

経済学と反事実分析 接触篇

Economics and Counterfactual Analysis: A Contact

ill-identified

Tokyo.R 83rd, 2020/1/25, Updated: 2020/2/23

目次

1	概要	3
2	イントロダクション	3
2.1	構造推定とは	3
2.2	因果推論と何が違うのか	4
2.3	今回取り組む問題	5
3	どう推定するか	6
3.1	最も古典的な方法	6
3.2	なぜ生産関数を知りたいのか	7
3.3	パネルデータ分析と観測されない効果	8
3.4	動学パネルデータ分析	9
3.5	企業内部の意思決定を考える	10
3.6	OP モデルの回帰モデルとしての意味	12
3.7	モデルの事後診断	14
3.8	OP 法以降の研究	15
4	R での実装	15
4.1	データのとり方	15
4.2	β_L, ϕ の識別	16
4.3	存続確率の推定	17
4.4	β_A, β_K の推定	18
5	結果	20
6	まとめ	21
	参考文献	22

補遺	25
A 数値計算上に関する実装上の注意	25
B ブートストラップ標準誤差	26
C kawaguchi の課題についての補足	28

1 概要

第 83 回 Tokyo.R のエントリ. 前回の戦後処理に時間を取られたので, 手抜き回.

前回にも言及したIgami (2018) の研究に関連して, **構造推定**についてももう少し詳しい話をしようと思っていたが, 残り時間の問題から**動学構造推定**は難しいと判断した. しかしアンケート至上主義なので経済学のネタは外せないから, 準備の簡単な**静学構造推定**の話を取りあえずやることにした. 静学構造推定は「動学」ではないので時間変化を考えないタイプの理論モデルを推定する一連のテクニックである. 今回の話は比較的簡単かつ昔から知られている話だから大学の経済学部でも授業で取り上げられることもあろう. 知ってる人も多いかもしれない.

具体的には, 生産関数を構造推定する OP 法 (Olley and Pakes, 1996) を R でやる.

2 イントロダクション

2.1 構造推定とは

では構造推定とは何だろうか. 構造推定で検索すると, 名前が似ているが違うものがいくつか出てくる.

■**構造方程式モデリング** 例えば構造方程式モデル (SEM) というものがある. これは変数間の関係構造を同時線形方程式で表したもので, 因果推論のフレームワークの原型ともいえるものだが, **構造推定とは違う**.

■**構造時系列モデル** 構造ベクトル自己回帰モデル (Structural VAR) とか, **ベイズ構造時系列モデル (BSTS)** というものがある.

通常の VAR は複数の変数の時系列変化を扱うモデルで, それぞれの変数が過去の値にのみ依存して決まる回帰モデルなのに対して, 構造 VAR は現在の別の変数にも依存する, **同時決定的な時系列モデル**である^{*1}. ベイズ構造時系列モデルの「構造時系列」は Harvey のいう構造時系列モデルのことを指す (Scott and Varian, 2014). これは 89 年とかなり古い教科書なので私も読んだことがないが, Lütkepohl (2007, p. 618) によればローカルトレンドモデルを基本とした時系列モデルらしいので, 多変量時系列モデルというわけでもなく, 状態空間モデルに属する. これらも**構造推定とは関係ない**.

■**構造型モデル** いちおう過去の Tokyo.R の発表ネタとかぶらないか確認したところ, こんなものを見つけた.

R で学ぶ『構造型モデル de 倒産確率推定』

「構造型モデル」という名前も構造推定と似ている. いちおう元になったMerton (1974) の論文も確認した. これも**構造推定ではない**.

構造推定の定義について, 多くの人はReiss and Wolak (2007) の紹介を挙げるが, この本はとにかく分厚くて高いので私は持っていない. 前回で触れたように, 観察されたデータにそのまま回帰式や他の関数を当ては

^{*1} 簡単な定義と特徴は<https://qiita.com/saltcooky/items/2d0119ea4a10bab6cff2>の最後のほうに書いてある. より詳しい話を知りたいなら, 沖本 (2010) が言及している. インターネットの日本語圏で出回っている情報の多くはこのあたりからだと思うが, もっと多変量時系列モデルを知りたいならLütkepohl (2007) の教科書がある (この本は約 700 ページの間, タイトル通りマジで多変量時系列モデルのことしか書いてない). 時系列モデルに関する本は北川 (2005) の教科書を始め和書でも多く出ているが, 経済学や数理ファイナンスの観点から書いたもの以外では, 構造 VAR について言及したものは私の知る範囲では見られない. 構造 VAR という考え方自体がパラメータの識別を前提としているからだろう. マクロ経済学における時系列モデルの利用を知りたいのなら, 渡部 (2016), 中島・渡部 (2012) が参考になる. 本格的な教科書は専門外なので知らない.

めても因果関係を知ることはできない。そこで、いくつかの要素が必要になる。そこで構造推定は仮定によって構造を特定しようとした。

北村 (2016) は一般的な型はないとしつつ、その解説を要約すると3つの特徴があるとしている。

1. 経済理論モデルであること
2. 確率モデルであること
3. データに基づいて推定されていること

(1) はいわゆるルーカス批判 (Lucas, 1976) を受けて、彼の批判を克服できるような経済理論に基づいたモデルづくりを意識するようになった。(2) はちょっとわかりづらい。回帰分析だって正規分布を暗に仮定しているから確率モデルになる。しかしここでは、観察できないものと、できるものの不確実性を考慮しているという意味である。例えば測定誤差や、分析者には観測できないが、経済活動の当事者は知っている情報というものがある。これらを見逃した推定はしばしばバイアスを引き起こすが、構造推定ではモデルにこのような仮定を取り込む。(3) は、完全な架空の話ではなく、なるべく現実と整合するために観測されたデータに基づいてパラメータを決定するということである。

よって、最初に挙げた「構造方程式モデリング」や「Merton の構造型モデル」は経済理論に基づいた構造を表現していないため構造推定ではない。ただし構造時系列モデルは、実は 80 年代にマクロ経済モデルとして利用された歴史がある。経済理論に基づかないという批判は当時からあったものの、単に予測するだけなら問題ないということで使われていた。しかし、「どうなるか」と「なぜそうなるのか」のどちらを求めているかと言えば前者であり、原因を分析することはできない^{*2}。

一般に定まった方法はないが、かといって完全な無手勝流ではない。大学院のコアコースで教えるモデルはだいたい決まっており、多くの研究も先行研究のフレームワークを応用していることが多い。

2.2 因果推論と何が違うのか

以前にも紹介したような RCT, 準実験 (操作変数法, 傾向スコア法, DID, RDD) なども、「反事実的な因果推論」と呼ばれる。しかしこれらは経済学の理論仮説に基づく因果関係ではなく、ランダム化と比較によって「平均処置効果」を求めるフレームワークである^{*3}。一方で構造推定は実際のデータと理論モデルを融合させた結果なので、本来の意味で現実には起こり得なかったこと (反事実) を推論するフレームワークである。

具体例を挙げよう。今回は経済学における生産関数の推定の例を紹介する。原材料を投入し、財を生産する、その変換ルールを関数に表したものを生産関数という。例えば、資本 (Capital^{*4}) と労働力 (Labor force), Y は出力、つまり生産量なので、生産関数を以下のように表せる

$$Y = F(K, L)$$

この生産関数の形状が分かれば、

- 実際のデータとの残差から、数値に表れない生産性 (TFP) を計測する (残差分析)

^{*2} そのため、VAR でも因果関係を定義できるということで Granger 因果性が注目を集めた。つまり、以前の投稿で問題にした、予測か因果かという二択は、経済学の発展のなかでもしばしば議論されてきた問題だということがわかる。

^{*3} 正確には少し違うが、重要なのはサンプルの平均的な差を見るという特徴が共通している

^{*4} ドイツ語でのスペル。C にしてしまうと消費, Consumption と重複するからか、昔からこの記号が使われている。また、正確には「資本ストック」であり、設備投資や固定資産などから計算する

- もし設備投資を増やしたら (K を増加させたら) どれくらい生産量が増えるか
- もしパラメータが変化したら, 生産力はどう変化するか (比較静学分析)

と言ったことがわかる。経済学部出身者ならば、ミクロ経済学の授業で聞いた覚えがあるだろう。しかし、データから推定した係数が「もし変わったら」というのは、本質的に現実に起こっていない事象について話していることになる。同等のことを因果推論でやるとすれば、両方のケースについて実際に観測したデータに基づいて差を比較しなければならない。つまり、単にデータを近似しただけの関数では、これらのことを言う根拠として弱い。

経済学では RCT や準実験アプローチは誘導形 (reduced form) 推定^{*5}とも呼ばれ、構造推定派と誘導形推定派は数年前までお互い相手のアプローチの問題点を指摘し激しく論争していた。その結果、現在ではそれぞれのアプローチでできることの限界がはっきりしてきたため、論争というほどのことはない^{*6}。

両者の使い分けとしては、構造推定は経済理論モデルというある意味「架空のもの」に基づいて未来に起こりうることについての分析もできるが、因果推論は実態としては実験計画法と本質的に同じで統計学の理論に基づいて観察できたデータから事後評価するアプローチである。よってしばしば、構造推定は反事実 (反実仮想) シミュレーション、因果推論はプログラム (政策) 評価とか介入 (intervention) といっって呼び分けられる^{*7}。そして構造推定は理論仮説に強く依存したフレームワークであるので、仮説が妥当なのかはよく注意が必要である。ただし、かといって因果推論が仮定に依存しないというのも完全な誤りである。前回でも言及したが、因果推論であっても必ず仮定が存在するので、それを無視して計算した平均値の差が因果効果を意味するとは限らない。

2.3 今回取り組む問題

動学構造推定はテクニカルでなんかすごいことをやってるという印象を簡単に与えられる。しかし、今回は時間がないのでもう少し簡単にできるほうをやる。前回話題にしたのは「動学的」構造推定であり、時間変化を取り入れたモデルを扱うジャンルである。一方で、今回は考慮しない「静学的」構造推定の具体例を紹介する。といっても昔からある研究なので、すでに日本語の講義スライドもネット上に転がっていたりするし、もしかしたら発表者の中にも授業でやったことがあると言う人がいるかもしれない。

今回は特に、生産関数の推定の問題を例にする。生産関数をどういう形にするかはいろいろだが、よく使われるものの1つにコブ = ダグラス型生産関数がある^{*8}。特に、一企業の実業関数を

$$Y = f(K, L) := \alpha K^{\beta_K} L^{\beta_L}$$

という指数関数で表す^{*9}。 α, β_K, β_L はパラメータである。 $\beta_K = \beta_L = 0.5$ の場合、図1のような形状になる。

^{*5} 構造推定に使うような、経済現象の構造を表したモデルを、平均の差の比較という因果推論のフレームワークで推定できるように単純な形に変形することからこう呼ばれる。この式変形のため、根本的な構造パラメータを特定しづらいという問題がある。

^{*6} 歴史的経緯は北村 (2016)、中嶋 (2016) などが少しだけ解説している。計量経済学の観点では統計モデルとして Hayashi (2000, Ch. 8) が言及している。また、Cameron and Trivedi (2005, Ch. 2) は教科書の最序盤で構造モデルと誘導モデルの特徴付けを行うという独特の構成になっている。

^{*7} 後者の因果推論フレームワークも、もともとは counterfactual causal model と呼ばれていたが、(おそらくはこの性質の違いのため) 少なくとも経済学の分野では最近あまり強調されていない気がする。

^{*8} 他にも色々あるが、例えば物理学のように法則がかなり明確になっている分野を勉強した人間からはおかしなことに聞こえるかもしれない。しかし、経済学ではむしろ、変に特定化すれば特定化の誤りを免れない。

^{*9} 生産関数のより詳しい数学的性質はミクロ経済学の教科書に書かれている。例えば、特に有名なのは西村 (1990)、『ミクロ経済分析』(Varian, 1992)、Mas-Colell et al. (1995) などである。それ以降にも評価の高いミクロ経済学の和書がいくつかでていますが、私はあまり読んでいないのでなんとも言えない。

