

貿易における重力モデル

貿易における重力モデルは、両辺の対数をとることで、線形モデルとして最小2乗推定することが可能になります。しかし近年では、対数化することの問題が認識されており、非線形モデルをポアソン回帰によって推定することが推奨されています。本稿では、重力モデルの推定方法を詳しく説明した後、データ分析を通じて理解を深めていきます。

こちらは藪友良『入門 実践する計量経済学』（2023年、東洋経済新報社）の補足資料です。ポアソン回帰については、サポートウェブサイトの追加資料「カウントデータ」をご覧ください。

1. 重力モデルの基本

貿易における重力モデル(gravity model)は、貿易理論から導出することができます。

$$\text{Trade}_{ij} = A \frac{GDP_i^{\beta_1} GDP_j^{\beta_2}}{\text{Distance}_{ij}^{\beta_3}}$$

ここで、 i 国と j 国の間の貿易額(Trade_{ij})は、両国の名目GDP(それぞれ GDP_i 、 GDP_j)、両国間の距離(Distance_{ij})に依存して決まるとしています(詳しくは6.3節参照)。パラメータ(β_1 、 β_2 、 β_3)は正の値をとります。両国のGDPが大きくなると需要や供給が増加することで貿易額は増加する一方、両国間の距離が遠くなると輸送コストが増加し貿易額が減少することになります。

2. 誤差項ありの重力モデルと OLS 推定の問題

これまで誤差項を考えてきませんでしたが、重力モデルは誤差を含んだ形でのみ成立します。

$$\text{Trade}_{ij} = A \frac{GDP_i^{\beta_1} GDP_j^{\beta_2}}{\text{Distance}_{ij}^{\beta_3}} + u_{ij}$$

ここで、誤差項 u_{ij} は他の説明変数とは無相関と仮定します。

このままでは、両辺の対数をとっても線形関数にならないため、次のように書き換えます。

$$\text{Trade}_{ij} = A \frac{GDP_i^{\beta_1} GDP_j^{\beta_2}}{\text{Distance}_{ij}^{\beta_3}} \left(1 + \frac{1}{A} \frac{\text{Distance}_{ij}^{\beta_3}}{GDP_i^{\beta_1} GDP_j^{\beta_2}} u_{ij} \right)$$

$$= A \frac{GDP_i^{\beta_1} GDP_j^{\beta_2}}{Distance_{ij}^{\beta_3}} \eta_{ij}$$

ただし、 η_{ij} は次のように定義しました。

$$\eta_{ij} = \left(1 + \frac{1}{A} \frac{Distance_{ij}^{\beta_3}}{GDP_i^{\beta_1} GDP_j^{\beta_2}} u_{ij} \right)$$

上式両辺の対数をとると、次の式が得られます。

$$\ln(\text{Trade}_{ij}) = \ln(A) + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) - \beta_3 \ln(\text{Distance}_{ij}) + \ln(\eta_{ij})$$

ここで、 $\ln(\eta_{ij})$ は新しい誤差項であり、上式を最小2乗(OLS)推定することでパラメータ(α 、 β_1 、 β_2 、 β_3)を推定することができます。しかし、Silva and Tenreyro (2006)では、対数対数モデルの OLS 推定に2つの問題があることを指摘しています¹。

第1に、説明変数(名目 GDP、距離)と誤差項 $\ln(\eta_{ij})$ の間に相関が生じるため、OLS 推定量はバイアスを持つという点です(内生性の問題は教科書 8.3 節を参照してください)。説明変数と誤差項との間に相関が生じる理由は、 η_{ij} が説明変数の関数になっていることから明らかでしょう。

第2に、貿易額は0をとることが多いデータであり、その場合に、貿易額の対数 $\ln(\text{Trade}_{ij})$ が定義できない点があります(定義から0の対数は存在しません)。このとき、貿易額の対数は、欠損値として扱われることになり、データから除外されてしまいます。これはサンプルサイズを小さくし、パラメータの推定精度を下げることになります。さらに言うと、貿易額が0の値をとるのは、両国間で貿易が存在しないケース、貿易額が小さすぎたことで0に丸められたケースが該当します。たとえば、1000ドル単位のデータで500ドルの貿易額なら0.5になりますが、これを丸めて0として記録される可能性があります。貿易額が0となることは、小国で頻繁に生じる現象であり、小国をサンプルから除くことで、サンプルセレクションバイアスが生じます。

¹ この論文は対数化の問題を指摘しており、その1例として貿易モデルに焦点を当てています。当然ながら、対数化したモデルなら、重力モデル以外でも、彼らの指摘が当てはまる可能性があります。詳しくは、以下の論文を参照してください。Silva, J. M. C. S., and Tenreyro, S. (2006) "The Log of Gravity," *Review of Economics and Statistics*, 88(4), 641-658.

