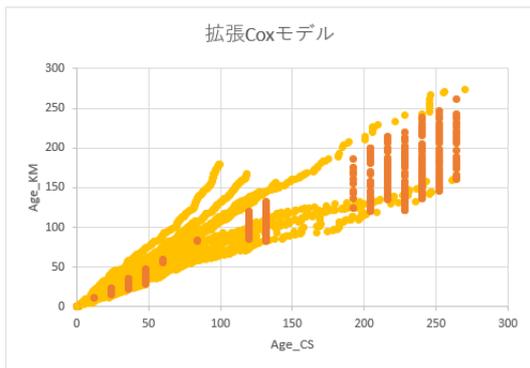
まず、経過時間への変換を行わない $\{(R_j^{CS}, R_j^{KM})\}_{j=0,1,\dots,J}$ について見る。図 2 の通り、全てのイベントにおいて、全体的に 45 度に近い角度は保っているものの、残差の高い領域において広く散在し、特にデフォルトではばらつきが広い。また、これも特にデフォルトにおいて顕著であるが、残差は無規則にばらついているわけではなく、いくつかの線上に分かれて並んでいるように見える。これは残差が過去のデータに依存して決まる累積値であるからである。これらの線上の残差が同じコーホートに属していることは、コーホート毎に推定モデルの当てはまり具合が異なることを示している。これに対応する分析として、次節以降では共変量にコーホートを表すダミー変数を加えた分析も行う。



(a) デフォルト



(b) 全額繰上返済



(c) 一部繰上返済

図 3: 時間変換された Cox-Snell 残差プロット (拡張 Cox モデル): 横軸に \hat{t}_j^{CS} , 縦軸に t_j を取りプロットしている.

$\{(R_j^{CS}, R_j^{KM})\}_{j=0,1,\dots,J}$ を経過時間に変換した散布図 (\hat{t}_j^{CS}, t_j) である図 3 についても, 上記と同様の傾向が見られる¹⁸. 特にデフォルトでは, 全線, 一線に比べると当初からばらつきが広く, 経過時間の長い部分では 45 度線上を中心にデータ点が幅広く散在している, これは, 生存時間を過大に見積もるにせよ過少に見積もるにせよ, モデルの当てはまりが良くないことを示唆している. 一方, 全線では 45 度線に近いコーホートが比較的多いように見える. 一線では, デフォルトほどではないが経過時間の比較的短いところから様々な傾きの直線が伸びて次第に広がっている. ここでも線上に並ぶ点は同じコーホートに属しており, 前述のコーホートダミー変数の必要性が示唆される.

3.4.3 評価

3.4.2 節の Cox-Snell 残差に関する分析により, 拡張 Cox モデルには改善の余地があることが分かる. これまで示した拡張 Cox モデルでは, イベントに対する共変量の影響度は経過期間を通じて一

¹⁸ グラフの中で垂直方向に点が並んでいる個所は $\hat{S}^{CS}(t) \leq \min_t \hat{S}^{KM}(t)$ となる点である. この場合, $\hat{R}^{CS}(t)$ に対する逆関数値 $(-\log \hat{S}^{KM})^{-1}(\hat{R}^{CS}(t))$ をイベント発生時刻の最大値あるいは観測打ち切り時間でしか返せないで, $\hat{R}^{CS}(t)$ の値に依らず同じ値になり, フラットになる. 3.4.1 節の最後の部分も参照されたい.

定であるという仮定を置いている。しかし、実際には同じ経済環境でも経過期間によってイベントの起こりやすさへの影響は異なると考えられる¹⁹。そこで、拡張 Cox モデルの仮定を緩めた、経過期間に応じて共変量の影響度が変化する区分 Cox モデルを導入し、拡張 Cox モデルとの比較を行う。なお、区分 Cox モデルの分析に使用する共変量は 3.2 節で示した拡張 Cox モデルの共変量と同じとする。これは、拡張 Cox モデルから区分 Cox モデルに変更することによる分析結果の改善効果を明らかにするためである。実際に最適な区分 Cox モデルを検討する際には、すべての共変量を使って分析をやり直すことを薦める。

区分 Cox モデルで共変量の感応度を全ての期間毎（本研究では 1ヶ月）に変化させることも可能であるが、モデルが複雑になり過ぎるため実用的ではない。そこで本研究では、経過期間を 3 区間に分割した区分 Cox モデルを検討した。3 区間に分ける際の分割点については、 $t = 1, 2, \dots, 15$ 年からの 2 組の組み合わせ（ ${}_{15}C_2 = 105$ 組）の全てでモデル推定を行い、AIC が最も低くなる分割点を採用する。

また、3.4.2 節で示したように、各イベントで拡張 Cox モデルの Cox-Snell 残差はコーホート毎に異なる傾向が見られる。これは次節以降で示す区分 Cox モデルでの Cox-Snell 残差でも同様である。そこで、本研究では共変量をマクロ経済変数とする区分 Cox モデルに更にコーホートダミー変数を加えた分析も行う。なお、コーホートダミー変数は 1997 年に組成された住宅ローンを基準とし、コーホートダミー変数は、マクロ経済変数とイベントに対する影響の度合いを比較できるように標準化して分析した結果と、イベントに対する影響の度合いを各コーホート間だけで比較するために標準化せず分析した結果の両方について示す。

さらに本研究では、コーホートダミー変数を加えた分析において、区間分割はダミー変数を使用しない区分 Cox モデルのときに揃えることにする。その理由は、コーホートダミー変数を加えることによる改善効果が、ダミー変数のないモデルと比較することで明確になるようにするためである。また、試みにダミー変数を加えた分析で分割点を変えながら推定を行ったところ、「デフォルト」において AIC が最も小さくなる分割点は $(\hat{t}_1, \hat{t}_2) = (1, 4)$ となったが、この第 1 区間ではデフォルト件数がごく僅かであり、パラメータ推定が正常に機能しなかったことも理由の一つである。

¹⁹例えば、借換は住宅ローンの残存期間が長いほど総返済額を減らすメリットを享受できるため、同じ金利差でも短い経過期間でより全線が起りやすいと考えられる。一方、こうした借換やローンの返済により相対的に信用度の高いローンがコーホートから離脱すること、あるいは逆に相対的に信用度の低いローンがデフォルトを起こすことでコーホートから離脱することによって、時間の経過とともにコーホート全体の信用度が変化する可能性なども考えられる。

4 デフォルト考察

4.1 拡張 Cox モデルと区分 Cox モデルの比較

本節では、拡張 Cox モデルと区分 Cox モデルを比較する。まず前述の通り、区分 Cox モデルの分割点について、AIC が最も低くなるのは $(\hat{t}_1, \hat{t}_2) = (1, 4)$ 年のときであった。しかし、 $t_1 \leq 1$ 年の第 1 区間ではデフォルト件数が少なく、後述で比較対象とするコーホートダミー変数を加えた区分 Cox モデルでは全てのコーホートダミーに対して有効な推定値が得られなかったことから、 $t_1 > 1$ 年で最も AIC が低くなる分割点 $(\hat{t}_1, \hat{t}_2) = (4, 11)$ 年を採用した。

表 2: 区分 Cox モデルによる分析結果 (デフォルト)。拡張 Cox モデルについても比較のため再掲した。

モデル名	Wald 検定統計量 (p 値) AIC	共変量		感応度	標準誤差	p 値
		(L : ラグの大きさ)				
拡張 Cox モデル	5,015.0(0.0) 1,296,254	失業率 (L12)	β	0.12157	0.00502	0.0
		金利差 (L3)	β	0.36808	0.00744	0.0
		鉱工業 生産指数 (L0)	β	-0.05852	0.00512	0.0
区分 Cox モデル (ダミーなし) 変更点 (4, 11) 年	7,522.0(0.0) 1,293,851	失業率 (L12)	β_1	0.25468	0.00846	0.0
			β_2	-0.05826	0.00716	0.0
			β_3	0.37821	0.01878	0.0
		金利差 (L3)	β_1	0.45309	0.01289	0.0
			β_2	0.31045	0.01002	0.0
			β_3	0.55038	0.03125	0.0
		鉱工業 生産指数 (L0)	β_1	-0.14440	0.00634	0.0
			β_2	0.11673	0.00901	0.0
			β_3	-0.13559	0.03141	0.0

表 2 の推定結果によると、拡張 Cox モデルより区分 Cox モデルは AIC が低くなった。各共変量の感応度を比較すると、区分 Cox モデルでは、「金利差」に対する感応度の符号はどの区間でも拡張 Cox モデルと同じく正だが、「失業率」と「鉱工業生産指数」に対する第 2 区間 (4-11 年) での感応度の符

号は拡張 Cox モデルと異なる。さらに、どの共変量の影響力も第 2 区間が他の区間に比べ低い。また、区分 Cox モデルの共変量間の比較では、どの区間でも、金利差は他の共変量より影響力が高い。

