House Price Indexes: A Comparison of Repeat Sales and Other Multilateral Methods

Jan de Haan (Delft University of Technology and ESCoE)
Rob van de Laar (Statistics Netherlands)

ESCoE Conference, 21-23 May 2025, London

Structure of paper and presentation

- Introduction
- Basic model and hedonic regression
- Time Product Dummy (TPD)
- Repeat Sales (RS)
- GEKS-Jevons
- Geary-Khamis (GK) and Arithmetic Repeat Sales (ARS)
- Revisions and other issues
- Data and empirical results
- Summary and conclusions

Introduction

Construction of house price indexes

Various methods available

Repeat Sales (Bailey, Muth and Nourse, 1963) widely used

Repeat Sales is a regression-based multilateral method

Multilateral price indexes

- Estimated simultaneously for multiple periods from pooled data
- Transitive, hence independent of choice of base period
- Revisable

Introduction

Other multilateral methods

- Time Dummy Hedonic
  Regression-based; quality adjustment using property attributes;
- Non-hedonic methods
  Time Product Dummy – regression-based; fixed property effects
  GEKS – “transitivizing” bilateral (Jevons) price indexes
  Geary-Khamis – arithmetic aggregation
  Arithmetic Repeat Sales

Basic model and hedonic regression

Basic model

$$p_i^t = P^t a_i$$

Stochastic model (log-linear)

$$\ln p_i^t = \delta^t + \gamma_i + \varepsilon_i^t$$

Fixed effects for time t and property i plus error term

Starting point for Time Dummy Hedonic (TDH), TPD and RS methods

Basic model and hedonic regression

TDH estimating equation

$$\ln p_i^t = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i^t$$

Sample period t=0,…,T; characteristics x (fixed, i.e., **time invariant**)

OLS regression on pooled data; $P_{TDH}^{0t} = \exp(\hat{\delta}^t)$

$$P_{TDH}^{0t} = \prod_{i \in S^t} \left(\frac{p_i^t}{\hat{p}_i^0} \right)^{1/N^t}$$

Imputation Jevons price index

Time Product Dummy (TPD)

TPD estimating equation (OLS regression)

$$\ln p_i^t = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{n=1}^{N-1} \gamma_i D_i + \varepsilon_i^t$$

$$P_{TPD}^{0t} = \prod_{i \in S^t} \left(\frac{p_i^t}{\tilde{p}_i^0} \right)^{1/N^t}$$

$$\tilde{p}_i^0 = \prod_{t \in S_i} \left(\frac{p_i^t}{P_{TPD}^{0t}} \right)^{1/N(S_i)}$$

Imputation Jevons price index

System of equations

Zeroing out "single observations"

Repeat Sales (RS)

Focus on price change (rather than price)

Identifies “first” and “second” sales, not multiple sales

RS estimating equation (OLS regression)

$$\ln\left(\frac{p_i^{s(i)}}{p_i^{f(i)}}\right) = \sum_{t=0}^{T} \delta^t D_{i(RS)}^t + \varepsilon_i^{f(i)s(i)}$$

Time dummy: 1 (second sale), -1 (first sale), or 0

Repeat Sales (RS)

$$P_{RS}^{0t} = \exp(\breve{\delta}^t)$$

$$P_{RS}^{0t} = \prod_{i \in S_f^t} \left(\frac{p_i^t}{\breve{p}_i^0} \right)^{1/N_{RS}^t} \prod_{i \in S_s^t} \left(\frac{p_i^t}{\breve{p}_i^0} \right)^{1/N_{RS}^t}$$

$$\breve{p}_i^0 = \frac{p_i^{s(i)}}{P_{RS}^{0s(i)}} \quad \text{for } i \in S_f^t$$

$$\breve{p}_i^0 = \frac{p_i^{f(i)}}{P_{RS}^{0f(i)}} \quad \text{for } i \in S_s^t$$

System of equations

Repeat Sales (RS)

TPD is a constrained version of RS

"TPD is a matched pairs method that, through averaging all available deflated selling prices, produces a unique base period price estimate for each property rather than two or more (unconstrained) estimates as RS does."

Advantage of RS over TPD: fewer parameters to estimate

GEKS-Jevons

Elements

Bilateral matched-model Jevons price indexes for all possible base (link) periods 0,…,T

$$P_{GEKS-J}^{0t} = \prod_{b=0}^{T} \left[P_J^{0b} \times P_J^{bt} \right]^{1/(T+1)}$$ GEKS-Jevons index, 0 to t

$$P_{GEKS-J}^{rt} = \prod_{b=0}^{T} \left[P_J^{rb} \times P_J^{bt} \right]^{1/(T+1)}$$ GEKS-Jevons index, r to t

GEKS-Jevons

Regression-based approach

$$\ln P_J^{rt} = \delta^t - \delta^r + \varepsilon^{rt}$$

OLS regression produces GEKS-Jevons index

WLS regression with number of repeat sales (matched properties) as weights yields RS index

RS is essentially **weighted GEKS**

Price comparisons with few observations are downweighted

Geary-Khamis (GK) and Arithmetic Repeat Sales (ARS)

Geary-Khamis

$$P_{GK}^{0t} = \frac{\sum_{i \in S^t} p_i^t / N^t}{\sum_{i \in S^t} \breve{p}_i^0 / N^t} = \frac{\sum_{i \in S^t} \breve{p}_i^0 \left(\frac{p_i^t}{\breve{p}_i^0} \right)}{\sum_{i \in S^t} \breve{p}_i^0}$$

Imputation Dutot price index

(system of equations)

“Value weighted” price index

GK is a constrained version of (value-weighted) ARS

Geary-Khamis (GK) and Arithmetic Repeat Sales (ARS)

ARS - arithmetic counterpart to (geometric) RS

$$P_{ARS}^{0t} = \frac{\sum_{i \in S_f^t} p_i^t + \sum_{i \in S_s^t} p_i^t}{\sum_{i \in S_f^t} \hat{p}_i^0 + \sum_{i \in S_s^t} \hat{p}_i^0} = \frac{\sum_{i \in S_{RS}^t} p_i^t / N_{RS}^t}{\sum_{i \in S_{RS}^t} \hat{p}_i^0 / N_{RS}^t} = \frac{\sum_{i \in S_{RS}^t} \hat{p}_i^0 \left(\frac{p_i^t}{\hat{p}_i^0} \right)}{\sum_{i \in S_{RS}^t} \hat{p}_i^0}$$

System of equations

$$\hat{p}_i^0 = \frac{p_i^{s(i)}}{P_{ARS}^{0s(i)}} \text{ for } i \in S_f^t$$

$$\hat{p}_i^0 = \frac{p_i^{f(i)}}{P_{ARS}^{0f(i)}} \text{ for } i \in S_s^t$$

Revisions and other issues

Revisions

Extending the sample period – multilateral indexes will be revised

Evidence of systematic revisions in RS indexes

Rolling window approach and splicing – probably needs a large window

Length of holding period

Increasing variance of price changes – weighted version of RS

Adding holding period as explanatory variable to check for (downward) quality change bias of RS (Diewert, Huang and Burnett-Isaacs, 2017)

Revisions and other issues

TPD computational difficulties

Huge number of property dummies (which was one of the reasons to propose RS)

Instead of running regression: iterative solution

Stratification reduces number of parameters

How to combine strata indexes (sales vs. stock weighting)?