3. ポアソン疑似最尤推定

ここでは近年、貿易における重力モデルの推定で用いられているポアソン疑似最尤推定について紹介します。

3.1 ポアソン回帰との関係

対数関数の性質から、 $y = e^{\ln(y)}$ が成立します(対数をとったら両辺は同じです)。この性質から、誤差を含んだ本当の貿易の重力モデルは、次のように記述できます。

$$\begin{aligned}\text{Trade}_{ij} &= A \frac{GDP_i^{\beta_1} GDP_j^{\beta_2}}{\text{Distance}_{ij}^{\beta_3}} + u_{ij} \\ &= e^{\ln(A) + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) - \beta_3 \ln(\text{Distance}_{ij})} + u_{ij}\end{aligned}$$

Silva and Tenreyro (2006)では、このモデルについて次の2点を指摘しています。

第1に、誤差項 u_{ij} に不均一分散がある点です。貿易額 Trade_{ij} は0以上の値となるため、 $e^{\ln(A) + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) - \beta_3 \ln(\text{Distance}_{ij})}$ が0に近い値だと、誤差項 u_{ij} の分散は0に近い値になります。これは、そうでないと貿易額がマイナスになってしまうためです。逆に、 $e^{\ln(A) + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) - \beta_3 \ln(\text{Distance}_{ij})}$ が大きくなると、誤差項 u_{ij} の分散も大きくなる傾向があります。以上から、誤差項 u_{ij} の分散が、 $e^{\ln(A) + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) - \beta_3 \ln(\text{Distance}_{ij})}$ に比例していることがわかります。

第2に、貿易モデルとポアソン回帰モデルには類似性がある点です。説明変数が確率変数ではないとすると、貿易額 Trade_{ij} の期待値は、

$$E[\text{Trade}_{ij}] = e^{\ln(A) + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) - \beta_3 \ln(\text{Distance}_{ij})}$$

となり、貿易額の期待値の対数は線形モデルになります。

$$\ln(E[\text{Trade}_{ij}]) = \ln(A) + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) - \beta_3 \ln(\text{Distance}_{ij})$$

なお、貿易額は0をとることがありますが、その期待値は正の値をとるので、期待値の対数は問題なく定義ができます。

被説明変数の期待値の対数が線形モデルになる点は、ポアソン回帰モデルと同じ特徴です。これは重力モデルが最尤法によって推定できることを意味しています。ポアソン回帰モデルの最尤推定量は、たとえポアソン分布の仮定が誤っていても、被説明変数の期待値の対数が線形である点が正しければ、一致性を持つことが知られています²。

当然ですが、貿易額はポアソン分布していないため(ポアソン分布の仮定が誤っている)、重力モデルの最尤推定は、**ポアソン疑似最尤推定(Poisson Pseudo Maximum Likelihood**

² 説明変数が確率変数なら、条件付き期待値 $E[Y_i|X_i]$ が $e^{\alpha + \beta X_i}$ となれば(同じことですが、条件付き期待値の対数が線形 $\alpha + \beta X_i$ となれば)、ポアソン回帰による推定は一致性を持つことになります。

Estimator、PPML 推定)と呼ばれています。ポアソン分布の仮定が誤っているため、通常の標準誤差ではなく、ロバスト標準誤差を用いることが推奨されます³。

3.2. シミュレーション

Silva and Tenreyro (2006)は、シミュレーションによって、さまざまな推定方法を比較しています(詳細は補足参照)。推定方法としては、通常対数対数モデルによる OLS 推定、貿易額に 1 を足したうえでの対数対数モデル、非線形最小 2 乗法、ポアソン回帰などがあります。

貿易額に 1 を足す方法は、貿易額が 0 であっても対数を定義できる利点がある一方、1 を足すという恣意的な操作を行っている欠点があります⁴。0.1 や 0.01 ではなく、なぜ 1 を足すのでしょうか？また、対数対数モデルと同様に、説明変数と誤差項に相関が生じる欠点は変わりません。