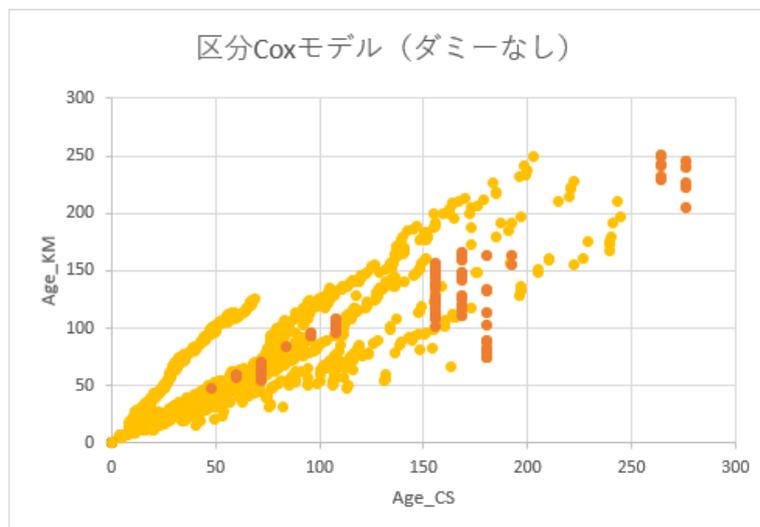


図 4: 時間変換された Cox-Snell 残差 (デフォルト, 区分 Cox モデル (ダミーなし))

図 4 に、区分 Cox モデルによる Cox-Snell 残差を示す。図 3 に比べると若干残差は 45 度線上に近付いているが、まだ両側に広く分布しており、更なる改善の余地があると考えられる。

4.2 コーホートダミー変数を加えた比較

本節では、4.1 節の区分 Cox モデルの共変量にコーホートダミー変数を加えて推定結果を考察する。なお、ダミー変数を含めて再度モデルの変数選択を行えば、最適な共変量の組み合わせが拡張 Cox モデルの場合と異なる可能性があるが、ここではダミー変数を加えることによる分析結果への影響に焦点を当てているため、共変量に使用するマクロ経済変数はタイムラグを含め同じとする。コーホートダミー変数を加えた区分 Cox モデルでは、第 2 区間や第 3 区間に入る経過期間で比較的新しいコーホートの観測データがないため、それら区間で感応度の推定ができないダミー変数があることに注意されたい²⁰。

²⁰例えば、c2009 の観測期間は 2009 年 4 月から 2020 年 3 月までの 11 年間 (132ヶ月) であり、 $t_3 > 11$ 年の第 3 区間ではパラメータの推定ができない。

表 3: コーホートダミー変数付区分 Cox モデルによる分析結果 (デフォルト). ダミー変数のない区分 Cox モデルについても比較のため再掲した.

モデル名	Wald 検定統計量 (p 値) AIC	共変量 (L : ラグの大きさ)		感応度	標準誤差	p 値
区分 Cox モデル (ダミーなし) 変更点 (4, 11) 年	7,522.0(0.0) 1,293,851	失業率 (L12)	β_1	0.25468	0.00846	0.0
			β_2	-0.05826	0.00716	0.0
			β_3	0.37821	0.01878	0.0
		金利差 (L3)	β_1	0.45309	0.01289	0.0
			β_2	0.31045	0.01002	0.0
			β_3	0.55038	0.03125	0.0
		鉱工業 生産指数 (L0)	β_1	-0.14440	0.00634	0.0
			β_2	0.11673	0.00901	0.0
			β_3	-0.13559	0.03141	0.0
区分 Cox モデル (ダミーあり) 変更点 (4, 11) 年	25,095.1(0.0) 1,278,142	失業率 (L12)	β_1	0.16276	0.02030	0.0
			β_2	0.11960	0.02097	0.0
			β_3	0.12435	0.05528	0.0245
		金利差 (L3)	β_1	0.28096	0.03097	0.0
			β_2	-0.07696	0.01846	0.0
			β_3	-0.26781	0.09482	0.0005
		鉱工業 生産指数 (L0)	β_1	-0.09866	0.00895	0.0078
			β_2	-0.07765	0.00827	0.0
			β_3	-0.10490	0.03324	0.0016

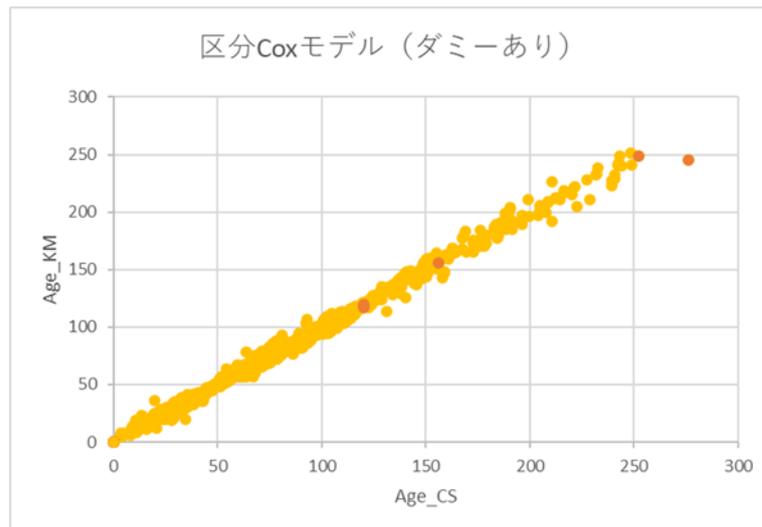
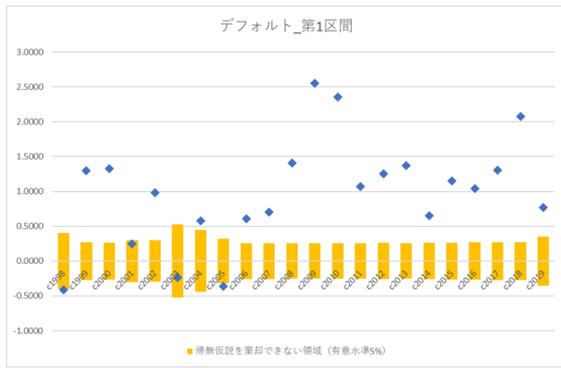


図 5: 時間変換された Cox-Snell 残差 (デフォルト, 区分 Cox モデル (ダミーあり))

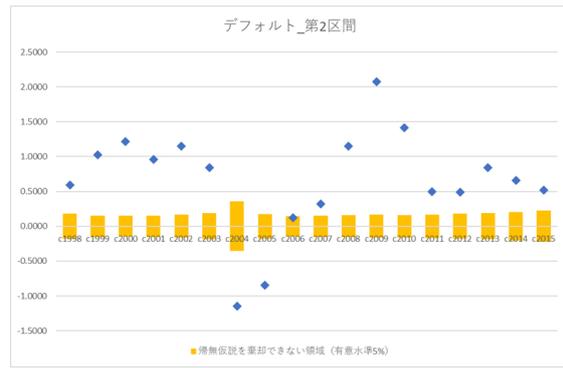
まず, 図 5 にコーホートダミー変数を加えた区分 Cox モデルによる時間変換された Cox-Snell 残差を示す.

コーホートダミー変数を加えることで図 3 や図 4 と比較して残差が 45 度線上に集まっており, モデルが大幅に改善した.

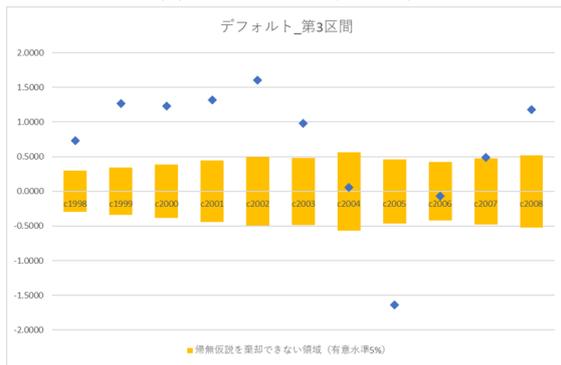
共変量にコーホートダミー変数を加えることで AIC はさらに低下し, 表 3 で, マクロ経済変数に対する感応度を見ると, 共変量にコーホートダミー変数を加えない区分 Cox モデルの感応度と符号が変わる区間が見られる. 特に, 「金利差」に対する感応度の変化幅は, 第 2 区間と第 3 区間 (4 年超) で大きい. 「失業率」と「鉱工業生産指数」に対する感応度の符号はどの区間でも同じとなったが, どの経過期間でも失業率が上昇すると, または鉱工業生産指数が低下するとデフォルトが起りやすくなることを示唆し, 直観にも整合的である. また, マクロ経済変数とコーホートダミー変数の感応度を比較すると, 特に 2008 年以降のコーホートダミー変数はマクロ経済変数より感応度が大きく, コーホート内の債務者の質自体がデフォルトに強く影響している. コーホートの質を表す代理変数を選定できれば, 将来のコーホートに対するイベント発生率の予測が可能で, より適したモデルを構築できるかもしれない.



