What do missing prices mean for the choice of index number method with alternative data?

Ben Hillman

Office for National Statistics, United Kingdom

8th June 2022

Office for National Statistics

Data and methods

|	Fox et al	ONS
Type of data	Household scanner	Retailer scanner
Relaunch linking	No	Yes
Unit standardisation	No	Yes
ID used	GTIN	SKU
Months used (number of "splices")	84 (59)	56 (31)
# elementary aggs (con segs*regions)	52 (=52*1)	2244 (=187*12)

Some key definitions

Churn – percentage of expenditure on products in a month that were not present in the previous month.

Benchmark – an index calculated from the same data with a pure equivalent of the same index method

Residual – the difference between the chained index and the benchmark in the final time period
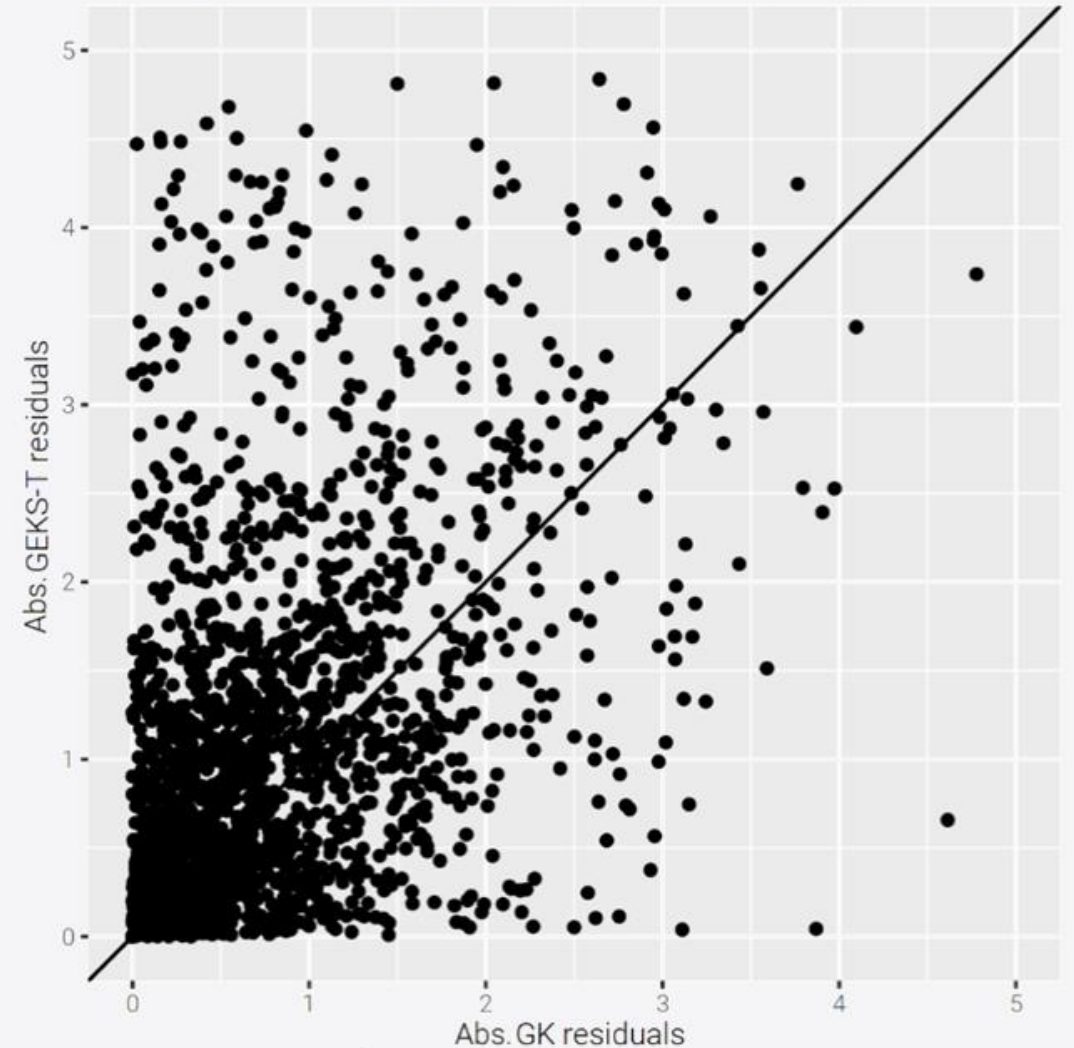
Multilateral methods

