VAR モデルを用いたドメイン名登録数の推移予測

Predicting Trends in Domain Name Registrations Using a VAR Model

宇井　隆晴/Takaharu UI[1]・池原　翔太/Shota IKEHARA[2]・森　健太郎/Kentaro MORI[2]

金子　明広/Akihiro KANEKO[2]・広瀬　啓雄/Hiroo HIROSE[1]

[1]公立諏訪東京理科大学　・　[2]（株）日本レジストリサービス

［Abstract］
This study proposes a method for forecasting the number of domain name registrations under country-code top-level domains (ccTLDs) using publicly available socio-economic indicators and a Vector Autoregressive (VAR) model. While prior approaches often rely on confidential registry data, our method enables prediction using open data, facilitating external analysis and international comparisons. We examined 11 economic indicators and evaluated various dimensionality reduction strategies to address multicollinearity, with principal component analysis (PCA) showing the best balance of stability and accuracy.
To enhance model learning under limited annual data, we introduced linear interpolation and tested preprocessing techniques such as differencing to enforce stationarity. Experimental results for JP domain registrations demonstrated that PCA-based dimensionality reduction and quarterly-level interpolation significantly improved prediction accuracy. We further extended the model to other ccTLDs and observed that differencing improved performance in some countries but degraded it in others, underscoring the need for flexible preprocessing strategies.
Notably, the model not only enables accurate forecasting, but also provides a reference baseline: deviations between predicted and actual values may reveal the impact of promotional campaigns, regulatory changes, or other non-economic factors affecting domain name trends. This dual role suggests broader applicability in registry operations, policy evaluation, and cross-country digital infrastructure analysis.

［キーワード］
ドメイン名、インターネット安定運用、VAR モデル、多重共線性、次元削減、時系列予測

１．はじめに

インターネットの基盤を構成する重要要素として、その黎明期から現在に至るまで、ドメイン名と IP アドレス、そしてこれらを関連付ける DNS（ドメインネームシステム）が存在する[1][2][3]。インターネット上でのコミュニケーションは、これらの要素に大きく依存している。ドメイン名と IP アドレスはインターネットのアドレス割り当てを行う国際組織である ICANN（Internet Corporation for Assigned Names and Numbers）を頂点とする階層構造によって管理されており、インターネットの運営における重要な役割を担っている[4][5]。

有限のインターネット資源とも言われる IP アドレスの分配においては公平性が求められる[6]。このため、大陸レベルや国・地域レベルで IP アドレスを管理する組織である「IP アドレスレジストリ」は非営利機関が担っていることが多い。

一方で、ドメイン名を管理する組織である「ドメイン名レジストリ」については、ドメイン名の商業価値の高まりとともに、民間のサービスという形も取り入れながら競争環境の中で発展してきた。ドメイン名レジストリを非営利機関や政府系組織が担っている場合でも、そのサービスはビジネス的な競争環境の中に置かれている。

ドメイン名レジストリは、サービスの公平・中立な提供はもちろん、利便性の追求も求められるが、何よりも最も重要なのは、そのような競争環境の中でサービスを安定的かつ継続的に提供することである。

DNS は IP アドレスを用いて通信を行うインターネットに、ドメイン名という名前空間を関連付けるもので、これが機能しなくなるとドメイン名によるあらゆるアクセスができなくなる。ドメイン名レジストリが運用する TLD（トップレベルドメイン）の DNS が障害により停止すると、その TLD に属するドメイン名を用いてのアクセスができなくなり、インターネット社会に大きな影響を与える。Web や電子メールだけでなく、多くの社会・経済基盤がインターネット上で実現されている現代において、この影響は計り知れない。

また、継続性の観点からも、ドメイン名レジストリが経営的に破綻し、サービスが停止・終了するような事態が発生すれば、DNS の障害以上に大きな社会的影響を及ぼすことになる。このため、ドメイン名レジストリにとっては、インターネット社会の成長の中で DNS を安定的に提供するための将来予測と、システムへの長期的な設備投資計画を持つことが経営上の重要課題となる。

ドメイン名レジストリのサービスは、ドメイン名の登録料や登録更新料からの収益によって成り立っている。一般にドメイン名はドメイン名レジストリに登録する際に登録料を支払い、1 年ごとの登録更新の際に登録更新料を支払う、という収益モデルとなっている。経営の観点では、ドメイン名の登録数がどのように変化していくのかを長期的に見通すことが重要である。ドメイン名の登録数は、日々新規に登録される数と、それらがどれだけの期間にわたって登録継続されるかという、二つの要素から成り立っている。

ドメイン名の登録数は、ミクロな観点から見れば、ドメイン名レジストリがサービス対象としている地域の状況や、展開しているサービス、新たに登録されるドメイン名の種類、登録継続期間など、多様な要素の積み重ねであると考えられる。一方で、マクロな観点からは、ドメイン名レジストリがサービス対象とする地域の社会発展やインターネット利用拡大の時間軸における現在の状況を反映したものであるとも考えられる。

登録数の予測手法については、各ドメイン名レジストリにおける取り組みがあると考えられるが、それらの手法が学術的な研究として公開されることは少ない。インターネットの発展が先行した国や地域の事例は、後進にとってはこれから起こる将来の事象の予測となるものでもあり、このような取り組みを研究成果として公開していくことには一定の価値があると考える。

２．先行研究と本研究の意義

ドメイン名の登録数の推移予測に関する研究として、筆者らによる先行研究が存在する[7]。この研究はコーホート分析の手法を用いている。1 か月ごとのドメイン名の新規登録集団をコーホートとして捉え、毎月の新規登録数と、それらの 1 年後の更新率のデータを用いて、季節性要素を考慮した時系列データ予測を行うことができる SARIMA（Seasonal AutoRegressive Integrated Moving Average）モデル[8][9][10][11]によってドメイン名登録数の推移予測を行う手法を提案した。そして実際に JP ドメイン名を事例として予測を試み、高い精度での予測を実現した。

しかし、ドメイン名の登録に関するこのような詳細なデータは、ドメイン名レジストリ自身は保有していても、それを外部に公開していないことが多い。このため、コーホート分析を用いたドメイン名登録数推移の予測手法は、実質的にはドメイン名レジストリが自らサービスするドメイン名についてのみ適用することが可能であり、外部の第三者の立場では予測に用いることが難しい。

ドメイン名レジストリが自らサービスするドメイン名登録数の予測を行うことができることはもちろん大きな意義がある。その一方で、ドメイン名レジストリ自身しか知りえない情報を用いることなく、一般に公開されている情報だけを用いてドメイン名登録数の推移を予測することができれば、ドメイン名レジストリ以外の第三者の立場での予測が可能となる。

この点に関して、上村による先行研究では、複数の ccTLD を対象とした統計分析が行われ、ドメイン名登録数は人口規模、GDP、インターネット普及率といった社会経済的要因の影響を受けることが示されており[12]、他にも経済成長動向とインターネット利用の関係を論じた研究も多い[13][14][15]。

本研究では、多くの国で一般に公開されている経済指標の統計量を用いて予測を行う方法を提案する。これにより、ドメイン名レジストリ自身しか持ちえない詳細なドメイン名登録と更新に関するデータがなくても、過去のドメイン名登録数の推移データと、容易に入手可能な当該国の経済指標の統計量をもとに客観的にドメイン名の登録数を予測することができるようになる。

また、経済指標の統計量は世界各国で同じ算出方法で計算され公開されており、これにより同じ手法

で各国のドメイン名登録数を予測し、比較研究することができるようになる。経済そしてインターネットの発展の状況は国ごとに異なり、また時間的な先行・後進も存在する。そのような状況の中で、自国と他国の実績と予測を比較検討できることは、ドメイン名の登録数の予測というだけでなく、それと因果関係を持つ経済状況の把握・予測の観点からも、意義深いものであると考える。