非線形最小 2 乗法(nonlinear least squares)は、非線形モデルの残差 2 乗和を最小にするようにパラメータを決める方法です。これはサンプルサイズが大きければ、一致性を持った推定になります。しかし、不均一分散があると、推定精度が低くなる欠点があります。補足では、非線形最小 2 乗推定とポアソン回帰との比較をしており、ポアソン疑似最尤推定は、不均一分散のもとで推定精度が優れていることを指摘しています。

シミュレーションの結果、通常対数対数モデル、貿易額に 1 を足した対数対数モデルでは、推定量のバイアスが大きいのを確認しています。これに対して、非線形最小 2 乗推定は不均一分散の仮定が正しければ推定精度は高い一方、不均一分散のもとでは推定精度が低くなっていました。そして、全般的に推定精度が優れていたのはポアソン疑似最尤(PPML)推定であることが確認されています。

例：横断面データを用いた実証分析

松浦寿幸氏のウェブサイトでは、1985 年の貿易データを使って、重力モデルを推定しています⁵。ここでは松浦氏の結果を再現して紹介します。データのサンプルサイズは 19602 ですが、貿易額が 0 をとるのは計 7540 となります。これは全体の 38%が 0 となることを意味しており、0 をとるサンプルを欠損値とすることの問題がわかります。

表 1 では、対数対数モデル、貿易額に 1 を足した対数対数モデル(表では対数対数(+1)と表記)、ポアソン回帰の推定結果を示しています。ここで、2 国間の貿易額を $Trade_{ij}$ 、輸出国の GDP を GDP_i 、輸入国の GDP を GDP_j 、2 国間の距離を $Distance_{ij}$ 、貿易協定ダミー(2 国間

³ Stata ならポアソン回帰は `poisson` や `ppmlhdfc` を使います。後者は高次の固定効果を考慮するのに便利です。

⁴ Tobit を使って推定する Eaton and Tamura (1994)の方法も検証されていますが、推定精度はあまり良くないことが確認されています。Eaton, Jonathan, and Akiko Tamura. "Bilateralism and regionalism in Japanese and US trade and direct foreign investment patterns." *Journal of the Japanese and international economies* 8.4 (1994): 478-510.

⁵ 松浦氏の note は「[R によるデータ分析入門]対数による回帰分析でエラーが出るときの対処法」では、データだけでなく、R による推定方法が解説されています。

に貿易協定があれば 1 となる)としています。貿易協定があれば、両国間の貿易額は増加するでしょうから、その係数はプラスになると予想されます。実証分析では、GDP だけでなく 1 人当たり GDP を追加するような推定もみかけますが、ここでは基本的なモデルだけを推定しています。

最初のモデルは対数対数モデルです。欠損値(貿易額が 0 をとる場合)が 7540 ありますので、サンプルサイズは 12062(=19602-7540)です。表 1 をみると、GDP の対数の係数は 1 に近く、距離の対数の係数は-1 に近くなっています。また、貿易協定ダミーの係数は 1.085 となります。しかし、OLS 推定はバイアスを持っていますから、この結果は信頼できないものです。

貿易額に 1 を足してから対数をとった対数対数モデルでは、GDP と距離の対数の係数は 0 に近くなる一方、貿易協定ダミーの効果は 1.474 と大きく評価されています。この推定では、欠損値はなくなっているため、サンプルサイズは 19602 に増加しています。しかし、この推定は、恣意的に 1 を足す操作によって得られたものであり、また、通常の OLS 推定と同様にバイアスを持つため、信頼できない推定結果になります。

ポアソン回帰の推定結果をみると、GDP の対数の係数は 1 より有意に小さく、距離の対数の係数も-1 より有意に 0 に近い値になっています。そして、貿易協定ダミーの効果は 0.442 と、他の推定結果よりも小さく評価されています。係数が 0.442 とは、貿易協定が結ばれることで貿易額が 56%($= 100 \times (e^{0.442} - 1)\%$)だけ増加することを意味します(詳しくは補足参照)。

表 1： 横断面データを用いた重力モデルの推定結果

	対数対数	対数対数(+1)	ポアソン
ln(GDPi)	0.902 *** (0.009)	0.475 *** (0.005)	0.763 *** (0.032)
ln(GDPj)	0.817 *** (0.009)	0.423 *** (0.005)	0.803 *** (0.040)
ln(Distance)	-1.118 *** (0.024)	-0.603 *** (0.014)	-0.737 *** (0.078)
貿易協定	1.085 *** (0.096)	1.474 *** (0.072)	0.442 *** (0.167)
定数項	-6.158 *** (0.246)	-1.763 *** (0.130)	-6.466 *** (0.864)
R2	0.5588	0.5491	0.8188
OBS	12062	19602	19602