GEKS-T has a wider spread of residuals than the GK, and the scatter plot suggests they drift for different reasons

Residuals, by index method
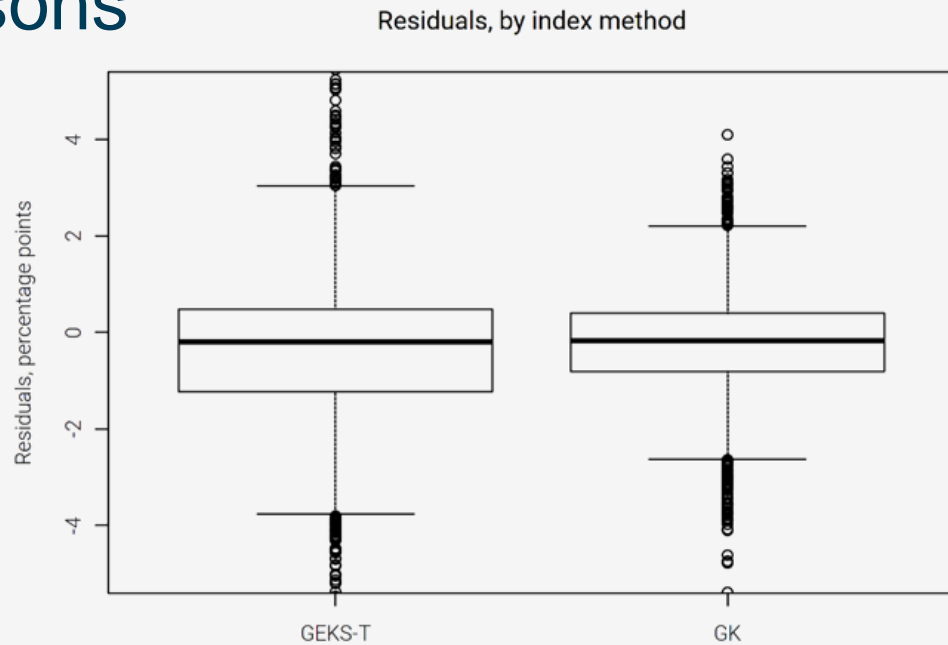
GEKS-T residuals against GK residuals
Points below the y=x line are where GK residuals are bigger
Points above the line are where GEKS-T residuals are bigger