Data and empirical results

Data

Residential property sales data from Dutch land registry

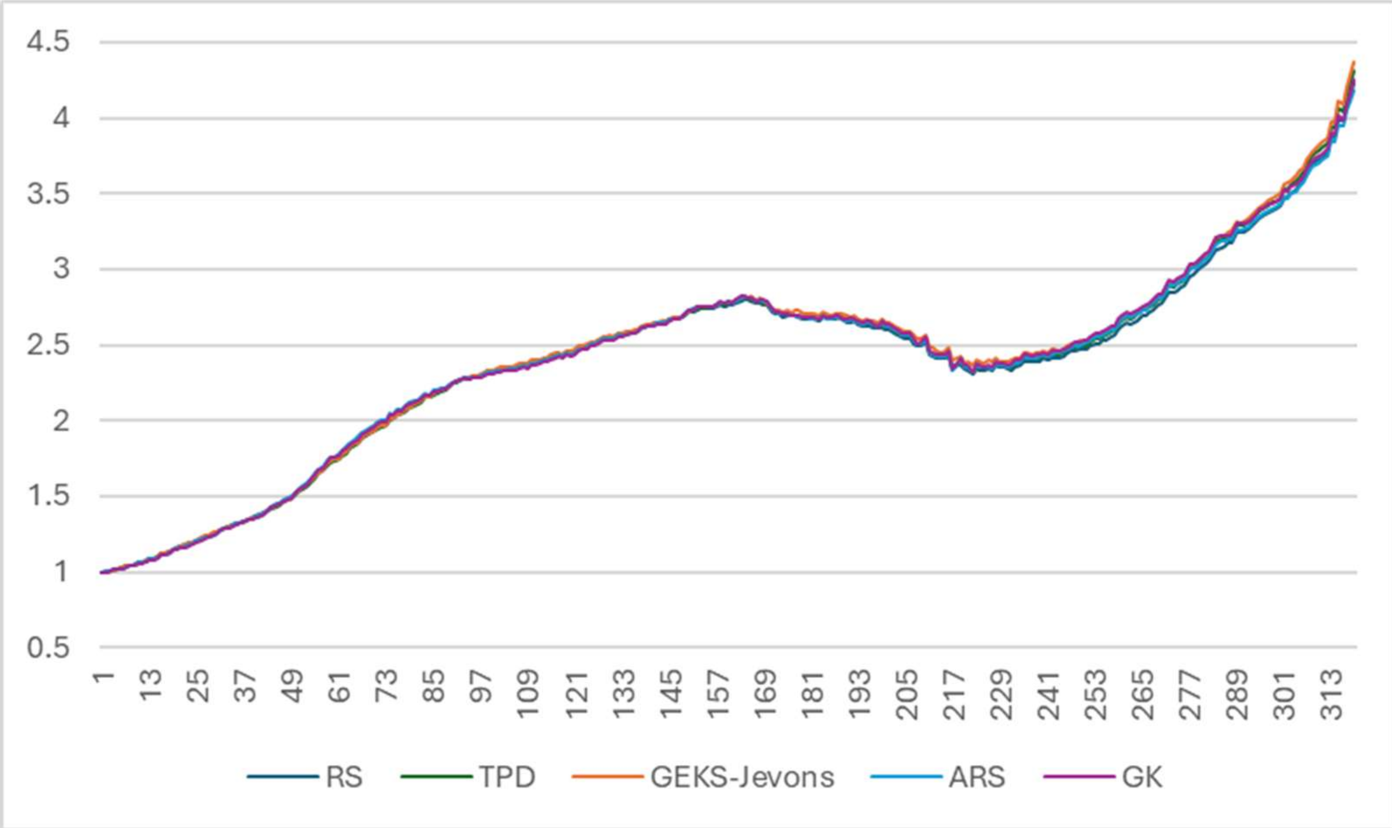
January 1995 – July 2021 - more than 26 years of data