(a) 時間区間 1 (0-4 年)



(b) 時間区間 2 (4 年超-11 年)



(c) 時間区間 3 (11 年超)

図 6: 標準化を行わない場合における各コーホートダミー変数に対する感応度（デフォルト，区分 Cox モデル（ダミーあり））

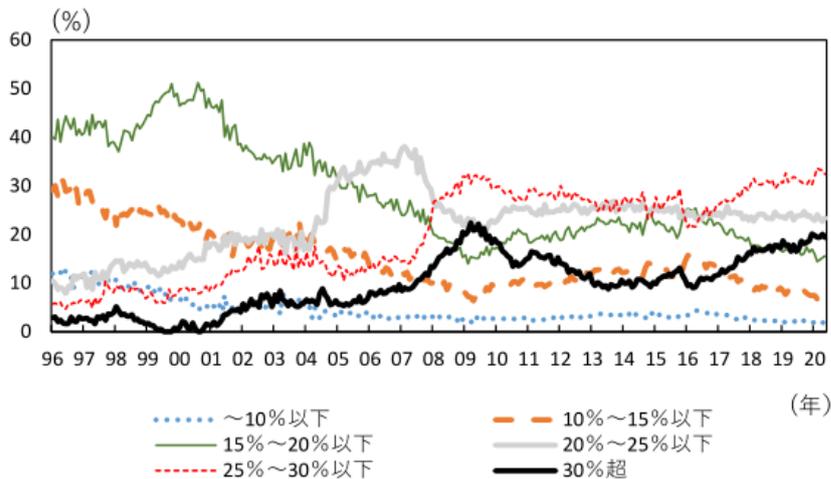


図 7: 住宅金融支援機構による住宅ローンの当初返済負担率（住宅ローンの年間返済額を申し込み時の年収で除したもの）の推移。藤原 [14] より引用。

次に，コーホートのデフォルトに対する影響力を比較するために，コーホートダミー変数を標準化

せずに感応度を推定した結果を図6にグラフで示す。なお図中の青い点は各コーホートダミー変数に対する感応度の値を表し、黄色の領域は各感応度が0であるという帰無仮説を棄却できない領域を示している。青い点が黄色の領域の外にある場合、 $\beta = 0$ という帰無仮説が有意水準5%で棄却されることを意味する（以降の節でも同様）。

特に、第1区間及び第2区間のグラフを見ると、2009-2010年のコーホートの感応度が比較的に高く、2009-2010年頃に貸付実行された住宅ローンはデフォルトが起きやすいことを示唆している。図7で示される通り、今回分析対象とした住宅金融支援機構の住宅ローンのうち、2009年頃に貸付実行されたものは高い返済比率の割合が多く、このことがデフォルトが起きやすい要因であると考えられる。この分析結果からも、多くの金融機関の住宅ローン審査で返済比率が重要視されていることは理に適っている。

なお、拡張Coxモデルの共変量にコーホートダミー変数を加えた推定結果を表4に示す。表10によると、AICは拡張Coxモデルに比べて低下しており、統計学的にはより尤もらしいモデルが得られたことを示唆している。共変量にダミー変数を加えることで「金利差」に対する感応度の符号が負となり、「失業率」が最も影響力のある変数となった。「金利差」に対する感応度の変化の要因については今後調査する必要があるが、「失業率」がデフォルトに対して安定して相応の説明力を有するということは納得できる結果である。

表4: コーホートダミー変数付拡張Coxモデルによる分析結果（デフォルト）。ダミー変数のない拡張Coxモデルについても比較のため再掲した。

モデル名	Wald 検定統計量 (p 値)	共変量		感応度	標準誤差	p 値
	AIC	(L: ラグの大きさ)				
拡張 Cox モデル (ダミーなし)	5,015.0(0.0) 1,296,254	失業率 (L12)	β	0.12157	0.00502	0.0
		金利差 (L3)	β	0.36808	0.00744	0.0
		鉱工業 生産指数 (L0)	β	-0.05852	0.00512	0.0
拡張 Cox モデル (ダミーあり)	23,724.3(0.0) 1,279,159	失業率 (L12)	β	0.22984	0.01013	0.0
		金利差 (L3)	β	-0.12497	0.01160	0.0
		鉱工業 生産指数 (L0)	β	-0.11000	0.00578	0.0

最後に、図8, 9各モデルにおけるベースラインハザード関数を示す。拡張Coxモデルと区分Coxモデルで比較すると、区分Coxモデルの第1, 第2区間内である経過期間11年までは概ね同水準だが、経過期間11年以降は拡張Coxモデルが区分Coxモデルの1.5~2倍程度高い。また、それぞれ

のモデルにダミー変数を加えた場合のベースラインハザード関数は経過期間の後半で高い値をとり、元のモデルとの乖離が大きい。ダミー変数を加えた方がベースラインハザード関数が高くなるのは次節以降で示す全線、一線の場合とは異なる結果であるが、その理由はまだ説明できていない。

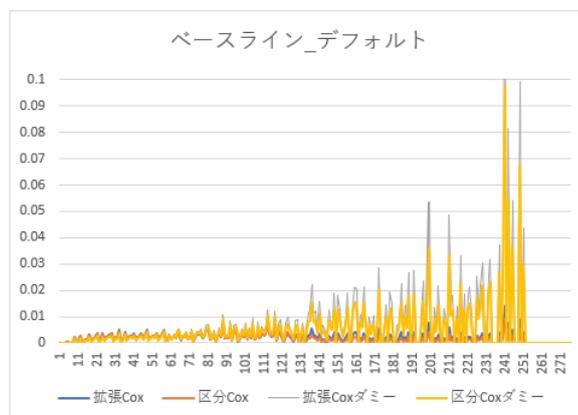


図 8: 各モデルにおけるベースラインハザード関数の比較 (デフォルト)

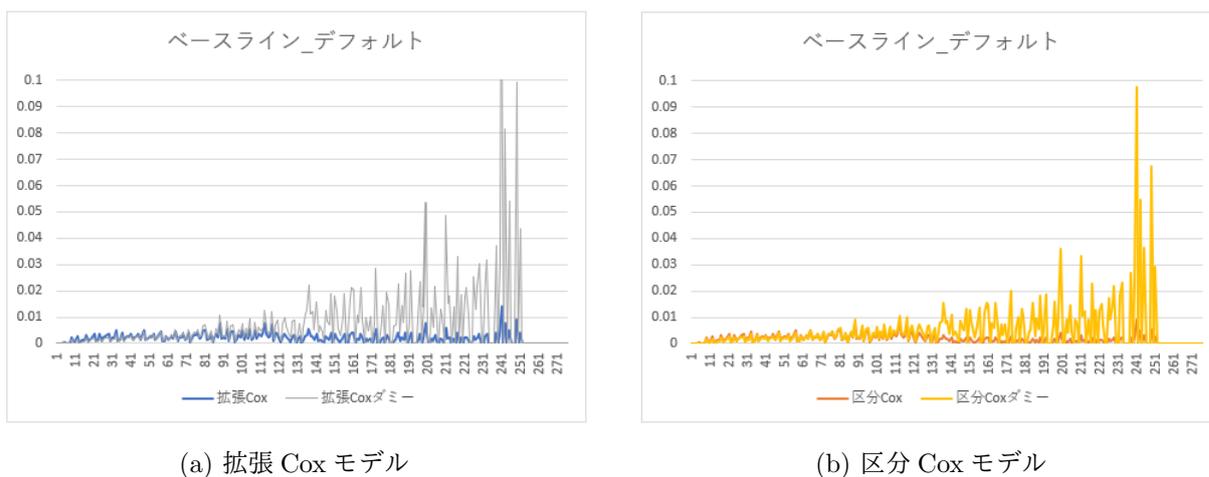


図 9: コーホートダミー変数の有無によるベースラインハザード関数の比較 (デフォルト)

5 全額繰上返済考察

5.1 拡張 Cox モデルと区分 Cox モデルの比較

本節では拡張 Cox モデルと区分 Cox モデルの比較を行う。区分 Cox モデルについては、分割点が $(\hat{t}_1, \hat{t}_2) = (4, 6)$ 年のとき AIC が最も低くなる。後述する通り、第 1 区間と第 2 区間では「金利差」の影響が大きく異なっており、このことが今回の分割点を選択する理由となった可能性がある。以降では、各変数の感応度について、モデル間での比較を行う。

表 5: 区分 Cox モデルによる分析結果（全線）。拡張 Cox モデルについても比較のため再掲した。

モデル名	Wald 検定統計量 (p 値) AIC	共変量 (L : ラグの大きさ)		感応度	標準誤差	p 値
拡張 Cox モデル	75,330.3(0.0) 31,618,306	失業率 (L0)	β	-0.06919	0.00107	0.0
		金利差 (L2)	β	0.36163	0.00154	0.0
		鉱工業 生産指数 (L6)	β	-0.11292	0.00117	0.0
区分 Cox モデル (ダミーなし) 変更点 (4,6) 年	127,320.5(0.0) 31,570,608	失業率 (L0)	β_1	-0.33069	0.00200	0.0
			β_2	-0.16131	0.00200	0.0
			β_3	0.11652	0.00179	0.0
		金利差 (L2)	β_1	0.90183	0.00318	0.0
			β_2	0.34434	0.00325	0.0
			β_3	0.21357	0.00218	0.0
		鉱工業 生産指数 (L6)	β_1	-0.13330	0.00171	0.0
			β_2	-0.17216	0.00251	0.0
			β_3	-0.09502	0.00205	0.0

まずデフォルト同様、表5の推定結果によると、拡張 Cox モデルより区分 Cox モデルにおいて AIC が低くなるのがわかる。

次に、拡張 Cox モデルにおいて最も影響の強かった「金利差」については、区分 Cox モデルでは、その影響の強さは経過期間に対して一定ではないことがわかる。具体的には第1区間が全変数・区間中で最も影響が強く、そこから経過期間が長くなるにつれて影響が逓減する様子が見られる（特に6年程度までにかかなりの低下が見られる）。これは所謂バーンアウト効果に相当するものと考えられ、従来の説と整合的である。