Causes of drift

Residuals increase in magnitude as churn increases

No imputation GEKS-T residuals against log mean monthly churn

GK residuals against log mean monthly churn
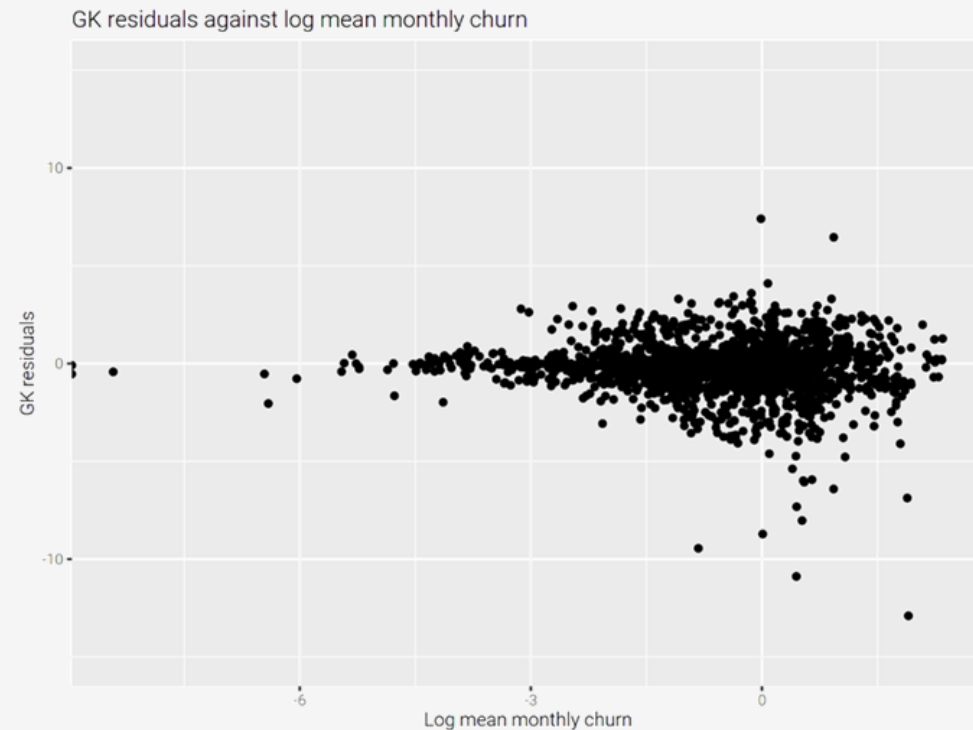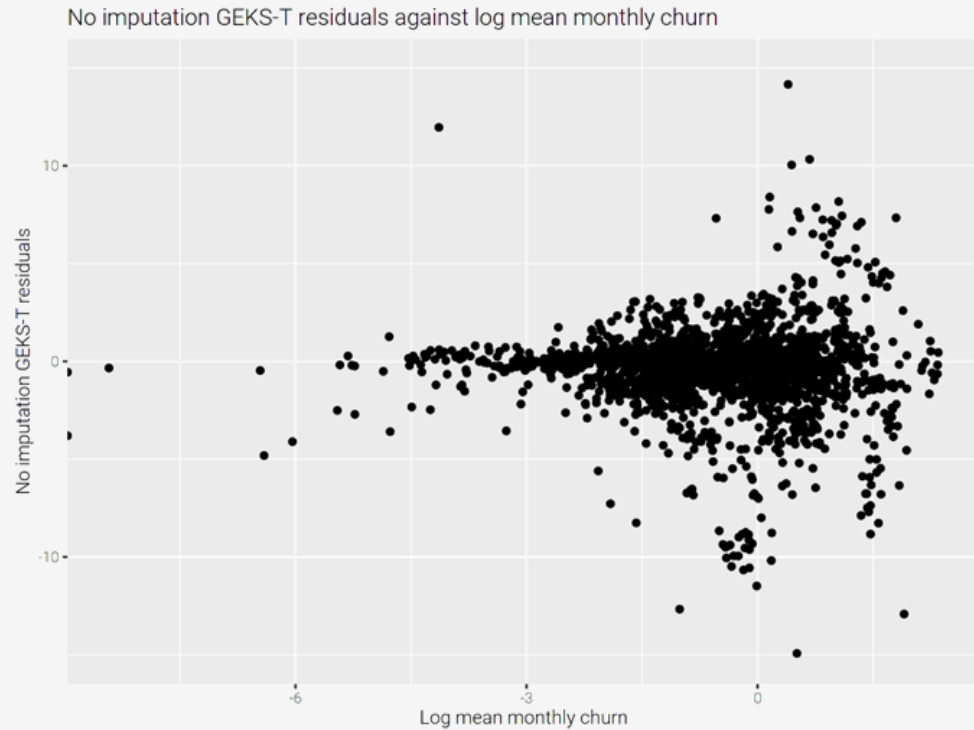
Imputation as a solution?
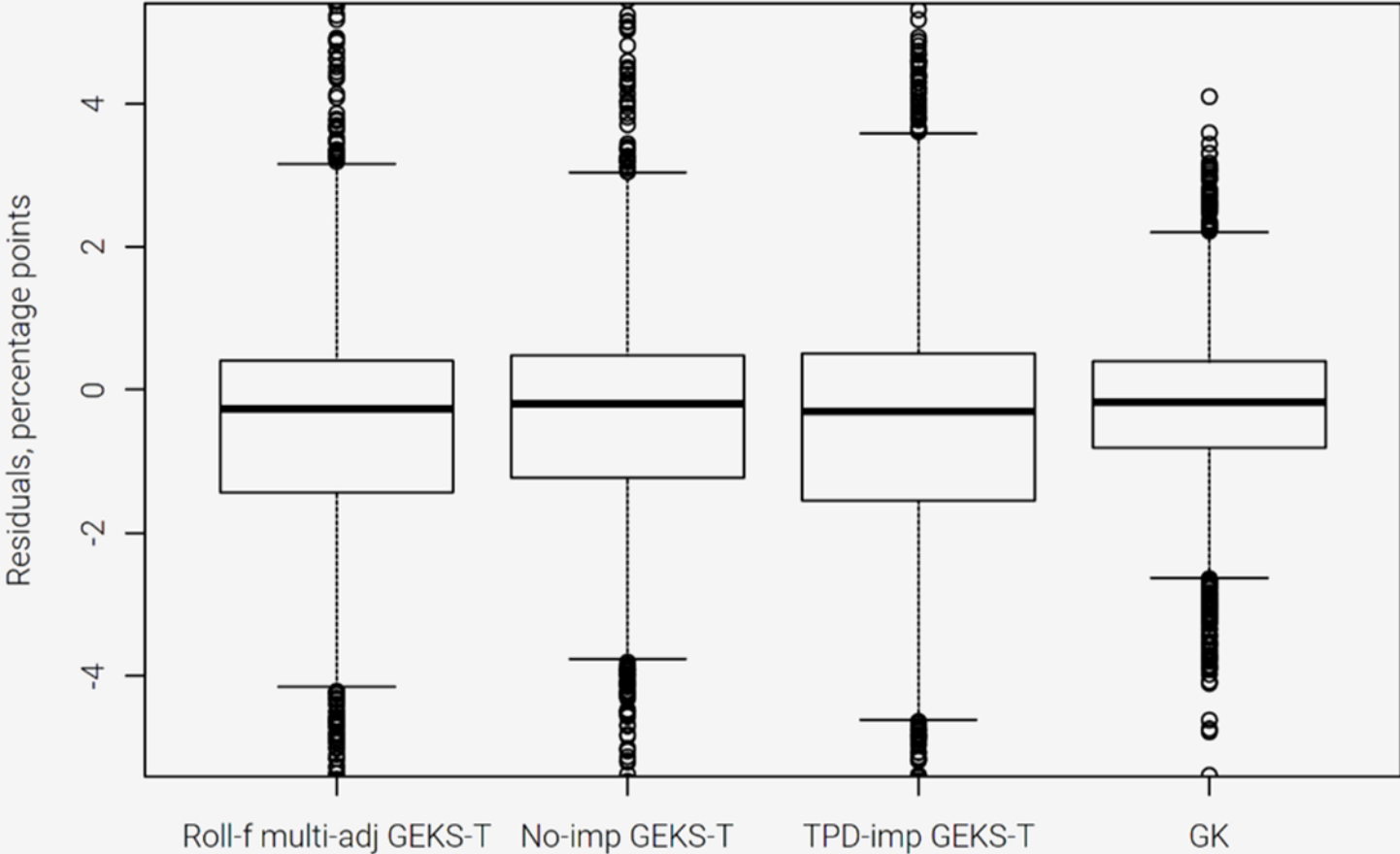
GK method does not support imputation