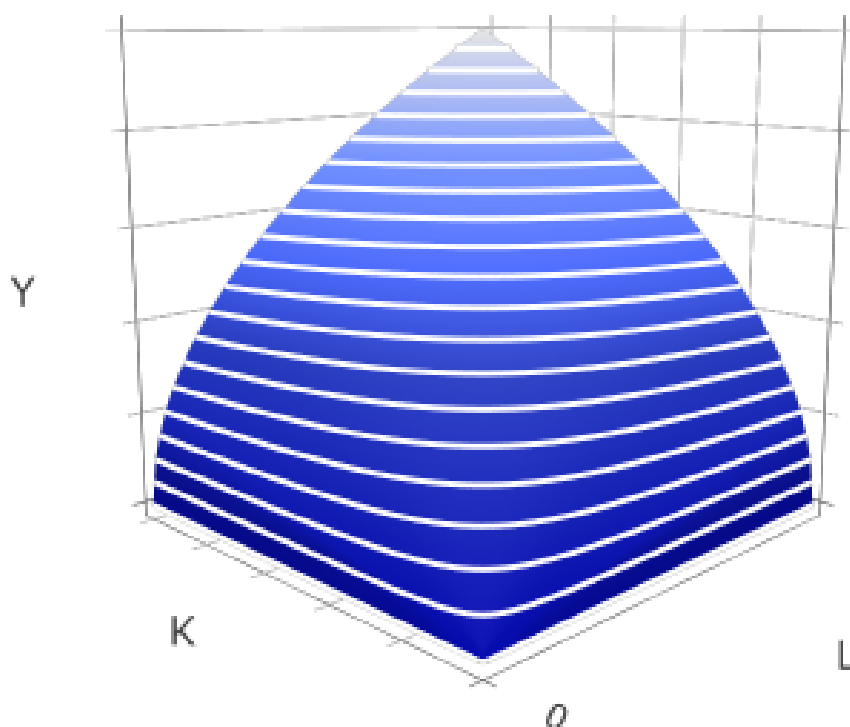


図1 コブ＝ダグラス型生産関数の例

3 どう推定するか

3.1 最も古典的な方法

経済学部 of 計量経済学の講義では、**回帰分析**で生産関数の推定をやるところも多いと思う。前節の生産関数は、 Y を**被説明変数**、 K, L を**説明変数**としても線形式ではない。しかし、 Y, K, L がいずれもゼロ以下になりえないことに着目すると、対数を取って以下ようになる（いわゆる**対数線形回帰モデル**）。

$$\ln Y = \ln \alpha + \beta_K \ln K + \beta_L \ln L$$

となる。つまり、**おなじみの最小二乗法 (OLS) で計算できる**。観測できなかった α も、切片パラメータとして推定できることになる。以降はそれぞれの対数を y, β_A, k, l に置き換えて、

$$y = \beta_A + \beta_K k + \beta_L l + \varepsilon \quad (3.1)$$

という線形回帰モデルを考える。しかし以前書いたように、経済学の文脈では非線形だろうが線形式だろうがデータに関数を当てはめただけでは用をなさない。経済理論に即して、発生しうる問題を排除した方法で推定しなければならない。つまり、現在では OLS では生産関数のパラメータを**推定するには不十分**と考えられている。

発表時間の問題から、スライドでは大幅に省略しているが、ここでは冗長気味にパネルデータ分析の発展の経緯も合わせて説明していく^{*10}。

現在の経済学の観点からすると、(3.1)をそのまま OLS で推定するのをためらう大きな理由は以下の 3 つである。

1. 企業特有の効果が存在する
2. 需要要因を考慮していない
3. セレクションバイアス

3.2 なぜ生産関数を知りたいのか

しかし、テクニカルなことを説明する前に、そもそもなぜこのようなことをするのかを表明したほうがよからう。今回、最後に紹介するのは [Olley and Pakes \(1996\)](#) らの研究で、彼らは法規制が産業の生産性にどう影響するかを知ることを研究目的としていた。では、生産性なるものをどう計測すればいいのだろうか。

経済学には全要素生産性 (TFP; total factor productivity) という概念がある。これは、投入したリソースの量では説明できない生産量の多さを生産性と定義した概念で、実際には推定した生産関数による生産量の予測値 (理論値) と実測値の誤差で計測する。例えば今回はコブ = ダグラス型関数で、モデルの仮定から、物的・人的リソースによる生産量は $K^{\beta_K} L^{\beta_L}$ の部分になる。よって TFP を計算する式は、

$$TFP := \frac{Y}{K^{\beta_K} L^{\beta_L}}$$

となる。もし TFP を毎年計測して値に変化があるのなら、それは技術進歩に由来する生産性の向上であると見なせる (正確にはこれはソロー残差とか成長会計モデルとか呼ばれる)^{*11}。

さらに、(3.1)からも分かるように、対数の性質には $\ln(x/y) = \ln x - \ln y$ というものがある (高校数学でも扱うはずだ)。よって、仮に(3.1)の係数パラメータ β_K, β_L を求められれば、

$$\begin{aligned}\ln(TFP) &:= y - \hat{\beta}_K k - \hat{\beta}_L l \\ \therefore TFP &= \exp(y - \hat{\beta}_K k - \hat{\beta}_L l)\end{aligned}$$

となる。よって回帰分析で推定した式を使うと、TFP は α だけでなく誤差項も含まれることになる。TFP は回帰モデルの残差に依存する。

ではここで、当てはまりの良さのために生産関数の形状を変えたらどうなるだろうか？ 当てはまりを良くするのなら、いろいろなモデルを試していくらでも生産量の予測誤差の少ないモデルを作り出すことができる。しかしそこから計算した残差は一体何を意味するのだろうか？ 機械学習や統計モデリングというと、MSE とか対数損失がどれくらい小さいかというのがモデルの良さの指標だと語られることがしばしばあるが、しかしこのようにそれが活用方法としての全てではない。

そして TFP や生産性といった概念は、このように生産構造をどう仮定するかに強く依存することに注意する。分析にはかならず「ある仮定のもとで...」という前置きが存在する。「生産性」という抽象的な概念が何を

^{*10} こういった初歩的な話は、私なんかよりよほど深く理解した大学の先生がいくつも教科書を書いており、正確さや詳しさでそちらのほうが遥かに良い。参考文献には一般公開されている [奥井 \(2016\)](#) のスライド、[Wooldridge \(2010, Ch. 10-11\)](#) の教科書などがある。私はほぼ Wooldridge の教科書で勉強したため、日本語の教科書に詳しくない……

^{*11} TFP の計測方法やその批判についてはググれば内閣府とか日銀とか経済産業研究所 (RIETI) の公開している日本語の論文がたくさん引っかかるので適当に勉強すれば良いだろう。

企業 ID	時期	生産量	資本	労働力
001	2019Q1	100	10	10
001	2019Q2	110	11	10
001	2019Q3	90	12	12
001	2019Q4	120	14	13
002	2019Q1	80	5	5
002	2019Q2	90	6	5

表1 パネルデータの例

意味するのか、何を言っているのかはよく考えるべきであろう^{*12}。

3.3 パネルデータ分析と観測されない効果

「企業特有の効果」とは、特定の企業だけが「特殊な技術を持っている」「優秀な人材が集まっている」など、単純な量だけでは捉えきれない情報を意味する。つまり、企業ごとに異なる ω_i が生産量に影響していると考えられる。

$$y_i = \beta_A + \beta_K k_i + \beta_L l_i + \omega_i + \varepsilon_i$$

ω_i は観測できないので、このまま (y_i, k_i, l_i) のデータで回帰すると、 ω_i は定数項 β_0 または誤差項 ε に吸収される。 ω_i が k_i, l_i と全く相関しない変数ならば、 β_A と ω_i を識別できないものの、少なくとも β_K, β_L の推定はできる。

このような問題をどう解決すれば良いだろうか？ まずは問題 1 だけを考えよう。 ω_i は観測できないが、では推定できないだろうか？ つまり、同一の企業の情報を何度も観測しすれば、そこから推定できないか、というアイデアである。これがパネルデータ分析のスタート地点になる。パネルデータとは、同一個体を追跡して複数時点で観測したデータである(表1)。よって、企業 $i = 1, \dots, N$ ごとに、 $t = 1, \dots, T$ の観測がある。そこで以降は回帰モデルを以下のように書く。

$$y_{i,t} = \beta_A + \beta_K k_{i,t} + \beta_L l_{i,t} + \omega_i + \varepsilon_{i,t} \quad (3.2)$$

ここではひとまず、 ω_i が企業 i ごとに異なるが、時間 t に対しては変化しないと考えておこう。また、問題 1 だけを考えるにあたって、 ω_i は $k_{i,t}, l_{i,t}$ と相関しないとしておこう。このような ω_i を観測されない個別効果(unobserved individual effect)という。

さらに追加の仮定として、 ω_i が定数である場合、ダミー変数を作成することで OLS であるかのように ω_i を推定できる^{*13}。これを LSDV 推定量という。また別の解き方としては、(3.2)を企業ごとに平均を取った場合を考える。

$$\bar{y}_i = \beta_A + \beta_K \bar{k}_i + \beta_L \bar{l}_i + \bar{\omega}_i + \bar{\varepsilon}_{i,t} \quad (3.3)$$

^{*12} 私が過去、残差を使って何かを言おうとする「データ分析」のやり方に対して、例えば『CausalImpact でできること、できないこと』で書いたように注意を喚起してきたのはこのように、強い仮定に依存しているにもかかわらず仮定の妥当性を確認せずに濫用されることを懸念したのが理由である。

^{*13} ただし、一致するのは期待値のみで、標準誤差の計算方法は変わる。仮説検定・区間推定をするならば clustering-robust な標準誤差と呼ばれるものを使う。

このとき、 ω_i は企業ごとに一定なので、 $\bar{\omega}_i = \omega_i$ である。よって、(3.2)から(3.3)を引いてから OLS で計算することでも同じ推定結果が得られる。こちらの方法は**固定効果 (FE; fixed effect)** 推定量または**グループ内 (within)** 推定量と呼ばれる^{*14}。

一方で、 ω_i が定数項ではなく確率変数の可能性もある。つまり、企業ごとに異なる ω_i は**実は期待値をとればゼロになる確率変数にすぎず**、定数項に見えたものは確率的なゆらぎに過ぎないかもしれない。この仮定の元に考案されたのが**変数効果 (RE; random effect)** モデルである。RE モデルは数学的に**分散不均一な回帰モデル**と同一のため、実質的に**一般化最小二乗法 (GLS)** と同じである。^{*15}

3.4 動学パネルデータ分析

このように、観察できない要因があってもパネルデータがあれば理論上は推定が可能になる。しかし、Mairesse (1990)^{*16}はこれまで紹介した方法とほぼ同じやり方で企業の生産関数を推定したところ、観測されない効果を識別しようと複雑に差分を取れば取るほど、係数の推定値が小さくなることを発見した。彼はこの原因の仮説をいくつか挙げている。

1. 内生性の問題。個別効果 ω_i は物理的なものではなく、環境、経営者の手腕といったものに由来するのなら、他の説明変数 k, l の決定に影響している。
2. 企業の意思決定のラグ。経営者が誤差項として現れる短期的な変化、つまり直前の $\varepsilon_{i,t-1}$ に対応して $k_{i,t}, l_{i,t}$ を決定しているなら、誤差項と説明変数が相関しており、単純な差分では推定できない。
3. 同時決定の問題。
4. 測定誤差。各変数に測定誤差があれば、差を取ることで打ち消される。

これらの問題への対応として考えられるのが、**操作変数**である。

(3.2)に、被説明変数 $y_{i,t}$ の自己回帰項を追加する。

$$y_{i,t} = \beta_A + \beta_K k_{i,t} + \beta_L l_{i,t} + \rho y_{i,t-1} + \varepsilon_{i,t} \quad (3.4)$$

こういう自己回帰項を含むモデルを扱う分野はパネルデータ分析のなかでも特に**動学パネルデータ分析**と呼ばれる。このあたりの問題も奥井 (2016) や Wooldridge (2010, Ch. 11) で丁寧に説明されている^{*17}ので冗長だが、一応簡単に述べておく。

前節のアイディアを流用すれば、Within 推定量の応用で時間平均 $\bar{y}_i = (T-1)^{-1} \sum_{t=2}^T y_{i,t}$ を取ることが考えられる。しかし、パネルデータでは N が大きい T は小さいことが多いので、一致推定量とみなせない。一方で、**1 階差分 (FD)** 推定量というものがある。これはある意味 **DID 推定量** と同じである。差分演算子を

^{*14} 逆に観測されない効果が時間ごとに異なるが、企業間では一定の場合も同じようにできるだろう。その場合は**グループ間 (between)** 推定量と呼ばれる。

^{*15} 久保 (2012) は生物学の文脈で異質効果のある場合の解法として**一般化線形混合モデル (GLMM)** を紹介している。これは観測されない効果の問題とよく似ている。生物の実験データではおそらく変数との相関がないため、定数項のバイアスと、当てはまりの良さだけが問題となるのだろう。

^{*16} 古い本なので大学図書館などでないと閲覧が難しそうだ。私も手元にないので学生の頃書いたメモを元にも書いている。また、川口先生の講義ノートではこのあたりの問題に関する参考文献として Griliches and Mairesse (1999) の論文を挙げている。タイトルからしてもこちらのほうがカバー範囲の広いサーベイのようなので、こちらを参照したほうがいいのかもかもしれない。

^{*17} GMM も以前ブログに定義を書いたが、詳しいことは Hayashi (2000)、また、他の教科書として動学パネルデータを重点的に解説した千木良他 (2011) などがある。Baltagi の教科書も授業でおすすめされていた気がするが私は読んでいないのでどこまで書いてあるか知らない。どうしても本を買いたくない/買えない人は奥井亮『**動学的パネルデータモデル**』の講義スライドを見ると良い。

$\Delta x_t := x_t - x_{t-1}$ というふうに定義する. (3.4)の両辺から $y_{i,t-1}$ を引く.

$$\Delta y_{i,t} = \beta_K \Delta k_{i,t} + \beta_L \Delta l_{i,t} + \rho \Delta y_{i,t-1} + \Delta \varepsilon_{i,t}$$

しかし, $\Delta y_{i,t-1} = y_{i,t} - y_{i,t-1}$ から, $\Delta \varepsilon_{i,t} = \varepsilon_{i,t} - \varepsilon_{i,t-1}$ と相関が存在する. そこで, $\varepsilon_{i,t-1}$ より過去の $y_{i,t-2}, \dots$ を操作変数として使う方法が提案されている. それが Anderson-Hsiao (1981; 1982) モデルや, (Arellano and Bond, 1991; Blundell and Bond, 1998) の GMM 推定法である (いわゆる階差 GMM とシステム GMM). しかし, これらは実際に操作変数として意味のある変数なのかということを深く議論せずに $\{y_{i,t-2}, \dots\}$ を操作変数に利用している. その場合, 上記で挙げたような意思決定が存在するかどうかを検証できないし, これらの推定量には過剰識別や弱相関操作変数の問題が発生することが指摘されており, 妥当性に疑問が残る.