Total number of sales (excl. newly built property): 4,884,394

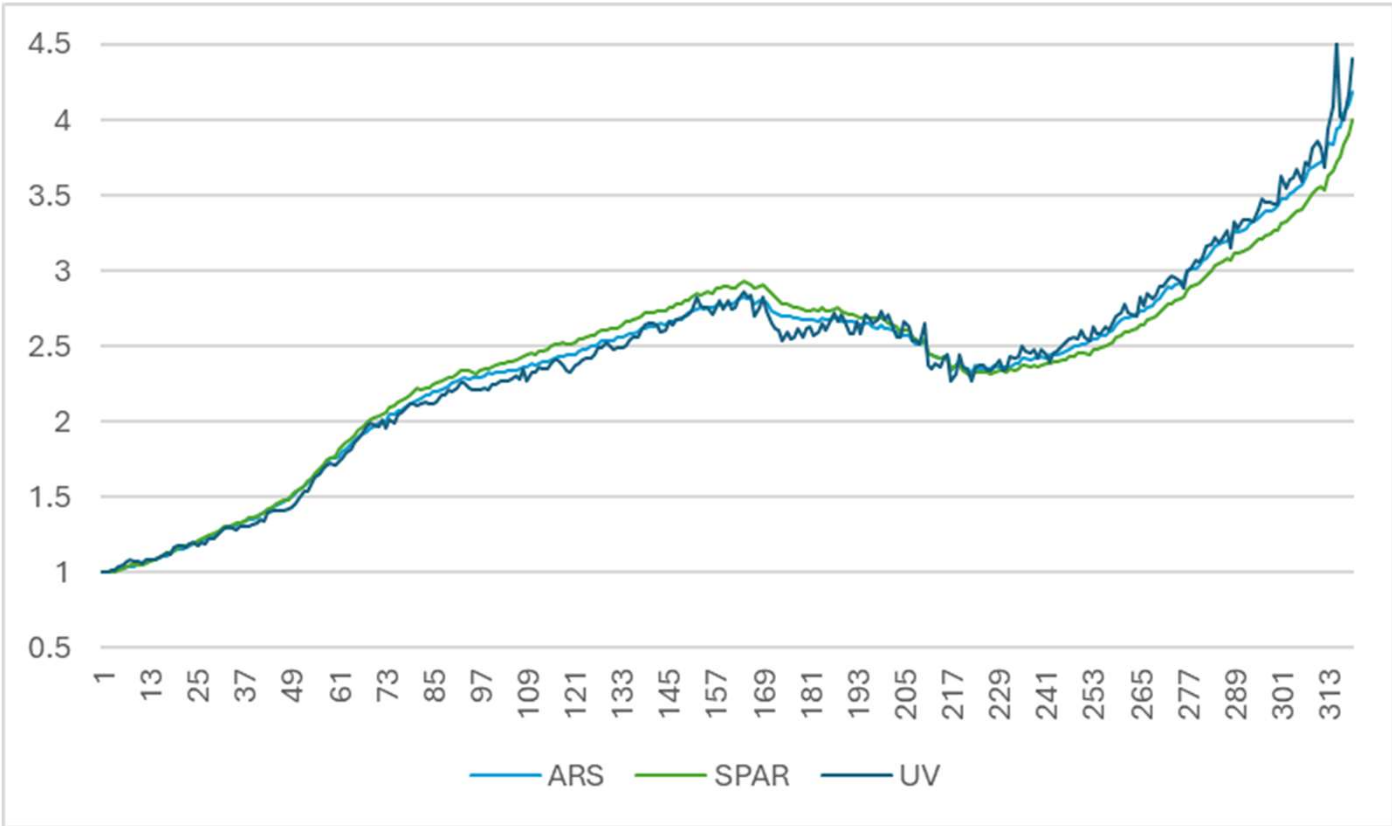
After data cleaning: 4,462,155

Repeat sales: 2,927,840 – 60% of total number of sales

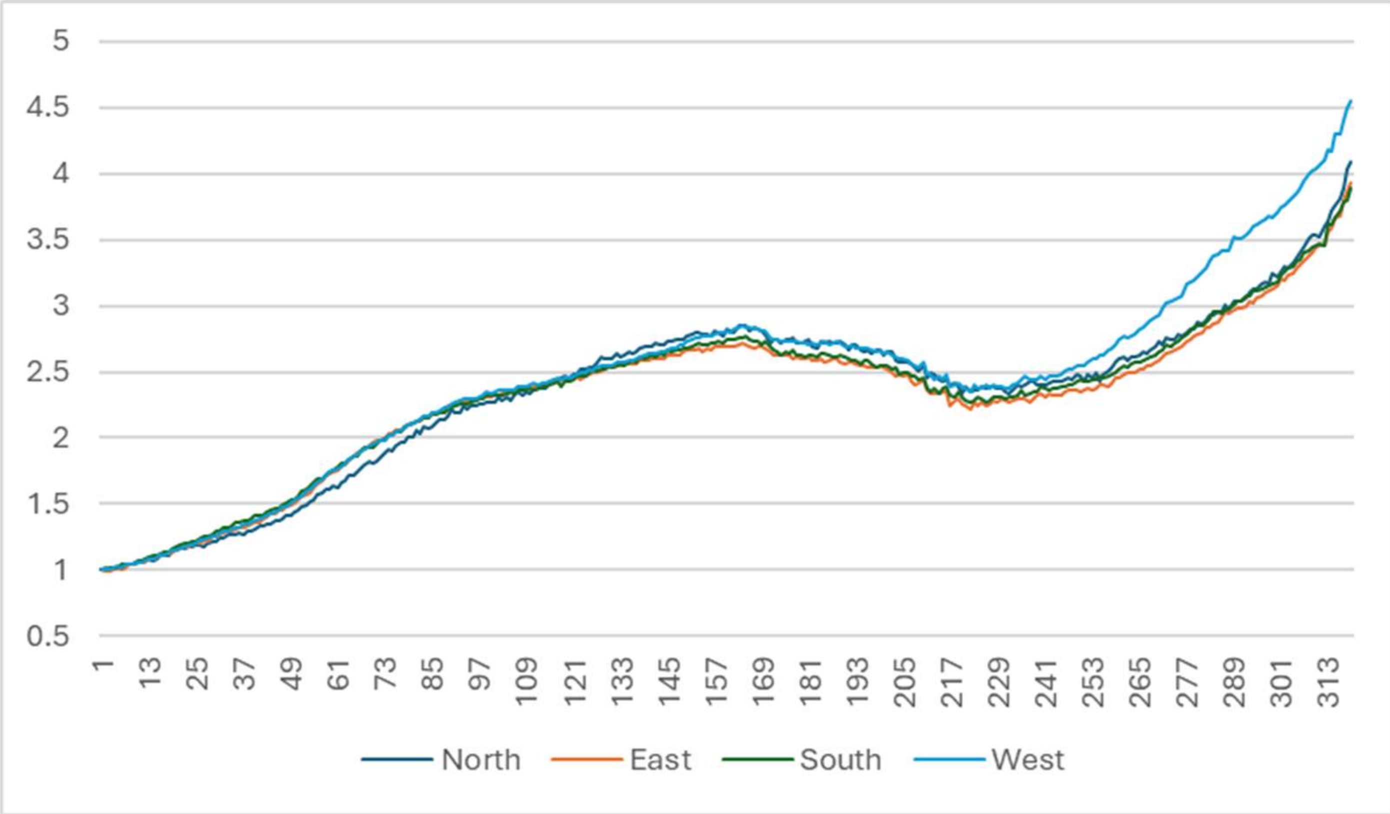
Five multilateral price indexes for the Netherlands



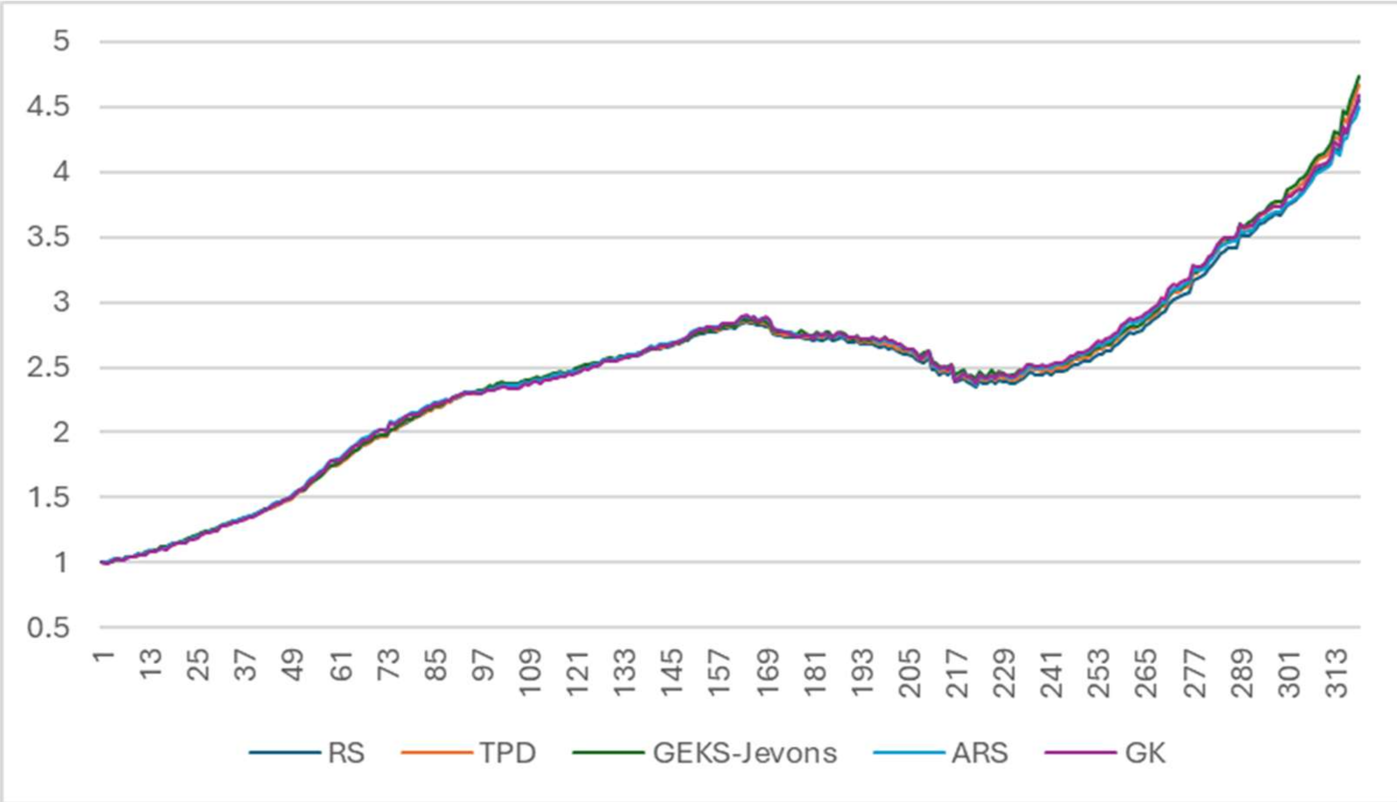
ARS, SPAR and UV indexes for the Netherlands