また「失業率」については、拡張 Cox モデルにおいてはそれほど影響力がなかったが、区分 Cox モデルにすることで第1区間（0-4年）の影響力が強くなる。ただ総じて負の値を取っている点、すなわち失業率が下がると全線が起きやすくなる、という傾向については変わらず、これも経済学的な理解とは矛盾しないと言える。

一方「鉱工業生産指数」については、区間により多少のばらつきはあるものの、拡張 Cox モデルと

区分 Cox モデルの双方で、それほど強い影響力は見られない。

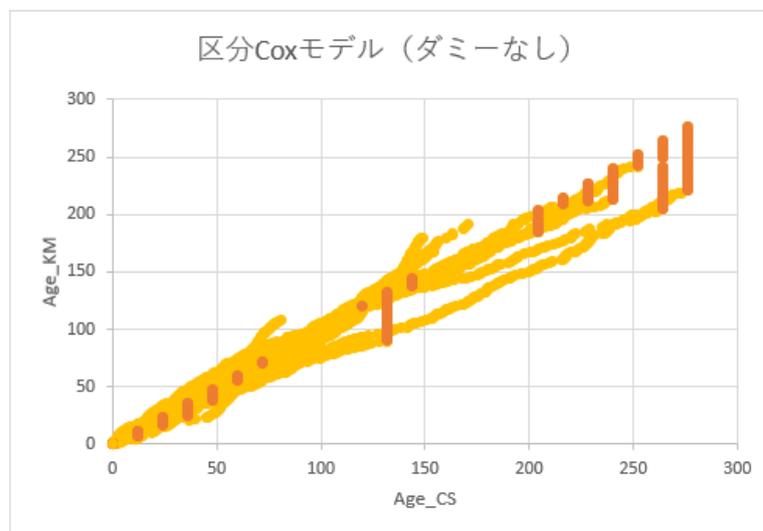


図 10: 時間変換された Cox-Snell 残差 (全線, 区分 Cox モデル (ダミーなし))

図 10 に, 区分 Cox モデルによる Cox-Snell 残差を示す. 図 3 と比較してそれほど大きな違いは見られず, こちらも更なる改善の余地があると考えられる.

5.2 コーホートダミー変数を加えた比較

本節では, 区分 Cox モデルにコーホートダミー変数を加えて推定結果を考察する.

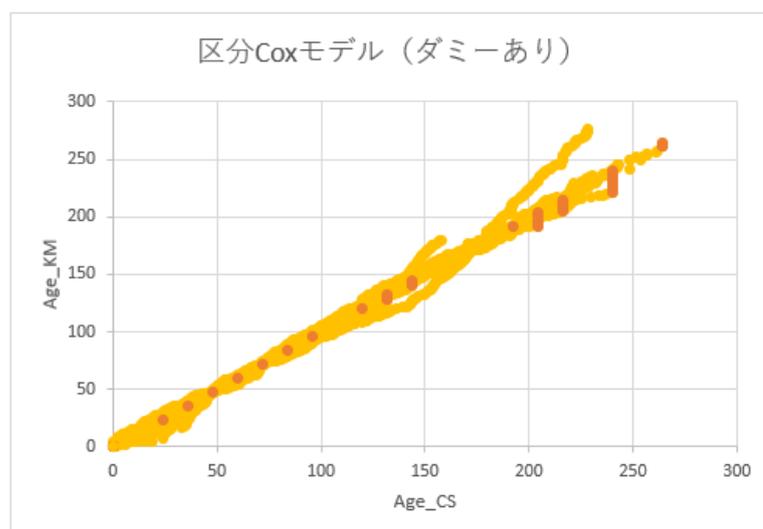


図 11: 時間変換された Cox-Snell 残差 (全線, 区分 Cox モデル (ダミーあり))

ダミー変数の有無のよる Cox-Snell 残差の違いについては、デフォルトと一線同様にダミー変数を加えたことによる改善が著しく、コーホートの違いを考慮することの重要性を示唆する結果といえる(図 11)。

まず、全線においてもデフォルトと同様、コーホートダミー変数を加えることで AIC の低下を観察できる(表 6)。

次にダミー以外の 3 変数の感応度について比較を行う。拡張 Cox モデルと区分 Cox モデルともに最も影響力の強かった「金利差」の影響はダミー変数を加えても変わらず、むしろ全区間でより強くなっていることがわかる。特に第 2 区間での感応度の変化は顕著であり、前節で指摘したバーンアウト効果が、ダミー変数のない区分 Cox モデルよりも遅く発現していることが示唆される。また「失業率」については最も影響力の強い区間が、第 1 区間から第 2 区間へと変わっており、ダミー変数を考慮すると期間構造の形状が変化することが示唆されるが、感応度の絶対値の大きさや符号等には大きな変化はない。最後に「鉱工業生産指数」についても、第 1 と第 2 区間での感応度がより小さくはなっているものの、総じて影響力が低いという結論は変わっていない。

表 6: コーホートダミー変数付区分 Cox モデルによる分析結果 (全線). ダミー変数のない区分 Cox モデルについても比較のため再掲した.

モデル名	Wald 検定統計量 (p 値) AIC	共変量 (L: ラグの大きさ)		感応度	標準誤差	p 値
区分 Cox モデル (ダミーなし) 変更点 (4,6) 年	127,320.5(0.0) 31,570,608	失業率 (L0)	β_1	-0.33069	0.00200	0.0
			β_2	-0.16131	0.00200	0.0
			β_3	0.11652	0.00179	0.0
		金利差 (L2)	β_1	0.90183	0.00318	0.0
			β_2	0.34434	0.00325	0.0
			β_3	0.21357	0.00218	0.0
		鉱工業 生産指数 (L6)	β_1	-0.13330	0.00171	0.0
			β_2	-0.17216	0.00251	0.0
			β_3	-0.09502	0.00205	0.0
区分 Cox モデル (ダミーあり) 変更点 (4,6) 年	248,445.1(0.0) 31,439,722	失業率 (L0)	β_1	-0.19491	0.00734	0.0
			β_2	-0.30450	0.01342	0.0
			β_3	-0.08438	0.00638	0.0
		金利差 (L2)	β_1	1.20919	0.00596	0.0
			β_2	0.98669	0.00721	0.0
			β_3	0.31487	0.00382	0.0
		鉱工業 生産指数 (L6)	β_1	-0.00693	0.00261	0.0078
			β_2	-0.00969	0.00362	0.0
			β_3	-0.10226	0.00286	0.0