3.5 企業内部の意思決定を考える

しかし, 経済学の理論では, 観察される経済データは市場が需要と供給のバランスを取った結果である. これが2つ目の「需要要因を考慮していない」という話になる. 企業が需要を全く考えずに事業計画を立てるとは考えにくいので, この ω_i が間接的に k_i, l_i に影響している可能性は十分ある. また, 企業は新規立ち上げだったり倒産したりする. 現実のパネルデータはしばしば欠落がある. そのようなデータでそのまま推定しようとしたらどうだろうか.

このようなパネルデータでの内生性に対処する推定方法に関する研究はいくつもあり, 特によく出てくるものとして FEIV モデル, がある.

さらに, Olley and Pakes (1996), Levinsohn and Petrin (2003) はこれらの動学パネルデータ分析の方法から一歩進んで, 企業の意思決定の構造をモデル化した生産関数を考え^{*18}, 操作変数を使わずに内生性の問題に対処した方法を提案している (一般には OP の方法はコントロール関数アプローチと呼ばれる)^{*19}.

そこで今回は Olley and Pakes (1996) のフレームワークを詳しく説明する. なお, 以降は(3.2)式をもとに説明するが, 各企業で同じことが言えるため i を省略して書く. OP では, 観察されない効果 ω_t は企業の経営者は知ることができるが, データとして現れないため分析者には観測できない, 企業固有の生産能力を表す成分である. ω_t は他の変数の影響を受けない (外生的) が, 自身の過去の値からは影響を受けると仮定する. つまり 1 次のマルコフ過程であるとする.

$$\omega_t \sim P(\omega \mid \omega_{t-1})$$

t 期の始めに経営者は ω_t を観測して, これをもとに今期の経営計画を決定, つまり, 純キャッシュフローの割引現在価値を最大化できるように k_t, l_t を決めていないと仮定する.

そこで, 動学的なモデルを想定する. 企業の t 時点現在での収益を $\pi_t(\omega_t, k_t, l_t)$ とし, 投資 inv_t に対する費用を $c(inv_t)$ とする. 経営者が最適化する現在の価値関数 (value function) を $V_t(\omega_t, k_t, l_t)$ とすると^{*20}, 動学モデルでは単に現在の価値関数を最大化することが最適化に繋がるとは限らない. 経営者は, これらを把握した

^{*18} より正確には, Olley らは工場単位でのデータを利用しているため彼らが推定したのは工場単位での生産関数である.

^{*19} 日本語では, たとえば北村他 (2009) が理論的根拠を詳しく解説している. 彼らの解説で使われているのは本来のものに加えて原材料費や純資産を考慮したモデルだが, 本質的な部分は変わらない. 統計モデルとしての説明なら彼らのものでも十分だが, 今回は経済モデルとの関係を強調したいため, あえて Olley らの論文をなぞって冗長に書いておく.

^{*20} Olley らの当初の仮定では, 企業年齢も生産性に影響があると考えていたが, 実際に推定してみるとほとんど影響がないことがわかった. モデルの特徴を説明する上でも企業年齢はあまり関与しないため, ここでは省略している.

上で、割引現在価値を最大化するように行動する。つまり、以下のような V^* の達成を目指す。

$$V^* = \max_{\{k_t, l_t\}} \sum_{t=0}^{\infty} \delta V_t(\dots)$$

これは無限和なのでそのまま計算するのは難しいが、一定の条件下では上記の最適化問題の解が以下のよう
なベルマン方程式 (Bellman equation) の解と同じであることが分かっている^{*21}。

$$V_t(\omega_t, k_t, l_t) = \max_{k_t, l_t} \left\{ \Phi, \sup_{i_t \geq 0} \pi_t(\omega_t, k_t, l_t) - c(inv_t) + \delta E[V_{t+1}(\dots)] \right\} \quad (3.5)$$

ベルマン方程式が意味するところは、 $t, t+1$ という 2 時点間の関係さえ分かれば上記の無限和の問題も解ける
ということである。

ただし、このモデルでは 2 つの仮定を追加していることに注意する。1 つは、(3.5) には、 Φ という選択肢がある
ことだ。これは退出、つまり工場の閉鎖や事業を売却することで得られるキャッシュフローである。よって、
このモデルでは採算が合わないと判断した経営者が事業撤退を選択する可能性も考慮している。もう 1 つは、
動学的な決定の必要があるのは k_t のみで、 l_t は每期独立して決定されるように仮定しているという点である。

では、 k_t の動学的な意思決定がどのようになされるかを考えていく。資本 k_t は、前期のストック k_{t-1} と、減
価償却率 δ 、そして新規の投資額 inv_{t-1} の和で表現できる。

$$k_t = inv_{t-1} + (1 - \delta)k_{t-1}$$

企業の退出ルールと投資関数を以下のように定義できる。 t 期に企業が存続するなら $x_t = 1$ 、そうでないなら
0 として、 x_t は以下のように決定される。

$$x_t = \begin{cases} 1 & \text{if } \omega_t \geq \underline{\omega}_t(k_t) \\ 0 & \text{otherwise} \end{cases}$$

$$inv_t = inv_t(\omega_t, k_t) \quad (3.6)$$

ここから、企業が t 期も退出せず操業を続ける確率は

$$p_t := P(x_t = 1 \mid \underline{\omega}_t(k_t))$$

$$= P(x_t = 1 \mid k_t, \omega_t, \underline{\omega}_t) \quad (3.7)$$

^{*21} これをベルマンの最適性原理という。しかしこれ以降の動学最適化問題の数理的に厳密な話をしているのはキリがないので、適当な
参考文献に投げることにする。ただし私は動学最適化を純粋に数理的側面で説明した教科書を読んだことはなく、経済学の文脈で
主に勉強しているので、挙げる参考文献も必然的にそういうものになる。日本語文献でかつ無料で一般公開されているものとして
は、マクロ経済学での応用の観点で書かれた工藤 (2007)、阿部 (2017)、蓮見 (2019) の講義ノートがある。教科書では A.C. Chang
の『動学的最適化の基礎』の邦訳は比較的簡単に解説しているが誤植が多いことで有名である。西村清彦の『経済学のための最適
化理論入門』という教科書もあるが、どこまでカバーしていたかの記憶が定かではない。もっとも詳しいのはやはり Stokey et al.
(1989)、Ljungqvist and Sargent (2018) の教科書あたりか？ 単に「ベルマン方程式」でググると強化学習に関連するものとして言
及する記事がよく引かれるが、これらはベルマン方程式というより分枝限定法の考え方の解説だったり、内容がメモ程度の断片的
だったりするので今回の話の補足説明には向かない。理論モデルでは連続時間を想定する場合もあり、ベルマン方程式の代わりに
Hamilton-Jacobi-Bellman 方程式を応用することもある。こちらは物理学でもよく使われるようだ。

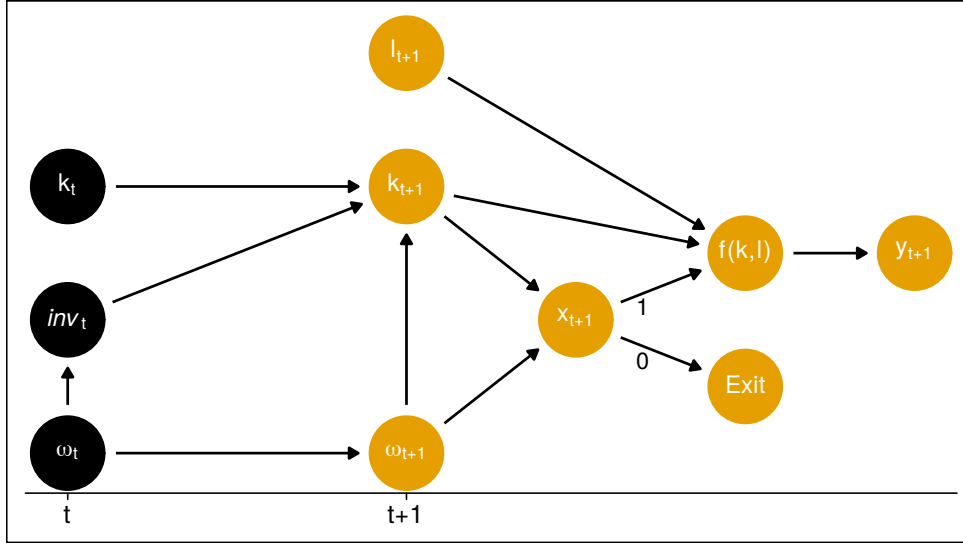


図2 OP モデルで仮定した変数の関係

のように $k_t, \omega_t, \underline{\omega}_t$ の条件に依存することを覚えておこう^{*22}. 退出による利益 Φ は観測できないため, 別の変数で表現する必要があるため, 退出を決定するしきい値を表す関数 $\omega_t(\cdot)$ を導入する. 利潤関数 π_t が k_t に対して減少関数なら, 価値関数にとっては増加関数, そして資本ストックの多い企業は一時的な生産性の変化 ω_t とはあまり関係なく, 将来に比較的大きな収益が見込まれるから, しきい値 $\omega_t(\cdot)$ は比較的低くなるはずなので $\omega_t(\cdot)$ は減少関数になる (退出確率 p_t の減少) はずだ. よってこの退出ルールから ω_{t+1} に $x_{t+1}, k_{t+1}, l_{t+1}, \omega_t$ を条件付けた ω_{t+1} の期待値

$$E[\omega_{t+1} \mid k_{t+1}, \omega_t, x_{t+1} = 1]$$

は k_{t+1} に対して減少関数であるとわかる (l_{t+1} は独立しているので ω_{t+1} に影響しない). これが観測できない生産性 ω_t と資本 k_t が相関していること理由付けになる.

ここまでで登場した変数間の関係をグラフィカルモデルで描くならば, 図2のようになる.

3.6 OP モデルの回帰モデルとしての意味

経営者が (分析者には) 観測されない効果 ω_t を参照して k_t, l_t を決めてしまうというのは, 計量経済学の文脈でいうならこれは欠落変数バイアスであり, ここまでで紹介したパネルデータの方法で推定してもバイアスが発生してしまう. さらに ω_t のマルコフ性を仮定しているため, もし ω_t が過去の値と相関しているならば, やはり内生性によるバイアスが発生する. また, OP モデルは経営者の撤退という選択肢も想定していることは, サンプルセレクションバイアスを考慮しているとも言える. 前節ではパネルデータには一切欠測がない (バランスドパネル) ことを暗黙の前提としていたが, 実際のパネルデータではしばしば欠測が起こる. もし欠測が企業の活動とは完全に無関係にランダムに発生している (MAR; missing at random) ならば推定に大きな影響はないが, すでに解説したように k_t は ω_t (の期待値) に相関し, ω_t の大きさ次第で企業経営者は撤退を決める. つまりモデル内の変数に依存するメカニズムで発生している (MNAR; missing not —) ことになるので推定にバ

^{*22} 元の論文では $p_t := P(x_{t+1} = 1 \mid \dots)$ としていることに注意.

イアスを引き起こす^{*23}.

