

修論進捗メモ

萱場 悠貴

2018 年 12 月 20 日

1 2018/11/16 発表

1.1 修正部分など

- 分析対象を拡張
分析対象を東京、大阪、福岡の 3 都市から東京、大阪、福岡、札幌、那覇の 5 都市に拡張した。
これは反実仮想シミュレーションにおいて北海道新幹線開通の影響を見るためである。
- 運賃パラメータの符号
関数 Objective で運賃パラメータの正負が入れ替わっているというミスを発見したので修正した。
このミスの発見により、正しく最適化できていたということが示唆されたがその一方で効用関数推定において運賃パラメータが正になっていることが重荷としてのしかかってくるようになった笑
- JR 部分の需要
シミュレーション結果において、鉄道経路の選択確率は実績値と乖離するようにプログラムが組まれているが、需要量については実績値のまま表示されているのでそこは修正する必要がある。
- JR 部分の運賃
シミュレーションを実行して simulationresults.csv が生成されるわけだが、それが生成される前に JR の運賃を 2 倍する必要がある。航空運賃が往復運賃、JR が片道運賃で入っていたところを半分にしてからシミュレーション行っているのでこの処置が必要となる。現状では別途組んだエクセルのマクロで対処されているが、それだとグラフを描くときに 2 倍が反映されないなのでその前の段階で修正する。
- 運賃の 2 次の項
前回の発表で運賃の項を 2 次で入れて上に凸の形で表現するのはどうかというアドバイスがあったが、Supply Analysis 4 のほうもそれに合わせた修正が必要になる。具体的には、効用関数が 9 つの項からなるのでそれを変更するのと、関数 Objective などにもそれを反映させる必要があるということに留意する必要がある。
→反映させた。やはり暴走する（均衡運賃は 7 万円台、二次関数のピークは 67875）

1.2 反実仮想シミュレーションについて考える

ここから反実仮想シミュレーションについて考える。ポイントをつらつらと書く。

- シミュレーション対象

とりあえず北海道新幹線札幌開業を考える。他の候補は新幹線延伸、LCC の参入くらいしか思いついていないのでなんか考えなければ。経済厚生の変化とかに落とせたりしないかな。

- 用意すべきもの
 - － 仮定の所要時間、運賃、滞在可能時間、アクセシビリティ（これは現実のをぶち込めばよし）
所要時間は東京－札幌間で 5 時間 14 分、指定席特急券が東京－札幌で 8090 円、運賃が 12910 円である。滞在可能時間は 24 時間－（5 時間 14 分）× 2 = 13 時間 32 分とする
 - － 需要をどんな感じで代替するか（OD は fix、それとも何らかの方法で増やす？）

1.3 中間報告を受けて

とりあえず新たに出てきた課題とか先生からのアドバイスとか。

- 課題はやはり運賃パラメータの符号条件
運賃パラメータの符号を負に差し替えれば基本的に Base Case、Counterfactual ともにシミュレーションはうまくいっているのどうにかして効用関数推定で運賃パラメータを負に出す必要がある。
考えられる方策として、
 - － 現在採用している変数を除く、あるいは新たな変数を採用する
これが一番やりやすそうではある。ただデータの収集が結構大変だし、仮に効用を運賃に単回帰したところで負に出るかどうかわからんから確かめるならまずそこからという感じはある。
 - － それでいて尚且つ符号制約をかける
先々週に符号制約をかけたときは 0 に飛んで行ってしまったが、変数選択等を変えれば負の領域で収まる可能性はある。
 - － 線形確率モデルを採用する
これよくわからんし先生もうまくいかなそうって言ってたからあまりやる意味はなさそう
 - － ロジットモデルをやめる
入れ子ロジットとかランダム係数ロジットに変更すれば論文としてはより高度になるみたいなどこはある。ただ時間かけてやった挙句正負が反転しないということは十分にありえてしまうのでリスクかなあ。もう一年あればやってもよかったけど、、
ぶっちゃけここまで来たら気合いしかないのでゴリゴリやるしかないですね。
- データの追加
北海道新幹線延伸開業で反実仮想をやったわけだけど東京札幌間ですら割と距離があるのであまりシミュレーションのうまみを感じられなかった。そこで東京札幌間の各都市（仙台、青森など）をシミュレーションに含めることが検討されるわけだが、一地点追加するだけで結構大変ということが発覚したのでこれはやるなら早めに動いたほうがよさそう。アクチュアリー試験あるからあんま修論やってる場合ではないんだがなあ、