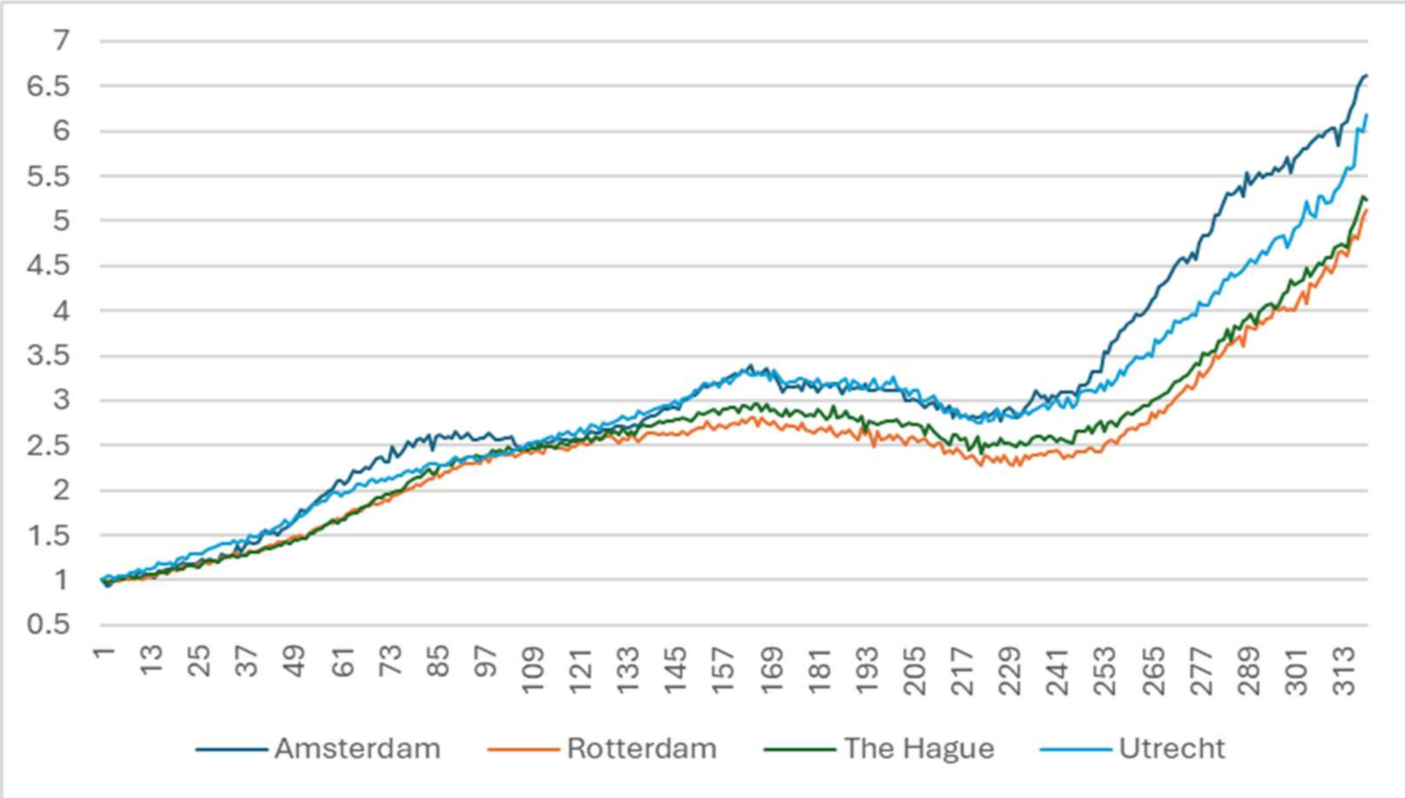
RS index for four parts of the Netherlands



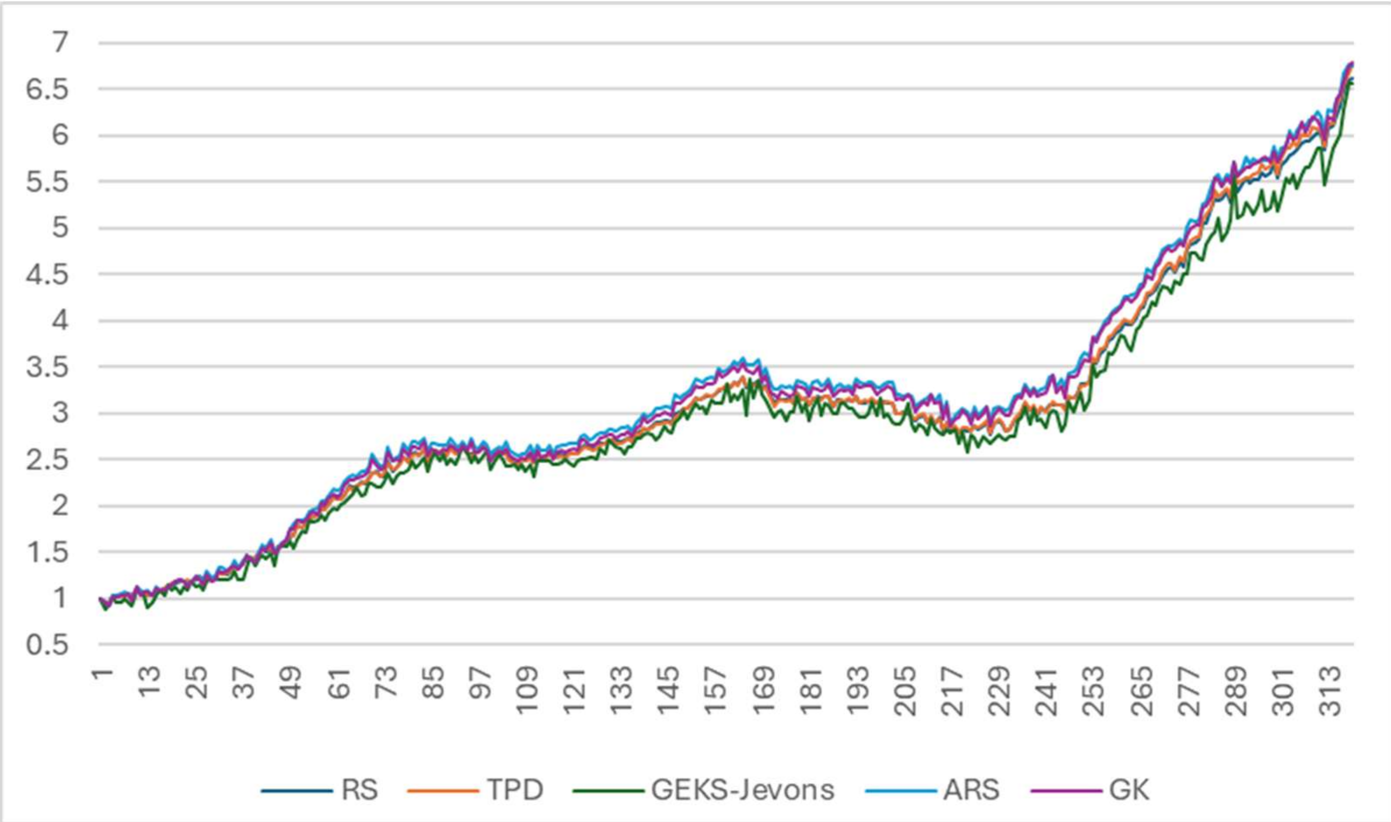
Multilateral indexes for the West of the Netherlands