推定のため、投資を $inv_t = inv(k_t, \omega_t)$ という関数で表現できるとして、 $inv_t > 0$ と仮定すると、投資関数 $inv_t(\cdot)$ が厳密に単調増加関数であれば^{*24}、逆関数 $\omega_t = h_t(k_t, inv_t)$ が存在することになる。言い換えるなら、 inv_t は操作変数である。これを(3.2)に代入すると、

$$y_t = \beta_L l_t + [\beta_A + \beta_K k_t + h_t(k_t, inv_t)] + \varepsilon_t \quad (3.8)$$

となり、観察されない効果を他の説明変数で表現できることがわかった。[...] の部分は関数形が特定されていない、いわゆるノンパラメトリックモデルである。しかし、今知りたいのは $\beta_A, \beta_K, \beta_L$ であって、投資関数を特定することは必須でない。そこで、OP 法では以下のような 2 段階の推定法で一致推定量を得る。

第 1 段階では k_t と ω_t という動学的に決定される部分をノンパラメトリック回帰することで、静学的に決まる β_L のみを識別する。そこで、相関のある部分をまとめて、以下のように ϕ_t と表す。

$$\phi_t(k_{i,t}, inv_{i,t}) := \beta_A + \beta_K k_{i,t} + h_t(k_{i,t}, inv_{i,t}) \quad (3.9)$$

ϕ_t を使うと(3.8)は以下のように書き換えられる。

$$y_t = \beta_L l_t + \phi_t(k_t, inv_t) + \varepsilon_t \quad (3.10)$$

このモデルの ϕ_t の部分をノンパラメトリック回帰で推定する^{*25}ことで、推定値 $\hat{\beta}_L, \hat{\phi}_t$ が識別された。これは別の見方をすれば部分線形モデル (partially linear —) である^{*26}。

$\hat{\beta}_L$ を得られたので、当初の式は以下のように書ける。

$$y_t - \hat{\beta}_L l_t = \phi_t(k_t, inv_t) + \varepsilon_t \quad (3.11)$$

よって残りは ω_t や k_t をひとまとめにした ϕ_t からどうやって β_A, β_K を識別するかという問題である。ここからが Olley らの研究の肝であり、込み入った説明になる箇所である。 $\hat{\phi}_t$ には $\beta_A + \beta_K k_t + \omega_t$ が含まれているが、 k_t は既に説明したように動学的に決定される。ここで、次の $t+1$ 時点の $y_{t+1} - \beta_L l_{t+1}$ の期待値を考えると、 ω_t のマルコフ性から以下のように書ける。

$$\begin{aligned} E[y_{t+1} - \beta_L l_{t+1} \mid k_{t+1}, x_{t+1} = 1] \\ &= \beta_A + \beta_K k_{t+1} + E[\omega_{t+1} \mid \omega_t, x_{t+1} = 1] \\ &= \beta_K k_{t+1} + g(\omega_t, \omega_{t+1}) \end{aligned}$$

^{*23} 欠測が推定結果に及ぼす影響については、星野 (2009)、高井他 (2016)、高橋・渡辺 (2017) のいずれかが参考になるだろう。ただしパネルデータでの問題については星野が少し触れている程度である。英語でもいならば Wooldridge (2010, Ch. 19) が解説している。

^{*24} k_t に対して単調増加である理由はすでに示唆した。一方で ω_t が大きければ、経営者はやはり将来の利益が大きくなると予測し、設備投資を増やす傾向にある、という理由でこの仮定を説明できる。しかし現実のデータがこの仮定を満たすとは限らない。実際には規模縮小のために資産を売却するという選択肢もあり、このとき設備投資はマイナス扱いになる。しかしこの仮定はかなり強い仮定であり、これ以降も Levinsohn and Petrin (2003) などがこの仮定がもたらす問題が取り上げられている。

^{*25} Olley らは 4 次の多項式補間で計算しているが、これは次数を増やして MSE および係数が変化しなくなるものを採用するというルールで推定したためである。一方で、年代で分割して試すと変化が見られたとしている。多項式に特定する理由もないので、部分線形モデルはカーネル平滑化による計算も可能だが、こちらは次元が増えると計算に時間がかかるという性質がある。

^{*26} 特定のパラメータだけを推定するために、操作変数の式を与えて内生性によるバイアスを排除する方法をコントロール関数アプローチという。Cameron and Trivedi (2005)、Wooldridge (2010) にもコントロール関数の解説があるが、日本語の教科書では見た記憶がない。ググってもコントロール関数のアイディアを解説しているのは明城 (2012) くらいで、それも需要関数推定の問題に限定したものなので少しわかりにくいかもしれない。解説したものは私が以前自分で書いたものくらいしかない(これもあまりおすすめできない)。Kawaguchi が部分線形モデルでの応用に関して参考文献として挙げる Todd と Ichimura の論文は Handbook of Econometrics に収録されているので研究機関に所属していない人間がアクセスするのは少し大変である。代わりに Blundell and Powell (2003) ならば一般公開されている。

ここで, $g(\omega_t, \omega_{t+1}) = \beta_A + E[\omega_{t+1} | k_{t+1}, x_{t+1} = 1]$ であり, 欠落バイアスを表している. g は観測できない ω_t で構成されるが, (3.7)から, $t+1$ 時点の存続確率 p_{t+1} は (k_{t+1}, ω_t) の関数で書けることが分かっている. さらに, 期首の k_{t+1} は前期の k_t, inv_t に依存する (この k_{t+1} は ω_{t+1} が決まる前の値なので, 経営者が投入量を決めるという話とは無関係). よって, $t+1$ 時点の $g(\cdot)$ は p_{t+1} と ω_t で表現できることがわかる (つまりこれは一種の傾向スコアである). 退出したかどうか (x_t) はパネルデータの欠測で知ることができるため, x_t に対してプロビット回帰^{*27}をすることで確率の推定値 \hat{p}_t を計算できる. さらに, ω_t は, 第1段階で推定した $\phi_t(k_t, inv_t) = \beta_A + \beta_K k_t + \omega_t$ から,

$$\hat{g}(\omega_t, \omega_{t+1}) = g(\hat{p}_{t+1}, \hat{\phi}_t - \beta_A - \beta_K k_t)$$

と書ける.

\hat{g} と ω_t のバイアスを $\hat{\xi}_{t+1} := \omega_{t+1} - \hat{g}(\omega_t, \omega_{t+1})$ として, 当初の式(3.11)に代入できる. 第2段階^{*28}では $\hat{\beta}_L$, $\hat{\phi}_t, \hat{p}_t$ を所与として, 以下のようなラグのある回帰式を推定する問題として扱える^{*29}.

$$(y_{t+1} - \hat{\beta}_L l_{t+1}) = \beta_K k_{t+1} + g(\hat{p}_t, \hat{\phi}_t - \beta_A - \beta_K k_t) + \hat{\xi}_{t+1} + \varepsilon_{t+1} \quad (3.12)$$

(プロビットモデルの推定がバイアスなくできているならば) $\hat{\xi}_{t+1}$ は平均ゼロでかつ k_{t+1} とも独立だが, l_{t+1} とは相関しているかもしれない. しかし β_L は第1段階で推定した $\hat{\beta}_L$ を代入できるため, もはや気にする必要はない. ($\hat{\xi}_{t+1} + \varepsilon_{t+1}$) を誤差項とみなすと, β_K は誤差項と相関しないので識別できる. そして $g(\cdot)$ は構造がわからない非線形関数である. そこで, Olley らは2通りの推定方法を試している. 1つは1段階目と同様に部分線形モデルとしてカーネル回帰を利用する方法, もう1つは, 多項式近似した上で非線形最小二乗法で解く方法である. つまり,

$$y_{t+1} - \hat{\beta}_L l_{t+1} = \beta_K k_{t+1} + \sum_{d=1}^q \sum_{m=0}^d \gamma_{m,d-m} \hat{h}_t^m \hat{p}_t^{d-m} + (\hat{\xi}_{t+1} + \varepsilon_{t+1})$$

である. ここで, $\hat{h}_t := \hat{\phi}_t - \beta_A - \beta_K k_t$ で, q は多項式の次数である. なお, OP 法では標準誤差を代数的に得られないため, ブートストラップシミュレーションで計算する必要がある. この方法で導出したモデルの統計的性質はPakes and Olley (1995) でより詳しく説明されている.

3.7 モデルの事後診断

さて, 反事実分析でのモデルの評価は単に当てはまりが良いかだけでなく, 仮定と現実のデータが矛盾していないかの確認も同等かそれ以上に重要である.

ここでも Olley らの研究を例にする. まず, 彼らの仮定が正しければ, 単なる最小二乗法や, IV などの方法では結果にバイアスが発生するはずである. よって, 提案する推定方法との比較が必要である. あるいは, サンプルセクションバイアスが実際に発生しているか確認するため, バランスドパネルでの結果と比較することも必要である. さらに, 市場の構造は法規制の変化によって変わりうる. 法改正のタイミングで年代ごとに区切って当てはまりを確認する必要がある.

^{*27} Olley らはこれも1段階目の推定と同様の方法で次数を決めた多項式展開でプロビット回帰をしている.

^{*28} Olley らの論文では p_t の推定を第2段階と呼んでいるので, これは原文では第3段階である. しかし後の解説では省略されることが多い.

^{*29} スライドでは余白節約のため(3.12)を $\hat{\omega}_t := \hat{\phi}_t - \beta_A - \beta_K k_t, \hat{\omega}_{t+1} := \hat{\xi}_t + \varepsilon_{t+1}$ と置き換えて表現している.

さらに、モデルのロバストネス分析として、モデルの仮定のうち特に重要な、投資関数の式(3.6)を検討している。投資が資本と観察できない生産性によって決まるという仮定が正しくない場合、 β_L の推定にバイアスが生じる。これを検証するには最後の式(3.12)の説明変数に、現在の l_t を加えることでできる (このようなタイプの検証は**特定化のテスト**と呼ばれる)。

3.8 OP 法以降の研究

Levinsohn and Petrin (2003) は OP の方法を改良したもの。多くの企業の会計情報を見ると、実際には設備投資は必ずしも発生せず、ゼロが多い。これは投資関数の単調性と矛盾するため、中間投入財 (材料費) の情報を利用することを提案している。現在はさらに Akerberg et al. (2015) が改良版を提案して、これがスタンダードになりつつあるらしいが今回は省略。

4 R での実装

4.1 データのとり方

企業の生産関数推定で一番むずかしいのは実はここかもしれない。というのも、理論上の K や L はどこにも記載されていないからだ。これはけっこう地道な計算が必要になる。例えば企業の決算報告書などから地道に計算する、あるいは日銀や DBJ のような政策金融機関が整備しているデータベースをなんらかの手段で見せてもらう、などである。あるいは、農業や工業など資源と生産物の関係が分かりやすい特定の産業だけを分析の対象とすることもできる。例えば農業で、 K を乳牛の頭数、原材料を飼料、 Y をミルクの生産量とした分析も可能だろう (昔見た気がするがすぐには見つけれなかった)。今回はとてもそんなことをしている暇も金もコネもない^{*30}ので、乱数データを使う。

`estprod` パッケージが、Olley-Pakes や Levinsohn-Petrin の方法を実装したという。さらに、このパッケージには練習用の乱数データセットも用意されている。この乱数がどのような方法で生成されたかがわからないので、実質的に動作確認にしかならないが、まずはこれを使ってみた。しかし、1 段階目の推定を検証してみたところ、残差プロットの形がかなり変になった。残差プロットが必ず完璧な 45 度線を引かなければならない、ということはないが、データの生成方法は分からないこともあって用例として不安である^{*31}。よって、Kawaguchi (2019) の課題 2 に沿って疑似データを生成してやり直すことにした。

しかし、去年の講義内容とはいえ課題の答えを勝手に一般公開してしまうのも無配慮すぎる気がするため、**乱数データを作成するプログラムは非公開**として、当初の課題から乱数シード値などを変更して作成したデータ `df_sample`, `df_ground_truth` のみ公開する。前者は標本データで、後者は ω_t, p_t など本来は観測できない変数も見えるようにしたデータである (表 2, 以下、後者を「正解データ」と呼ぶ)。以降はこの `df_sample` を使ってパラメータを推定する。推定したい 3 つのパラメータは元の課題どおり、 $\beta_A = 1, \beta_K = 0.7, \beta_L = 0.2$ としている。

^{*30} いくつかの計量経済学の教科書の著者は練習問題用にデータセットを公開している。例えば "Wooldridge data sets", "Datasets for empirical exercises in Econometrics" など。しかし、今回の問題に適した企業別パネルデータは見つけられなかった。

^{*31} このパッケージは CRAN に登録されているものの、作者のプロフィールがよくわからない。また、github リポジトリにもほとんど関心が集まっていないため、動作がどの程度適切なのかあまり確認が進んでいなさそう。一応ソースコードを大まかに確認はしたが、その点は留意してほしい。あと、`coefs()`, `predict()`、といった stats の基本関数にも対応していない。もちろん `stargazer` に対応していないし、`estimatr` にも収録されていない。他にも、企業退出を考慮しているように見えて実は不均衡パネルデータを想定した実装になっていないなどの問題があった。

	i	t	k	k_lag	inv	inv_lag	x	y	y_true	omega	p_x
df_sample	企業 ID	t	k_t	k_{t-1}	inv_t	inv_{t-1}	x_t	y_t	NA	NA	NA
df_ground_truth	企業 ID	t	k_t	k_{t-1}	inv_t	inv_{t-1}	x_t	y_t	退出しない場合の y_t	ω_t	p_t

表2 乱数データの変数対応表

なお、乱数生成したため、偶然にも一部の企業が全期間で退出している。データとして不自然だが計算には問題ないので修正せずそのままにしている。

4.2 β_L, ϕ の識別

1 段階目の推定は部分線形カーネル回帰で推定する。これは `np` パッケージで提供されている。今回は `npplregbw()` でデータから最適バンド幅を決定し、決定したバンド幅を `npplreg()` に与えて推定する^{*32*33}。この推定結果を残差プロット等で確認したら、 p_t の推定を行う。1 段階目のモデル(3.10)は、誤差項と相関のある部分線形モデルであり、単に多項式やスプライン近似を適用しただけでは正しく ϕ を推定できないため、`npplregbw(formula = y ~ l | k + inv)` としても誤差項の相関のために β_L は正しく推定されず、全く異なる値になる。そこで部分線形モデルとコントロール関数アプローチの併用によって β_L, ϕ を識別する。このためには、`formula = y ~ l | k + inv` と指定する^{*34}。これで、 β_L, ϕ を識別できる。ここで `y ~ l + k | k + inv` と指定することで β_K も推定できるが、この時点では正しく推定できない。

```
1 bw_1st <- npplregbw(formula = y ~ l | k + inv, data = df_sample %>% drop_na(y, k, l, inv))
2 fit_1st_np <- npplreg(bw_1st)
3 summary(fit_1st_np)
```

なお、`np` パッケージのノンパラメトリック回帰は計算負荷がけっこうある。そこで、面倒だったら Olley らと同様に多項式近似でごまかしても良い

```
1 lm(y ~ l + poly(k, inv, degree = q), data = df_sample)
```

