

Estymacja bayesowska modelu objaśniającego wolumen sprzedaży ogórków w USA

Dane

Do estymacji zostały użyte te same dane, co w poprzedniej pracy domowej. W skrócie modelowany jest wolumen sprzedaży ogórków w zależności od ceny ogórków, pomidorów i cebuli. Dane zostały zlogarytmowane w celu interpretacji jako elastyczności i obejmują lata 1970-2021.

Elicytacja parametrów a priori

Strona departamentu rolnictwa umożliwia dostęp do oszacowań elastyczności w formie przejrzystej tabeli¹, to z niej zostały zaczerpnięte poniższe parametry a priori:

Towar	Rodzaj elastyczności	Oszacowanie
Ogórki	Własna	-0,57
Pomidory	Krzyżowa	-0,04
Cebula	Krzyżowa	-0,11

Badania były robione w latach 1970-1992 i 1960-1993, czyli lata się na siebie nakładają. Dodatkowo po zgłębieniu prac okazuje się, że dane o cenach były zbierane z kilku źródeł i w paru latach nie było danych. Biorąc pod uwagę, że od badań minęło też 30 lat należy zachować pewną dozę ostrożności.

Wydaje się, że nie da się zdobyć wiedzy a priori o stałej w modelu, jednak jest możliwe pewne ograniczenie. Strona departamentu rolnictwa wskazuje, że parametry regresji będą ujemne. Dodatkowo oryginalne dane są dość duże (w sensie liczby), czyli logarytmy będą dodatnie. Sama zmienna objaśniana też będzie dodatnia. W związku z tym bardziej prawdopodobne jest, że stała powinna przyjąć wartości dodatnie. Oczywiście może się okazać, że pomidory i cebula są substytutami i mocno reagowałyby na zmiany ceny. W związku z tym dla stałej zostanie przyjęty rozkład jednostajny na przedziale $[-5, 25]$.

Ogórki prawie na pewno nie są dobrem luksusowym oraz nie spodziewamy się elastyczności w okolicy -2.5. Ponadto sensownym wydaje się, aby znacząca część gęstości znajdowała się w przedziale $[-1.2, 0]$ (przyjmujemy symetryczny rozkład). Aby znacząco ograniczyć prawdopodobieństwo w ogonach trzeba je „wyszczuplić”. Można to zrobić używając rozkładu, który ma większą kurtozę od rozkładu normalnego, dodatkowo automatycznie zwiększy się masa prawdopodobieństwa wokół wartości oczekiwanej. Biorąc pod uwagę wcześniejszy wywód zastosowany zostanie rozkład logistyczny. Aby w przedziale $[-1.2, 0]$ znajdowało się 95% masy parametr skali musi wynosić ~ 0.15 (średnia zostaje ustawiona na -0.57, zgodnie z literaturą).

Oszacowania elastyczności wskazywałyby, że pomidory i cebula są dobrami komplementarnymi, jednak sensowne wydaje się również patrzenie na te produkty jak na substytuty (w szczególności pomidory). W związku z tym, a także z faktem, że oszacowania empiryczne są bliskie 0, niech więc

¹ https://data.ers.usda.gov/reports.aspx?ID=17825#P3809bf1ae6a54082a55a1a41eb7ac5f9_4_214, dostęp: 20.04.2023

95% masy gęstości zawiera się w przedziale $(-0.7, 0.7)$ (odpowiednio przesuniętym dla produktu). Nie mając pewności co do postaci rozkładu, a jednocześnie chcąc wyrazić pewne zaufanie do literatury, dobrym kompromisem wydaje się użycie rozkładu normalnego (nie ma żadnego powodu, aby go zmieniać). W związku z tym dla parametrów regresji pomidorów i cebuli zostanie użyty rozkład normalny z odpowiednią średnią i odchyleniem standardowym 0.36.