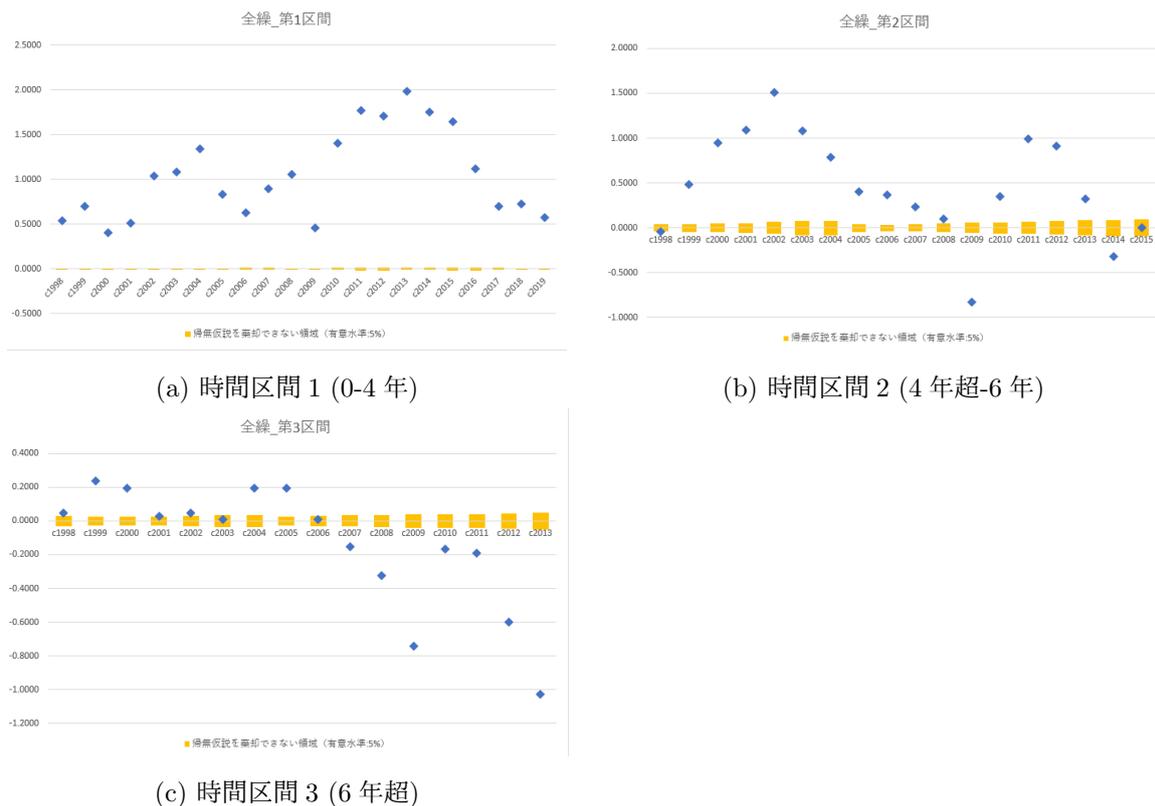


図 12: 標準化を行わない場合における各コーホートダミー変数に対する感応度（全線，区分 Cox モデル（ダミーあり））

次に，コーホートダミー変数について標準化を行わずに推定した結果を用いて，コーホートダミー変数に対する感応度を比較し，どの貸付実行年が全線の発生に影響を与えているのか考察を行う（図 12）．第 1 区間と第 2 区間については 2002-2004 年頃，及び 2011-2012 年頃に全線が行われる傾向の強いコーホートがあり，2009 年については大きな落ち込みが見られる．これは時期から見て金融危機の影響と考えられるが，この時期に貸付実行された住宅ローンは全線が起きにくいということの解釈は難しい．また第 3 区間については 2009 年に加えて 2012，2013 年の感応度も大きな負の値となっている．ただし，この 2 年については第 3 区間（6 年超）としてカウントされるデータ数がそれぞれ 2 年分，1 年分しかなく，このため 2019-2020 年頃の市場環境の影響を強く受けている可能性がある．この区間については，その他のコーホートの影響は限定的なものとなっている．

次に，拡張 Cox モデルにおいてダミー変数を加えた場合について，推定結果を表 7 に示す．まずデフォルト同様，拡張 Cox モデルにおいても AIC は低下しており，モデルとして適切であると示唆される．また各共変量に対する感応度に注目すると，まず「失業率」については，符号は変わる²¹ものの，絶対値はダミー変数の有無に関わらず小さく，影響度が小さいという結論は変わらない．一方，

²¹正確には，ダミー変数を加えた場合は 5%水準で有意ではない．

「金利差」についてはダミー変数を加えることで感応度が大きくなっており、これは区分 Cox モデルにおけるダミー変数の有無の比較と整合的である。また「鉱工業生産指数」については、符号は負のまま、絶対値は（元々それほど大きくはないが）半分になっている。すなわち全線については、拡張 Cox モデルにダミー変数を加えることで「金利差」の影響だけが強まり、一方で他の 2 つの共変量の影響は相対的にかなり小さいものとなることが分かった。

表 7: コーホートダミー変数付拡張 Cox モデルによる分析結果（全線）。ダミー変数のない拡張 Cox モデルについても比較のため再掲した。

モデル名	Wald 検定統計量 (p 値) AIC	共変量 (L: ラグの大きさ)		感応度	標準誤差	p 値
			β			
拡張 Cox モデル (ダミーなし)	75,330.2(0.0) 31,618,306	失業率 (L0)	β	-0.06919	0.00107	0.0
		金利差 (L2)	β	0.36163	0.00154	0.0
		鉱工業 生産指数 (L6)	β	-0.11292	0.00117	0.0
拡張 Cox モデル (ダミーあり)	123,324.4(0.0) 31,558,891	失業率 (L0)	β	0.00627	0.00321	0.0508
		金利差 (L2)	β	0.61490	0.00262	0.0
		鉱工業 生産指数 (L6)	β	-0.06391	0.00150	0.0

最後に、図 13, 14 において、各モデルにおけるベースラインハザード関数についても確認すると、モデルによらず基本的な形状は同じである。すなわち、初期はベースラインハザード関数の値が小さく、そこから経過時間が増えるにつれて関数の値も増加し、60 か月あたりをピークに緩やかに低下しつつ、120 か月で一度大きく上昇している。金利差、及びダミー変数の影響を取り除いても、こうした「バーンアウト効果」に近い形状が残ることは、特筆すべきと言えよう。

さらに、モデルごとの違いを詳細に見てみると、60 か月以降はダミー変数を加えたモデルの方がベースラインハザード関数の値が小さくなっていることがわかる。これは、ダミー変数を加えることにより全体のハザードのうち共変量によって説明される部分が大きくなることを示しており、ダミー変数を加えたことの妥当性を示唆している可能性がある。

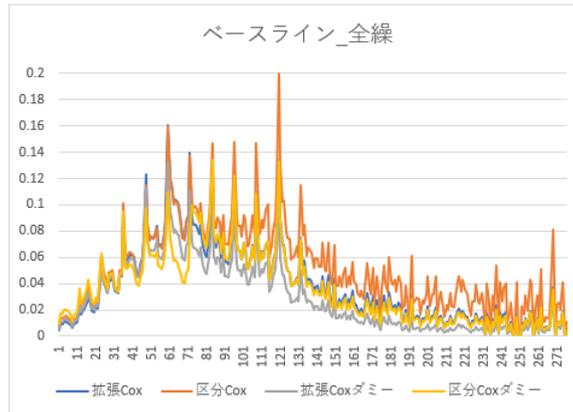
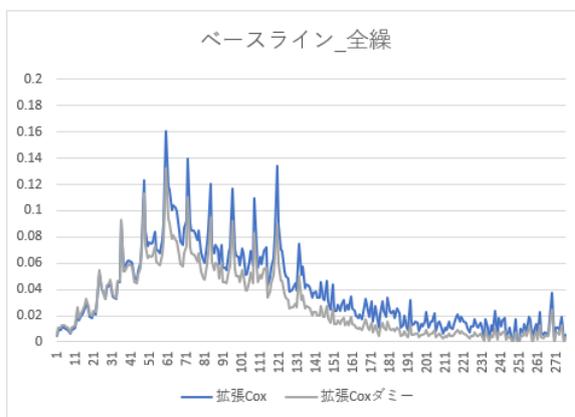
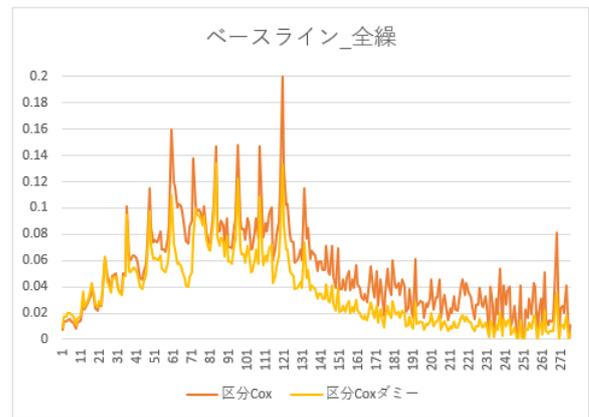


図 13: 各モデルにおけるベースラインハザード関数の比較（全線）



(a) 拡張 Cox モデル



(b) 区分 Cox モデル

図 14: コーホートダミー変数の有無によるベースラインハザード関数の比較（全線）

6 一部繰上返済考察

6.1 区分 Cox モデルによる推定結果と拡張 Cox モデルとの比較

本節では拡張 Cox モデルと区分 Cox モデルの比較を行う。

区分 Cox モデルについては、分割点が $(\hat{t}_1, \hat{t}_2) = (4, 10)$ 年のとき AIC が最も低くなる。表 8 によると、AIC は拡張 Cox モデルに比べて若干低下しており、統計学的にはより尤もらしいモデルが得られたことを示唆している。3 変数とも感応度はどの区間でも有意であり、第 1 区間（0-4 年）における感応度が正で最も絶対値が高く、第 2 区間（4-10 年）では正であるが絶対値は低下し、第 3 区間（10 年超）では負になった。このことは、マクロ経済変数に対する感応度の高い債務者層は早いうちに繰上返済を進めるため、時間とともに比較的感応度の低い債務者層の比率が高まるという債務者層の変

質を示唆している。また、第1区間と第2区間では「完全失業率」の感応度が3変数の中で一番高く、影響力が大きい。これは拡張Coxモデルの結果とも整合的である。このように、一線の感応度を経過時間依存させることで、拡張Coxモデルの結果と整合的で、かつ債務者層の経年的な変質を示唆する結果が得られた。ただし、第3区間でどの変数の感応度も有意に負になることは経済学的に説明できていない。

図15に、区分Coxモデルによる時間変換されたCox-Snell残差を示す。図3に比べると残差プロットは45度線に近付いているが、まだ両側に広く分布しているため、更なる改善の余地があると考えられる。

表8: 区分Coxモデルによる分析結果（一線）。拡張Coxモデルについても比較のため再掲した。

モデル名	Wald 検定統計量 (p 値) AIC	共変量 (L: ラグの大きさ)		感応度	標準誤差	p 値
拡張 Cox モデル	43,159.1(0.0) 5,546,755	失業率 (L2)	β	0.42998	0.00253	0.0
		金利差 (L12)	β	0.14796	0.00382	0.0
		鉱工業 生産指数 (L9)	β	0.18350	0.00198	0.0
区分 Cox モデル (ダミーなし) 変更点 (4, 10) 年	44,935.9(0.0) 5,544,190	失業率 (L2)	β_1	0.49418	0.00332	0.0
			β_2	0.34862	0.00437	0.0
			β_3	-0.06266	0.01576	0.0001
		金利差 (L12)	β_1	0.12514	0.00510	0.0
			β_2	0.09311	0.00667	0.0
			β_3	-0.30265	0.02296	0.0
		鉱工業 生産指数 (L9)	β_1	0.21155	0.00221	0.0
			β_2	0.06640	0.00500	0.0
			β_3	-0.10753	0.01737	0.0