３．モデルの構築と実験方法

３．１ VAR モデルの採用

ドメイン名は、企業等の組織や個人によって登録され、主に Web や電子メールのアドレスとして利用される。したがって、ドメイン名の登録数は、企業の数や人口、その国のインターネットの活用度、経済の状況などに関係していると考えられる。そして、多くの国においてこのような経済指標は過去から現在にわたって継続的に計測・公開されてきている

これらの時系列データとドメイン名の登録数推移という時系列データの間に因果関係があるのであれば、複数の時系列データの因果関係を分析する多変量の統計モデルである VAR モデル（ベクトル自己回帰モデル：Vector Autoregressive Model）[16][17][18]を用いることで、ドメイン名の登録数推移を予測することができると考えた。

VAR モデルでは、時間軸に沿って集められたデータを基に、同時点のみならず、各変数のラグ変数間にある関係性も考慮して定式化することができる。VAR モデルは、単変量自己回帰モデル（AR モデル）を多変量に拡張したモデルで、利用する変数がラグを伴って相互に影響を与えあっているような状況を分析することができ、経済分野での応用研究も多い[19][20][21]。

VAR モデルの一般的な形は以下の式で表される。

$$Y_t = C + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \epsilon_t, \qquad (1)$$

ここで、Y_t は複数の時系列変数（GDP や人口など）を並べたベクトル、Cは定数項、$A_1, A_2, \ldots, A_p$は遅延パラメータで、過去のデータが現在の値に与える影響を表す。ϵ_tは誤差項である。

近年では深層学習に基づく LSTM などの手法が時系列予測に応用されているが、それらはデータ量を多く必要とし、結果の解釈性に乏しい。一方、VAR モデルは経済学で広く使われる手法であり、変数間の相互影響を明示的に捉えつつ、少量の時系列データでも学習が可能であるという利点がある。

３．２ 予測に用いる特徴量としての経済指標

多くの国で一般に公開されている経済指標の中から、ドメイン名の登録数と相関があると考えた以下の 11 個の経済指標の推移データを特徴量として用いることとした。

なお、本研究では特徴量としてどの経済指標を用いるべきかというところにフォーカスするのではなく、後に述べるように、目的変数であるドメイン名登録数に相関がありそうな特徴量が多数あるときに、それらを用いてどのように予測精度を上げるか、という手法の構築にフォーカスしている。そのため、今回選択した 11 個の経済指標が予測に適したものであるかどうかや、他に組み入れるべき経済指標があるのではないか、ということについては本研究の中では扱わない。

1. population（人口）
   インターネット利用者やドメイン名登録数の潜在的な規模を示す基本的な指標である。人口が多いほど、インターネットやデジタルサービスの需要が高まるため、ドメイン名登録数に影響を与えると考えられる。
2. GDP（国内総生産）
   国全体の経済規模を示し、ドメイン名登録数の経済的背景を評価するための重要な指標である。GDP が高い国ほど、インターネットやオンラインビジネスの発展が期待され、ドメイン名登録数が増加する可能性がある。

3. perGDP（ひとりあたり GDP）
   国民の平均所得水準を示し、個人や企業がドメイン名を登録・維持する経済的余裕を反映する指標である。
4. internet（インターネット普及率）
   人口の中でインターネットにアクセス可能な割合を示す。インターネット利用者が増加することで、ドメイン名登録数が直接的に増加すると考えられる。
5. phone（携帯電話普及率）
   デジタル通信技術の普及状況を示す指標である。特に、モバイルインターネットの利用が進む国では、オンラインプレゼンスを確保するためにドメイン名が必要となるため、ドメイン名登録数に影響を与える可能性がある。
6. pdensity（人口密度）
   地域ごとの都市化やインフラの集中度を表す指標である。人口密度の高い地域では、インターネットアクセスが容易であり、ドメイン名登録の動向に影響を及ぼす可能性がある。
7. GNI（国民総所得）
   国全体の所得を示す指標であり、経済全体の購買力を反映する。GNI が高い国では、企業や個人がインターネットプレゼンスを拡大するための投資を行いやすく、ドメイン名登録数の増加につながる可能性がある。
8. unprate（失業率）
   労働市場の状況を表し、経済的安定性を示す指標である。失業率が高い国では、新しいビジネスや起業が制限される可能性があり、ドメイン名登録数に影響を与えると考えられる。
9. pbalance（国際収支）
   貿易や資本移動を含む経済活動全体を表し、国の経済的な健全性を示す指標である。経済が活発な国ほど、企業のオンライン活動が盛んであり、ドメイン名登録の増加が見込まれる。
10. perGNI（ひとりあたり GNI）
    国民一人あたりの所得水準を示し、インターネット関連サービスへの支出能力を測る指標である。個人レベルでのドメイン名需要に影響する可能性がある。
11. perpbalance（ひとりあたり国際収支）
    国際経済活動が個人レベルでどの程度影響を及ぼしているかを示す指標である。国際取引が盛んな国では、オンラインプレゼンスが重要視され、ドメイン名登録数に影響を与える可能性がある。

３.３ 予測に用いる時系列データと実験方法

本研究では、JP ドメイン名の登録数推移の予測を試みた。

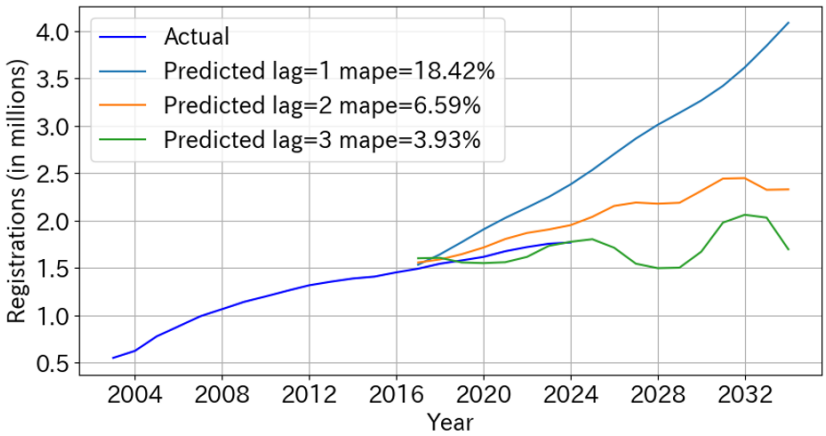
この予測に用いる時系列データとして、予測対象でもある JP ドメイン名の登録数については 2003 年から 2024 年までの推移データ[22]を、特徴量となる 11 個の経済指標については日本における 2003 年から 2023 年までの 1 年ごとの時系列データを収集した[23]。

実験方法は以下の通りとした。

- 時系列データのうち 2003 年から 2016 年までを学習データとして利用
- 2017 年から 2034 年までを VAR モデルで予測
- 実データと比較できる 2017 年から 2024 年について MAPE を算出して予測精度を評価

本研究では、予測精度の評価指標として MAPE（平均絶対パーセント誤差）を採用した。MAPE は予測誤差を実際の値に対するパーセンテージで表すため、スケールに依存せず、異なるデータセット間で容易に解釈できる。

本研究の目的は、VAR モデルに基づくさまざまな変数選択および次元削減手法の予測性能を評価することであり、MAPE は実験間で直接的かつ比較可能な指標を提供できる[24][25]。



図－1 11 個の特徴量を用いて VAR モデルで予測した結果

4．JP ドメイン名の登録数予測実験

4．1 実験 1：11 個の特徴量を用いた予測実験

最初に、11 個の経済指標のすべてを特徴量として用いて VAR モデルで JP ドメイン名の登録数推移の予測を行った結果を図－1 に示す。

ラグ（lag）の値によって異なる傾向を示していることが見て取れる。実際の JP ドメイン名の登録数推移は全体に対数的な曲線を描いているが、ラグ 1 の時の予測曲線は指数関数的な増加傾向となっており、MAPE も 18.42%と大きな乖離を示している。一方でラグ 3 の時に MAPE は最も小さく 3.93%となったが、予測曲線の振動が大きく、同様の振動を示すラグ 2 の曲線とともに、いずれも正しく予測できているとは言い難い。