Do oszacowania modelu potrzebne jest jeszcze oszacowanie odchylenia standardowego, jednak okazuje się, że bardzo często w ekonometrii bayesowskiej używany jest rozkład odwrotny gamma. Biorąc pod uwagę postać modelu, można przypuszczać, iż odchylenie standardowe nie powinno wynieść wartości większej niż 1 (ze względu na skalę pomiaru zmiennej objaśnianej). Sensownym wydaje się zatem użycie rozkładu odwrotnego gamma z parametrami 3 i 1.

Finalnie parametry a priori charakteryzują następujące rozkłady:

- stała $\sim U(-5, 25)$
- cena ogórków $\sim \text{Logistyczny}(-0.57, 0.15)$
- cena pomidorów $\sim \text{Normalny}(-0.04, 0.36)$
- cena cebuli $\sim \text{Normalny}(-0.11, 0.36)$
- odchylenie standardowe $\sim \text{Odwrotny Gamma}(3, 1)$

Dodatkowo przy estymacji założony został normalny rozkład składnika losowego, podobnie jak w przypadku dokonywania prognozy.

Wyniki estymacji

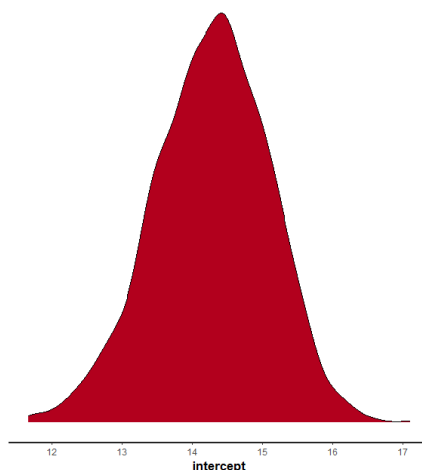
Wartości oczekiwane a posteriori

Wartości oczekiwane parametrów a posteriori prezentują się następująco:

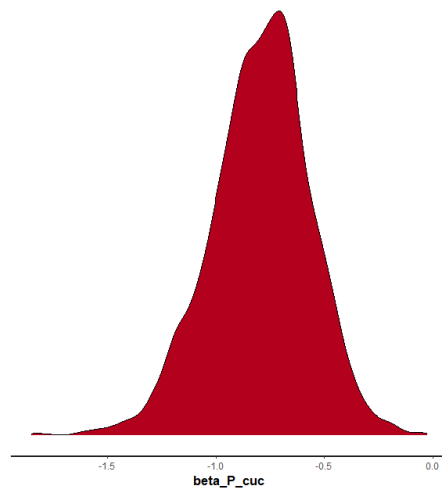
Stała	14.27
Cena ogórków	-0.79
Cena pomidorów	-0.66
Cena cebuli	-0.66

Rozkłady brzegowe parametrów a posteriori

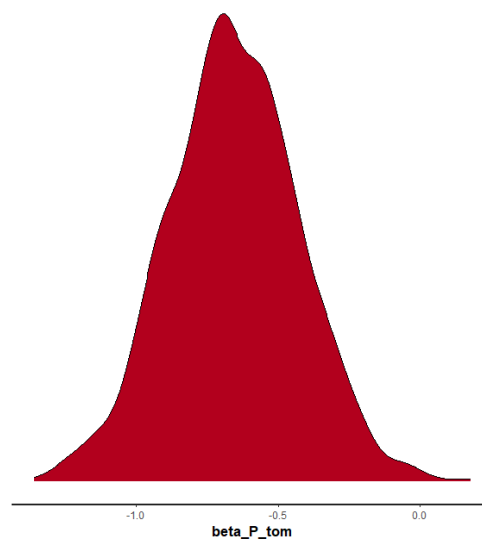
Stała:



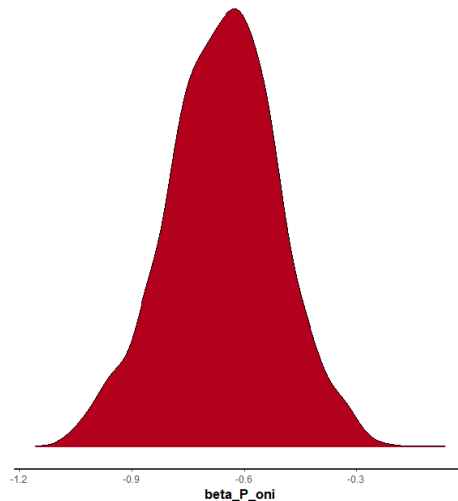
Cena ogórków:



Cena pomidorów:



Cena cebuli:



Domyślnie Stan nie umieszcza osi Y, ale przy rozkładach gęstości nie jest ona aż tak ważna, bardziej chodzi o kształt rozkładu.

Znaczenie zmiennych

HPDI

HPDI można odczytać z wydruku ze Stana:

```
Inference for Stan model: ogorki_full.
4 chains, each with iter=1000; warmup=500; thin=1;
post-warmup draws per chain=500, total post-warmup draws=2000.
```

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
beta_P_cuc	-0.79	0.01	0.23	-1.27	-0.94	-0.78	-0.65	-0.38	1118	1
beta_P_tom	-0.66	0.01	0.23	-1.11	-0.81	-0.66	-0.50	-0.21	927	1
beta_P_oni	-0.66	0.00	0.15	-0.96	-0.75	-0.65	-0.56	-0.36	1314	1
intercept	14.27	0.03	0.81	12.60	13.74	14.31	14.84	15.72	932	1
sigma	0.26	0.00	0.03	0.21	0.24	0.26	0.28	0.32	1284	1
lp__	40.03	0.06	1.72	35.63	39.15	40.45	41.30	42.18	843	1

95% przedział ufności dla parametrów regresji odczytujemy patrząc na kolumny 2.5% i 97.5%, widać że wszystkie parametry regresji są istotnie równe od 0 na tym przedziale.

Czynniki Bayesa

Czynniki Bayesa dla poszczególnych zmiennych prezentują się następująco*:

Zmienna	Czynnik Bayesa
Cena ogórków	12
Cena pomidorów	15
Cena cebuli	836

*Oszacowania mogą się różnić w zależności od wywołania Stana, wartości są orientacyjne

Podsumowanie znaczenia zmiennych

Czynniki Bayesa dla 2 pierwszych zmiennych są predyktorami mającymi znaczenie (Kass i Raftery), a cena cebuli polepsza model ponad wszelką wątpliwość. Przedziały HDPI są tutaj zgodne z czynnikami Bayesa. (dodatkowo widać, że parametr ceny cebuli ma najmniejsze odchylenie standardowe)

Prognozy

Do prognozy została wybrana 20 obserwacja, która prezentuje się następująco:

Zmienna	Wolumen ogórki	Cena ogórki	Cena pomidory	Cena cebula
Wartość	7,14	3,32	3,99	2,92

Prognoza z modelu dla 95% przedziału ufności wygląda następująco:

Dolny próg	Prognoza punktowa	Górny próg
6,56	7,09	7,61

Prognoza z modelu dla 50% przedziału ufności wygląda następująco:

Dolny próg	Prognoza punktowa	Górny próg
6,91	7,09	7,26

Dla obu przedziałów ufności prawdziwa wartość zmiennej się w nich zawiera. Na podstawie jednej obserwacji trudno się wypowiadać o zdolności predykcyjnej modelu, ale wynik jest obiecujący.

Podsumowanie modelu*

Wydruk ze Stana sugeruje, że model został oszacowany poprawnie. Zbieżność osiągnięta jest już dla 1000 iteracji, co jest bardzo dobrym wynikiem.

Model obliczony numerycznie wypada znacząco lepiej niż szacowany analitycznie, mimo że zmieniły się tylko rozkłady stałej i parametru przy cenach ogórków oraz odchylenie standardowe było szacowane. Odchylenie standardowe wszystkich parametrów się zmniejszyło, podobnie jak przedziały prognozy, ponadto wartość prognozy jest bliższa prawdziwej wartości.