`degree = q` は多項式の次数で、`formula` を評価する際には $\sum_{r \leq q} k_t^r inv_t^{q-r}$ と認識される。何も考えずに `degree = 4` としてもそこそこそれらしい値を推定できるが^{*35}、Olley のやるように推定値と MSE が安定する次数を選ぶのがスマートだろう。しかし残念ながら `stats::step()` 関数には `poly()` の次数を調整する機能がないので、ループ処理か予め決めた次数までを並列処理して比較する必要がある。そこで、プログラム1のように 10 次までの多項式あてはめを一度に実行する処理を書いた^{*36}。

^{*32} Olley ら本来のやり方を踏襲するなら、時点ごとに分けて当てはめるのが正統だが、今回はプール推定する。

^{*33} なお、`np` パッケージに与えるデータを `tibble` で与えるとエラーが出る (しかも一見関係ないメッセージ) ので注意する。

^{*34} その他、デフォルト設定である `na.action = stats::na.omit` で欠損値を含むレコードを除外して計算するはずだが、なぜか機能しない。そのため入力データから欠損値を除外する処理を自分で書かなければならない。

^{*35} 今回これでもうまくいくのは、データの生成過程が単純だからである。実際のデータに応用するときは、データの特性、つまりどういう現象から取得したデータなのかをよく考えて方法を柔軟に変えねばならない。

^{*36} `estimatr` や `tidymodels` などならもっと簡単に書けないかとも考えたが、これらのパッケージは今回使う推定方法に意外と対応していない。

プログラム 11 段階目の多項式近似

```
1 df_1st_poly <- tibble(q= 1:10) %>%
2   mutate(model = map(q, function(q){
3     if(q == 1){
4       poly_term <- "k_+inv"
5     } else{
6       poly_term <- paste("poly(k_+inv,degree=", q, ")_poly(k_+inv,degree=", q -1, ")
7     }
8     lm(as.formula(paste("y_~l_+k_+inv_", poly_term )), data = df_sample)
9   })) %>%
10  mutate(
11    rmse = map_dbl(model, function(x) sqrt(mean(x$residuals^2))),
12    coef = map(model, function(x) enframe(coef(x)) %>% filter(name %in% c("k", "l")))
13  ) %>% unnest(coef) %>% pivot_wider(names_from = name, values_from = value)
```

ところで、現在はノンパラメトリックな回帰モデルとして**スプライン回帰**の研究が進んでいる (基本的な事項は [Hastie et al., 2009](#), Ch. 5, [坂本他, 2009](#)などを参照). スプライン回帰にもバリエーションがあるが、典型的なものは区間ごとに異なる次数の多項式で表現するため、単純な多項式近似よりも表現の自由度の高い非線形近似の方法である. さらに `SemiPar::spm()` は部分線形モデルに対しても罰則付き最尤推定という方法で当てはめたスプライン回帰を計算することができる. 表現の自由度で言えばカーネル回帰のほうがさらに優れているが、計算時間はこちらのほうが少ない. 多項式近似である程度うまくいくのだからスプライン回帰でもうまくいくと思ったのだが、なぜかうまく行かない. 原因を調べるのがめんどくさくなったので読者の課題とする.

```
1 tmp <- select(df_sample, y, k, l, inv)
2 f_spm <- spm(form = tmp$y ~ tmp$l + f(tmp$k, tmp$inv), omit.missing = T)
3 summary(f_spm)
```

4.3 存続確率の推定

Olley は、理論モデルの仮定から p_{t+1} を決定する関数は k_t, inv_t の 2 変数関数で表現できることから、この 2 変数を使ったプロビットモデル (ロジットリンク関数を標準正規分布に置き換えたもの) と多項式近似またはカーネル回帰を組み合わせることで確率を推定している.

今回は正解が分かっているため、 p_{t+1} と k_t, inv_t の関係を知ることができる. 2 変数の散布図とカーネル回帰で近似曲線を当てはめたものが以下になる. プロビットモデルならば、 p_{t+1} を標準正規分布の逆分布関数 (分位点関数) で変換した値と説明変数が線形相関していれば、その説明変数を使うことで当てはまりの良いモデルができると予想できる. しかし、図3からわかるように、2 つの説明変数との相関関係はいずれもあまり直

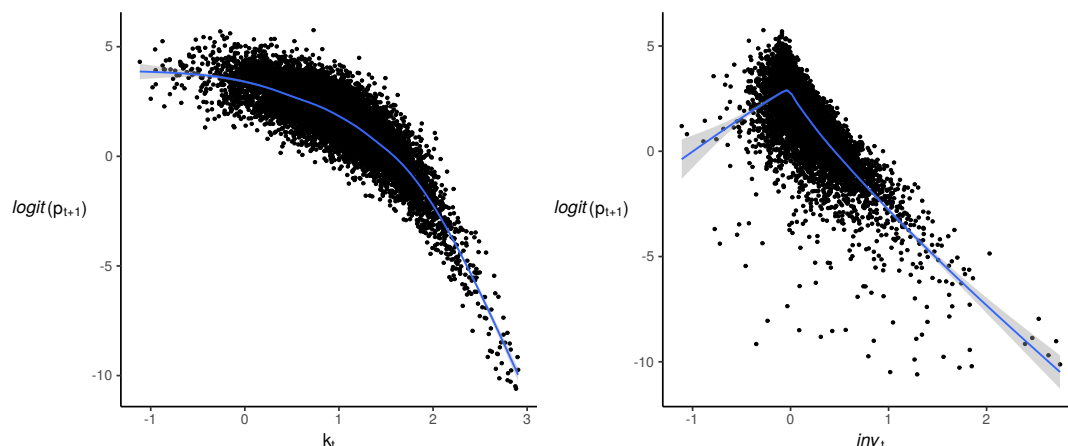


図3 真の確率と説明変数の相関

線的でないことがわかる。よって、単なる一般化線形回帰ではなく、より複雑な非線形近似方法を使う必要があるとわかる。Rでは多項式プロビット回帰は `glm()` と `poly()` の組み合わせで、例えば以下のようにして計算できる。ここでも次数をデータの当てはまりで決定するため `beta_L$` の推定と同様にこの関数を次数を変えて並列処理している。

```
1 glm(x~poly(k_lag, inv_lag, q), family = binomial(link = "probit"), data = df_sample)
```

1段階目で使用した `np` パッケージは線形回帰のみ対応しているため、プロビットモデルの推定はできない。他の R のパッケージでは、例えば各変数をスプライン補間などで非線形変換した一般化加法的モデル (GAM^{*37}) を計算できる `mgcv` パッケージ、一般化ガウシアンカーネル回帰のできる `bkmr` パッケージや、一般化部分線形モデルを計算できる `gplm` などがある。これらはいずれもノンパラメトリックあるいはセミパラメトリックなモデルと呼ばれるが、それぞれ表現方法が異なるため、当てはまりの良さも変わってくる^{*38}。また、このパートではそもそも係数の推定が不要なので、確率さえ推定できているならランダムフォレストのような自由度の高いモデルでも問題ないだろう^{*39}。今回は `mgcv`, `bkmr`, `gplm`, `ranger` などを使って推定してみた。なお Kawaguchi の課題では退出行動を省略しているのでこのパートは存在しない。

4.4 β_A, β_K の推定

最後の推定は、 ω_t を推定し、さらに企業の退出行動によるサンプルセレクションバイアスを補正するための計算である。

まず、計算の無駄をなくするためにプログラム2の関数のように推定に必要なラグ項をデータフレーム側で作

^{*37} 参考: [Hastie et al., 2009, Ch. 9](#), [鈴木大慈の講義スライド](#), あるいは言語研究者からの観点だが『[GLMM/GA\(M\)M の文献案内](#)』で挙げられているリストも興味深い。また、[以前紹介](#)した `prophet` のモデルも一般化加法的モデルの一種である。

^{*38} この場でこういったノンパラメトリック回帰の方法まで細かく説明するのは面倒なのでしない。GAM やカーネル回帰はとりあえず [Hastie et al. \(2009\)](#) の教科書を読んで欲しい。または [Li and Racine \(2007\)](#) の教科書か。それ以外の細かい話は各パッケージのマニュアルで参照されている論文を読んでほしい。

^{*39} ただし、ここで重要なのは退出・存続の分類ではなく確率の推定であることに注意する: 結果変数 x_t の条件分布ではなく p_t の分布を知る必要があるため、関数形の特定に依存する問題である。

プログラム 2 2 段階目の推定のためのデータフレーム

```
1 make_2nd_sample <- function(data, beta_l, phi, p_x){
2   df_sample %>%
3     mutate(
4       bl = l * beta_l * l,
5       y_bl = y - bl,
6       phi = phi,
7       p_x = p_x
8     ) %>%
9     arrange(i, t) %>% group_by(i) %>% mutate(
10      phi_lag = lag(phi),
11    ) %>% ungroup }
```

成しておく。mutate() 関数内で dplyr::lag() を使えばそれぞれのラグ項を得られる^{*40}。元のパネルデータを df_sample, 1 段階目の推定結果を fit_1st として, 2 段階目の推定のためのデータセット df_sample_2nd を作成している。y_bl, phi, p_x はそれぞれ $y_t - \hat{\beta}_L l_t, \hat{\phi}_t, p_t$ に対応する。

ここで一旦、実際には観測できない β_L, ω_t の情報を使って β_K を正しく推定できることを確認してみる。以下のように $y_t - \beta_L l_t - \omega_t$ に対して線形回帰をした結果、 $\beta_A = 1.005, \beta_K = 0.697$ とほぼ正確な値が出ている。

```
1 lm(formula = y_bl - omega ~ k, data = df_ground_truth_2nd)
```

しかし、実際には ω_t は不明であり、1 段階目で推定した h_t と確率 p_t から推定する^{*41}。Olley らの方法と同様に多項式近似による非線形最小二乗法で計算する。

なお再び正解データを覗き見すると、 $\hat{\phi}_t, \hat{p}_{t+1}$ と ω_{t+1} の関係はそれぞれ図4のようにになっている。

非線形最小二乗法の計算は stats::nls() 関数が提供されており、この関数は formula で記述できたり、lm() のように標準誤差を計算してくれたりする。しかし今回は多項式部分を formula で記述するのがやや面倒なのと、nls() の計算する標準誤差が意味をなさないため、optim() 関数で計算する。目的関数はプログラム3のように、任意の次数の(3.12)の MSE を計算できるような関数を作った。

p_t, h_t の交差項を含む q 次多項式の第 d 次項は $d+1$ 個存在するため、1 次から $d-1$ 次までの項の総数は $\sum_{i=1}^d i$ 、項の総数は $\sum_{d=1}^q (d+1)$ 個であるから、上記のように purrr::map_dfc(), purrr::reduce() を入れ子にして計算した^{*42}。なお、数値計算上の観点から初期値にも注意する必要がある。これは補遺Aで詳しく説明している。また、Olley らはさらにここでもカーネル回帰も試しているが、パラメータが含まれる非線形式のカーネル回帰は自分で 1 から書かなければならず面倒になったので今回は省略する。

^{*40} plm と stats にも lag() 関数があるが、こちらは mutate() 内で使用しても評価されないし、エラーも発生しないため、呼び出し順に注意。あるいは conflicted パッケージも併用すると良いだろう。このパッケージの解説は『名前空間の衝突を conflicted パッケージで防ぐ』でなされている。

^{*41} 欠測メカニズムが分かっている場合、セレクションバイアスを補正する Heckman の選択モデル (Heckit, or Type-II Tobit) モデルが応用できる場合がある。しかし今回は欠測確率 p_t が k_t に影響して決まるため、Heckit は意味をなさない。

^{*42} 実際にはこういう複雑な計算に出くわしたら何も考えずに rcpp に切り替えて for 文で書いたほうが効率が良い気もする。

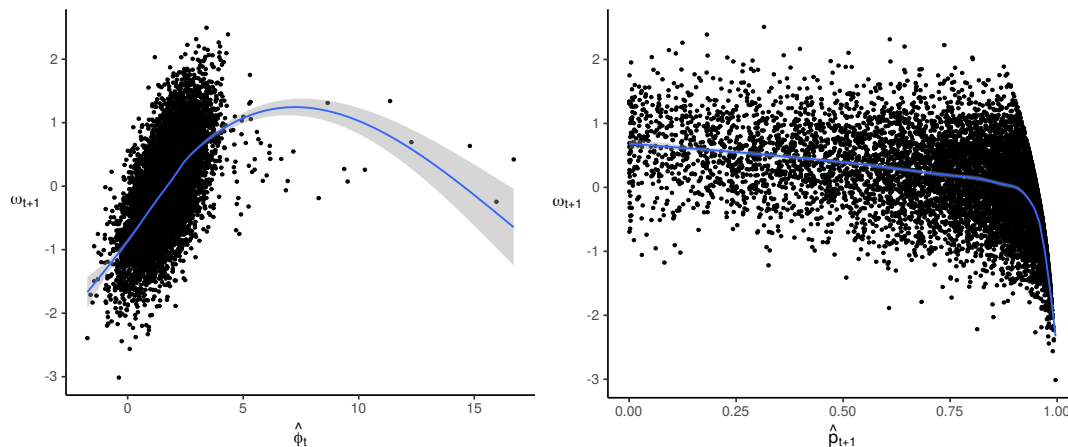


図4 $\hat{\phi}_t, \hat{p}_{t+1}$ と ω_{t+1} の正解データとの相関関係

プログラム 3 OP 法 2 段階目の目的関数

```

1 obj_op_2nd <- function(param, data, degree){
2   # param = [beta_a, beta_k, gamma1, gamma2, ...]
3   h <- with(data, phi_lag - exp(param[2]) * k_lag)
4   poly <- map_dfc(
5     1:degree,
6     function(d) {reduce(0:d, ~.x + param[2 + sum(1:d + 1) - d - 1 + 1 + .y] * with(data, h^(d
7       - .y) * p_x_lag^(.y)), .init = 0)}
8   ) %>% rowSums
9   mean((with(data, y_bl - param[1] - exp(param[2]) * k) - poly)^2, na.rm = T)
10 }