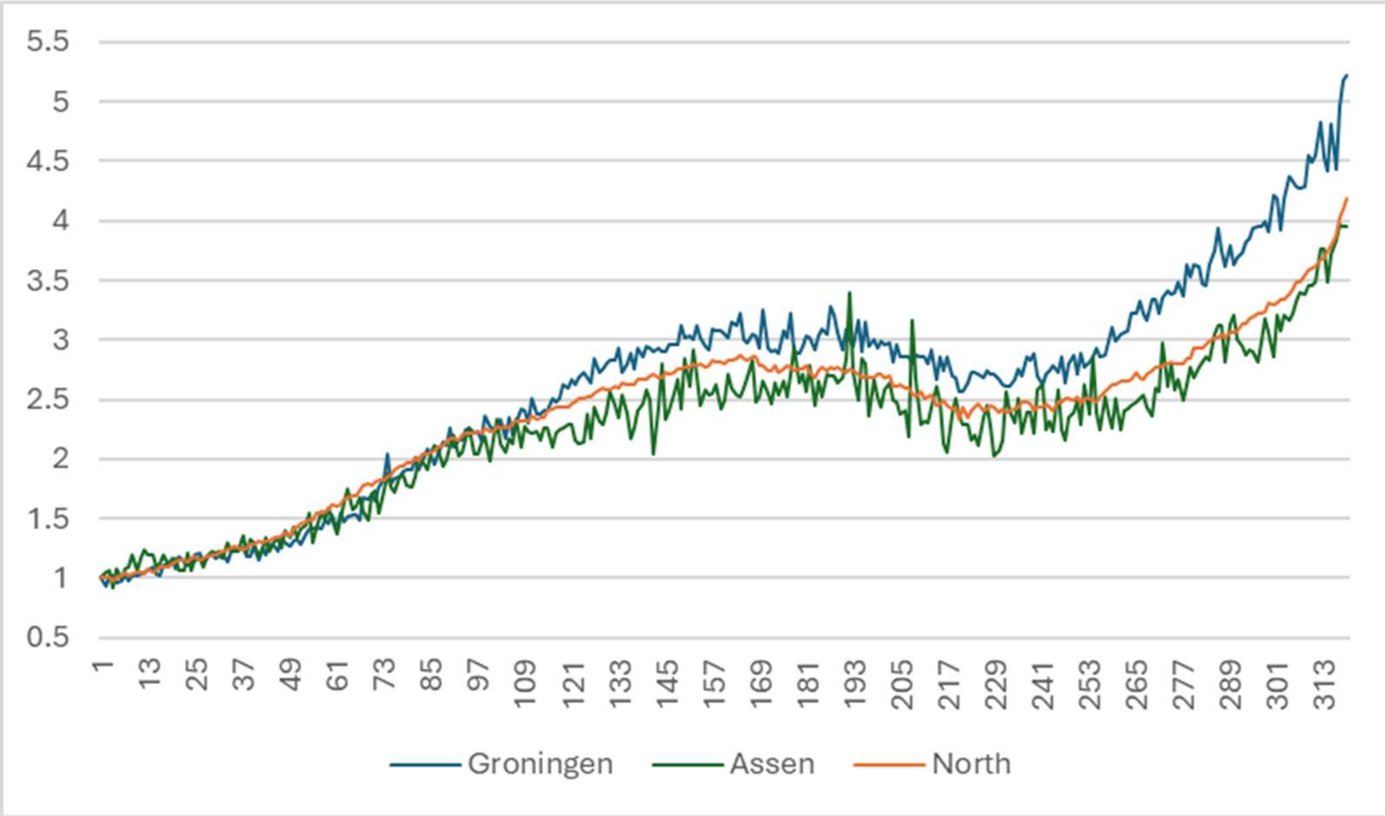
RS index for the four biggest cities of the Netherlands