注) また、***は 1%有意、**は 5%有意、*は 10%有意を表す。R2 は、対数対数と対数対数(+1)では決定係数、ポアソンでは疑似決定係数となる。また、OBS はサンプルサイズを表す。

この表から、どの推定方法を用いるかで、推定結果がかなり違うことが確認できました。

既に議論したとおり、対数対数モデルや貿易額に 1 を足す操作はバイアスを持っており、ポアソン回帰モデルの推定結果が最も信頼できるものになります。

4. 重力モデルを拡張する

基本の重力モデルでは、欠落変数が存在しており、それが推定結果にバイアスを生じさせる可能性があります。本節では、欠落変数として多角的貿易抵抗指数の存在を指摘し、それを考慮するために国別固定効果を含めた推定を紹介します。また、これまで横断面データを考えてきましたが、パネルデータの分析についても説明します。

4.1. 多角的貿易抵抗指数

Anderson and van Wincoop (2003)では、広い意味の物価を表す**多角的貿易抵抗**(multilateral resistance)**指数**の重要性を指摘しています⁶。たとえば、日本とアメリカとの間の貿易を考えましょう。日本は、中国から低価格な商品を輸入しており、それが日本の物価を引き下げるため、アメリカからの輸入が阻害されます。こうした多角的貿易抵抗指数を考慮しないなら、重力モデルの推定に欠落変数バイアスが生じることになります。

多角的貿易抵抗指数を考慮するため、次のモデルが正しいとします。

$$\text{Trade}_{ij} = A \frac{GDP_i^{\beta_1} GDP_j^{\beta_2}}{Distance_{ij}^{\beta_3}} e^{\theta_i + \varphi_j} + u_{ij}$$

多角的貿易抵抗指数の効果は $e^{\theta_i + \varphi_j}$ によって表されます。ここで、 θ_i は輸出国 i の固定効果、 φ_j は輸入国 j の固定効果となります。

このとき、モデルは次のようにも表現できます。

$$\text{Trade}_{ij} = e^{\ln(A) + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) - \beta_3 \ln(Distance_{ij}) + \theta_i + \varphi_j} + u_{ij}$$

つまり、貿易額の期待値の対数は、次の線形モデルとなります。

$$\ln(E[\text{Trade}_{ij}]) = \ln(A) + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) - \beta_3 \ln(Distance_{ij}) + \theta_i + \varphi_j$$

国別固定効果(θ_i , φ_j)は、ダミー変数によって置き換えることができます(この点はパネル分析の固定効果モデルと同じです)。輸出国の固定効果 θ_i は、ダミー変数(DE_{1i} , DE_{2i} , ..., DE_{ni})を使って次のように表せます。

$$\theta_i = \theta_1 DE_{1i} + \theta_2 DE_{2i} + \dots + \theta_n DE_{ni}$$

ただし、 DE_{1i} は輸出国が $i = 1$ なら 1 となるダミー変数、 DE_{2i} は輸出国が $i = 2$ なら 1 となるダミー変数です。同様に、輸入国の固定効果 φ_j は、ダミー変数(DI_{1j} , DI_{2j} , ..., DI_{nj})を使って次のように表現できます。

$$\varphi_j = \varphi_1 DI_{1j} + \varphi_2 DI_{2j} + \dots + \varphi_n DI_{nj}$$

⁶ Anderson, James E., and Eric Van Wincoop. "Gravity with gravitas: A solution to the border puzzle." *American economic review* 93.1 (2003): 170-192.

ただし、 DI_{1j} は輸入国が $j = 1$ なら 1 となるダミー変数、 DI_{2j} は輸入国が $j = 2$ なら 1 となるダミー変数です。

以上から、推定式は次のようになります。

$$\ln(E[\text{Trade}_{ij}]) = \beta_3 \ln(\text{Distance}_{ij}) + \text{輸出国ダミー} + \text{輸入国ダミー}$$

輸出国ダミーは、ダミー変数(DE_{1i} 、 DE_{2i} 、... DE_{ni})、輸入国ダミーは、ダミー変数(DI_{1j} 、 DI_{2j} 、... DI_{nj})を表しています。ここで GDP の対数は推定式から除かれています。これはダミー変数の線形関数として、時間を通じて一定の変数を表すことができるためです(多重共線性が生じていることは練習問題 2 を参照してください)。

4.2. パネルデータを用いた推定

4.1 節では、多角的貿易抵抗指数は時間を通じて一定としました。しかし、多角的貿易抵抗指数は時間を通じて変化すると考えるのが自然であり、パネルデータであれば、多角的貿易抵抗指数が時間を通じて変化することを考慮することができます。