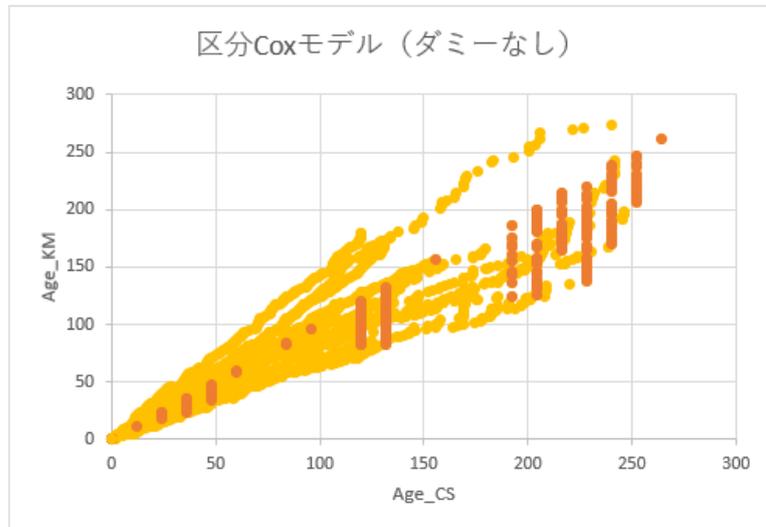


図 15: 時間変換された Cox-Snell 残差 (一線, 区分 Cox モデル (ダミーなし))

6.2 コーホートダミー変数を加えたモデル

まず, 図 16 にコーホートダミー変数を加えた区分 Cox モデルによる時間変換された Cox-Snell 残差を示す. 図 3 や図 15 に比べると残差プロットが顕著に 45 度線に近付いていることから, コーホートダミー変数の追加によりモデルの説明力が格段に向上したことがわかる. 以下ではそのモデルの内容を詳細に説明する.

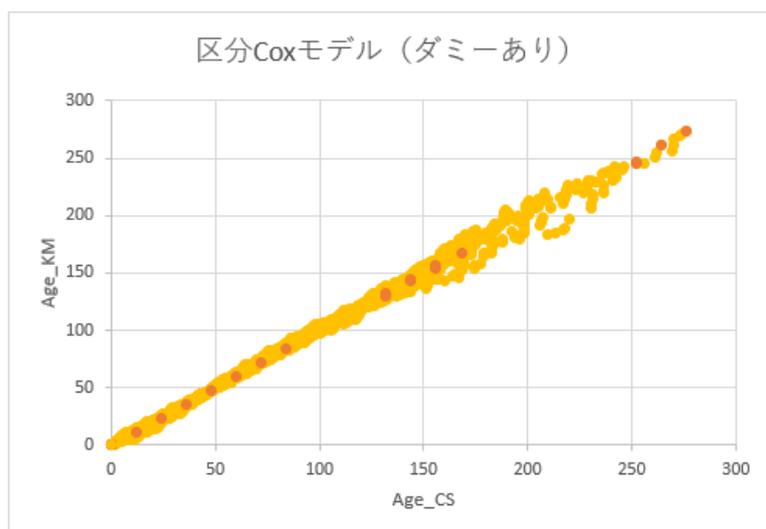


図 16: 時間変換された Cox-Snell 残差 (一線, 区分 Cox モデル (ダミーあり))

表 9: コーホートダミー変数付区分 Cox モデルによる分析結果（一線）. ダミー変数のない区分 Cox モデルについても比較のため再掲した.

モデル名	Wald 検定統計量 (p 値) AIC	共変量		感応度	標準誤差	p 値
		(L : ラグの大きさ)				
区分 Cox モデル (ダミーなし) 変更点 (4, 10) 年	44,935.9(0.0) 5,544,190	失業率 (L2)	β_1	0.49418	0.00332	0.0
			β_2	0.34862	0.00437	0.0
			β_3	-0.06266	0.01576	0.0001
		金利差 (L12)	β_1	0.12514	0.00510	0.0
			β_2	0.09311	0.00667	0.0
			β_3	-0.30265	0.02296	0.0
		鉱工業 生産指数 (L9)	β_1	0.21155	0.00221	0.0
			β_2	0.06640	0.00500	0.0
			β_3	-0.10753	0.01737	0.0
区分 Cox モデル (ダミーあり) 変更点 (4, 10) 年	49,836.7(0.0) 5,535,707	失業率 (L2)	β_1	0.19016	0.00820	0.0
			β_2	0.01426	0.01884	0.4491
			β_3	0.00672	0.06232	0.9141
		金利差 (L12)	β_1	0.05269	0.00853	0.0
			β_2	0.09234	0.01276	0.0
			β_3	0.02503	0.05897	0.6712
		鉱工業 生産指数 (L9)	β_1	0.05368	0.00353	0.0
			β_2	0.03453	0.00732	0.0
			β_3	-0.02191	0.02503	0.3814

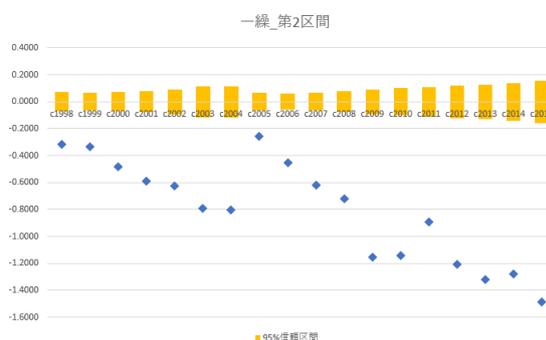
表 9 にはコーホートダミー変数を加えた区分 Cox モデルの推定結果を示す。まず、AIC は区分 Cox モデルに比べて低下しているため、これがより尤もらしいモデルであることを示唆している。また、区分 Cox モデルの結果に比べるとマクロ経済変数の感応度の絶対値が小さくなり、統計的に有意な変数が減少している。区間毎に見ると、第 1 区間における感応度はすべて有意である。「完全失業率」の影響力は低くなったが、なお 3 変数の中で最も影響力が大きいことはこれまでのモデルと同様である。第 2 区間では、これまで最も影響力が大きかった「完全失業率」の推定値が 1 桁低下して有意でなくなったが、他の 2 変数は依然として有意であった。「鉱工業生産指数」の感応度は区分 Cox モデルのときより低下したが、「金利差」の感応度はほとんど変わらず、第 2 区間で最も影響力のある変数となった。なお、区分 Cox モデルで感応度の符号が反転して負になった第 3 区間では、どのマクロ経

済変数も有意にならなかった。

図 17 にコーホートダミー変数を加えた区分 Cox モデルによるコーホートダミー変数に対する感応度の期間構造を示す。図 17 より、第 1 区間では 2000 年代中頃まで感応度はゼロ付近あるいは正負に変動するが、その後絶対値の大きな負値をとる。また、第 2 区間の感応度はマイナスで、2000 年代末頃からは絶対値の大きな負値をとるようになる。一方、第 3 区間の感応度は 2000 年代半ばまで有意に負であるが、その後は上昇し、明確なパターンは見られない。上記のうち、第 1 区間と第 2 区間で共通に見られる傾向、2000 年代末頃からの有意な負の感応度は、経済学的には 2009 年頃からの債務者の返済比率の上昇に対応していると考えられる。返済比率が高いと、一線に充てるための余裕資金の捻出が困難になると考えられるからである。



(a) 時間区間 1 (0-4 年)



(b) 時間区間 2 (4 年超-10 年)



(c) 時間区間 3 (10 年超)

図 17: 標準化を行わない場合における各コーホートダミー変数に対する感応度（一線，区分 Cox モデル（ダミーあり））

以上のように、モデルにコーホートダミー変数を加えたことで、理論モデルと観測値の残差は明確に縮小した。これまでのモデルに比べると、マクロ経済変数の感応度の推定値は第 1 区間と第 2 区間では低下し、逆に第 3 区間では有意ではなくなったが上昇した。つまり、ダミー変数の追加により感応度の絶対値は低下し、マクロ経済変数の影響力は小さく評価されることになった。また、ダミー変数の期間構造が債務者の返済比率の上昇という現象に対応する可能性を指摘することができた。一

方、どの区間でもマクロ経済変数の感応度の値が（有意でなくなる区間はあるものの）正で統一されるようになった。しかし、「完全失業率」と一線の発生率が正の関係にあることの経済学的解釈は依然として難しい。これが本稿の一連の一線に関するモデルにおいて最も気になる点である。

なお、ダミー付拡張 Cox モデルによる経済変数の感応度の推定値を表 10 に、ダミー変数の期間構造を図 18 に示す。表 10 によると、AIC は拡張 Cox モデルに比べて低下しており、統計学的にはより尤もらしいモデルが得られたことを示唆している。どの感応度も有意に正であり、かつ「完全失業率」の影響力が最も大きいことは拡張 Cox モデルの結果と同様で、コーホートダミーを入れたことでどの感応度の値も拡張 Cox モデルよりゼロに近くなっている。また、図 18 によると、ダミー変数の期間構造は図 17 の第 1 区間と第 2 区間に似た傾向を示している。