この結果の原因としては、以下の要因が考えられる。

- 特徴量間に強い相関関係がある（多重共線性がある）
- 各特徴量の系列上の個数に対して特徴量の数（次元）が多い

多重共線性とは、回帰分析において説明変数同士が強い相関を持っている状態である。VAR モデルを用いる場合、多重共線性は係数推定の不安定化や過剰適合を引き起こし、予測精度を低下させる要因となる。また、VAR モデルでは扱う系列データが多いほど、各系列上のデータ個数は多く必要となり、これが不足することで安定的な予測が行えなくなる[26][27]。

以降のステップでは、複数の手法で多重共線性を排除しつつ次元削減を行い、予測精度の向上を試みた。

4．2 実験 2：多重共線性を回避した特徴量の選択

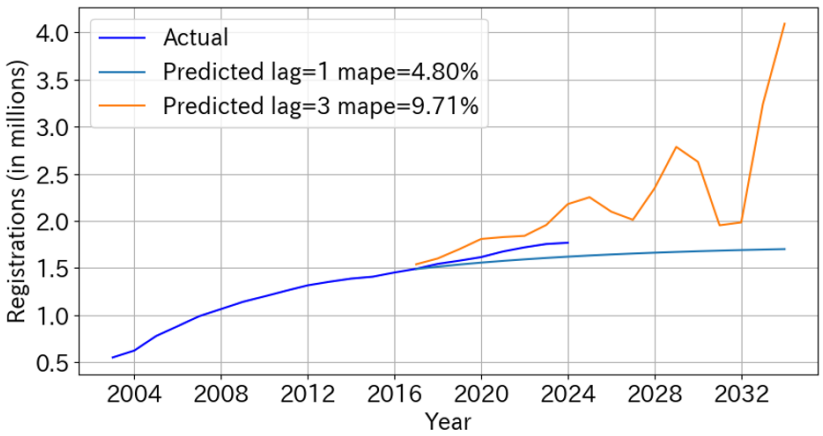
4．2．1 実験 2-1：重回帰分析による優位な特徴量の選択

各特徴量が目的変数にどれだけ有意な影響を与えているのかということを評価するため、11 個の経済的指標を説明変数、ドメイン名登録数を目的変数として重回帰分析[28][29]を行った。その結果、決定係数（R-squared）は 0.988 と非常に高く、説明変数が目的変数を非常によく説明していることが示された。決定係数は、目的変数の分散のうち、モデルによって説明される割合を示す指標であり、1.0 に近い値は、モデルがデータの変動をほぼ捉えていることを意味する。

さらに、回帰全体に対する F 統計量の P 値は 4.21e-10 と非常に小さく、モデル全体が統計的に有意

表―1 説明変数ごとのP値

Feature	t	P > |t|
const	-0.023	0.982
population	0.021	0.984
GDP	1.918	0.079
perGDP	-1.913	0.080
internet	1.903	0.081
phone	2.977	0.012
pdensity	-0.023	0.982
GNI	-1.247	0.236
unprate	0.952	0.360
pbalance	-0.189	0.853
perGNI	1.271	0.228
perpbalance	0.196	0.848



図―2 重回帰分析により選択した優位な特徴量を用いて予測した結果

であることが確認された。F 統計量は、少なくとも 1 個の説明変数が目的変数の分散を説明するのに有意な寄与をしているかどうかを検定するものであり、ごく小さい P 値は、観測された関係が偶然によって生じた可能性が極めて低いことを示す。

一方で、デザイン行列のコンディション数は 2.84e+16 と極めて高く、説明変数間に強い多重共線性が存在する可能性が強く示唆される。この多重共線性は、分散の膨張や係数推定値の不安定化を引き起こす可能性がある［30］。

説明変数ごとの P 値を表－1 に示す。一般に、重回帰分析においては説明変数の P 値が 0.05 未満であるものが目的変数に有意な影響を与えていると評価される。表－1 の結果では、P 値が 0.05 未満のものは"phone"のみであるが、P 値の分布を鑑みると"GDP"、"perGDP"、"internet"の 3 個の説明変数も目的変数に有意な影響を与えていると判断できる。

そこで、これらの 4 個の説明変数を選択して、VAR モデルにより JP ドメイン名の登録数推移の予測を行った結果を図－2 に示す。(ラグ 2 の予測曲線は振動幅が非常に大きく描画範囲に収まらない結果となったため省略している)

予測曲線の全体の傾向が実績曲線に沿う形になっていることは、目的変数に有意な影響を与えている説明変数のみを抽出した結果、ノイズ成分が除去されたことによるものと考えられる。一方で、ラグが 2 以上の場合に大きな振動が現れているのは、多重共線性による過剰適合によるものと考えられる。

この説明変数間での多重共線性は重回帰分析の P 値による評価では取り除くことができないため、他の手法を検討する必要がある。

表－２ 11 個の説明変数の相関行列

	1	2	3	4	5	6	7	8	9	10	11
1. population	1.000	0.531	0.474	−0.472	−0.896	0.999	0.316	0.766	−0.180	0.228	−0.209
2. GDP	0.531	1.000	0.997	0.015	−0.285	0.531	0.754	0.421	−0.248	0.720	−0.260
3. perGDP	0.474	0.997	1.000	0.054	−0.226	0.474	0.759	0.376	−0.242	0.731	−0.253
4. internet	−0.472	0.015	0.054	1.000	0.778	−0.473	0.168	−0.752	−0.048	0.218	−0.034
5. phone	−0.896	−0.285	−0.226	0.778	1.000	−0.898	−0.083	−0.885	0.068	0.001	0.095
6. pdensity	0.999	0.531	0.474	−0.473	−0.898	1.000	0.314	0.768	−0.177	0.225	−0.206
7. GNI	0.316	0.754	0.759	0.168	−0.083	0.314	1.000	0.202	−0.616	0.995	−0.622
8. unprate	0.766	0.421	0.376	−0.752	−0.885	0.768	0.202	1.000	−0.160	0.131	−0.181
9. pbalance	−0.180	−0.248	−0.242	−0.048	0.068	−0.177	−0.616	−0.160	1.000	−0.617	0.999
10. perGNI	0.228	0.720	0.731	0.218	0.001	0.225	0.995	0.131	−0.617	1.000	−0.620
11. perpbalance	−0.209	−0.260	−0.253	−0.034	0.095	−0.206	−0.622	−0.181	0.999	−0.620	1.000

表－３ 7 個の説明変数の VIF 計算結果

Feature	VIF
population	609.569051
GDP	4.037708
internet	3.955024
phone	4.884636
GNI	4.711627
unprate	6.848410
pbalance	2.114877

表－４ 6 個の説明変数の VIF 計算結果

Feature	VIF
GDP	281.215362
internet	168.068950
phone	56.268741
GNI	390.769220
unprate	55.290928
pbalance	11.455180