取りあえずパッと思いつくのはこのくらい。1.1 節に書いたこともまだ治ってなかったりなのでそれはちゃんとやったほうが良い。

2 2018/11/21 発表

2.1 修正進め方

もはや符号条件に合致した効用関数パラメータを得るしかない状況なので、いろいろ頑張っていきたい。その前に、どんな変数が入っていてどのような形でデータが採られているかをまず整理したい。

- 所要時間

航空、鉄道経路のそれぞれの所要時間（分）。航空に関しては、JTB 時刻表から代表できそうな所要時間をとってきている。エアラインごとに丸めているが、エアラインごとに所要時間は変わっていないので特にその部分で問題はなさそう。ただ効用関数推定においては東京エリアは成田空港、大阪エリアは関西空港というように大都市圏は代表空港を定めてしまっているその部分は確か単純平均をとっていたような気がする。航空便数を反映して加重平均をとったほうが良いかも。鉄道の方はよくわからんサイトからとってきた標準的な所要時間を入れているのであまり、というか本当に良くない。ただ割と所要時間に関しては大嘘というわけではなく、いくつか実際の時刻表と照合してもだいたい同じだったのでいったん問題はないかと。

- 費用

航空の方は所要時間と同じ方法で処理している。すなわちエアラインごとに異なる運賃を単純平均して算出している。問題の根幹はここなので修正すべきであろう。少なくとも航空便数で加重平均をとるくらいのはしたほうがよさそう。鉄道の方はもっとひどくて、駅間距離を測定し、その距離によって普通運賃を計算し、そんでそれを 2 倍している笑（笑い事ではない）。というわけでここは絶対に修正しなければならないのだが、どうやって修正すべきかはちょっとまだ見えていない。最悪の場合ひと駅ひと駅運賃を計算しないといけないかもしれない。

- 航空便数の対数

航空便数は丸める部分は総和するより外に仕方がないのでいったんオクケーといえる。ただ、年間の航空便数を入れたいがために単純に 365 倍しているのをどうすべきかは考えなくてはならない。あとそのせいで鉄道の方の航空便数が 365 本になってしまっている問題を解決しないといけない。

- 滞在可能時間

航空の方はちゃんとデータをとってるのでこれは良さそう。でも鉄道のほうは始発が 6 時、終電の到着が 24 時とし、そこから所要時間の 2 倍を引いてるので割とカスである。これも調べるの大変なのでどうしようか、という感じである