GEKS-T can use imputation

Two methods examined (multi-adj. roll-f and TPD)

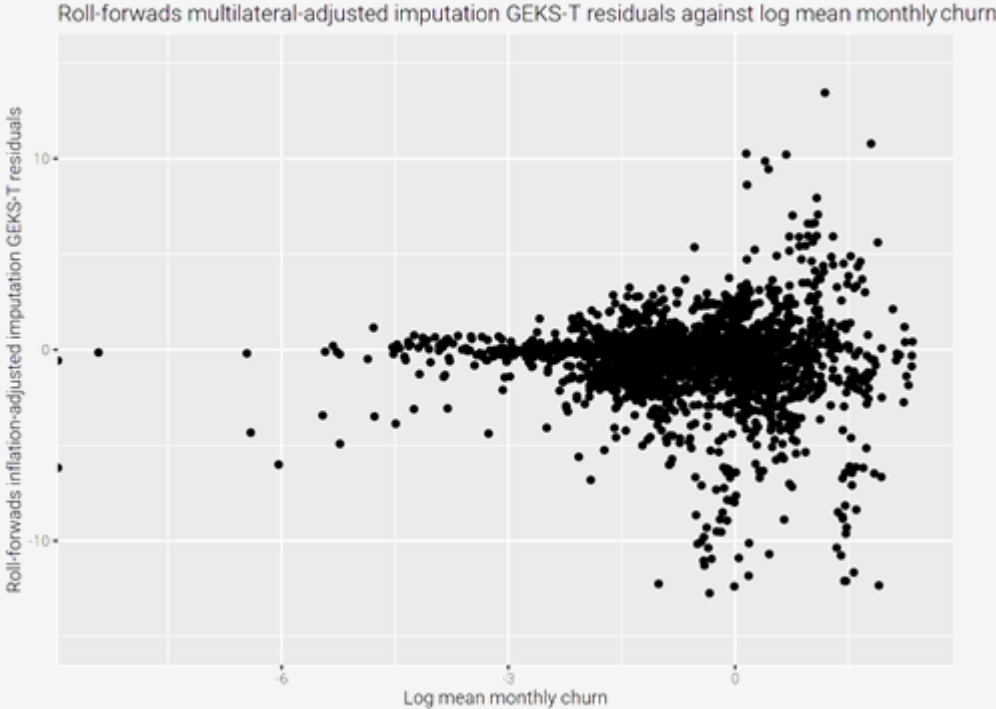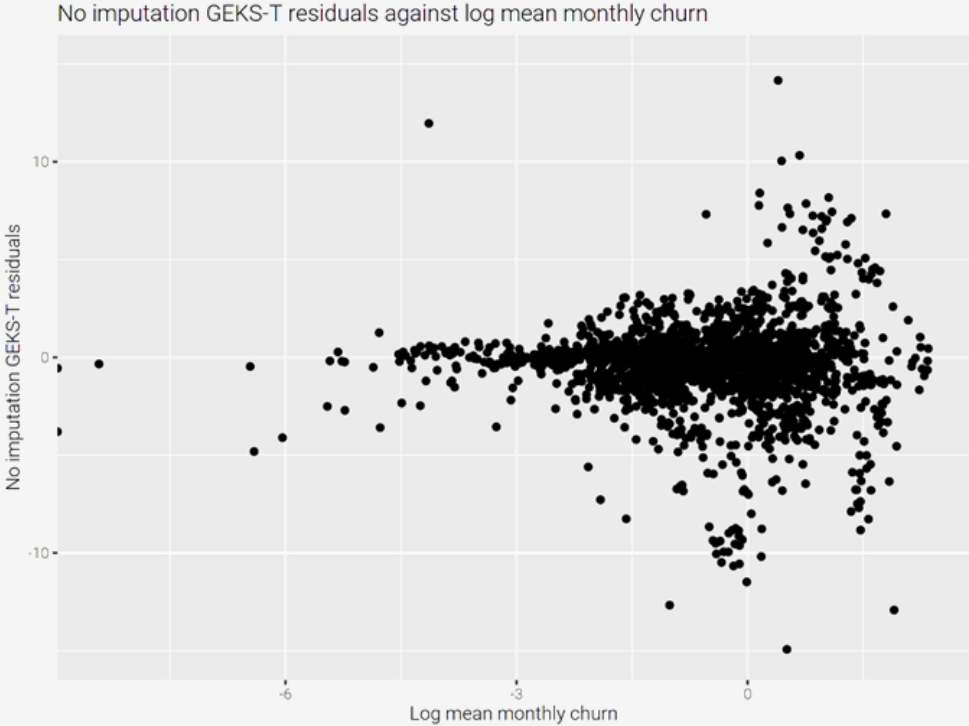| Period	Product 1 price	Product 1 quantity	Product 2 price	Product 2 quantity
1	£1.49	205	£2.05	62
2	£1.49	212	£2.05	58
3	£1.49	199	£2.07	46
4	£1.55	201	£2.03	31
5	£1.55	204	-	0
6	£1.55	197	-	0
7	£1.60	196	-	0

Imputation as a solution?

Residuals, by index method
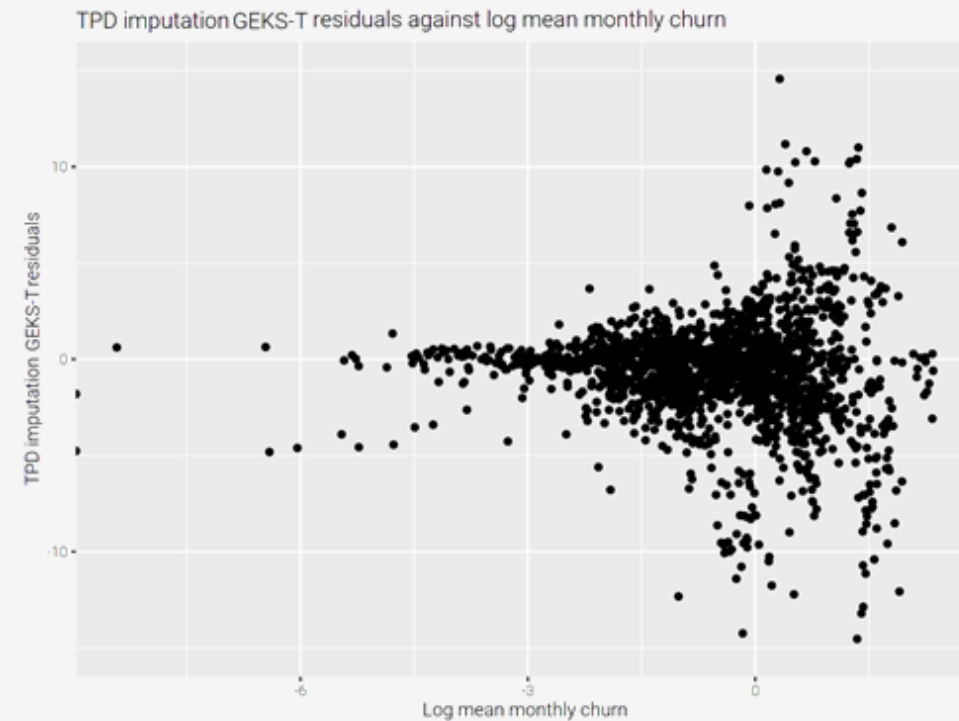
Roll-forwards multilateral-adjusted imputation