```

そして標準誤差を代わりに (ノンパラメトリック) ブートストラップ法で計算する。ブートストラップ法による標準誤差の計算方法は補遺Bで補足説明している。

以上で説明した内容は, Github リポジトリの all_public.R にまとめてある。

5 結果

FE, RD などの内生性を想定しないパネルデータの方法と, Arellano-Bond の階差 GMM などの動学パネルデータモデル, OP の方法, そしてCで詳しく紹介する Kawaguchi の課題での方法のそれぞれの推定結果を図5に示す^{*43}。OP の方法は, 多項式回帰とカーネル回帰の両方を試したいところだが, カーネル回帰は既存の

^{*43} estprod は OP 法の実装とiiつつ企業の途中退出を考慮していない上, 退出のないデータに当てはめても全く違う結果が帰ってきたのでなかったことにした。また, 今回書いた OP 法の各バリエーションは結果がほとんど変わらなかったため多項式の場合のみを代表して掲載し, AB/BB 推定量はあまりにも誤差が大きいため掲載を省略した。

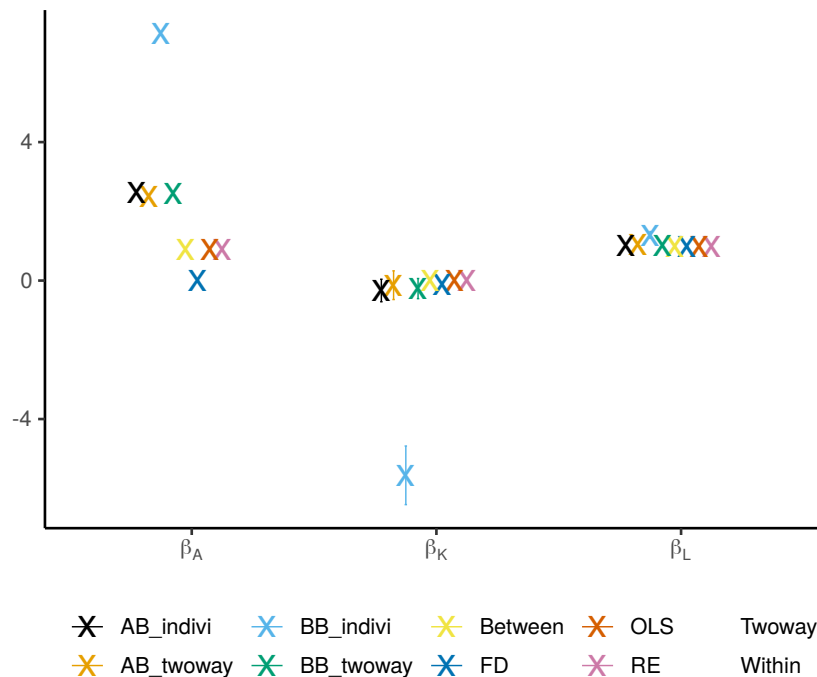


図5 推定結果

パッケージを使いまわせず実装にも時間がかかるので省略した。また, 確率推定を GAM やランダムフォレストに置き換えたケースについても, 結果にほとんど違いがなかったため省略した。Kawaguchi の課題では一般化モーメント法 (GMM) で計算しており, その条件の特定は今回のデータとも合致しているが退出行動を省略しているためバイアスが発生することが予想できる (やる気があれば本来の課題との結果の違いを比較すると面白いだろう)。

6 まとめ

今回は, 因果推論と呼ばれる Rubin 的な介入効果の推定メソッドが経済学では誘導形と呼ばれること, そして誘導形と対比される形で構造推定と呼ばれる方法論が存在することを説明した。今回具体例として取り上げた Olley と Pakes の研究は, 変数の内生性によって, 動学パネルデータや個別効果モデルといった従来の推定方法では生産関数を適切に推定できないことを指摘した上で, 経済モデルを想定した推定方法によってパラメータを識別したというものである。そして, 提案される計算手順を R でデモンストレートした。

今回は取り掛かった段階で準備期間が 1 週間を切っていたため, かなり手を抜かざるを得なかった。以前の発表の続編ということで, 動学構造推定を R で実演したかったのだが, まず説明を書くのに時間がかかるので, 静学構造推定に, そして R のプログラムを用意する時間もなくなってきたので実装もなし, となし崩し的に縮退していった。さらには, タイトルで「反事実分析」と書いたわりに「反事実」的な要素のあまりない研究を挙げることになってしまった。

今回挙げた Olley-Pakes や Levinsohn-Petrin の研究は 10 年以上前の話なので, すでに大学院, あるいは学部の授業でも取り上げられることすらあるようだ。そういう意味でもあまりおもしろみのない発表だったとは

思う。

しかし、実際に乱数データで OP 法本来のやり方を試してみると、簡単な生成ルールであっても初期値に依存して標準誤差が大きくなるなど、意外と扱いづらいことがわかった。これは再現性の問題にも関わってくると思うので、より安定した推定方法を議論するのもおもしろいかもしれない。

参考文献

- Akerberg, Daniel A., Kevin Caves, and Garth Frazer (2015) “Identification Properties of Recent Production Function Estimators,” *Econometrica*, Vol. 83, No. 6, pp. 2411–2451, DOI: [10.3982/ECTA13408](https://doi.org/10.3982/ECTA13408).
- Anderson, T. W. and Cheng Hsiao (1981) “Estimation of Dynamic Models with Error Components,” *Journal of the American Statistical Association*, Vol. 76, No. 375, pp. 598–606, September, DOI: [10.1080/01621459.1981.10477691](https://doi.org/10.1080/01621459.1981.10477691).
- (1982) “Formulation and Estimation of Dynamic Models Using Panel Data,” *Journal of Econometrics*, Vol. 18, No. 1, pp. 47–82, DOI: [10.1016/0304-4076\(82\)90095-1](https://doi.org/10.1016/0304-4076(82)90095-1).
- Arellano, Manuel and Stephen Bond (1991) “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations,” *The Review of Economic Studies*, Vol. 58, No. 2, p. 277, April, DOI: [10.2307/2297968](https://doi.org/10.2307/2297968).
- Blundell, R and Steven Bond (1998) “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models,” *Journal of Econometrics*, Vol. 87, No. 1, pp. 115–143, November, DOI: [10.1016/S0304-4076\(98\)00009-8](https://doi.org/10.1016/S0304-4076(98)00009-8).
- Blundell, Richard and James L. Powell (2003) “Endogeneity in Nonparametric and Semiparametric Regression Models,” in Dewatripont, Mathias, Lars Peter Hansen, and Stephen J. Turnovsky eds. *Advances in Economics and Econometrics*, Cambridge: Cambridge University Press, pp. 312–357, DOI: [10.1017/CBO9780511610257.011](https://doi.org/10.1017/CBO9780511610257.011).
- Cameron, AC and PK Trivedi (2005) *Microeconometrics: Methods and Applications*, Cambridge: Cambridge University Press, DOI: [10.1017/CBO9781107415324.004](https://doi.org/10.1017/CBO9781107415324.004).
- Efron, B. (1979) “Bootstrap Methods: Another Look at the Jackknife,” *The Annals of Statistics*, Vol. 7, No. 1, pp. 1–26, January, DOI: [10.1214/aos/1176344552](https://doi.org/10.1214/aos/1176344552).
- Griliches, Zvi and Jacques Mairesse (1999) “Production Functions: The Search for Identification,” in Strom, Steinar ed. *Econometrics and Economic Theory in the 20th Century*, Cambridge: Cambridge University Press, pp. 169–203, DOI: [10.1017/CCOL521633230.006](https://doi.org/10.1017/CCOL521633230.006).
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*: Springer, 2nd edition, retrieved from [here](#), (杉山将・井出剛・神寫敏弘・栗田多喜夫・前田英作・井尻善久・岩田具治・金森敬文・兼村厚範・鳥山昌幸・河原吉伸・木村昭悟・小西嘉典・酒井智弥・鈴木大慈・竹内一郎・玉木徹・出口大輔・富岡亮太・波部斉・前田新一・持橋大地・山田誠訳, 『統計的学習の基礎 — データマイニング・推論・予測 —』, 共立出版, 2014 年) .
- Hayashi, Fumio (2000) *Econometrics*, Princeton: Princeton University Press.
- Igami, Mitsuru (2018) “Artificial Intelligence as Structural Estimation: Economic Interpretations of Deep Blue, Bonanza, and AlphaGo,” March, arXiv: [1710.10967](https://arxiv.org/abs/1710.10967).
- Kawaguchi, Kohei (2019) “ECON 6120I Topics in Empirical Industrial Organization,” retrieved from

[here](#).

- Levinsohn, James and Amil Petrin (2003) “Estimating Production Functions Using Inputs to Control for Unobservables,” *Review of Economic Studies*, Vol. 70, No. 2, pp. 317–341, April, DOI: [10.1111/1467-937X.00246](#).
- Li, Qi and Jeffrey Scott Racine (2007) *Nonparametric Econometrics: Theory and Practice*, Princeton, N.J: Princeton University Press, OCLC: ocm67346329.
- Ljungqvist, Lars and Thomas J. Sargent (2018) *Recursive Macroeconomic Theory*, Cambridge, Massachusetts: MIT Press, 4th edition.
- Lucas, Robert E. (1976) “Econometric Policy Evaluation: A Critique,” *Carnegie-Rochester Confer. Series on Public Policy*, Vol. 1, No. C, pp. 19–46, DOI: [10.1016/S0167-2231\(76\)80003-6](#).
- Lütkepohl, Helmut (2007) *New Introduction to Multiple Time Series Analysis*, Berlin: Springer, 1st edition, OCLC: 254802195.
- Mairesse, Jacques (1990) “Time-Series and Cross-Sectional Estimates on Panel Data: Why Are They Different and Why Should They Be Equal?” in Hartog, Joop, Geert Ridder, and Jules Theeuwes eds. *Panel Data and Labor Market Studies*, North-Holland: Elsevier Science Publishers B.V., pp. 81–95.
- Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green (1995) *Microeconomic Theory*: Oxford University Press.
- Merton, Robert C. (1974) “On the Pricing of Corporate Debt: The Risk Structure of Interest Rates,” *The Journal of Finance*, Vol. 29, No. 2, p. 449, May, DOI: [10.2307/2978814](#).
- Olley, G. Steven and Ariel Pakes (1996) “The Dynamics of Productivity in the Telecommunications Equipment Industry,” *Econometrica*, Vol. 64, No. 6, pp. 1263–1297, DOI: [10.2307/2171831](#).
- Pakes, Ariel and Steven Olley (1995) “A Limit Theorem for a Smooth Class of Semiparametric Estimators,” *Journal of Econometrics*, Vol. 65, No. 1, pp. 295–332, January, DOI: [10.1016/0304-4076\(94\)01605-Y](#).
- Reiss, Peter C. and Frank A. Wolak (2007) “Structural Econometric Modeling: Rationales and Examples from Industrial Organization,” in Heckman, James J. and Edward E. Leamer eds. *Handbook of Econometrics*, Vol. 6 of Handbook of Econometrics: Elsevier, Chap. 64, pp. 4277–4415, DOI: [10.1016/S1573-4412\(07\)06064-3](#).
- Scott, Steven L. and Hal R. Varian (2014) “Predicting the Present with Bayesian Structural Time Series,” *International Journal of Mathematical Modelling and Numerical Optimisation*, Vol. 5, No. 1/2, p. 4, DOI: [10.1504/IJMMNO.2014.059942](#).
- Stokey, Nancy L., Robert E. Lucas, and Edward C. Prescott (1989) *Recursive Methods in Economic Dynamics*, Cambridge, Mass: Harvard University Press.
- Varian, Hal R. (1992) *Microeconomic Analysis*, New York: Norton, 3rd edition, (佐藤隆三・三野和雄訳, 『ミクロ経済分析』, 勁草書房, 1986 年) .
- Wooldridge, Jeffrey M (2010) *Econometric Analysis of Cross Section and Panel Data*: The MIT Press, 2nd edition, retrieved from [here](#).
- 阿部修人 (2017) 「上級マクロ経済学講義ノート動的計画法」, retrieved from [here](#).
- 汪金芳・大内俊二・景平・田栗正章 (1992) 「ブートストラップ法最近までの発展と今後の展望」, 『行動計量学』, 第 19 卷, 第 2 号, 50–81 頁, DOI: [10.2333/jbhmk.19.2_50](#).
- 沖本竜義 (2010) 『経済・ファイナンスデータの計量時系列分析』, 朝倉書店.