ここで、モデルは次のようになります。

$$\ln(E[\text{Trade}_{ij}]) = \ln(A) - \beta_3 \ln(\text{Distance}_{ij}) + \theta_{it} + \varphi_{jt}$$

ここで、 θ_{it} は t 年の輸出国 i の固定効果、 φ_{jt} は t 年の輸入国 j の固定効果です。固定効果が時間を通じて変化することに注意してください。

各時点の固定効果(θ_{it} 、 φ_{jt})は、ダミー変数によって置き換えることができます。なお、輸出国の固定効果 θ_{it} を考慮するために、ダミー変数は計 $n \times T$ 個だけ必要です(n は国の数、 T は時点の数)。同様に、輸入国の固定効果 φ_{jt} を考慮するには、ダミー変数はやはり計 $n \times T$ だけ必要です。

重力モデルの推定では、多角的貿易抵抗以外の要因も重要でしょう。たとえば、貿易コストに影響するものとして、貿易協定ダミー(両国に貿易協定があれば 1 となる)、WTO ダミー(両国が世界貿易機関に加盟していたら 1 となる)、EU ダミー(両国が欧州連合に加盟していたら 1 となる)、植民地ダミー(過去に両国間で植民関係があったら 1 となる)、隣接ダミー(両国の国境が地続きなら 1 となる)、共通言語ダミー(両国の言語が同じなら 1 となる)などが挙げられます。

上記以外の変数化できないような要因も、貿易コストを決めるうえで重要かもしれません。たとえば、文化的な共通要因などが挙げられます。こうした要因は時間を通じて一定と考えられるため、ペア(i , j)に対応する固定効果 ϑ_{ij} によって表すことができます。

$$\ln(E[\text{Trade}_{ij}]) = \ln(A) - \beta_3 \ln(\text{Distance}_{ij}) + \theta_{it} + \varphi_{jt} + \vartheta_{ij}$$

ペア固定効果は、 i と j のペアなら 1 となるペアダミー D_{ij} を用いて表すことができます。ペアダミーは、 i と j の組み合わせの数だけ必要となります。

ペアダミーを含めるなら、説明変数として、距離の対数、植民地ダミー、接合ダミー、共

通言語ダミーを含めることはできません。これらの変数は時間を通じて一定であり、ペアダミーの線形関数として表すことができるからです。これに対して、貿易協定、WTO や EU の加盟国は時間を通じて変化するため、推定式に追加することが可能です。

以下では、実証分析を通じて理解を深めていきましょう。

例：パネルデータを用いた実証分析

Herman (2023)では、2000 年から 2018 年までの 150 カ国のデータを分析しています。データは 3 年毎に観察されるため、計 7 年分のデータになります(2000、2003、2006、2009、2012、2015、2018)⁷。これまでと同様に、2 国間の貿易額を Trade_{ij} 、輸出国の GDP を GDP_i 、輸入国の GDP を GDP_j 、2 国間の距離を Distance_{ij} とします。

表 2 は、ポアソン回帰で推定した結果を示しています。OLS 推定にはバイアスがあるため、ここでは推定していません。また、標準誤差はクラスターロバスト標準誤差を用いています。

(1)式では、ポアソン回帰によって重力モデルを推定しています。表をみると、すべての係数が 1 もしくは-1 の近くで推定されること、また、疑似決定係数が 0.8434 と高いことから、重力モデルがデータを上手に説明していることがわかります。

(2)式では、貿易協定ダミー、WTO ダミー、EU ダミー、植民地ダミー、隣接ダミー、共通言語ダミーを新たに追加しています。貿易協定ダミーを除いて、これらの係数は正の値をとっています。貿易協定の存在は、貿易コストを方向に働くため、係数が負という結果は予想と異なります。その原因として、多角的貿易抵抗指数が考慮されていない点が考えられます。

(3)式では、輸出国ダミーと輸入国ダミーを含めることで、多角的貿易抵抗指数を考慮しています。表をみると、貿易協定ダミーが正に有意になっており、予想どおりの結果にかかわっています。これに対し、植民地ダミーは有意ではなくなっています。これは過去に植民地の関係にあったとしても、貿易額は増加するとはいえないことを示しています。これに対し、共通言語ダミーの係数は 0.258 であり、両国の言語が同じなら貿易額が 26%ほど増加することがわかります。また、隣接ダミーの係数は 0.263 であり、国境が地続きで接しているなら、貿易額が 26%ほど増加します。なお、多重共線性のため、この推定では GDP の対数が除かれています。