- アクセシビリティ指標

これを距離で妥協するのは仕方ないとして、イグレスの方のアクセシビリティを入れていないのは問題といえる。修正しよう。

- 鉄道ダミー

これはさすがに大丈夫。

- エアライン参入数

これどうやったか忘れたけどちゃんとやってた気がするのでいったん大丈夫でしょう。

こうしてみるとなかなかひどい状況なので修正大変そう。頑張って 4 時くらいには帰るようにしたい。

3 2018/11/30 発表

3.1 運賃パラメータ問題についての検討

運賃パラメータが正に出る問題についていろいろ考える。

まず運賃と選択確率の関係を可視化する。グラフは以下：

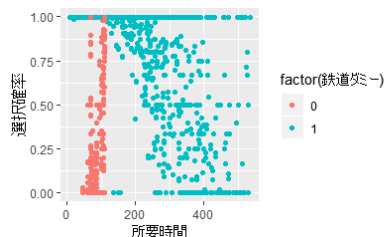


図1 所要時間と選択確率

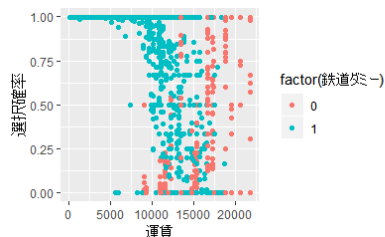


図2 運賃と選択確率

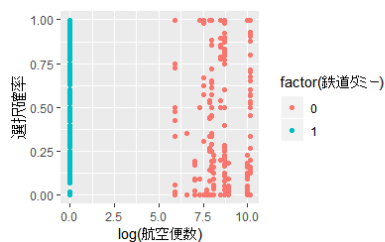


図3 log(航空便数)と選択確率

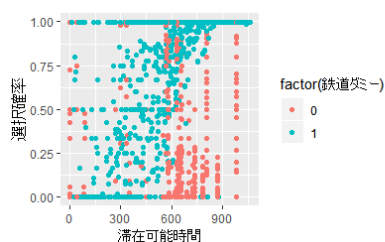


図4 滞在可能時間と選択確率

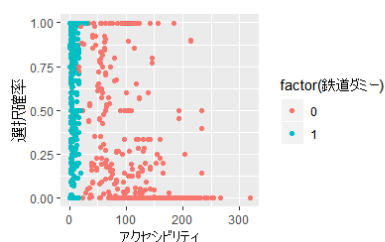


図5 アクセシビリティと選択確率

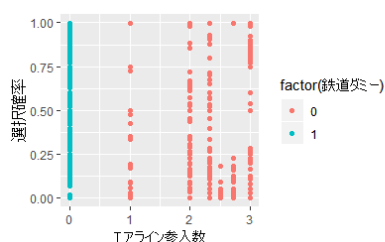


図6 エアライン参入数と選択確率

図1をみると、運賃と選択確率は正の相関を持つことが示唆される。この原因として以下が考えられる。

- 近距離においては自動車の選択確率が高い

本論文ではOD間の移動需要のうち、航空と鉄道の選択確率の実績値を求めている。残りの交通手段に関しては、いわばoutside optionという扱いにしている。この場合、比較的近距離の移動に関しては航空便が供給されていないことが多い。また鉄道に関しても選択確率が低く、自動車を含めたoutside optionの選択確率が高くなってしまっていることが示唆される。

相関をみる

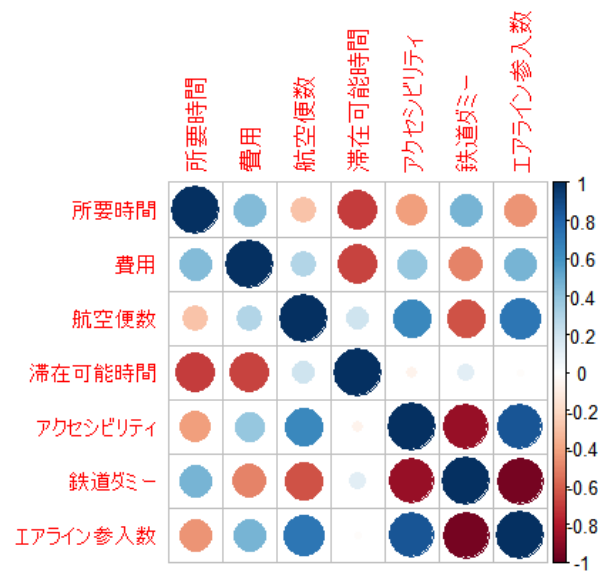


図 7 相関行列

| | 所要時間 | 費用 | 航空便数 | 滞在可能時間 | アクセシビリティ | 鉄道ダミー | エアライン参入数 |
|----------|--------|--------|--------|--------|----------|--------|----------|
| 所要時間 | 1.000 | 0.439 | -0.287 | -0.691 | -0.412 | 0.469 | -0.442 |
| 費用 | 0.439 | 1.000 | 0.291 | -0.679 | 0.382 | -0.491 | 0.469 |
| 航空便数 | -0.287 | 0.291 | 1.000 | 0.209 | 0.642 | -0.631 | 0.730 |
| 滞在可能時間 | -0.691 | -0.679 | 0.209 | 1.000 | -0.061 | 0.124 | -0.018 |
| アクセシビリティ | -0.412 | 0.382 | 0.642 | -0.061 | 1.000 | -0.866 | 0.858 |
| 鉄道ダミー | 0.469 | -0.491 | -0.631 | 0.124 | -0.866 | 1.000 | -0.953 |
| エアライン参入数 | -0.442 | 0.469 | 0.730 | -0.018 | 0.858 | -0.953 | 1.000 |