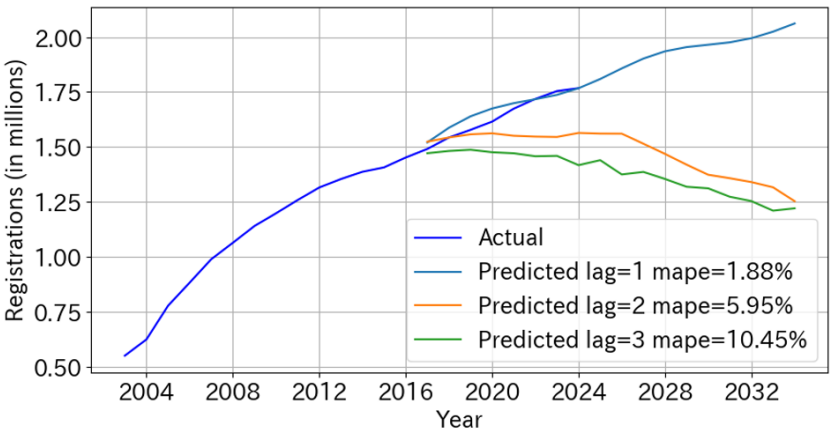
４．２．２ 実験 2-2: 多重共線性を回避した特徴量の選択

11 個の説明変数間における多重共線性を確認するため、相関行列を求めた結果が表－２である。

この中から相関係数が高い(0.9 以上)説明変数の組み合わせ(表－２中における赤字の組み合わせ)について、その片方を説明変数から外すことで相関を持つ状態を解消することを試みた。本実験では縦軸側の説明変数である"perGDP", "pdensity", "perGNI", "perpbalance"の 4 個を外すこととした。そのうえで、さらに残る 7 個の説明変数間における多重共線性を確認するために分散拡大係数（VIF: Variance Inflation Factor）を計算した結果が表－３である。

VIF は多重共線性の程度を数値的に評価するために用いられる指標であるが、一般に、VIF の値が 5 または 10 以上の変数は多重共線性の疑いがあるとされる[31][32][33]。今回はこの結果より非常に大きな VIF 値を示した"population"を説明変数からさらに外すこととした。残る 6 個の説明変数を選択して、改めて VAR モデルにより JP ドメイン名の登録数推移の予測を行った結果を図－３に示す。

ラグ 1 の時の予測結果の MAPE は 1.88%と高い精度を示しているが、ラグが 2 以上の時の予測曲線は



図－３ 多重共線性のある特徴量を外して予測した結果

大きく外れてしまっている。

ここで改めて予測に用いた 6 個の説明変数の多重共線性を確認するために VIF を計算した結果が表－４であるが、多重共線性の疑いのある説明変数を外した結果としての 6 個の説明変数の間にも強い多重共線性の疑いがあることがわかる。

相関係数に基づく特徴量の選択は、2 変数間の線形関係に注目する手法であるため、多変数間にまたがる複雑な共線構造を十分に捉えることができない場合がある。今回、相関係数が高い特徴量の一方と、VIF の高かった ″population″ を除いたことで、一見多重共線性は解消されたかに見えたが、残った説明変数間で新たな構造的共線性が強まった可能性がある。その結果、再度 VIF を計算すると、全体的に高い値を示した。これは、削除した変数が分散の一部を分担していた役割を失ったことにより、残された変数同士の冗長性が高まったことが一因と考えられる。

なお、VIF の高い変数を順に除外することで表面的には多重共線性を低減できるものの、この方法は同時に予測に有用な説明変数を失うリスクもはらんでいる。特に、VIF は予測の不安定性を評価する指標であって、モデルの予測性能や説明力を直接評価するものではない。そのため、説明変数の選定にあたっては、単純に VIF の閾値に従って変数を削除するのではなく、目的変数に対する説明力や予測精度とのバランスを考慮する必要がある。次の実験では、意味のある変数構成を維持するアプローチを試みる。

４．３ 実験３：多重共線性を考慮した重要度の高い特徴量の選択

ランダムフォレストは多数の決定木から構成されるアンサンブル学習手法である［34][35]。ランダムフォレストはその構造上、入力特徴量間に多重共線性が存在していても比較的頑健であり、冗長な情報による過学習の影響を受けにくいという特性がある［36]。このような多重共線性に対する耐性を活かし、本研究ではランダムフォレストを特徴量の重要度評価による次元削減の支援ツールとして活用することを試みた。

ランダムフォレストを用い、11 個の経済的指標について重要度評価を行った結果を図－４に示す。

この重要度の分布に基づき、以下の 2 つの特徴量セットを選定し、VAR モデルによるドメイン名登録数の予測を行った。

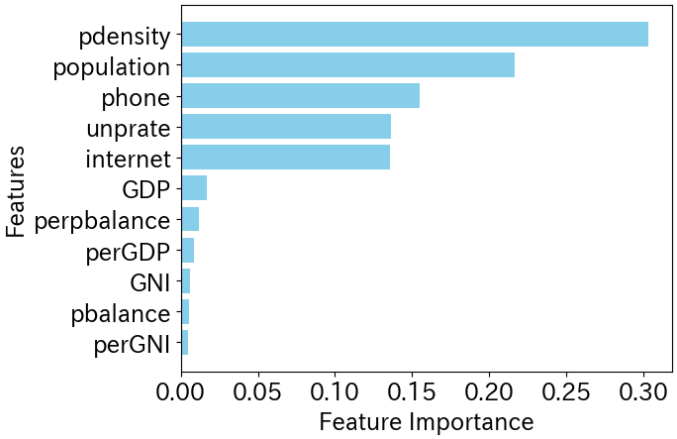
- 最も重要度の高かった ″pdensity″、″population″、″phone″ の 3 個の特徴量セット
- 次に重要度の高かった ″unprate″、″internet″ を加えた 5 個の特徴量セット

それぞれの予測結果を図－５および図－６に示す。

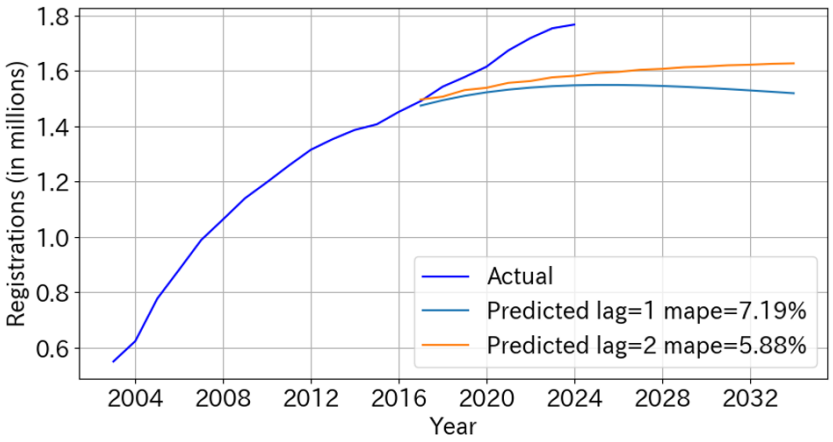
2 つの結果を比較すると、3 個の特徴量セットによる予測は予測曲線がなめらかである一方、MAPE に基づく予測精度は 5 個の特徴量セットの方が優れていた。このことは、追加された 2 つの特徴量

（"unprate"と"internet"）が予測精度の向上に寄与する有益な情報を含んでいることを示唆している。

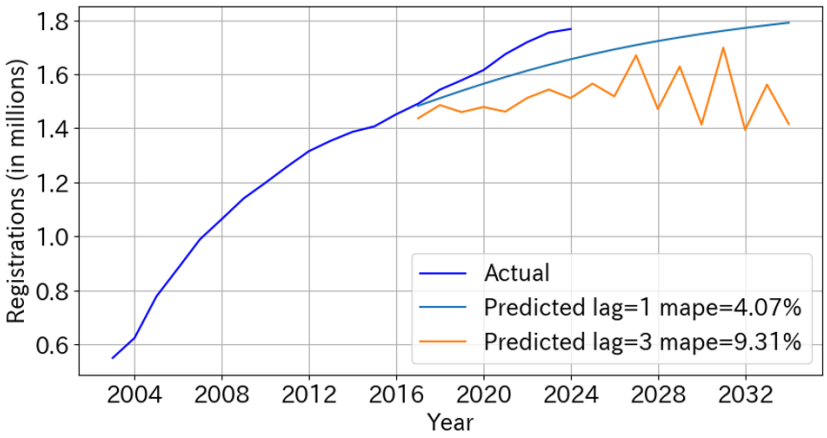
しかしながら、5 個の特徴量セットによる予測では、ラグが大きくなるにつれて予測曲線に振動が現れる傾向が見られた。これは、情報として有用な特徴量を追加したものの、それによって多重共線性の影響が再びモデルに入り込んだことが原因である可能性が高い。