最後に、(4)式では、ペア固定効果を考慮しており、貿易協定ダミー、WTO ダミー、EU ダミーはすべて有意に正となっています。この推定では、GDP の対数だけでなく、時間を通じて一定となる変数(植民ダミー、近隣ダミー、共通言語ダミー)はすべて除かれています。貿易協定ダミーの係数は 0.081 であり、貿易協定によって 8%ほど貿易額が増加します。同様に、EU ダミーの係数は 0.121 であり、両国が欧州連合に加盟することで貿易額が 12%ほ

⁷データの出所や再現プログラムは以下の論文を確認してください。Peter R. Herman (2023) “Gravity Estimation: Best Practices and Useful Approaches,” Economic Working Paper 2023-10-C.

ど増加しています。WTO ダミーの係数は 0.120 であり、両国が世界貿易機関に加盟することで貿易額が 12%ほど増加していることがわかります。この推定結果が、表 2 の中で最も信頼できるものになります。

表 2： パネルデータを用いた重力モデルの推定結果

	(1)式	(2)式	(3)式	(4)式
ln(GDPi)	0.954 *** (0.026)	0.923 *** (0.022)		
ln(GDPj)	0.953 *** (0.028)	0.923 *** (0.025)		
ln(Distance)	-1.040 *** (0.032)	-0.832 *** (0.036)	-0.733 *** (0.031)	
貿易協定		-0.047 (0.074)	0.249 *** (0.053)	0.081 *** (0.029)
EU		0.482 *** (0.089)	0.156 * (0.080)	0.121 *** (0.039)
WTO		0.730 *** (0.052)	0.542 *** (0.154)	0.120 ** (0.054)
植民		0.279 ** (0.136)	0.101 (0.064)	
隣接		0.269 ** (0.108)	0.263 *** (0.058)	
共通言語		0.375 *** (0.067)	0.258 *** (0.052)	
国別固定効果	No	No	Yes	Yes
ペア固定効果	No	No	No	Yes
Pseudo R2	0.8434	0.8634	0.9411	0.9917
OBS	117445	117445	150544	148251

注) また、***は 1%有意、**は 5%有意、*は 10%有意を表す。国別固定効果では、輸出国ダミーと輸入国ダミーを用いており、ペア固定効果では、さらにペアダミーを追加した推定となる。標準誤差はクラスターロバスト標準誤差を用いた。ポアソン回帰は最尤法で推定されている。Pseudo R2 は疑似決定係数である。OBS はサンプルサイズを表す。(1)式、(2)式の推定では定数項も含まれているが、表には記載されていない。

補足

非線形最小 2 乗推定とポアソン回帰の比較

記述の簡単化のため、被説明変数を y_i とし、説明変数を X_i とします。

$$y_i = e^{\alpha + \beta X_i} + u_i$$

上式の推定は、**非線形最小 2 乗法**で行うことができます。つまり、残差の 2 乗和である

$$\sum_{i=1}^n (y_i - e^{\alpha + \beta X_i})^2$$

を最小にするようにパラメータを決めるわけです。

α に関して偏微分して 0 と置くと、合成関数の微分の公式により、

$$\frac{\partial}{\partial \alpha} \sum_{i=1}^n (y_i - e^{\alpha + \beta X_i})^2 = -2 \sum_{i=1}^n (y_i - e^{\alpha + \beta X_i}) e^{\alpha + \beta X_i} = 0$$

となり、さらに β に関して偏微分して 0 と置くと次の式が得られます。

$$\frac{\partial}{\partial \beta} \sum_{i=1}^n (y_i - e^{\alpha + \beta X_i})^2 = -2 \sum_{i=1}^n (y_i - e^{\alpha + \beta X_i}) X_i e^{\alpha + \beta X_i} = 0$$

上式の両辺に -2 を割ると、2 つの条件式が得られます。

$$\begin{aligned} \sum_{i=1}^n (y_i - e^{\alpha + \beta X_i}) e^{\alpha + \beta X_i} &= 0 \\ \sum_{i=1}^n (y_i - e^{\alpha + \beta X_i}) X_i e^{\alpha + \beta X_i} &= 0 \end{aligned}$$

両式を満たしたパラメータが**非線形最小 2 乗推定量**となります。これは解析的には解けないため、数値探索法を用いて解を見つける必要があります(数値探索法は 12 章の補足参照)。