- 奥井亮 (2016) 「動学的パネルデータモデル」, retrieved from [here](#).
- 北川源四郎 (2005) 『時系列解析入門』, 岩波書店.
- 北村行伸 (2016) 「応用ミクロ計量経済学の手法と論点」, 『経済セミナー増刊進化する経済学の実証分析』, 日本評論社, 28–36 頁, retrieved from [here](#), 個人サイトにリンクないんだけど落として良かったのか?.
- 北村行伸・西脇雅人・村尾徹士 (2009) 「不完全資本市場下での生産関数の推定について」, 一橋大学ディスカッションペーパー 70, 193–204 頁, retrieved from [here](#).
- 金明哲・汪金芳・桜井裕仁 (2011) 『ブートストラップ入門』, 共立出版, 東京, OCLC: 769347675.
- 工藤教孝 (2007) 「動学的最適化入門」, Technical report, retrieved from [here](#).
- 久保拓哉 (2012) 『データ解析のための統計モデリング入門』, 岩波書店.
- 坂本亘・井筒理人・白旗慎吾 (2009) 「罰則付きスプラインによる非線形回帰構造の推測」, 『計算機統計学』, 第 21 巻, 第 1-2 号, 55–94 頁, retrieved from [here](#).
- 高井啓二・星野崇宏・野間久史 (2016) 『欠測データの統計科学』, 岩波書店.
- 千木良弘朗・早川和彦・山本拓 (2011) 『動学的パネルデータ分析』, 知泉書館, 東京, OCLC: 751998792.
- 中島上智・渡部敏明 (2012) 「時変ベクトル自己回帰モデル-サーベイと日本のマクロデータへの応用-」, 『経済研究』, 第 63 巻, 第 3 号, 193–208 頁, DOI: [10.15057/25864](#).
- 中嶋亮 (2016) 「「誘導型推定」 v.s. 「構造推定」」, 『経済セミナー増刊進化する経済学の実証分析』, 日本評論社, 52–62 頁.
- 西村和雄 (1990) 『ミクロ経済学』, 東洋経済新報社.
- 蓮見亮 (2019) 「動学マクロ経済学入門」, retrieved from [here](#).
- 星野崇宏 (2009) 『調査観察データの統計科学-因果推論・選択バイアス・データ融合』, 岩波書店.
- 渡部敏明 (2016) 「ルーカス批判とマクロ計量分析」, 『経済セミナー増刊進化する経済学の実証分析』, 日本評論社, 37–41 頁.
- 明城聡 (2012) 「製品差別化財の需要関数推定における内生問題について」, 『国民経済雑誌』, 第 206 巻, 第 5 号, 83–99 頁, nov 月, retrieved from [here](#).
- 高橋将宜・渡辺美智子 (2017) 『欠測データ処理 - R による単一代入法と多重代入法 -』, 統計学 One Point, 第 5 号, 共立出版.

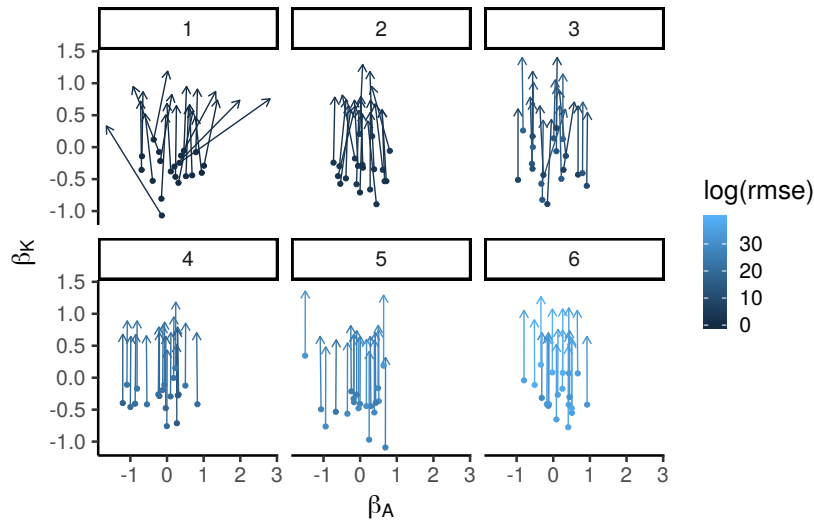


図6 解の初期値と収束先

補遺

A 数値計算上に関する実装上の注意

非線形最小二乗法は一般に解が一意とは限らない。プログラム3では $h_t := \hat{\phi}_t - \beta_K k_t$ として計算しており、 β_A を除いていない。この方法では β_A, ω_t を識別できないからである (Olley らの目的は TFP の計算なので、 β_K, β_L さえ識別できれば問題にならない)。例えば $E\omega_t = 0$ と分かっているならば識別できるが、今回の仮定ではそれはわからないため、数値計算の安定のために β_A の減算を省略している。

また、非線形最適化は初期値を変えると結果が大きく変わることがある。一般的な解決方法はないが、今回は生産関数の意味から考えて $0 < \beta_K$ という制約を与えている。もし β_K がゼロや負なら、資本投入量を増やしてもまったく生産量が増えないか、むしろ減少することになる。これは通常ならありえないことだろう。optim() 関数には lower というオプションがあるが、ここでは計算速度を損なわないように β_K のみ指数変換して計算するように修正しているさらに、初期値を $1 - \hat{\beta}_L$ とした。これは、 $\beta_K + \beta_L = 1$ ならば生産関数が規模に対して収穫一定になるからである。もちろんそのようになるという仮定をしていないが、生産関数のパラメータが極端な値にならないという前提ならば中庸な設定であり、初期値に向いていると考えられる。この初期値に乱数でばらつきを与えて複数回の結果を確認すると、初期値と収束した解の対応関係は図6のようになり、次数 q ごとの誤差 $\pm\sigma$ は図7のようになった。このように、 β_K は初期値によって結果にばらつきがあり、 β_A は初期値に完全に依存していることがわかる。そのため、今回は各次数について 100 回試行し、平均値から次数を決定し、その後再びブートストラップ法で反復計算した結果を最終的な推定値として採用することにした。

また、乱数データを生成するパラメータを変更すれば当てはまり易さも変わってくることに注意する。例えば p_t の推定には inv, k_t を使用しているが、もしこれらと x_t との相関が弱ければ、弱相関操作変数と同様にかえってバリエーションが増加する原因になると予想できるため、企業のデータを取れば常に OP 法をそのまま適用すればいいということはあるまいだろう。このような観点からも構造推定は決まったパッケージの決まった構文を実行すれば終わりというものではなく、各段階で仮説が正しいかを検証する必要がある。

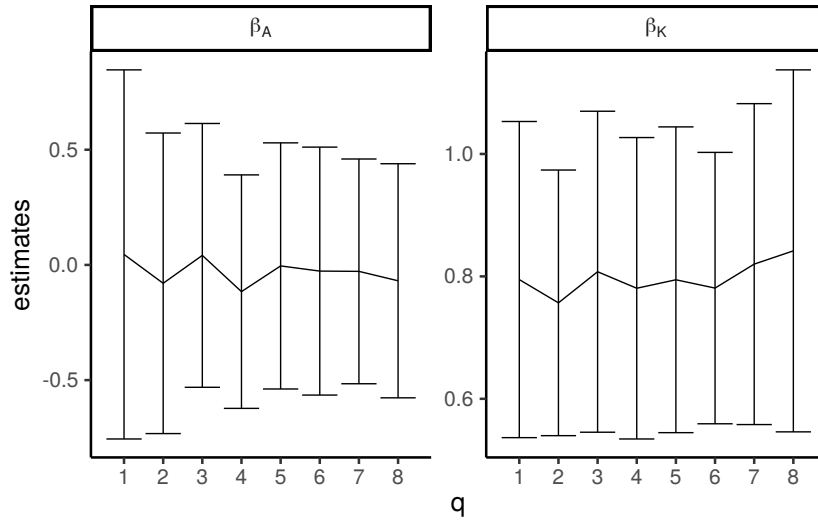


図7 次数ごとの初期値の違いによる誤差

B ブートストラップ標準誤差

Efron (1979) は再標本化によって標準誤差や分布の分位点を推定する方法をいくつも提案し, これらをブートストラップ法と総称した. ブートストラップ法はパラメトリックなものとノンパラメトリックなものに大別されるが, 今回はノンパラメトリックなブートストラップ法を利用する. 最も単純なブートストラップ法は次の通り^{*44}. N 件のデータ $D = \{z_1, \dots, z_N\}$ を取得できたとして, ここから復元抽出した N 件の標本 (ブートストラップ標本) を B 個作成する.

$$\begin{aligned} D_1 &= \{z_1^{(1)}, \dots, z_N^{(1)}\}, \\ &\vdots \\ D_B &= \{z_1^{(B)}, \dots, z_N^{(B)}\} \end{aligned}$$

もし, この擬似的に作られたブートストラップ標本 D_b が母集団から無作為抽出された標本と同等の分布に従うならば, それぞれの標本平均 $\bar{z}^{(b)}$ はすべて同一の分布にしたがう. よって, 以下のように計算した標本標準偏差を標準誤差の推定値とすることができる.

^{*44} 日本語で書かれ, かつネット上で公開されているブートストラップ法の基本的かつ有用な解説は少ない. 標準誤差には言及していないが, R を使った簡単な実演が『[ブートストラップ入門](#)』で紹介されている. pdf ファイルが不鮮明だが, 汪他 (1992, 第 5 節) で最も詳しい説明がなされている. これらの記述を総合すればだいたいのはわかるが, いちおうここで要約しておく. それ以外にもいくつかの研究論文が見つかるが, 回帰モデルのブートストラップ標準誤差を求める方法の解説としては使いづらい. Efron の論文は半世紀も前のものなので, 日本語で書かれた有用な教科書ならば多く存在する. ただし私は金他 (2011) のものしか読んでいない.

$$\hat{SE}(\bar{z}) := \sqrt{\frac{1}{B} \sum_{b=1}^B (\bar{z}^{(b)} - \bar{z}_B)^2},$$

$$\bar{z}_B := \frac{1}{B} \sum_{b=1}^B \bar{z}^{(b)}$$

これが標本平均でなく回帰分析の係数に置き換わっても同じことが言える。つまり今回の OP 法のように推定量の分布を解析的に導出するのが難しい場合にブートストラップ法を利用する場面がある。

しかし、「ブートストラップ標本が母集団から無作為抽出された標本と同等の分布に従う」という前提を忘れてはならない。上記の単純なケースではどの標本も等確率に抽出されており、言い換えるなら母集団の分布がそのようになっていると暗黙に仮定していることになる。[Efron](#)は、推定した回帰モデルの残差から誤差項の経験分布を構成し、これを誤差項 ε の分布とみなしてサンプリングした乱数 $\hat{\varepsilon}_i^*$ とモデルの予測値の和 $\hat{y}_t^* := \hat{y}_t + \hat{\varepsilon}_t^*$ を生成することを提案している。

標本からの計算のため、大数の法則や中心極限定理からの類推で直観的に B が大きいほど推定の精度がよくなると予想できる。しかしながら、具体的にどれくらいなら十分であるかを示す理論は存在しないようだ。ネット上では具体的な数字を挙げる例も見られるが、具体的な根拠は挙げられていないことが多い^{*45}。これに関しては "[Rule of thumb for number of bootstrap samples - Cross Validated](#)" の AdamO の回答が参考になる。いわく、十分な回数は問題に応じて変わる (例えばコーシー分布に従うデータならば何万、何千万回やろうとそもそも存在しない平均を求めるのに十分ではない) ため、ブートストラップ計算の結果をヒストグラムに表し分布が適切な形になっているかを確認せよ、ということである。とはいえ、彼の回答の追加コメントでも指摘されるように、多くの場合は回数が多いほど精度が増すはずなので、例えば先行研究と比較できるようにいったん同じ回数で計算した上でヒストグラムを確認し、もう少しうまくいってないなら回数を増やす (またはバグを疑う)、といった方法が現実的ではないかと考える。

ブートストラップ法の計算に関する R のパッケージは `simpleboot` と `boot` パッケージの 2 つが代表的である^{*46}。`simpleboot` は名前の通り簡単なブートストラップ法の計算だけに限定したパッケージで、OP 法のような特殊なモデルには対応していないから `boot::boot()` を使う。ただし、`np` パッケージによるカーネルベースの推定方法はただでさえ時間がかかるため、今回は 1 段階目を多項式近似でやることにした。

`boot()` の使い方はパッケージのヘルプを見ればすぐ分かるが、一応簡単に解説しておく。第 1 引数 `data` にデータを、第 2 引数 `statistic` に計算したい統計量 (のベクトル) を返す関数を与える。`boot()` は `data` をリサンプリングしてブートストラップ標本を作り、それぞれを `statistic` に与えてブートストラップ推定値を計算する。`statistic` に与える関数の第 1 引数には `data` が、第 2 引数には生成ブートストラップ標本が `data` のどこかを表すインデックスベクトルが与えられる。さらに関数ヘルプには補外データに対してブートストラップ法を適用する例が掲載されている。この場合、予測値の計算にももう 1 つインデックスベクトルが必要になる。これが第 3 引数である。第 4 引数以降は、ユーザーが自由に定義できる引数である。そのため、既に行ったように回帰モデルに対してノンパラメトリックブートストラップ法を適用する場合は、予め残差を計算して入力データに含めておく必要がある。

^{*45} 例えば標準誤差の計算なら 100 程度、分位点の計算なら 1000 程度、という記述がよく見られる。この数値の根拠は[Efron](#)の論文に掲載されている実験でそのように設定されていたからではないかと思う。しかし Efron はこの数字の根拠を述べていない。