図－４ ランダムフォレストによる特徴量の重要度評価



図－５ 重要度の高い 3 つの特徴量で予測した結果



図－６ 重要度の高い 5 つの特徴量で予測した結果

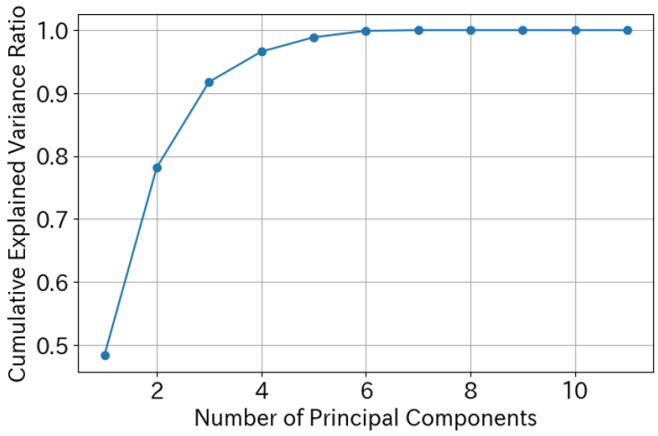
４．４ 実験４：PCA による次元削減

これまでの実験では、相関係数や VIF（分散拡大係数）、ランダムフォレストによる重要度評価などを用いて、説明変数の選択や多重共線性の緩和を試みてきた。これらの手法はそれぞれ一定の効果を示したものの、多重共線性を完全に除去しつつ予測精度を維持することは容易ではなかった。

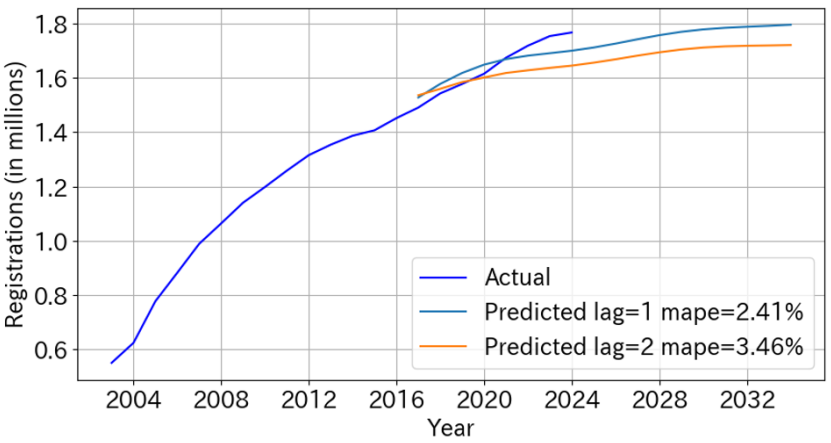
そこで本節では、次元削減のための新たな手法として主成分分析（PCA: Principal Component Analysis）を導入する。PCA は、多数の相関を持つ元の変数群を、互いに直交する新たな変数（主成分）へと変換することで、情報の損失を最小限に抑えつつ、次元を削減する方法である。この変換によって得られる主成分は、元の変数の線形結合でありながら、互いに相関を持たない（＝多重共線性が存在しない）という特長を持つ。これにより、回帰分析や時系列予測モデルにおいて、係数推定の安定性が向上し、モデルの解釈や予測性能に好影響を与えることが期待される［37][38]。

PCA を用いた分析では、保持する主成分の数を適切に決定する必要がある。その基準として一般的に用いられるのが累積寄与率（cumulative contribution ratio）である。これは、主成分が元のデータの分散をどの程度説明できているかを表す指標であり、累積寄与率が高いほど、情報をよく保持していると判断できる。本研究では、11 個の説明変数に対して PCA を実行し、累積寄与率を算出した。その結果を図－７に示す。

今回は、累積寄与率が 90%以上に達するまでの主成分を保持することを方針とし、その結果、第１主成分から第３主成分までの３つを選定した。これらの主成分は、元の変数情報の大部分を保持しつつ、



図－７ 累積寄与率



図－８ PCA で３次元に次元削減した主成分で予測した結果

互いに直交しているため、多重共線性の影響を受けない特徴量として利用できる。

この 3 つの主成分を説明変数として用い、VAR モデルによって JP ドメイン名の登録数推移を予測した結果を図－８に示す。これまでの実験では、ラグが増えるにつれて予測曲線に振動が生じる傾向が見られたが、今回の PCA を適用した予測結果では、そのような振動は顕著に抑えられ、より安定した予測曲線が得られている。

また、MAPE による予測精度の評価においても、高い精度が示されており、PCA による次元削減が本研究の目的である安定かつ高精度な予測に有効であることが確認できた。

５ 予測手法の評価と改善の検討

実験 1 から実験 3 の結果に共通に見られた現象として、ラグ次数を増やした時の予測曲線に振動が現れたことがあげられる。実績曲線においてはこのような振動は現れておらず、この振動は予測精度を下げる方向に影響している。この予測曲線に現れた振動の原因としては、以下の 2 つが考えられる。

１．特定の経済指標が振動しており、この影響を受けて予測も振動している
２．経済指標の数（特徴量の次元数）が多いことと、観測時点数が少ないことが、経済指標間の多重共線性の影響を強く発現させている

まず、１の特定の経済指標の振動が影響している可能性については、実験 1 から実験 3 で見られる予測曲線の振動が、それぞれ時間軸上の異なる場所に現れていることから、特定の経済指標の影響を受けたものとは考えにくい。そこで、以降では上記２が原因である可能性について論じていく。

本研究で用いた 11 個の経済指標の間には顕著な多重共線性が存在していた。これは実験 2-2 の結果において明らかであり、相関行列および VIF の分析によって、変数間に強い依存関係があることが示された。これにより、説明変数の選定は容易ではなく、慎重な検討が求められる課題であることがわかった。

表－５に、各実験において選択された特徴量をまとめている。特徴量の選定傾向は用いた手法によって異なっており、これは一つの変数が強調されると、それと高い相関を持つ他の変数の評価が相対的に抑制されるという現象によるものと考えられる。つまり、使用する手法によって重要とされる変数の順序が変わりうるということであり、経済変数同士が相互に依存しているような多変量時系列の予測においては、手法の選択によって結果が大きく左右される可能性があることを示している。

VAR モデルによる予測では、すべての変数が他のすべての変数のラグ（過去値）に依存するため、多数のパラメータを推定する必要がある。そのため、ラグ次数を増やすと必要なデータ量も急激に増加する。また、多重共線性の高い時系列変数を VAR に使用すると、ノイズの増幅や過学習のリスクが高まり、モデルの汎化性能や予測の安定性が損なわれる可能性がある。

表－５ 各実験で選択された特徴量

Feature	1	2-1	2-2	3
population	○			○
GDP	○	○	○	
perGDP	○	○		
internet	○	○	○	○
phone	○	○	○	○
pdensity	○			○
GNI	○		○	
unprate	○		○	○
pbalance	○		○	
perGNI	○			
perpbalance	○			

表－6 各実験の MAPE 値の比較

Experiment	Lag=1	Lag=2	Lag=3
1	18.42%	6.59%	3.93%
2-1	4.80%	-	9.71%
2-2	1.88%	5.95%	10.45%
3	4.07%	6.18%	9.31%
4	2.41%	3,46%	-