Imputation does not appear to reduce the effect of churn

No imputation GEKS-T residuals against log mean monthly churn


Roll-forwads multilateral-adjusted imputation GEKS-T residuals against log mean monthly churn

Time-Product Dummy imputation

Imputation does not appear to reduce the effect of churn

No imputation GEKS-T residuals against log mean monthly churn

TPD imputation GEKS-T residuals against log mean monthly churn

Office for National Statistics

Churn
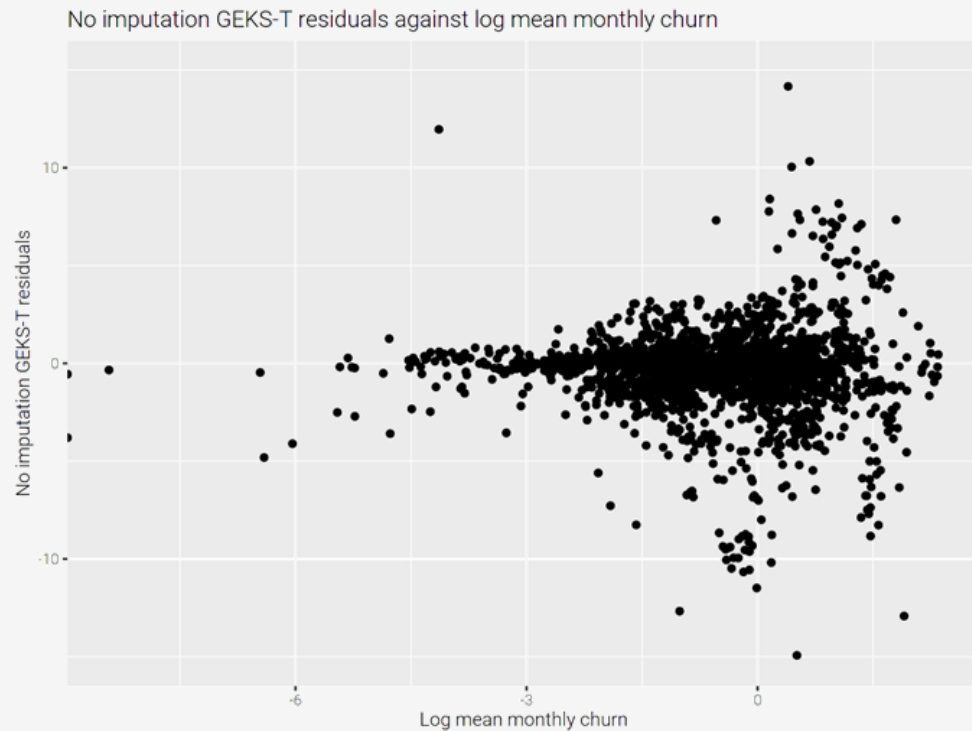
Imputation and drift

When compared to GEKS-T without imputation, neither method shows an improvement in their residuals

Abs. TPD imputation GEKS-T residuals against abs. no-imp GEKS-T residuals
Points below the y=x line are where the imputation residuals are bigger
Points above the line are where the no-imputation residuals are bigger

Abs. roll-f multi-adj GEKS-T residuals against abs. no-imp GEKS-T residuals
Points below the y=x line are where the imputation residuals are bigger
Points above the line are where the no-imputation residuals are bigger