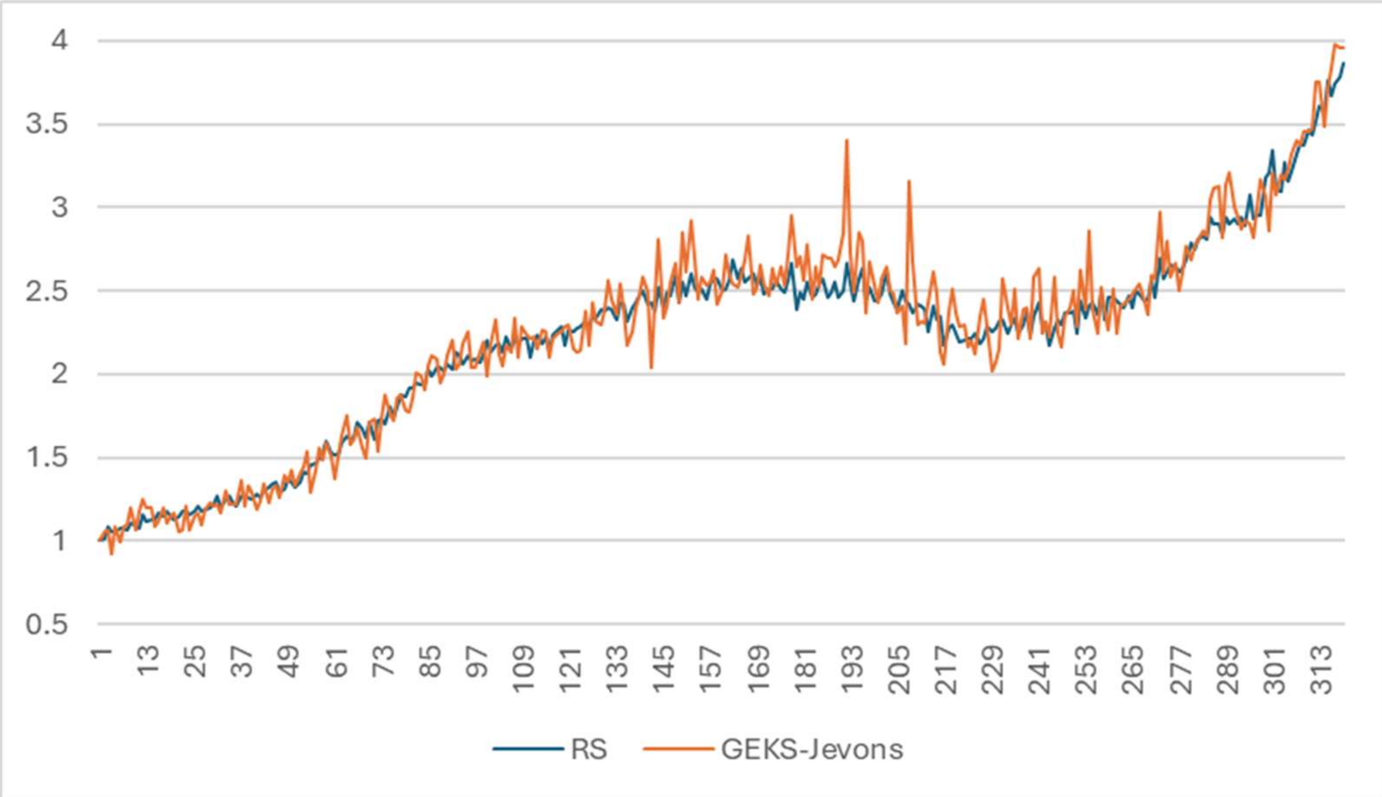
Multilateral indexes for Amsterdam



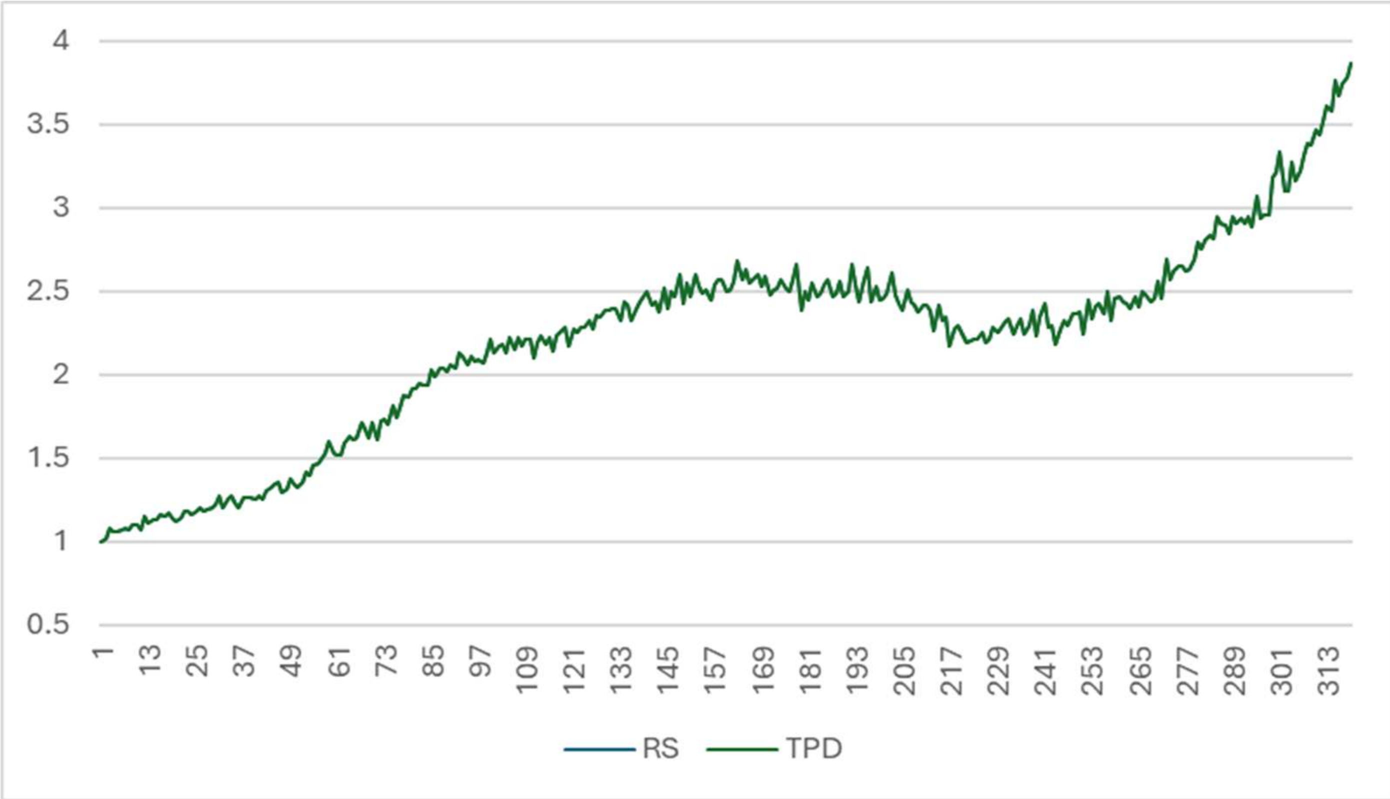
GEKS-Jevons index for the North of the Netherlands and two cities