このような背景から、次元削減は重要な手順となる。実験 3-1 では、特徴量を 3 個から 5 個に増やしたことで、予測精度（MAPE）は向上したが、一方でラグ次数が大きくなるにつれて予測曲線に振動が見られた。これは、ランダムフォレストが多重共線性には比較的頑健であっても、VAR モデル自体はそうではないために生じたと考えられる。すなわち、ランダムフォレストで高い重要度を示した変数であっても、VAR モデルに適用すると再び多重共線性が影響を及ぼし、モデルの安定性が損なわれる可能性がある。これは、機械学習における特徴量選択が必ずしも時系列モデル（特に VAR）に適合するとは限らず、方法論的な整合性が問われる事例である［27］。

こうした問題に対して、実験 4 で採用した PCA による次元削減は、従来の変数選択法では完全には解消できなかった多重共線性を効果的に除去することに成功した。PCA は、元の変数を互いに直交する主成分に変換することで、予測に有効な分散構造を保ちつつ、冗長または重複した情報を削減する。これにより、VAR モデルにおける予測精度の向上および予測曲線の安定性の確保に寄与した。

表－6 には、すべての実験における MAPE の比較結果を示している。最も低い MAPE は、相関係数および VIF に基づいて特徴量を選定した実験 2-2 で得られた。次いで精度が高かったのは、PCA によって次元削減を行った実験 4 であった。ただし、これらの結果を評価する際には注意が必要である。実験 2-2 では確かに多重共線性は軽減されたものの、VIF 分析からは依然として残存する共線性の可能性が示唆されていた。また、本研究は日本の JP ドメイン（ccTLD）に限定して登録数推移の予測を行ったため、高い予測精度が得られた背景には、日本の経済構造とドメイン利用の特性がたまたま一致していた可能性も否定できない。日本以外の国におけるデータを用いた検証を行い、予測手法の評価を行う必要がある。

６ 予測手法の改善

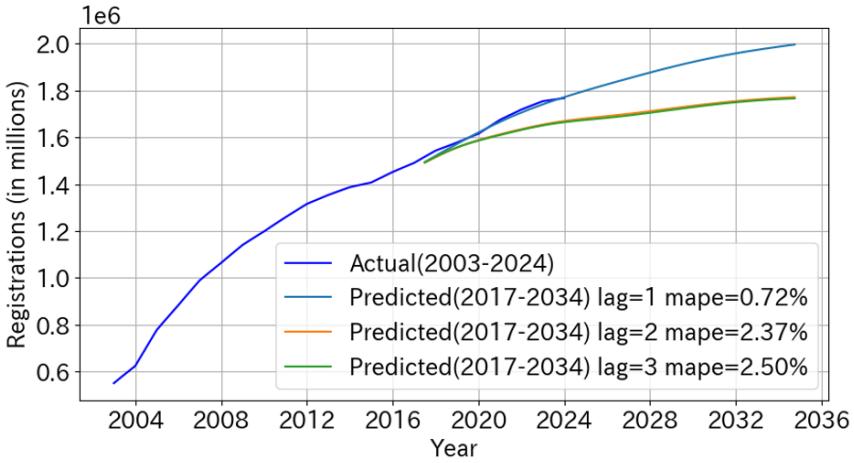
６．１ 実験５：経済指標データの補間による VAR モデルの安定化

ここまでの実験では、2003 年から 2016 年までの 14 年間の年次経済指標データを用いて、JP ドメイン名の登録数推移を予測するための VAR モデルを構築した。しかし、非常に多くの学習データを必要とする LSTM などの深層学習アルゴリズムに比して少ないデータで予測が可能な VAR モデルとはいえ、年次単位で 14 時点分のデータしか得られない状況では十分な学習が困難である可能性がある。特に、VAR モデルでは、すべての説明変数の過去の値（ラグ）を用いて各変数を同時に予測する構造上、推定すべきパラメータの数が急増しやすく、観測時点数が少ない場合、モデルが不安定になりやすいという問題がある。

先の実験 4 では、説明変数として 11 個の経済的指標をそのまま用いるのではなく、これらを PCA によって次元を 3 つに削減している。このとき、VAR モデルの推定に必要なパラメータ数[39][40]は以下の式で表される。

$$\text{パラメータ数} = K^2 \times p\,, \qquad (2)$$

ここで、$K = 4$（3 つの主成分＋ドメイン名登録数）、$p = 1$（ラグ次数）とすれば、$4^2 \times 1 = 16$ のパラメータが必要となる。加えて、各系列の定数項や誤差項も含めると、実際には 20 前後の自由度が必要となる。一方、年次データ 14 点のうち、ラグを考慮して使用可能なデータ点数は 13 点であり、こ



図－９ 時系列データを４倍に補完予測した結果

れはパラメータ数を下回っているため、モデルの安定性に欠け、過学習のリスクが高い。

このような背景から、本節では、時系列の時間解像度を高めるために、四半期相当の線形補間によるデータ拡張を導入する。具体的には、各年の間に等間隔で 3 つの補間点を追加し、1 年あたり 4 時点の時系列を構成する。この処理により、学習データ期間（14 年間）のデータ点数は 56 点に拡張される。ラグ 1 を除いても、有効なデータ点数は 55 点に達し、16 のパラメータを安定して推定できるだけの十分なデータ数を確保できる。さらに、この補間による 4 倍のデータ拡張により、ラグ 2 とした場合でも、$4^2 \times 2 = 32$ のパラメータ数に対して十分なデータ数を確保できている。

図－９は、この 4 倍補間処理を施した時系列データを PCA で 3 次元に次元削減し、VAR モデルによる JP ドメイン名の登録数予測を行った結果である。図－８と比べて予測精度の向上が見て取れる。

このような補間処理は、年次経済指標に特有の性質とも整合的である。多くの経済指標は年単位で比較的滑らかに変化する傾向があり、短期間で急激に変動することは少ない。このため、隣接する年のデータ間を線形で補完することは、経済的にも統計的にも合理的な仮定といえる。

また、補間による拡張は、特定の年のデータが欠落している場合にも有効である。たとえば、ある国や機関で特定年次の統計が未公表である場合であっても、前後のデータから合理的に推定可能であり、分析対象からデータ全体を除外せずに済む柔軟な対応が可能となる。このように、補間は欠測値の補完とモデル学習のための時点数確保という 2 つの目的を同時に満たす手法として有用である。

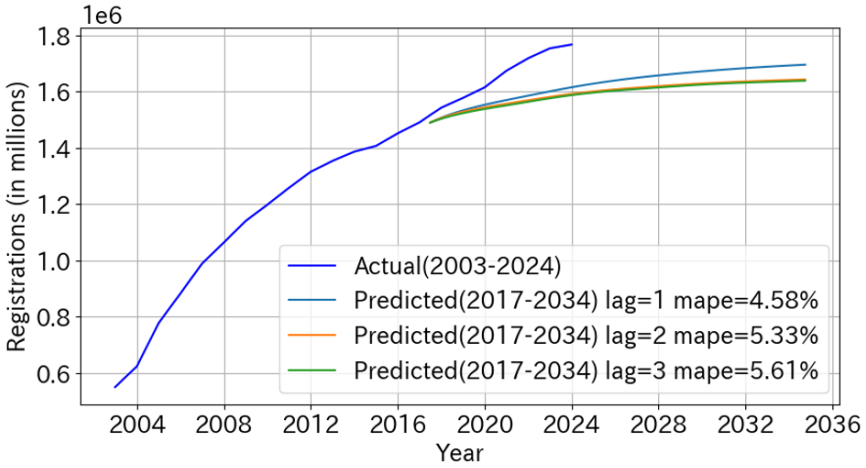
以上のように、PCA による次元削減と四半期相当の線形補間を組み合わせることで、VAR モデルにおけるパラメータ推定の安定性と予測精度の向上を図ることができた。