これに対し、ポアソン回帰では、対数尤度を最大にするようにパラメータを決めます。 α と β に関して対数尤度を偏微分すると、次の 2 式が得られます(式の導出はサポートウェブサイト追加資料「カウントデータ」参照)。

$$\begin{aligned} \sum_{i=1}^n (y_i - e^{\alpha + \beta X_i}) &= 0 \\ \sum_{i=1}^n (y_i - e^{\alpha + \beta X_i}) X_i &= 0 \end{aligned}$$

両式を満たすパラメータが最尤推定量となります。なお、残差を $y_i - e^{\alpha + \beta X_i}$ と考えると、第 1 式は残差の和が 0、第 2 式は残差と説明変数の積和が 0 を意味します。

非線形最小 2 乗法と最尤推定の条件式の違いは、非線形最小 2 乗法では、 $e^{\alpha + \beta X_i}$ が大きい観測値に高い加重を付けた推定になっている点です。しかし、既に確認したとおり、 $e^{\alpha + \beta X_i}$ が大きい観測値ほど、誤差項 u_i の分散が大きくなり、ノイズが多く含まれた観測値となっています。つまり、「非線形最小 2 乗推定量はノイズが多く含まれたデータに高い加重を課した不効率な推定になっている」わけです。これに対し、ポアソン回帰では、「すべての観測値に同じ加重を付けた推定になっており」、より良い推定方法であることがわかります。

シミュレーション

シミュレーションとは、分析者が真のモデルを知っているとし、そのモデルからデータを発生させることで、統計量の性質を明らかにする方法です。

Silva and Tenreyro (2006)では、真のモデルを次のように設定しました。

$$Y_i = e^{\alpha + \beta_1 X_{1i} + \beta_2 X_{2i}} \eta_i$$

ここで、 X_{1i} は標準正規分布 $N(0,1)$ 、 X_{2i} はベルヌーイ確率変数(1となる確率 p は0.4)、 η_i は対数正規確率変数(期待値1、分散 σ_i^2)としています。また、パラメータは $\alpha = 0$ 、 $\beta_1 = 1$ 、 $\beta_2 = 1$ と設定しています。 η_i は対数正規分布なので正の値をとり、分散 σ_i^2 は均一分散だけでなく、3 パターンの不均一分散を考えています⁸。

第1に、これらの分布から、PC を用いて X_{1i} 、 X_{2i} 、 η_i を生成します。そして、モデル

$$Y_i = e^{X_{1i} + X_{2i}} \eta_i$$

を用いて Y_i を生成します。このモデルからサンプルサイズ 1000 のデータを生成します。

第2に、生成したデータを使って、様々な推定方法でパラメータ(α 、 β_1 、 β_2)を推定し、それぞれの推定結果を記録します。推定方法としては、通常対数対数モデルによる OLS 推定、貿易額に1を足したうえでの対数対数モデル、非線形最小2乗法、ポアソン回帰などがあります。

第1と第2のプロセスを10000回繰り返します。そして、各推定方法について、パラメータの推定精度を求めます。推定精度を測る指標として、バイアスの程度(推定値と真の値との差)、推定量の標準偏差を用います。

Silva and Tenreyro (2006)の Table 1 には、シミュレーションの結果がまとめられており、それをみると、ポアソン疑似最尤推定(PPML)のパフォーマンスが優れていました。興味がある方は論文をチェックしてください。

ポアソン回帰の係数の解釈

重力モデルのポアソン回帰は、説明変数が対数をとるか、実数をとるかによって解釈が異なります。説明変数が対数なら対数対数モデル、実数なら対数線形モデルと同じように解釈ができます。

説明変数が対数としましょう(GDPや距離に該当する)。このとき、次式が成立します。

$$\ln(E[Y]) = \alpha + \beta \ln(X) + u$$

係数 β の意味を考えます。 X が X' に変化したとき、 Y は Y' に変化するとします。このとき、モデルから、次式が成立します。

⁸ 対数正規分布は、その対数が正規分布する確率変数となります。たとえば、 u が正規分布とすると、 e^u は対数正規分布に従います(e^u の対数 $\ln(e^u)$ は正規確率変数 u になります)。対数正規確率変数は、指数関数になっているため、必ず0より大きな値を取ります。

$$\begin{aligned}\ln(E[Y']) - \ln(E[Y]) &= \underbrace{(\alpha + \beta \ln(X'))}_{Y'} - \underbrace{(\alpha + \beta \ln(X))}_Y \\ &= \beta(\ln(X') - \ln(X))\end{aligned}$$