Residuals compared
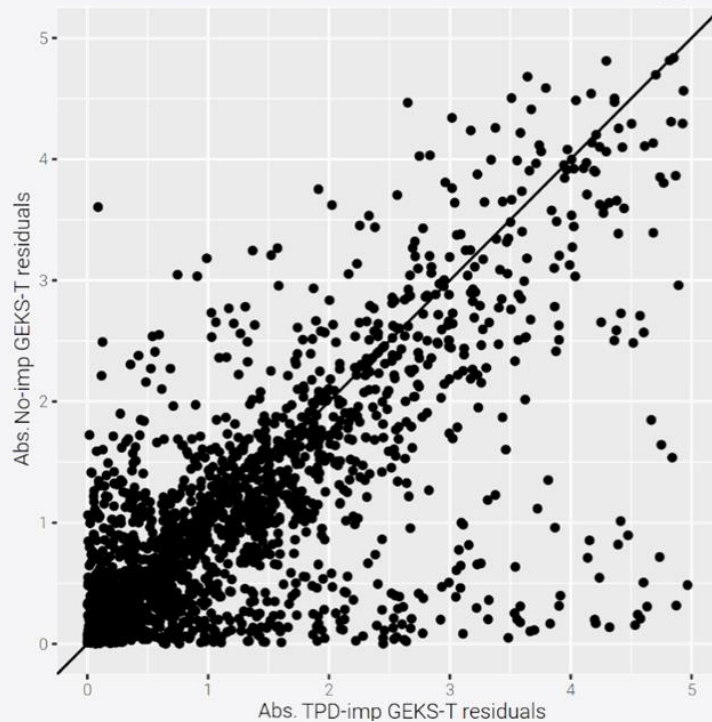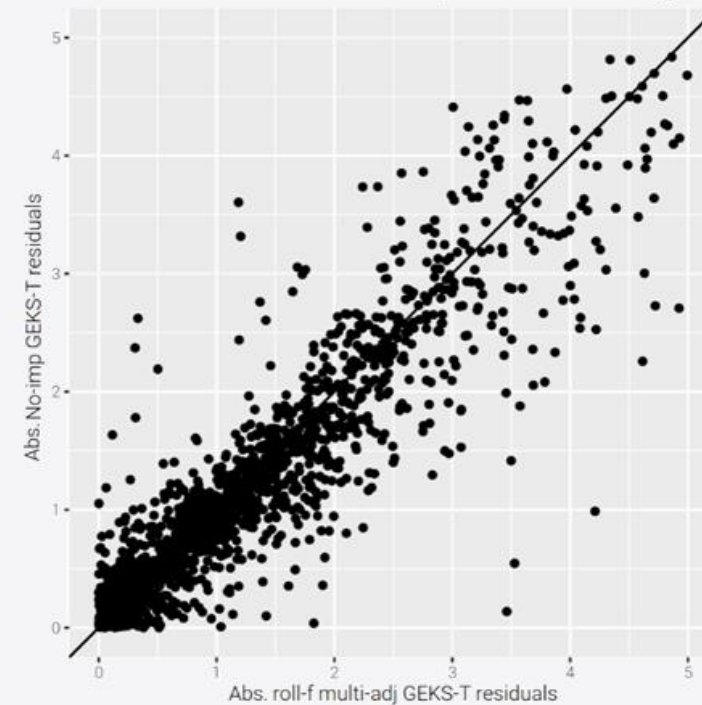
Why no improvement?