６．２ 実験６：非定常データ系列の定常化による VAR モデルの安定化

VAR モデルは、理論的にすべての系列が定常であることを前提とした時系列モデルである。定常性とは、平均や分散などの統計的性質が時間によらず一定であることを指し、VAR モデルの安定なパラメータ推定と予測性能の信頼性にとって重要な前提となる。これに対して非定常な系列、すなわちトレンドや構造的変化を含む系列をそのまま VAR モデルに含めた場合、モデルのパラメータ推定が不安定になり、推定値がデータに過度に依存することで過学習やスプリアス（見かけ上の）な相関のリスクが高まることが知られている。

本実験では、VAR モデルの入力として用いる 4 系列（PCA により次元削減された主成分 3 系列および目的変数であるドメイン名登録数系列）について、それぞれの定常性の有無を ADF（Augmented Dickey-Fuller）検定によって個別に評価した。その結果、非定常であると判定された系列に対しては、1 階差分をとることで定常化を行った。差分処理は、系列の傾向成分や長期的トレンドを除去し、系列の平均や分散を安定化させる効果がある。これにより、系列間の構造的な不均衡を解消し、VAR モデルが本来想定する定常性の条件を満たすように前処理が施される。

図－１０はこのような定常化の処理を行った上で JP ドメイン名の登録数予測を行った結果である。



図－１０ 時系列データの定常化処理を行い予測した結果

予測精度の改善を狙った定常化処理であるが、これを行っていない図－９と比べて予測精度は低下してしまった。この結果についての考察は次節で述べる。

７ 実験７：他の ccTLD への適用による予測手法の検証

これまでの実験を通じて、PCA による次元削減と四半期相当の線形補間を組み合わせることで、VAR モデルのパラメータ推定に必要な時系列長を確保しつつ、予測の安定性と精度を向上させることが可能であることが確認された（実験 5)。一方で、非定常なデータ系列に対して ADF 検定に基づく差分処理を行うことで VAR モデルの理論的前提を満たす試み（実験 6）は、JP ドメイン名登録数の推移予測においてはかえって予測精度を低下させる結果となった。

このような結果が、ドメイン名登録数の時系列特性や経済指標との関係性が国（ccTLD）によって異なることに起因するのか、それともモデル構造や前処理手法に固有の性質によるものなのかを検証するために、本実験では複数の ccTLD を対象として同様の予測手法を適用した。その目的は、PCA による次元削減、補間によるデータ拡張、差分による定常化という一連の処理が、他国のドメイン名登録数予測においても有効に機能するかを実証的に評価することである。

表－７に、各 ccTLD に対して VAR モデルを適用した際の予測精度（MAPE）を、定常化の有無で比較した結果を示す。これより、定常化の有無によって予測精度の改善が見られる国と、むしろ悪化する国が存在し、国ごとに異なる傾向が明らかとなった。たとえば、kr（韓国)、fr（フランス)、cn（中国)、it（イタリア）では、定常化により予測精度が大幅に向上した。これは、これらの国のドメイン名登録

表－７ 各実験の MAPE 値の比較

ccTLD	Lag	MAPE (%)	
		定常化なし	定常化あり
jp	1	0.72	4.58
	2	2.37	5.33
	3	2.50	5.61
cn	1	153.20	76.13
	2	20.74	59.44
	3	19.85	62.25
de	1	1.81	5.43
	2	1.17	1.14
	3	1.12	1.66

ccTLD	Lag	MAPE (%)	
		定常化なし	定常化あり
it	1	7.79	8.59
	2	18.18	9.42
	3	21.17	10.41
fr	1	8.04	2.45
	2	9.42	2.58
	3	9.91	2.74
kr	1	25.33	4.70
	2	15.09	4.51
	3	13.61	3.91

数や経済指標の系列において、トレンドや構造的変動といった非定常成分が強く、これが予測性能を妨げていた可能性がある。差分処理によってこれらの影響が除去され、VAR モデルが本来の定常性前提に近い構造で機能するようになった結果、予測の安定性が向上したと考えられる。

一方で、jp（日本）や de（ドイツ）のように差分後に MAPE が増加する国も存在した。これらの系列では、もともと相対的に安定した動きを示しており、差分によってレベルや傾向といった予測に有効な情報が損なわれた可能性がある。VAR モデルの構造上、系列の差分化は自己回帰的な情報構造を変化させるため、場合によっては短期予測精度の低下を招くことがある。

このような結果は、時系列データに対する前処理の効果が、対象系列の統計的性質や対象国の経済構造に大きく依存することを示唆している。つまり、VAR モデルにおける定常性の確保は理論的に重要である一方で、すべての系列に対して機械的に差分処理を適用することは、かえって予測性能を損なう場合がある。したがって、定常性の検定と差分処理は、各系列の特性に応じて柔軟に適用すべきであると言える。

8 考察

本研究では、ccTLD におけるドメイン名登録数の推移を、公開されている経済指標をもとに VAR モデルによって予測する手法を提案し、その有効性を検証した。複数の変数選定・次元削減手法を比較した結果、PCA による次元削減が予測精度とモデルの安定性の両面で最も効果的であった。また、時系列データの線形補間によって学習データ数を拡張することは、VAR モデルが必要とする自由度を確保する上で有効であり、特に年次データしか得られない現実的な条件下では現実的かつ有力な手法であることが明らかになった。

一方で、VAR モデルが理論的に要求する定常性に着目し、差分処理によって定常化を試みた実験では、かえって予測精度が低下するケースも確認された。このことは、時系列データの性質がモデルの仮定と一致しているかどうかを機械的に判断するのではなく、予測対象や実務的な目的に応じた柔軟な前処理が求められることを示唆している。特にドメイン名登録数のように、トレンド成分が予測対象として本質的に重要な場合、その情報を捨ててまで定常性を追求することが適切とは限らない。

また、今回の研究では、予測精度を高めることが主目的であったが、その過程で得られた副次的な知見として、予測と実績の乖離自体に意味がある可能性を考えたい。つまり、モデルが経済指標から導き出す「期待されるドメイン名登録数」の推移と、実際の登録数に大きな差がある場合、それは予測不能な外部要因、あるいはモデルに含まれていない変数の影響が存在することを示している可能性がある。

例えば、ドメイン名登録数は、国や地域の経済活動に加え、ドメイン名レジストリによる料金の変更やキャンペーン施策、新ドメインの解放、行政機関による規制の強化・緩和などの制度的要因によって大きく変動することがある。これらの要因は通常の経済指標には含まれないため、モデルからすれば予測不可能な変動となる。しかし、こうした乖離の発生自体を「異常値」あるいは「注目すべき変化」として捉えることで、施策の影響を定量的に分析したり、レジストリ運営上の重要な意思決定材料としたりすることが可能となる。

すなわち、本研究で構築した VAR ベースの予測モデルは、単に未来を予測するためのツールにとどまらず、「何が通常で、何が異常か」を判断するための基準線（ベースライン）としても活用できる。これは、ドメイン名レジストリの中長期的な計画策定に加え、実施した施策の効果測定や、外的ショック（規制、災害、政策変動など）に対する定量的評価にもつながる。

さらに本研究の枠組みは、他の ccTLD にも適用可能であり、国際的な比較研究の基盤としても有効である。国ごとの乖離傾向を分析することで、インターネット政策の有効性や経済成長との連動性、デジタルインフラの整備状況などを間接的に評価することもできる可能性がある。今後、国際機関や政策立案者による比較分析やモニタリングツールとしての発展も視野に入る。

9 まとめ

本研究では、ccTLD におけるドメイン名登録数の将来予測に対して、公開可能な経済指標を用い、VAR モデルを基盤とした予測手法を構築した。主な取り組みと成果は以下の通りである。