この関係式を β について解くと、次の式となります。

$$\beta = \frac{\ln(E[Y']) - \ln(E[Y])}{\ln(X') - \ln(X)} = \frac{\ln\left(1 + \frac{E[Y'] - E[Y]}{E[Y]}\right)}{\ln\left(1 + \frac{X' - X}{X}\right)} \approx \frac{\frac{E[Y'] - E[Y]}{E[Y]}}{\frac{X' - X}{X}}$$

最後の近似($\approx$)は、変化率は 0 に近い値をとることが多いためです。これは、「 X が 1%変化すると、 Y の期待値は $\beta\%$ 変化する」ことを意味します。係数 β は、「 X に対する Y の期待値の弾力性(elasticity)」です。

次に、説明変数が実数とします。

$$\ln(E[Y]) = \alpha + \beta X + u$$

X が0から 1 に変化したとき、 Y は Y' に変化するとします。このとき、

$$\ln(E[Y']) - \ln(E[Y]) = (\alpha + \beta) - \alpha = \beta$$

となり、この式を β について解くと、次の式となります。

$$\beta = \ln(E[Y']) - \ln(E[Y]) = \ln\left(1 + \frac{E[Y'] - E[Y]}{E[Y]}\right) \approx \frac{E[Y'] - E[Y]}{E[Y]}$$

最後の近似($\approx$)は、変化率は 0 に近い値をとることが多いためです。

上式は、 Y の変化が小さいときの近似ですが、 Y の変化が大きいとき近似は不正確です。近似を使わないでも、 $\beta = \ln(E[Y']) - \ln(E[Y])$ は成立するため、

$$e^\beta = \frac{E[Y']}{E[Y]}$$

となります(この式が正しいことは両辺の対数をとったら明らかです)。つまり、 X が 1 単位変化すると、期待値が $100 \times (e^\beta - 1)\%$ 変化します。なお、 Y の変化が 30%程度までは近似でも問題ありませんが、それを超えたら、こちらを使うことをお勧めします。

練習問題

1. 特殊ケースでは、説明変数と誤差項は無相関になる。これは、

$$u_{ij} = A \frac{GDP_i^{\beta_1} GDP_j^{\beta_2}}{Distance_{ij}^{\beta_3}} v_{ij}$$

とし、確率変数 v_{ij} が説明変数と独立とした場合となる。このとき、

$$\eta_{ij} = 1 + v_{ij}$$

となり、 $\ln(\eta_{ij})$ は説明変数と無相関になる。これは特殊ケースであり、一般には成立しないことを述べよ。

2. $\ln(GDP_i)$ がダミー変数(DE_{1i} 、 DE_{2i} 、... DE_{ni})の線形関数になることを示せ。これは 5 章で学習した多重共線性を意味する。

問題 1 の答え

これは誤差項 u_{ij} の分散が、次のように書けることを意味する。

$$\left(A \frac{GDP_i^{\beta_1} GDP_j^{\beta_2}}{Distance_{ij}^{\beta_3}} \right)^2 V(v_{ij})$$

これは分散が次のようになることを意味する(上式と下式は対数をとったら同じことを確認できる)。

$$e^{2[\ln(A) + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) - \beta_3 \ln(Distance_{ij})]} V(v_{ij})$$

これは特殊ケースであり、一般には成立しない。議論の中で、誤差項 u_{ij} の分散が、 $e^{\ln(A) + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) - \beta_3 \ln(Distance_{ij})}$ に比例していると指摘しているが、これは乗数に 2 を掛けた $e^{2[\ln(A) + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) - \beta_3 \ln(Distance_{ij})]}$ に比例することとは異なる。

問題 2 の答え

GDP の対数は次のように表すことができる。

$$\ln(GDP_i) = \ln(GDP_1) DE_{1i} + \ln(GDP_2) DE_{2i} + \dots + \ln(GDP_n) DE_{ni}$$

ただし、 DE_{1i} は輸出国が $i = 1$ なら 1 となるダミー変数、 DE_{2i} は輸出国が $i = 2$ なら 1 となるダミー変数です。ここで、 $c_1 = \ln(GDP_1)$ 、 $c_2 = \ln(GDP_2)$ 、...、 $c_n = \ln(GDP_n)$ とすれば、 $\ln(GDP_i)$ はダミー変数の線形関数である。

$$\ln(GDP_i) = c_1 DE_{1i} + c_2 DE_{2i} + \dots + c_n DE_{ni}$$