^{*46} 金明哲『[R とブートストラップ](#)』と奥村晴彦『[ブートストラップ](#)』はいずれもこれらの R パッケージを使った計算方法も含めて言及している。『[ブートストラップのための boot パッケージ](#)』にも `boot` パッケージの言及がある。

今回はプログラム4のように書いた。回帰モデルのブートストラップ統計量の計算のため、`data` には通常の推定に必要な変数に加え、ブートストラップ標本を生成するための $\hat{y}_{i,t}, \hat{\varepsilon}_{i,t}$ の列を追加する必要がある。

今回 `statistic` が出力すべき統計量は3つの係数パラメータ $\beta_A, \beta_K, \beta_L$ だが、確認のため各段階で推定したモデルの当てはまりについても出力するようにしている^{*47}。

なお、既にかいたように OP 法では β_A を識別できないが、他の方法との比較のため $\hat{\beta}_A := (NT)^{-1} \sum_{i,t} (y_{i,t} - \hat{\beta}_L l_{i,t} - \hat{\beta}_K k_{i,t})$ として計算した。これは本来の仮定に加えて $E\omega_t = 0$ が成り立ってなければ適切な推定量とは言えないが、今回の乱数データでは ω_t が平均ゼロの定常分布に収束するような方法で生成している。

引数 `R` はブートストラップ標本の生成数である。本文での解説に沿って 100 に設定し、各パラメータのヒストグラムが不自然でないか確認した。`strata` は層別抽出をするためのインデックスである。今回は y_t に欠損があり、OP 法は欠損していないデータだけで推定しているから、ブートストラップ標本 \hat{y}_t^* も本来のデータの欠損メカニズムを再現する必要がある。ここでは `strata` に元のデータの欠損判定 x_t を与えることでこれを表現している^{*48}。また、`boot.ci()` で `boot()` の結果から信頼区間を計算できるが、今回は利用しない。

C kawaguchi の課題についての補足

Olley らの方法に倣えば非線形最小二乗法として係数 β_A, β_K を求める必要があるが、課題では退出行動を省略し、 ω_t の構造もシンプルのため Olley らのような複雑な式の導出はなく、非線形一般化モーメント法 (非線形 GMM) で推定することを課している。GMM の解説はすくなく、非線形 GMM となるとなおさらなので^{*49}、ここで少し脱線して課題では何をやっているのかを書いておく。

練習用データは ω_t を AR(1) 過程の乱数として生成しているの、課題では係数 α は未知であるものの AR(1) であることは分かっているという前提である。つまり、以下のように仮定している。

$$\begin{aligned} h(\omega_{t+1}) &= \alpha \omega_t + \nu_t, \\ \nu_t &\sim \mathcal{N}(0, \sigma_\nu^2) \forall t \end{aligned}$$

であると仮定している^{*50}ので、

$$\hat{\xi}_{t+1} := \omega_{t+1} - \alpha \left(\hat{\phi}(k_t, inv_t) - \beta_A - \beta_K k_{t+1} \right)$$

とすると、

$$y_{t+1} = \beta_L l_{t+1} + \beta_A + \beta_K k_{t+1} + \alpha \left(\hat{\phi}(k_t, inv_t) - \beta_A - \beta_K k_{t+1} \right) + \hat{\xi}_{t+1} + \varepsilon_{t+1}$$

となり、これを変形すると以下ようになる。

$$\hat{\xi}_{t+1} + \varepsilon_{t+1} = y_{t+1} - \hat{\beta}_L l_{t+1} - \beta_A - \beta_K k_{t+1} - \alpha \left(\hat{\phi}(k_t, inv_t) - \beta_A - \beta_K k_{t+1} \right)$$

^{*47} サンプルプログラムでは返り値に名前を付けているが、`boot()` は名前の情報を保存してくれないようだ

^{*48} 欠損値を含む場合のブートストラップ法に詳しくないが、 y_t に対して x_t は先決 (predetermined) 変数だから、単純に x_t の層化で欠損を条件付けるだけで十分だと思う。

^{*49} 非線形 GMM は Hayashi (2000) を始め大学院向け標準レベルの教科書なら必ずと言っていいほど載っているが、日本語の教科書では線形 GMM までしか書いていないことが多い。私が以前書いたブログ記事でも肝心の非線形 GMM の説明がなかった。どんなモーメント条件であっても求められるわけではなく、ニュートン法は目的関数が凹関数でなければ一意な解を得られない。実際には一貫性や漸近分布を保証するためにさらにいくつかの十分条件を考える必要がある。それらの議論も詳しくは教科書を参照してほしい。

^{*50} 課題文にも書いてあるように、本来は ω_t の決定メカニズムを知ることは難しいため現実の問題に対してこのような強い仮定を置くことは好ましくない。


```

1 op_params_poly <- list(beta_k = exp(fit_2nd_poly$par[2]), beta_l = unname(coef(fit_1st_poly)
  ["l"]))
2
3 estimate_op_ordinary <- function(data, degree_1, degree_p, degree_2nd){
4   fit_1st <- lm(y ~ l + poly(k, inv, degree = degree_1), data = data, na.action = na.omit)
5   beta_l <- unname(coef(fit_1st)["l"])
6   rmse_1st <- sqrt(mean(resid(fit_1st)^2))
7   fit_exit <- glm(x ~ l + poly(k_lag, inv_lag, degree = degree_p), data = data %>% drop_na(k_
  lag, inv_lag), family = binomial(link = "probit"))
8   std_Brier <- std_Brier(fit_exit$data$x, fitted(fit_exit))
9   df_2nd <- make_2nd_sample(
10    data, beta_l,
11    phi = predict(fit_1st, newdata = data) - beta_l * data$l,
12    p_x = predict(fit_exit, newdata = data, type = "response")
13  )
14  fit_2nd <- optim(
15    par = c(rnorm(n = 1, sd = .5), log(ifelse(1 - beta_l <= 0, .5, 1 - beta_l)) + rnorm(n =
  1, sd = .3), rnorm(n = sum(1:degree_2nd + 1), sd = .5)),
16    fn = obj_op_2nd, data = df_2nd, degree = degree_2nd)
17  beta_a <- mean(with(df_2nd, y_bl - exp(fit_2nd$par[2]) * k), na.rm = T)
18  return(c(beta_a = beta_a, beta_l = beta_l, beta_k = unname(exp(fit_2nd$par[2])), rmse_1st
  = rmse_1st, std_Brier = std_Brier, rmse_2nd = sqrt(fit_2nd$value)))
19 }
20
21 get_stat_boot <- function(data, indices, indices_, estimate, params){
22   # estimate = estimate function
23   # params = model hyper parameter (e.g., degrees of polynomials)
24   data$y <- with(data, fitted + resid[indices])
25   do.call(estimate, args = c(list(data = data), params)) }
26
27 boot_ordinary <- df_sample_2nd_poly %>%
28   mutate(., fitted = if_else(is.na(y), NA_real_, do.call(get_fitted, args = c(list(data =
  .), op_params_poly))),
29     resid = y - fitted) %>%
30   boot(., get_stat_boot, R = 100, strata = .$x, estimate = estimate_op_ordinary, params = list(
  degree_1 = 3, degree_p = 2, degree_2nd = 2))

```

既書いた仮定では $\hat{\xi}_{t+1} + \varepsilon_{t+1}$ は同時点のいくつかの変数と相関する可能性が残るが同時点の k_{t+1} および 1 時点前の $k_t, inv_{i,t}$ と無相関なので以下のようなモーメント条件が成り立つ。

$$E(\hat{\xi}_{t+1} + \varepsilon_{t+1}) \begin{bmatrix} k_{t+1} \\ k_t \\ inv_t \end{bmatrix} = 0$$

$\hat{\xi}_t + \varepsilon_t$ にはパラメータに対して非線形な項が含まれるので非線形モデルということになる。線形モデルの場合は単に GMM と呼ばれ、今回のような非線形モーメント条件を特定した場合を非線形 GMM と呼び、ニュートン法などの非線形最適化の方法で計算することになる。なお計算には新たに $y_t - \hat{\beta}_L l_t, \hat{\phi}_t, k_t, inv_t$ のラグ項が必要になるが、これらは既にプログラム 2 で作成している。

R で GMM 推定法を計算するパッケージには `gmm`, `plm` がある^{*51}。後者の `plm` は GMM だけでなく、FE/RE モデルなどパネルデータを想定したモデルの計算用関数が用意されている。今回もパネルデータを利用しているが、個別効果を想定していないため `plm` を必ず使わなければならないわけではない。実際のところ、`plm::pgmm()` ができる GMM は既に少しだけ言及した Arellano-Bond/Blundell-Bond の動学パネルデータモデルの推定だけであるので、今回のような非線形 GMM には使えない。一方で `gmm` パッケージは非線形 GMM にも対応しているので、今回はこちらを使う。

また、GMM を使う場合も標準誤差はブートストラップ法で計算する必要がある。しかしカーネル回帰は非常に時間がかかるため、100 回も 1000 回も繰り返すのは現実的ではない。今回時間がかかるのは主に最適バンド幅を計算する `npplregbw()` の実行なので、最初の計算で得たバンド幅を全てのブートストラップ標本で共用することにした。しかし、`np` パッケージは引数に対する挙動がかなり意味不明で混乱を招く。所与のバンド幅でデータを変えて再計算したい場合の方法が不明瞭である。いろいろ試したが、`data` に与えるだけではバンド幅計算時の結果がそのまま出力されるようだ。`txdat`, `tydat`, `tzdat` にそれぞれ新しい数値を与える必要があるらしい。

gmm パッケージの解説

`gmm` の使い方はヘルプを見れば十分なのだが、タダで見られてかつ日本語の GMM の解説は少なく、非線形 GMM の解説は一層少ない。そういうものを求める人のためにここで少しだけ解説しておく。まず、線形の場合は `lm()` などと同じように `formula` オブジェクトでモーメント条件を指定できる。例えば、

$$E(y_i - \alpha - \beta x_i) \begin{bmatrix} z_{i,1} \\ z_{i,2} \\ z_{i,3} \end{bmatrix} = 0$$

のようなモーメント条件をもとに GMM 推定したい場合は、

```
1 gmm(y ~ 1 + x | z1 + z2 + z3, x = data)
```

^{*51} `plm` の使い方は開発者による解説 "Estimation of error components models with the `plm` function" が参考になるだろう。一方で日本語の情報では慶應義塾大の長倉大輔氏? 作の pdf 『R によるパネルデータモデルの推定』があるが、それ以外では参考になるものは少ない。また、公式ドキュメントを見ても `plm()` がデフォルトでどのように個体・時間インデックスを認識するのがよくわからない。念のため `index=` を指定したほうが良いだろう。

となる^{*52}. `data` はデータフレームまたは `matrix` で与える.

非線形モデルの場合はこのようにモデルを `formula` で表現できないため, 代わりに標本モーメントを返す関数を与える. この関数関数は第 1, 第 2 引数にそれぞれパラメータと入力データを与て $N \times q$ 行列を返すものが要求される. N は入力データの件数, q はモーメント式の本数である. つまり上記の期待値の部分で,

$$m_i(\alpha, \beta; x_i) := (y_i - \alpha - \beta x_i) \begin{bmatrix} z_{i,1} \\ z_{i,2} \\ z_{i,3} \end{bmatrix}$$

という部分であり, 返す行列は

$$\begin{bmatrix} m_1 z_{1,1} & m_1 z_{1,2} & m_1 z_{1,3} \\ m_2 z_{2,1} & m_2 z_{2,2} & m_2 z_{2,3} \\ \vdots & \vdots & \vdots \\ m_N z_{N,1} & m_N z_{N,2} & m_N z_{N,3} \end{bmatrix}$$

という形になる.

非線形 GMM とは $m_i(\theta, x_i)$ の部分について, 例えば $m_i(\theta, x_i) := (y_i - \alpha - \gamma \beta x_i) \begin{bmatrix} z_{i,1} & \cdots & z_{i,3} \end{bmatrix}^\top$ のような非線形関数の場合を指す. この m_i の標本平均の重み付きノルム F_n を目的関数として最小化するのが GMM である^{*53}.

$$F_n(\alpha, \beta) = \left[\frac{1}{N} \sum_{i=1}^N m_i(\theta, x_i) \right]^\top \hat{\mathbf{W}}^{-1} \left[\frac{1}{N} \sum_{i=1}^N m_i(\theta, x_i) \right]$$

非線形 GMM はニュートン法でも計算できるため, 目的関数 $F_n(\cdots)$ を計算する関数さえ作れば, 係数だけなら `gmm` パッケージを使わなくとも `optim()` で簡単に推定できる. そこで, 念のため `optim()` も使って検算するのも良いだろう.

^{*52} `formula` の構文はデフォルトで定数項を加えるため, 1 は必須ではない. そもそも `formula` の構文が分からない場合は『71. 回帰分析と重回帰分析 - R-Source』や私が以前書いた『予測モデルを作るには `formula` を活用せよ』などを参照

^{*53} 典型的な GMM では重み行列 $\hat{\mathbf{W}}$ に共分散行列の推定値を与えるが, 課題では簡略化のため $\hat{\mathbf{W}}$ に単位行列を与えるよう指示がある. 単位行列の場合, 線形モデルという 2 段階最小二乗法 (2SLS) と等価な計算になる. 重み行列の役割は主に標準誤差の縮小なのでブートストラップ法で計算する今回のケースではさほど重要ではない.