表 10: コーホートダミー変数付拡張 Cox モデルによる分析結果（一線）。ダミー変数のない拡張 Cox モデルについても比較のため再掲した。

モデル名	Wald 検定統計量 (p 値)	共変量		感応度	標準誤差	p 値
	AIC	(L : ラグの大きさ)				
拡張 Cox モデル (ダミーなし)	43,159.1(0.0) 5,546,755	失業率 (L2)	β	0.42998	0.00253	0.0
		金利差 (L12)	β	0.14796	0.00382	0.0
		鉱工業 生産指数 (L9)	β	0.18350	0.00198	0.0
拡張 Cox モデル (ダミーあり)	48,359.4(0.0) 5,536,721	失業率 (L2)	β	0.15019	0.00559	0.0
		金利差 (L12)	β	0.08985	0.00598	0.0
		鉱工業 生産指数 (L9)	β	0.06546	0.00287	0.0

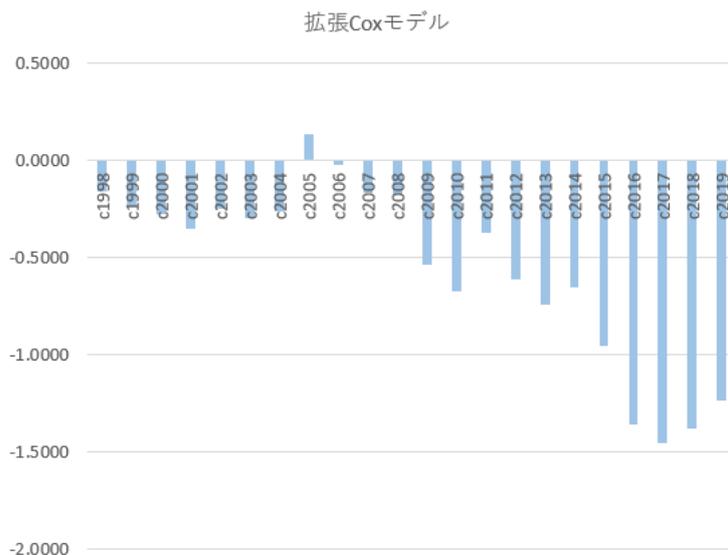


図 18: 標準化を行わない場合における各コホートダミー変数に対する感応度（一線，拡張 Cox モデル（ダミーあり））

なお，図 19 によると，ベースラインハザード関数は最初の 10 年頃まではどのモデルでもあまり変わらず，1 年周期の期間構造を持ち，およそ同水準で変動している．しかし，10 年以降を比較すると，図 20 に見られるように，拡張 Cox モデルでも区分 Cox モデルでもコホートダミー変数を導入すると，ベースラインハザード関数の期間構造の形状は変わらないが，水準は低下する．このことは，ダミー変数の導入の妥当性を暗示している．また，10 年以降では拡張 Cox モデルよりも区分 Cox モデルのベースラインハザード関数の方が水準が高まり，変動性も激しくなる．このことは，区分 Cox モデルの 10 年以降で共変量の感応度の符号がどれも負になるという結果と関連しているかもしれないが，経済学的に辻褃のあう説明はできていない．

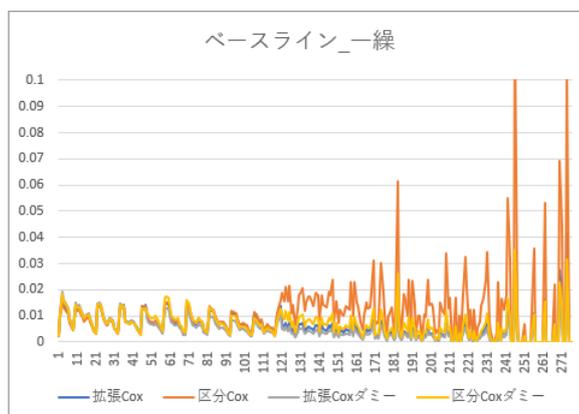


図 19: 各モデルにおけるベースラインハザード関数の比較（一線）

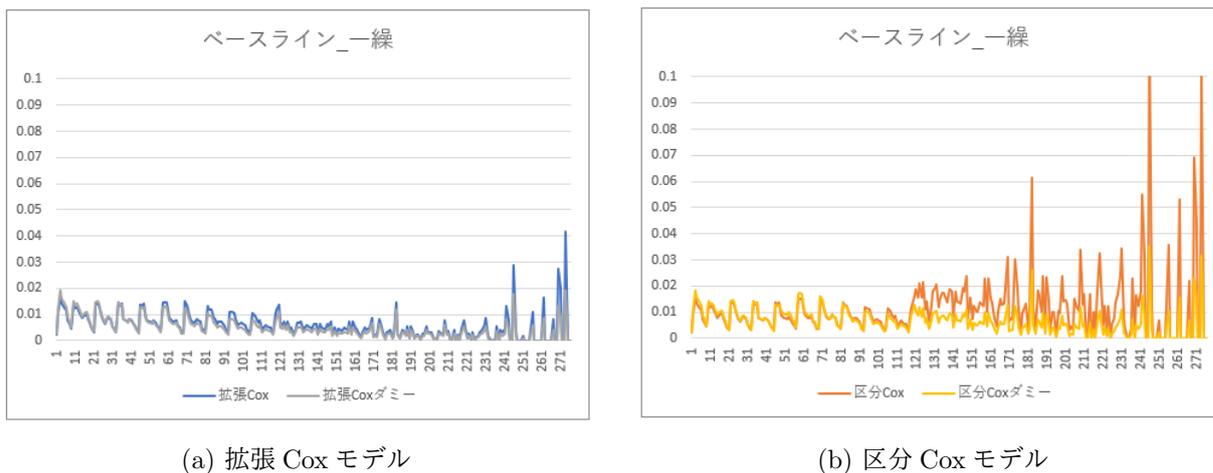


図 20: コーホートダミー変数の有無によるベースラインハザード関数の比較（一線）

7 おわりに

本稿では、住宅ローンに内在する「デフォルト」「全額繰上返済」「一部繰上返済」の3つの競合リスクをタイムラグを考慮したマクロ経済変数を共変量とする拡張Coxモデルを用いて分析し、デフォルトでは「完全失業率（ラグ12か月）」が、全繰では「金利差（ラグ2か月）」が相応の説明力を有し、かつ、住宅ローンの貸付実行時期に依らず高い説明力を維持することを示した。

加えて、Cox-Snell 残差によるモデルの評価では拡張Coxモデルに改善の余地があることが示唆されたため、拡張Coxモデルの仮定を緩めた、経過期間に応じて共変量の影響度が変化する「区分Coxモデル」を導入し、それぞれのモデルを比較した。どのイベントに対しても区分Coxモデルの方がAICは低く、より適したモデルであることが示された。

区分Coxモデルで推定された共変量の影響度に注目すると、経過期間に応じてイベントに対する影響度が大きく変化するマクロ経済変数があった。特に、全繰に対する「金利差」は経過期間が長くなるにつれて影響が逓減する、所謂「バーンアウト効果」が区分Coxモデルによる分析で示された。さらに、拡張Coxモデルでも区分Coxモデルでも、Cox-Snell 残差をみると、コーホート毎でモデルの当てはまりに差異が示されたことから、区分Coxモデルにコーホートダミー変数を加えて分析を行った。ダミー変数を加えることでAICは低下し、特にCox-Snell 残差ではより残差プロットが45度線上に集まっており、大幅にモデルが改善した。

推定されたダミー変数の影響度の比較から、2009年頃に貸付実行された住宅ローンについてデフォルトは起きやすく、全繰及び一繰は起きにくいということが示唆された。この世界的な金融危機の影響を受けた時期に貸付実行された住宅ローンは、他の時期と比較して返済比率が高く、債務者の信用度の質がイベントの発生確率に大きく影響していると考えられる。本稿では主に外的変数であるマク

口経済変数を共変量として分析を行ったが、多くの先行研究で共変量として用いられる内的変数と組み合わせることで、より適したモデルの構築が期待できる。

本研究では、デフォルトやプリペイメントのイベント発生確率とタイムラグを考慮した経済変数の間の関係をいろいろなモデルで分析したが、まだまだ考えるべき点が残されている。例えば、本研究ではそれぞれの経済変数で最も影響力の強いラグを総当たり法で選択し、より詳細なモデルを検討の方がよいかもしれない。具体的には、同じ経済変数への依存性でも、異なるタイムラグの複数の変数、例えばラグ0から6までの変数との関係を同時に考慮して、その係数（感応度）をタイムラグの多項式で表現するといったアーモンラグモデルの適用も考えられる。

同様に、区分Coxモデルにおける区間数についても、今回の分析では感応度の有意性やモデルの単純さ等の規準から3と仮定したが、感応度に対する区間数を増やす、あるいはアーモンラグモデルのように感応度自体を何らかのパラメータを持つ経過期間の関数として推定することも考えられる。このような分析により、感応度の期間構造についてより深い知見が得られる可能性もあり、実際試験的に行った区間数を4とした推定において、全線における「金利差」の感応度が、はじめに上昇し4年程度をピークとして低下する、いわゆる「山型」の期間構造を示したことは興味深い。