The method

Other methods, especially fully hedonic ones, might better estimate prices
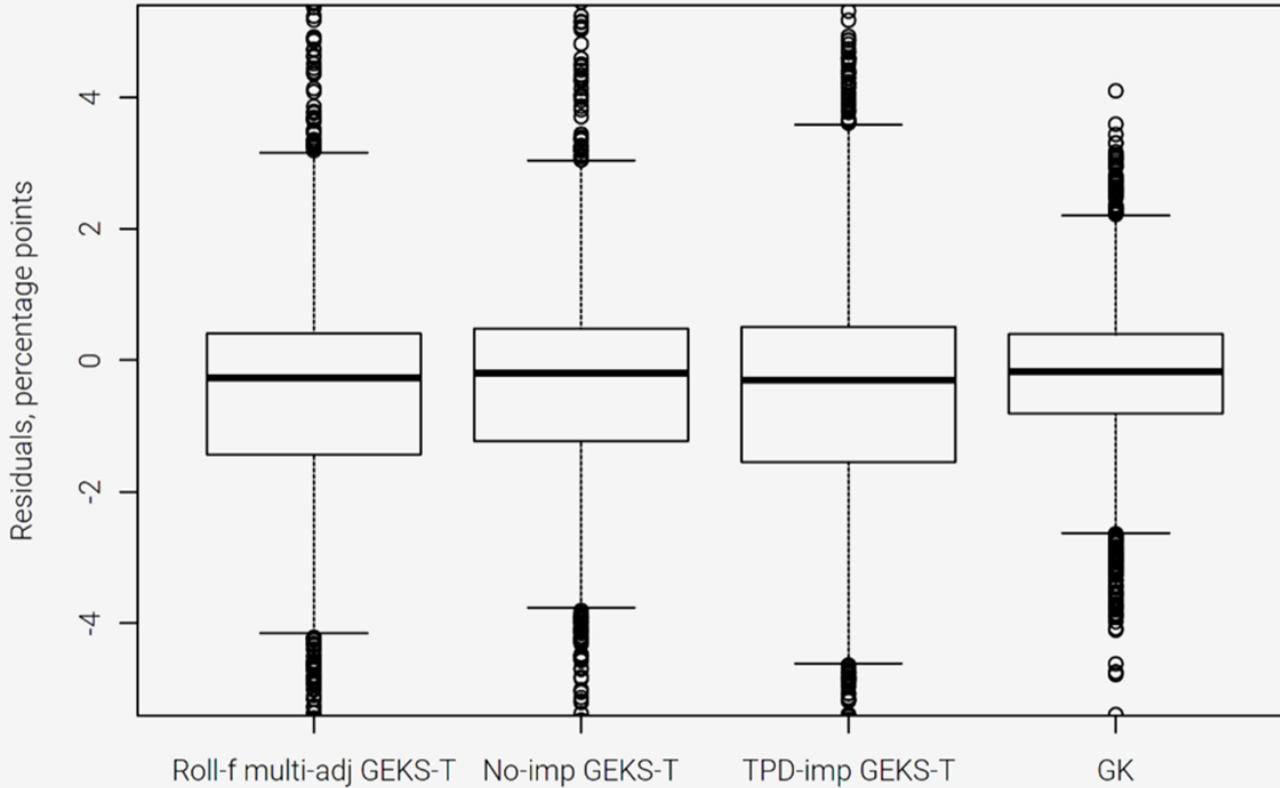
The metric

A pure index will 'look ahead', which may be unreasonable as a target
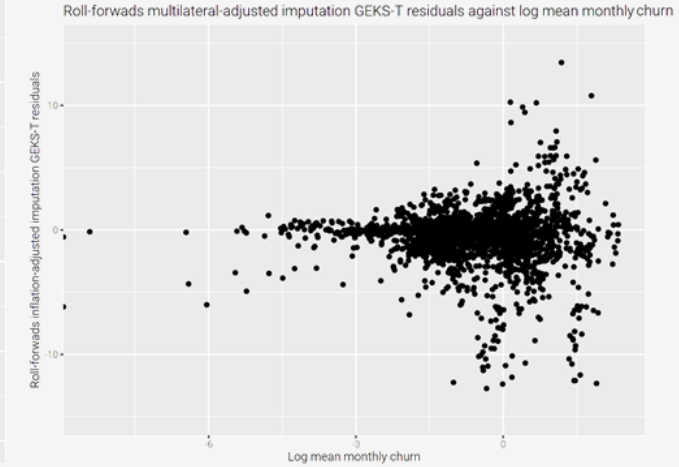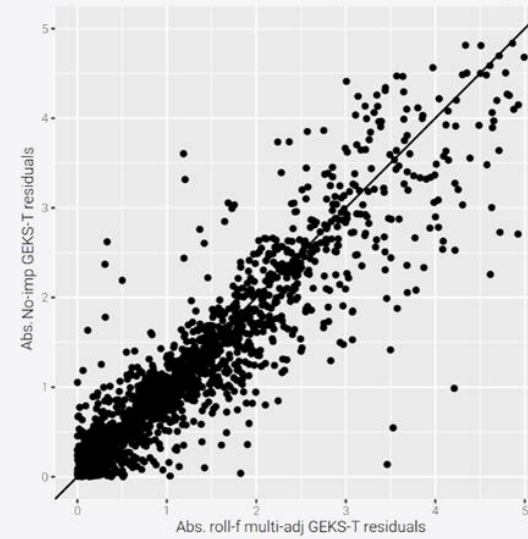
The index

The Törnqvist index will only 'half weight' missing prices, limiting overall effect

Any Questions?

Residuals, by index method

Abs. roll-f multi-adj GEKS-T residuals against abs. no-imp GEKS-T residuals
Points below the y=x line are where the imputation residuals are bigger
Points above the line are where the no-imputation residuals are bigger

Roll-forwads multilateral-adjusted imputation GEKS-T residuals against log mean monthly churn

Abs. TPD imputation GEKS-T residuals against abs. no-imp GEKS-T residuals
Points below the y=x line are where the imputation residuals are bigger
Points above the line are where the no-imputation residuals are bigger

TPD imputation GEKS-T residuals against log mean monthly churn