- VAR モデルでの予測に悪影響を及ぼす多重共線性の影響を排除するため、重回帰分析、相関係数、VIF、ランダムフォレスト、主成分分析（PCA）による次元削減手法の比較を行い、PCA を用いた次元削減が最も予測精度とモデルの安定性を両立することを確認した
- 年次データを線形補間することでデータ数の制約を補い、モデルの予測精度を向上させた
- 系列データの前処理として定常化（差分処理）を試みたが、これは予測精度の向上に必ずしも寄与せず、対象系列ごとに慎重な判断が必要であることが示された。
- 他の ccTLD への適用実験により、本手法の一定の汎用性が確認された。国ごとの乖離傾向は、それぞれの経済構造や政策環境の違いを反映している可能性がある。

本研究の成果として、予測手法の構築、予測精度の向上だけでなく、予測と実績の乖離を注視することで、実社会の変化を検出・解釈するためのツールとしての応用可能性が考えられる。この枠組みは、ドメイン名レジストリの経営支援だけでなく、制度設計や国際比較、さらにはデジタル経済のモニタリングにも応用が可能であると考える。

今後の発展に向けて以下の課題を挙げる。

- 使用する経済指標とドメイン名登録数の因果関係を明確化し、変数選定の理論的妥当性を高めること
- ICT 投資や法人設立件数など、ドメイン名登録と関連の深い新たな指標の導入を検討すること
- 登録者の属性に応じたカテゴリ別・セグメント別の予測モデルを構築すること
- 時系列データの自動取得やモデルの定期的再学習を通じたリアルタイム予測体制の構築を進めること
- VECM、Bayesian VAR、LSTM など他の予測手法との比較・統合で、精度と柔軟性を高めること
- モデルの予測と実績の乖離を活用し、異常検知や政策効果の分析ツールとして展開すること
- 他国の ccTLD への適用を通じて、ドメイン名登録と経済・制度の関係を国際的に比較すること

本研究は多くの実務的・学術的可能性を有しており、今後の研究ではこれらの課題を段階的に解決することで、より実用的かつ汎用的な予測・分析基盤として成熟させていくことが期待される。

［参考文献］

[1] Mockapetris, P. (1983). RFC 883: Domain names - Implementation and specification. Internet Engineering Task Force. Retrieved from https://www.ietf.org/rfc/rfc883.txt

[2] Mockapetris, P. (1987). RFC 1034: Domain names - Concepts and Facilities. Internet Engineering Task Force. Retrieved from https://www.ietf.org/rfc/rfc1034.txt

[3] Mockapetris, P. (1987). RFC 1035: Domain names - Implementation and specification. Internet Engineering Task Force. Retrieved from https://www.ietf.org/rfc/rfc1035.txt

[4] Mueller, M. (2002). Ruling the Root: Internet Governance and the Taming of Cyberspace. Cambridge: MIT Press.

[5] National Research Council. (2005). Signposts in Cyberspace: The Domain Name System and Internet Navigation. Washington, DC: The National Academies Press.

[6] K. Hubbard, M. Kosters, D. Conrad, D. Karrenberg, J. Postel (1996). RFC 2050: Internet Registry IP Allocation Guidelines. Internet Engineering Task Force. Retrieved from https://www.ietf.org/rfc/rfc2050.txt

[7] 宇井隆晴, 池原翔太, 森健太郎, 尾崎剛, 広瀬啓雄. (2024). コーホート分析による JP ドメイン名登録数の推移予測. 情報社会学会誌, 19(1), 77-90.

[8] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed.). Wiley.

[9] Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice (3rd ed.). OTexts. https://otexts.com/fpp3/

[10] Shahin, M. A., & Wahab, M. H. A. (2019). Forecasting monthly electricity consumption using SARIMA

model: A case study in Egypt. Energy Reports, 5, 1405-1412. https://doi.org/10.1016/j.egyr.2019.10.028
[11] Alzahrani, S. I., & Mousa, H. (2021). Time series analysis and modeling to forecast: A survey. Journal of King Saud University - Computer and Information Sciences, 33(6), 695-705. https://doi.org/10.1016/j.jksuci.2018.09.014
[12] 上村圭介. (2013). 国別トップレベルドメイン名の利用促進要因の推定と統治体制の特徴抽出. 情報社会学会誌, 7(2), 23-40.
[13] Changkyu Choi, Myung Hoon Yi. (2009). The effect of the Internet on economic growth: Evidence from cross-country panel data. Economics Letters, 105(1), 39-41.
[14] Matthieu Pélissié du Rausas, James Manyika, Eric Hazan, Jacques Bughin, Michael Chui, and Rémi Said. (2011). Internet matters: The Net's sweeping impact on growth, jobs, and prosperity. McKinsey Global Institute. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/internet-matters
[15] Noorhamizah Abdul Wahab, Nurul Shuhada Salleh, & Faridah Syed Alwi. (2020). Internet user and economic growth: Empirical evidence from panel data. Journal of Emerging Economies and Islamic Research, 8(3), 17-25.
[16] Y. Morikawa. (2020). Non-stationary time series analysis on the effects of monetary policy using VAR model (Doctoral dissertation, Meiji University). Meiji University Institutional Repository. https://meiji.repo.nii.ac.jp/records/14790
[17] Ghorbani, M., & Chong, E. K. P. (n.d.). Stock price prediction using principal components. Department of Systems Engineering, Colorado State University. https://journals.plos.org/plosone/article/authors?id=10.1371/journal.pone.0230124
[18] T. Hayashi. (2015). Long-term prospects for the Japanese economy (APIR Discussion Paper Series No. 39). Asia Pacific Institute of Research. https://www.apir.or.jp/uploads/files/DP39_jp_lonterm_prospects1.pdf
[19] Hamilton, J. D. (1994). Time series analysis. Princeton University Press.
[20] Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. Journal of Economic Perspectives, 15(4), 101-115. https://doi.org/10.1257/jep.15.4.101
[21] Sims, C. A. (1980). Macroeconomics and reality. Econometrica, 48(1), 1-48. https://doi.org/10.2307/1912017
[22] JPRS. (n.d.). Number of JP domain name registrations. Japan Registry Services Co., Ltd. Retrieved November 1, 2024, from https://jprs.jp/about/stats/domains/
[23] World Bank. (n.d.). *World Development Indicators*. The World Bank. Retrieved November 1, 2024, from https://data.worldbank.org/
[24] Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679-688.
[25] Armstrong, J. S. (Ed.). (2001). Principles of forecasting: A handbook for researchers and practitioners. Springer.
[26] Sims, C. A. (1980). Macroeconomics and reality. Econometrica, 48(1), 1-48. https://doi.org/10.2307/1912017
[27] Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. Journal of Economic Perspectives, 15(4), 101-115. https://doi.org/10.1257/jep.15.4.101
[28] Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
[29] Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). Applied linear statistical models (5th ed.). Boston, MA: McGraw-Hill Irwin.
[30] Gujarati, D. N., & Porter, D. C. (2009). Basic econometrics (5th ed.). McGraw-Hill Education.
[31] Robert M. O'brien (2007). A caution regarding rules of thumb for variance inflation factors. Quality & Quantity, 41(5), 673-690. https://doi.org/10.1007/s11135-006-9018-6
[32] Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). Applied linear statistical models (5th ed.). Boston, MA: McGraw-Hill Irwin.

[33] O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. Quality & Quantity, 41(5), 673-690.
[34] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.
[35] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.
[36] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.
[37] Jolliffe, I. T. (2002). Principal component analysis (2nd ed.). Springer.
[38] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202.
[39] Lütkepohl, H. (2005). New introduction to multiple time series analysis. Berlin: Springer.
[40] Hamilton, J. D. (1994). Time series analysis. Princeton, NJ: Princeton University Press

(2025年8月27日受理)