また、本研究では競合モデルにおいてイベントの発生は互いに条件付独立と仮定しているが、最近では条件付独立性を仮定しない分析についても議論されているので、それらの適用も今後の課題である。

さらに、デフォルトやプリペイメントのイベント発生件数を明示的に扱わなかったが、実際にはそれぞれのイベントは一件ごとに異なるエクスポージャで発生する。そのため、本来であれば各イベントごとに確率的なエクスポージャを持つマーク付き点過程としてデフォルトやプリペイメントを取り扱うことが望ましい。機構債に関しては各件のエクスポージャに関するデータまでは公開されていないが、月次のイベント発生件数は公開されているので、このデータも同時に分析することにより、これまでとは異なる知見が得られることも期待できる。

あるいは、本稿で分析したローンのデフォルトや繰返などのイベントの発生率は債務者属性等にも依存すると考えられるが、最近では、機械学習による大規模データを用いた解析技術が進展し、多くの属性データとイベントとの関係が分析できるようになってきた。今後は、本稿のような研究をそれらの研究と結合し、経済変数と大量の属性等のデータを同時に用いて分析することで、より現実を反映した数理モデルの構築も期待できる。加えて、本稿のような分析は金融機関が保有するさまざまな資産・負債クラス、例えば、定期性預金や流動性預金における解約・預入などのイベントに対しても適用できるので、同様の分析をさまざまな資産負債クラスに対して行うことで、金融機関の資産負債管理の更なる高度化、例えば、金利依存性を資産負債の特性変化まで考慮した形で評価できるようになることが期待できる。それらの分析結果は、例えば、相対的なリスクの高低を利率に反映させるといった形で戦略的な活用に結び付けることも可能であろう。

また、本研究では経済変数だけでなくコーホートダミー変数を追加したところ、説明力が向上する

という結果が得られた。コーホートダミー変数の導入は、過去データの分析および既存コーホートに対するリスク分析としては問題ないが、将来構築されていくコーホートを含むポートフォリオのリスク分析には使い難いと考えられる。なぜならば、将来構築されるコーホートの特性（ダミー変数に対する感応度）を現時点で知ることはできないからである。保険では、既存コーホートだけからなるポートフォリオを対象とする評価モデルをクローズドモデル、将来構築されるコーホートを含むポートフォリオを対象とするモデルをオープンモデルと呼ぶ。これらの用語を使うと、コーホートダミー変数の導入は、クローズドモデルの分析としては問題ないが、オープンモデルの分析にはやや問題が残る。もしもコーホートダミー変数の代替となる効果を持つ経済変数を見つけられれば、その変数の時系列的特性を考慮したシミュレーションを行うことでオープンモデルの将来のリスク分析が可能になるのでたいへん望ましいが、残念ながら我々はそのような変数をまだ発見できていない。あるいは、本研究で得られたコーホートダミー変数に対する感応度の経験分布を使い、将来構成されるコーホートの感応度を確率的に与えることでもオープンモデルの将来のリスク分析を行えるが、その場合、ダミー変数に対する感応度の時系列的特性は考慮できない。このように、過去データの実証分析で終わらずにオープンモデルの将来のリスク分析まで行う場合、コーホートダミー変数のモデルへの導入の是非は、どこまでのリスク分析を行いたいかに依存するので、深い検討が必要であろう。

A Cox-Snell 残差の分析に用いる逆関数に関する補遺

本節は、3.4節で用いた $(-\log \hat{S}^{KM})(\cdot)$ の逆関数を正確に定義することを目的とする。

まず、(2.4)で導入した生存関数に対する Kaplan-Meier 推定量

$$\hat{S}^{KM}(t) = \prod_{j: t_{(j)} \leq t} \left(1 - \frac{d_j}{r_j}\right)$$

は、 $[t_{(j)}, t_{(j+1)})$ で一定の値を取り、 $t = t_{(j+1)}$ でジャンプする階段型の関数であることに注意する。すなわち、 $(-\log \hat{S}^{KM})(\cdot)$ に対して通常の意味での逆関数は定義できない。このため、本稿では次のような定義を行った。

1. ある j が存在して、 $t \in [t_{(j)}, t_{(j+1)})$ に対して $-\log \hat{S}^{KM}(t) = x$ となるような x について、

$$(-\log \hat{S}^{KM})^{-1}(x) := t_{(j)} \tag{A.1}$$

とする。

2. 1. のような j が存在しない x について、

$$-\log \hat{S}^{KM}(t_{(j)}) < x < -\log \hat{S}^{KM}(t_{(j+1)}) \tag{A.2}$$

をみたす j が存在する場合,

$$\overline{Inv}(x) := t_{(j+1)} = \inf\{t; -\log \hat{S}^{KM}(t) > x\}, \quad \underline{Inv}(x) := t_{(j)} = \sup\{t; -\log \hat{S}^{KM}(t) < x\} \quad (\text{A.3})$$

とし, これを

$$(-\log \hat{S}^{KM})^{-1}(x) := w(x)\overline{Inv}(x) + (1 - w(x))\underline{Inv}(x) \quad (\text{A.4})$$

と線形補間して逆関数を定義した. ただし, 補間のウェイト $w(x)$ は,

$$w(x) := \frac{(-\log \hat{S}^{KM}(t_{(j)})) - x}{(-\log \hat{S}^{KM}(t_{(j)}) - (-\log \hat{S}^{KM}(t_{(j+1)})))} \quad (\text{A.5})$$

で与えた.

3. 1.2. のいずれでもない場合,

$$x < -\log \hat{S}^{KM}(t_{(1)}) \quad (\text{A.6})$$

のときは, $\hat{S}^{KM}(0) = 1$ に注意して,

$$\underline{Inv}(x) := 0, \quad w(x) := \frac{0 - x}{0 - (-\log \hat{S}^{KM}(t_{(1)}))} \quad (\text{A.7})$$

を (A.4) 式に代入し定義した. また,

$$-\log \hat{S}^{KM}(t_{(j)}) < x \quad (\text{A.8})$$

のときは,

$$(-\log \hat{S}^{KM})^{-1}(x) := t_{(j)} \quad (\text{A.9})$$

と定義した.

参考文献

- [1] David Collett, (宮岡悦良監訳). 医療統計のための生存時間データ解析 原著第2版. 共立出版, 2013.
- [2] David Roxbee Cox. Regression models and life-table. Journal of the Royal Statistical Society, Series B (Methodological), Vol. 34(2), pp. 18–22, 1972.

- [3] Ran Deng and Sherminah Haghani. Fha loans in foreclosure proceedings: Distinguishing sources of interdependence in competing risks. Journal of Risk and Financial Management, Vol. 11(1), p. 2, 2017.
- [4] David G. Kleinbaum, Mitchel Klein, (神田英一郎・藤井朋子訳). エモリー大学クラインバウム教授の生存時間解析 原書第3版. サイエントティスト社, 2015.
- [5] Roberto G. Quercia, Anthony Pennington-Cross, and Chao Yue Tian. Differential impacts of structural and cyclical unemployment on mortgage default and prepayment. Journal of Real Estate Finance and Economics, Vol. 53(3), pp. 346–367, 2016.
- [6] Eduardo S. Schwartz and Walter N. Torous. Prepayment and the valuation of mortgage-backed securities. The Journal of Finance, Vol. 44(2), pp. 375–392, 1989.
- [7] 一條裕彦, 森平爽一郎. 住宅ローンのプリペイメント分析. ジャレフ・ジャーナル 2006『不動産金融工学の展開』, Vol. 2006, pp. 221–246, 2006.
- [8] 沓澤隆司. 住宅ローンの期限前償還, 借り換え行動, 延滞の分析. 都市住宅学, Vol. 2007, No.58, pp. 66–75, 2007.
- [9] 江村剛志, 道前洋史. コピュラを用いた生存時間解析 – 関連のあるエンドポイントとメタ分析の活用-. 統計数理, Vol. 68(1), pp. 147–174, 2020.
- [10] 松尾竜悟. マクロ経済変数を用いた住宅ローンのデフォルト及びプリペイメント分析, 東京都立大学経営学研究科修士論文, 2021.
- [11] 杉村徹. 住宅ローンのプリペイメント・モデルと実証分析: 返済タイプ別モデル・アプローチ. ジャフイージャーナル 2003『金融工学と資本市場の計量分析』, Vol. 2003, pp. 115–148, 2003.
- [12] 西川正子. 生存時間解析における競合リスクモデル. 計量生物学, Vol. 29(2), pp. 141–170, 2008.
- [13] 青沼君明, 木島正明. 定期預金のプリペイメント・リスク評価モデル. 日本応用数理学会論文誌, Vol. 8(1), pp. 45–66, 1998.
- [14] 藤原翼. 家計の住宅ローンを点検する-近年の動向とコロナショックによる現時点での影響-. 大和総研レポート, https://www.dir.co.jp/report/research/capital-mkt/securities/20210122_022038.html, 2021.
- [15] 独立行政法人住宅金融支援機構市場資金部. 住宅金融支援機構償還履歴データ《コアデータ》に関する項目説明. https://www.jhf.go.jp/investor/shisan_tanpo/kihatsu.html, 2014.