最終的な推定結果は以下の通り：

表 1

| | <i>Dependent variable:</i> | | | | | |
|-------------------------|----------------------------|------------------------|-----------------------|-------------------------|-------------------------|-------------------------|
| | y | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 費用 | −0.0004*** (0.00002) | 0.0001*** (0.00003) | 0.0001** (0.00003) | −0.0004*** (0.00002) | −0.0004*** (0.00002) | −0.0004*** (0.00002) |
| log(航空便数) | −0.649*** (0.024) | −1.308*** (0.044) | −0.890*** (0.025) | −0.093** (0.044) | 0.234* (0.121) | −1.351*** (0.106) |
| 所要時間 | | −0.023*** (0.001) | | | | |
| 滞在可能時間 | | | 0.010*** (0.0004) | | | |
| アクセシビリティ | | | | −0.041*** (0.003) | | |
| 鉄道ダミー | | | | | 7.673*** (1.035) | |
| エアライン参入数 | | | | | | 2.643*** (0.387) |
| Constant | 7.636*** (0.183) | 8.177*** (0.176) | −2.517*** (0.488) | 8.169*** (0.180) | −0.093 (1.058) | 7.622*** (0.182) |
| Observations | 2,588 | 2,588 | 2,588 | 2,588 | 2,588 | 2,588 |
| R ² | 0.455 | 0.514 | 0.542 | 0.498 | 0.467 | 0.465 |
| Adjusted R ² | 0.455 | 0.514 | 0.542 | 0.497 | 0.466 | 0.464 |

Note:

*p<0.1; **p<0.05; ***p<0.01

表 2

| <i>Dependent variable:</i> | | | | | | |
|----------------------------|-------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| | y | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 費用 | −0.0004*** (0.00002) | 0.0001*** (0.00003) | 0.0002*** (0.00003) | 0.0002*** (0.00003) | 0.0002*** (0.00004) | 0.0002*** (0.00004) |
| log(航空便数) | −0.649*** (0.024) | −1.308*** (0.044) | −1.203*** (0.042) | −0.751*** (0.056) | 0.247 (0.263) | −0.742** (0.310) |
| 所要時間 | | −0.023*** (0.001) | −0.013*** (0.001) | −0.012*** (0.001) | −0.024*** (0.003) | −0.021*** (0.003) |
| 滞在可能時間 | | | 0.008*** (0.0005) | 0.007*** (0.0005) | 0.002* (0.001) | 0.003** (0.001) |
| アクセシビリティ | | | | −0.031*** (0.003) | −0.031*** (0.003) | −0.031*** (0.003) |
| 鉄道ダミー | | | | | 10.413*** (2.680) | 6.421** (2.745) |
| エアライン参入数 | | | | | | 2.179*** (0.366) |
| Constant | 7.636*** (0.183) | 8.177*** (0.176) | −0.068 (0.551) | 1.195** (0.547) | −4.130*** (1.475) | −1.223 (1.545) |
| Observations | 2,588 | 2,588 | 2,588 | 2,588 | 2,588 | 2,588 |
| R ² | 0.455 | 0.514 | 0.556 | 0.580 | 0.582 | 0.588 |
| Adjusted R ² | 0.455 | 0.514 | 0.556 | 0.579 | 0.581 | 0.587 |

Note:

*p<0.1; **p<0.05; ***p<0.01

表 3

| | <i>Dependent variable:</i> | | | |
|-------------------------|----------------------------|-----------------------|------------------------|------------------------|
| | y | | | |
| | (1) | (2) | (3) | (4) |
| 費用 | 0.0001** (0.00003) | −0.00000 (0.00003) | 0.0001*** (0.00003) | 0.0001*** (0.00003) |
| log(航空便数) | −0.890*** (0.025) | −0.441*** (0.045) | −1.349*** (0.149) | −2.292*** (0.203) |
| 滞在可能時間 | 0.010*** (0.0004) | 0.009*** (0.0004) | 0.011*** (0.001) | 0.011*** (0.001) |
| アクセシビリティ | | −0.031*** (0.003) | −0.031*** (0.003) | −0.031*** (0.003) |
| 鉄道ダミー | | | −7.368*** (1.159) | −9.890*** (1.208) |
| エアライン参入数 | | | | 2.475*** (0.366) |
| Constant | −2.517*** (0.488) | −1.134** (0.488) | 4.146*** (0.961) | 6.512*** (1.015) |
| Observations | 2,588 | 2,588 | 2,588 | 2,588 |
| R ² | 0.542 | 0.567 | 0.573 | 0.581 |
| Adjusted R ² | 0.542 | 0.566 | 0.572 | 0.580 |