RS and GEKS-Jevons indexes for Assen

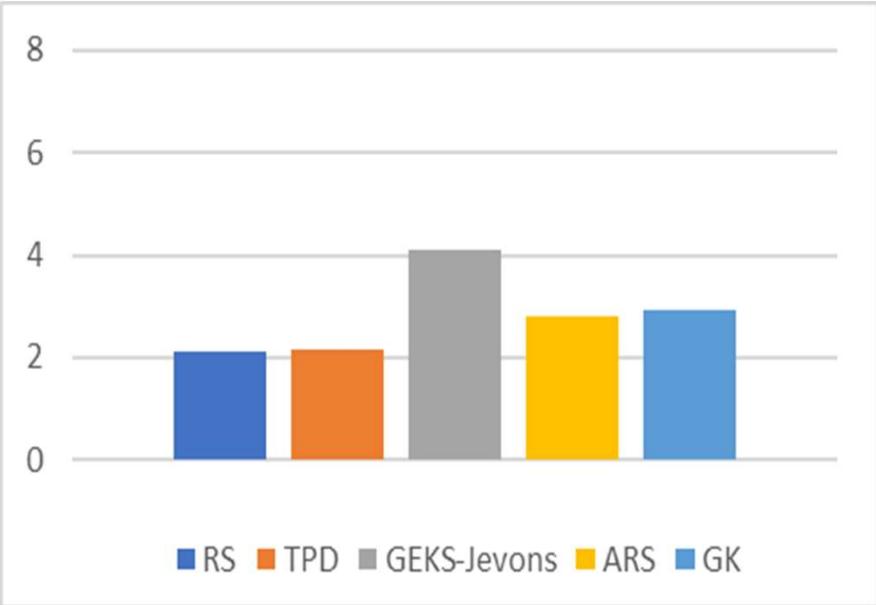


RS and TPD indexes for Assen

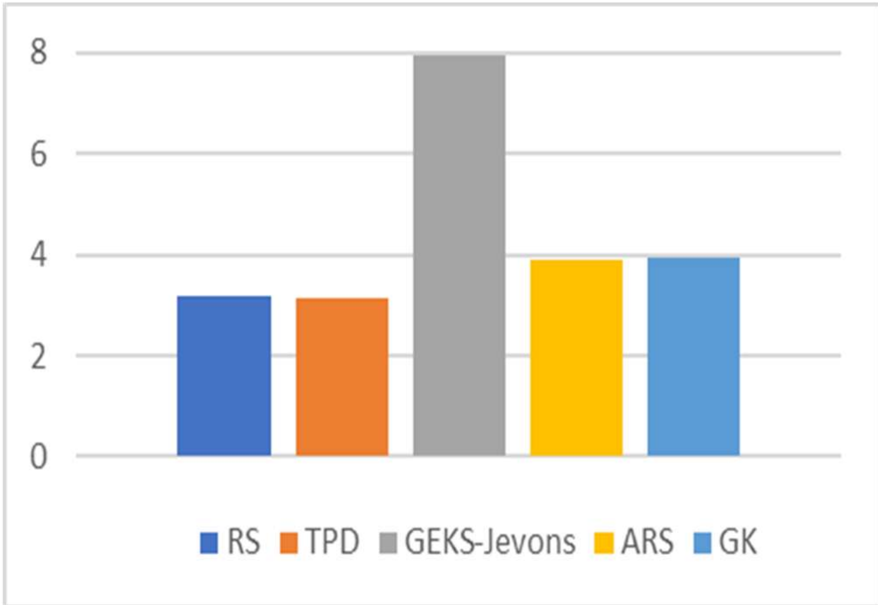


Volatility: standard deviation of monthly percentage index changes

Amsterdam

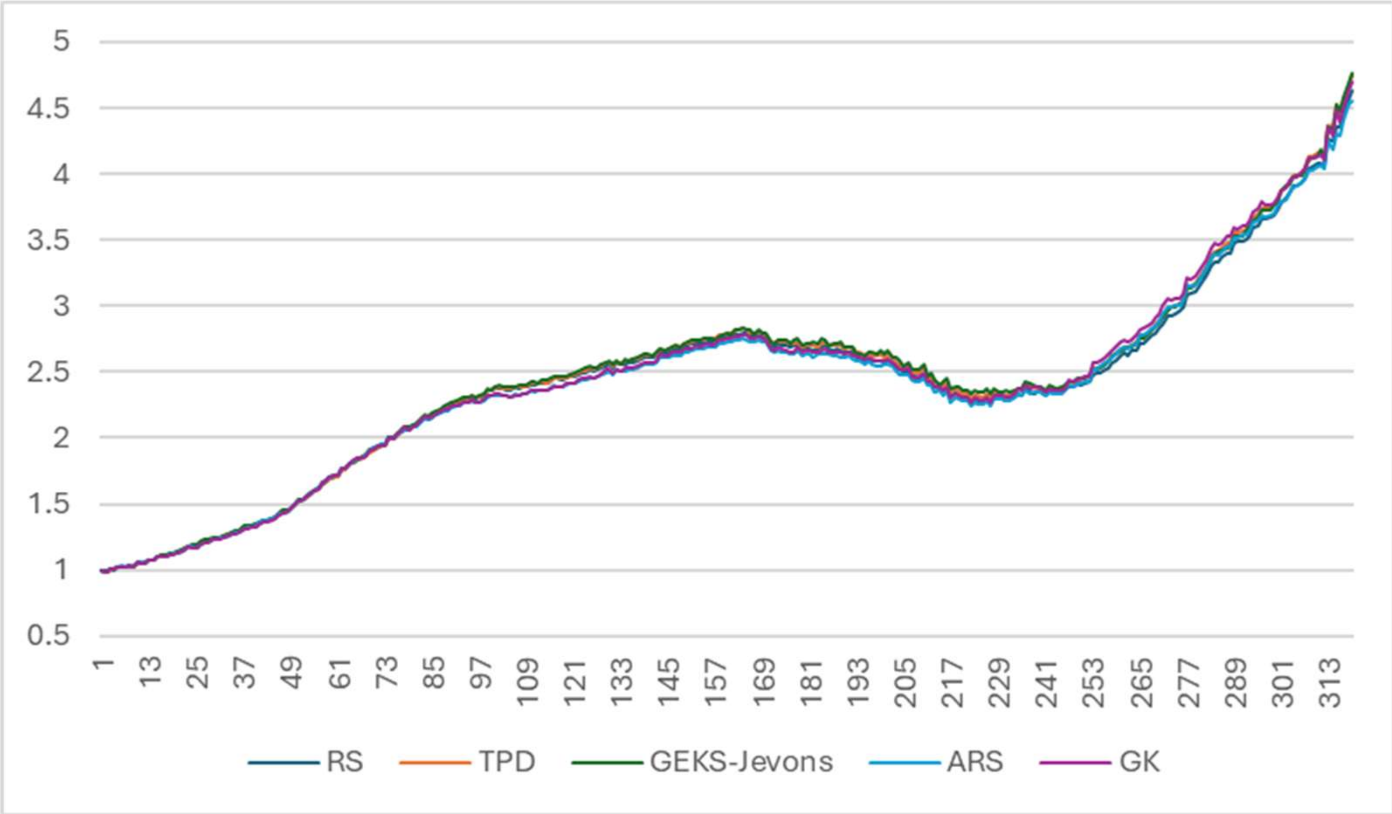


Assen

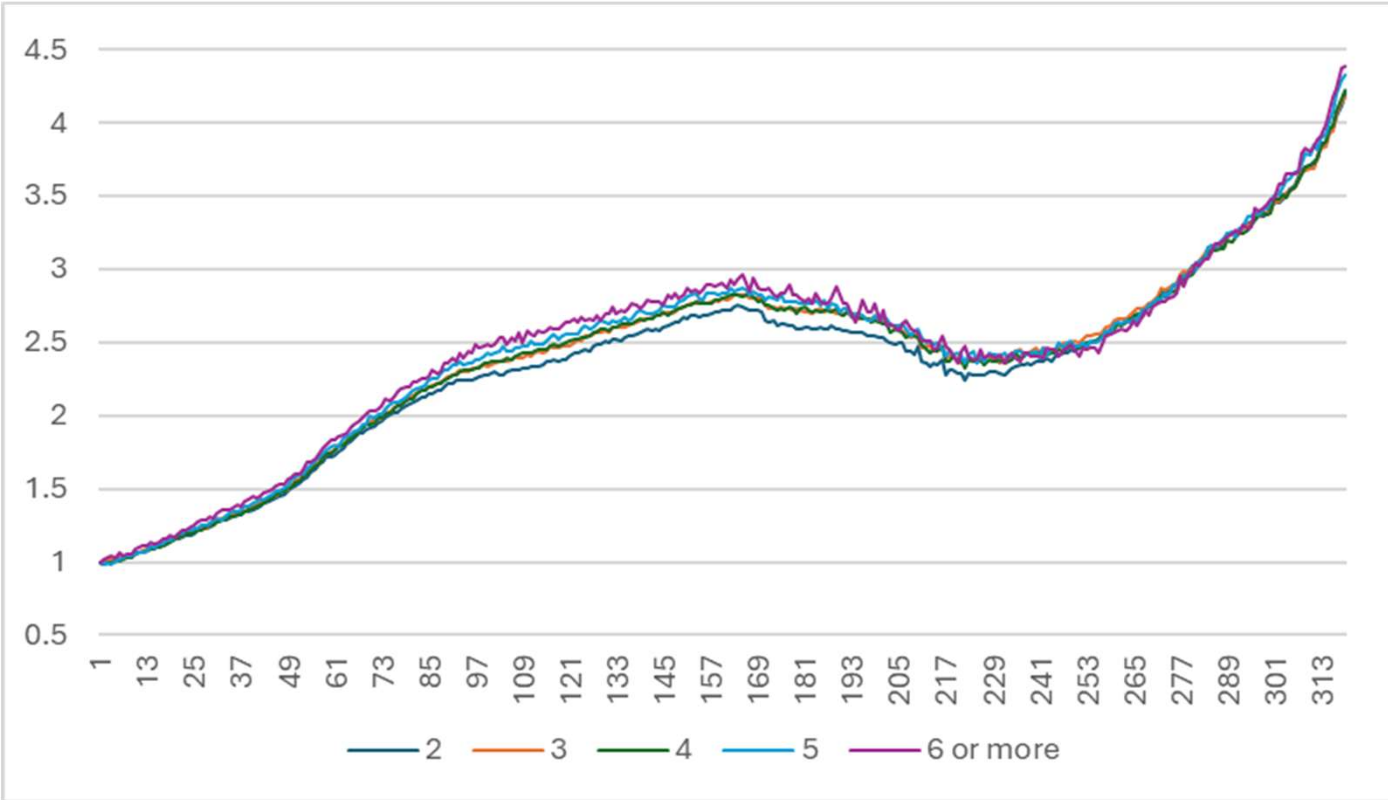


RS index for different types of dwelling in the Netherlands

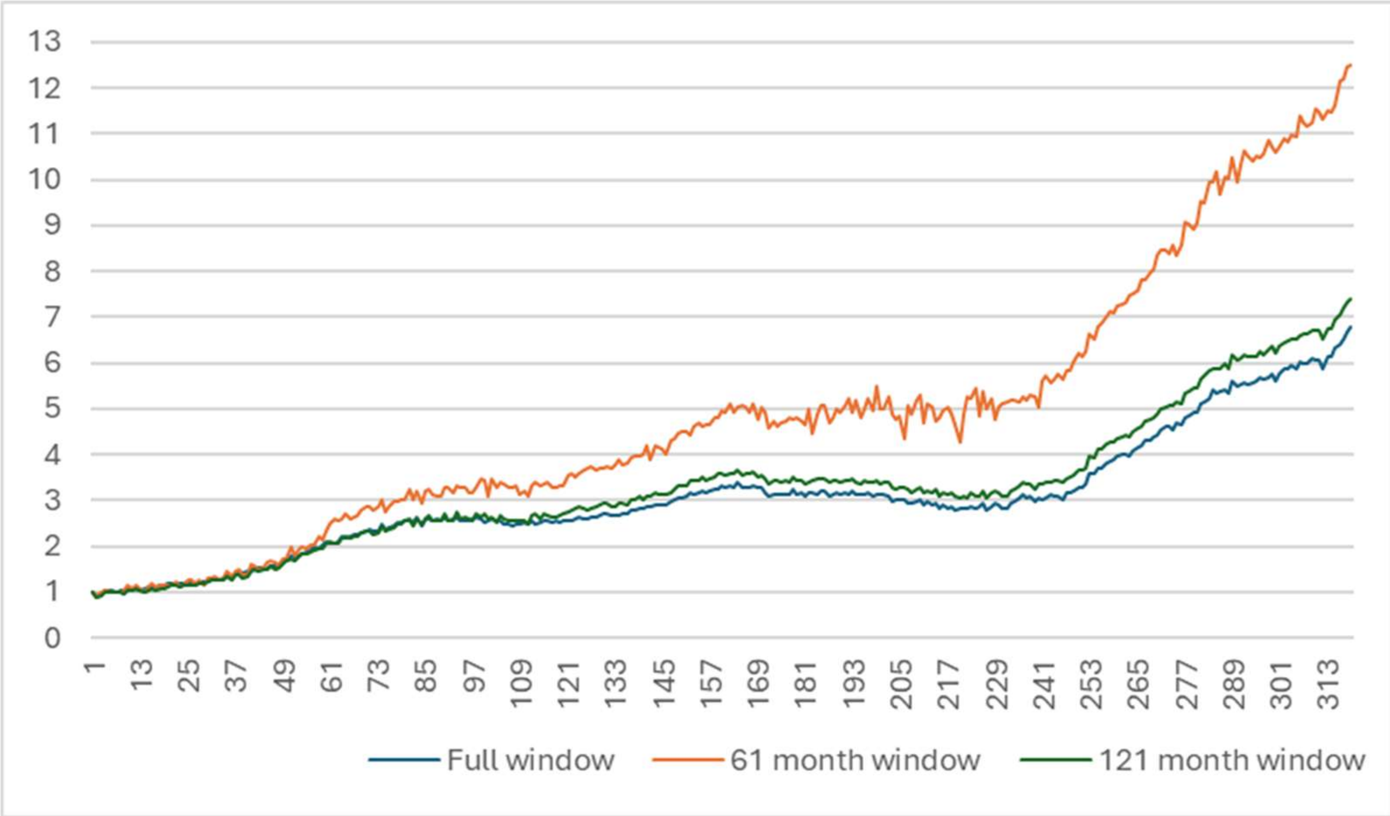
Multilateral indexes for apartments in the Netherlands



RS index according to number of times sold for the Netherlands



Full-window and chained TPD indexes for Amsterdam



Summary and conclusions

(Non-hedonic) multilateral methods

- are all matched-pairs (repeat sales) methods; adjust for mix change
- do not adjust for quality change of individual properties

Theoretical findings

- TPD and GK: constrained versions of RS and ARS; produce a single base period price estimate for each dwelling
- RS is essentially weighted GEKS-Jevons that down weights price comparisons with few observations; GEKS-Jevons likely more volatile

Summary and conclusions

Empirical findings

- For the whole country: choice of multilateral method not very important
- For subsections of data set: bigger differences – GEKS-Jevons quite volatile, as expected
- Downward revisions – long window required