Note:

*p<0.1; **p<0.05; ***p<0.01

表 4

| | <i>Dependent variable:</i> | | |
|-------------------------|----------------------------|-------------------------|-------------------------|
| | y | | |
| | (1) | (2) | (3) |
| 費用 | −0.0004*** (0.00002) | −0.0004*** (0.00002) | −0.0004*** (0.00002) |
| log(航空便数) | −0.093** (0.044) | 0.649*** (0.120) | −0.139 (0.188) |
| アクセシビリティ | −0.041*** (0.003) | −0.039*** (0.003) | −0.040*** (0.003) |
| 鉄道ダミー | | 6.608*** (0.999) | 4.615*** (1.060) |
| エアライン参入数 | | | 2.133*** (0.394) |
| Constant | 8.169*** (0.180) | 1.496 (1.024) | 3.501*** (1.084) |
| Observations | 2,588 | 2,588 | 2,588 |
| R ² | 0.498 | 0.506 | 0.511 |
| Adjusted R ² | 0.497 | 0.505 | 0.511 |

Note:

*p<0.1; **p<0.05; ***p<0.01

表 5 推定結果

| coef | 定数項 | 費用 | log(航空便数) | アクセシビリティ | 鉄道ダミー |
|------|--------|---------|-----------|----------|--------|
| OLS | 1.4962 | -0.0004 | 0.6485 | -0.0393 | 6.6077 |
| IV | 3.9594 | -0.0006 | 0.6813 | -0.0421 | 5.6476 |

表 6 Estimation Results

| coef | 定数項 | 費用 | log(航空便数) | アクセシビリティ | 鉄道ダミー | 一便あたり座席数 | 一便あたり貨物 | 所要時 |
|------|---------|---------|-----------|----------|---------|----------|---------|-------|
| OLS | -2.5210 | 0.0001 | 0.1591 | -0.0298 | 11.4224 | 0.0090 | 0.0015 | -0.02 |
| IV | 1.6419 | -0.0003 | 0.0777 | -0.0334 | 7.8231 | 0.0151 | 0.0017 | -0.01 |

表 7 Estimation Results

| coef | OLS | IV |
|----------------|---------|---------|
| intercept | -2.5210 | 1.6419 |
| price | 0.0001 | -0.0003 |
| log(frequency) | 0.1591 | 0.0777 |
| acccessibility | -0.0298 | -0.0334 |
| railwaydummy | 11.4224 | 7.8231 |
| seat | 0.0090 | 0.0151 |
| freight | 0.0015 | 0.0017 |
| duration | -0.0236 | -0.0113 |

表 8

| | <i>Dependent variable:</i> | |
|-------------------------|----------------------------|-------------------------|
| | y | |
| | OLS | IV |
| | (1) | (2) |
| 費用 | 0.0001** (0.00004) | −0.0003*** (0.00005) |
| log(航空便数) | 0.159 (0.129) | 0.078 (0.129) |
| アクセシビリティ | −0.030*** (0.002) | −0.033*** (0.002) |
| 鉄道ダミー | 11.422*** (1.163) | 7.823*** (1.191) |
| 一便あたり座席数 | 0.009*** (0.002) | 0.015*** (0.002) |
| 一便あたり貨物 | 0.002*** (0.0003) | 0.002*** (0.0003) |
| 所要時間 | −0.024*** (0.001) | −0.011*** (0.002) |
| Constant | −2.521** (1.176) | 1.642 (1.215) |
| Observations | 2,588 | 2,588 |
| R ² | 0.601 | 0.606 |
| Adjusted R ² | 0.600 | 0.604 |

Note: *p<0.1; **p<0.05; ***p<0.01