Survey Data versus Administrative Data in the CPI Rent Price index

Statistical center of Iran

June 2023

Mostafa Farrokhfal, Abdorahim Ira, Mohsen Saadatpour Moghaddam

CPI in Iran – base year 2021
Meta data

	Samples	Items	Outlets	Price Quotes
Urban index	394 urban area	460	67,000	260,000
Rural index	671 rural area	336 (151 items collected in rural area and 185 items from urban areas)	9,000	120,000
Rent index	30,000 households in 6 symmetric panels	2		
National index		474		
Deciles index		474		

3

CPI in Iran – base year 2021
from collecting to disseminating

- Executive categories respectively:
    1. 600 enumerators
    2. 100 reviewers
    3. 50 supervises in provinces
    4. 5 statistician at Iran's NSO
- Classification: COICOP 1999
- Weights: from Households Income and Expenditure Survey (HIES)
    - About 19,000 households in urban area in 2021.
    - About 18,000 households in rural area in 2021
- Collection data:
    - 1st until 27th month for urban area using tablet.
    - 7th until 27th month for rural area using tablet.
    - Some prices collect by SCI centrally like gasoline, mobile calls and etc.
- Imputation: usually from upper level and for some items from same level.
- Dissemination: 1st of next month.

4

Dwelling market in Iran

- Rent market in Iran include landlord, tenant and real estate agency.

- both in selling market and rent market, special loan for housing is very small share.

- The little share of rent transaction register as administrative data and many of them write manipulate and don't register in administrative system.

- Total monthly rent = (0.03 * mortgage value) + monthly rent

Rent index in current method
2021=100

➡ **Actual rent index**

- ➡ **Sampling** – Target sample of 30,000 renters using probability proportional to size for primary sampling units

- ➡ **Collection** – Divide sample into six balanced panels and price each panel every six months (t and t-6)

- ➡ **Estimation** – Produce 394 urban regions rent indexes and aggregates monthly by taking sixth roots of six month rent changes to obtain 31 province rent indexes.

➡ **Owner occupy Household (OOH) index**

- ➡ **Weights** – calculate based on owner equivalent rent approach.

- ➡ **Estimation** – using 394 actual rent index and weights of OOH. In fact, equal indices for actual rent and OOH, and differ in province and country level.

Methodology
registered data source

- Ministry of roads and city planning via web services regular and monthly

- Include rental and selling data for dwelling market in all provinces

- Real estate agencies don't have any Legal compulsion to prohibit entering incorrect data

- About 30,000 until 180,000 records in rent market per month.

Methodology
outliers detection

- Dwelling data are right skewed
  - Double MAD (Median Absolute Deviation)
  - $MAD = C \cdot median(|X_i - median(X)|)$. $C = 1.4826$
  - With the MAD approach, outlier thresholds define as:

    $Lower = Median - K \cdot MAD$, $Upper = Median + K \cdot MAD$, $K_{default} = 3$.

- The classic MAD approach defines a symmetric interval around the median.

- Double MAD approach ([Rosenmai2013]). One deviation should be calculated for the numbers below the median and one for the numbers above the median
  - once on total monthly rent
  - once again on total monthly rent per square meter

Preparing data set
register data Vs survey data

Provinces	Survey sample (Monthly)	Register base sample (min)	Register base sample (max)	Register base sample (Average)
1	435	7,401	39,038	21,115
2	410	819	9,125	4,387
3	360	781	10,986	5,007
4	240	1,925	9,854	6,454
5	240	1,314	11,385	5,834
27	110	133	2,469	965
28	105	95	1,281	494
29	105	58	2,366	992
30	100	83	1,144	532
31	95	589	1,861	1,249

Preparing data set variables

- Logarithm of total monthly rent
- Logarithm of building area: Building area limited between 30 and 1000 square meter.

- Building age:

Group	Interval ages
1	< 3 years
2	4 - 9 years
3	10 - 20 years
4	21 - 35 years
5	> 35 years

- Dwelling type: apartment, villa and building.
- Skeleton type: concrete, metal, concrete and metal, brick and no skeleton.
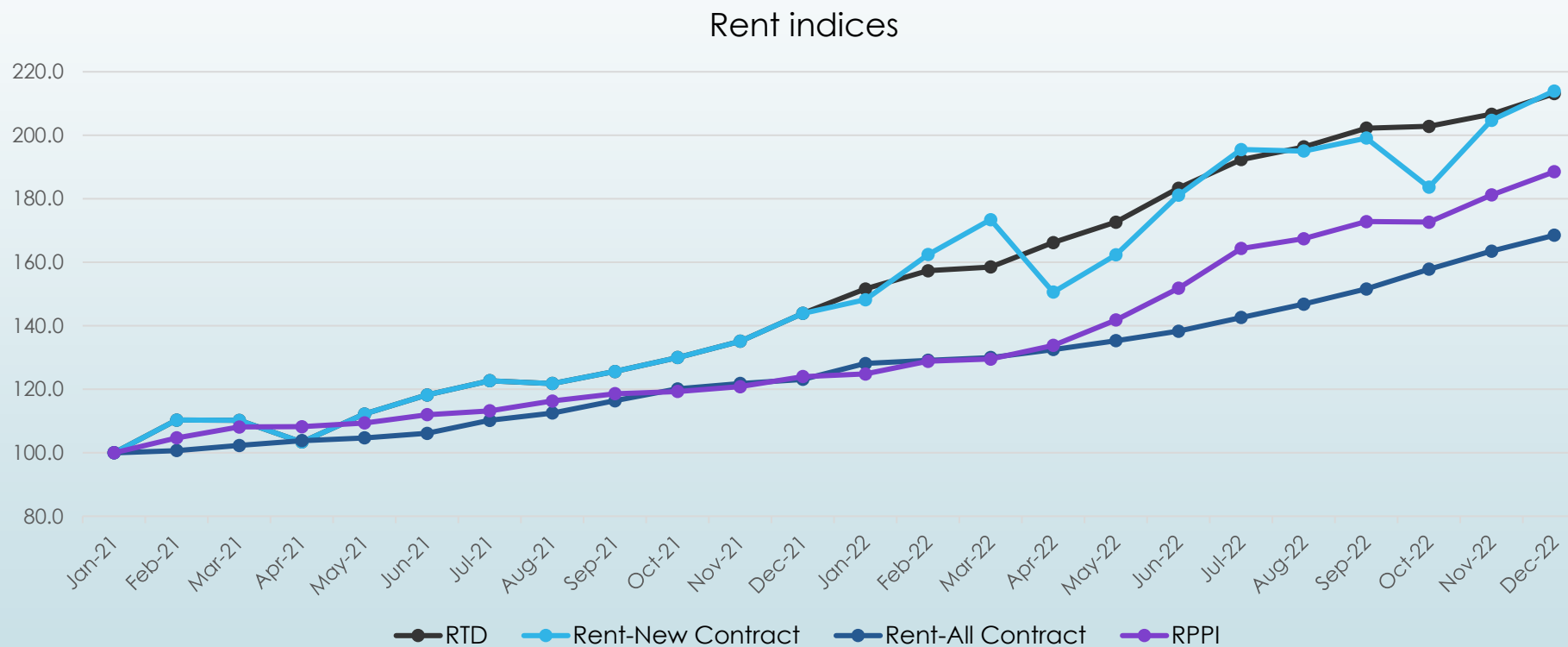- Provinces (30) and times (11) dummy variables

Regression result

Our result using R software in first rolling window:

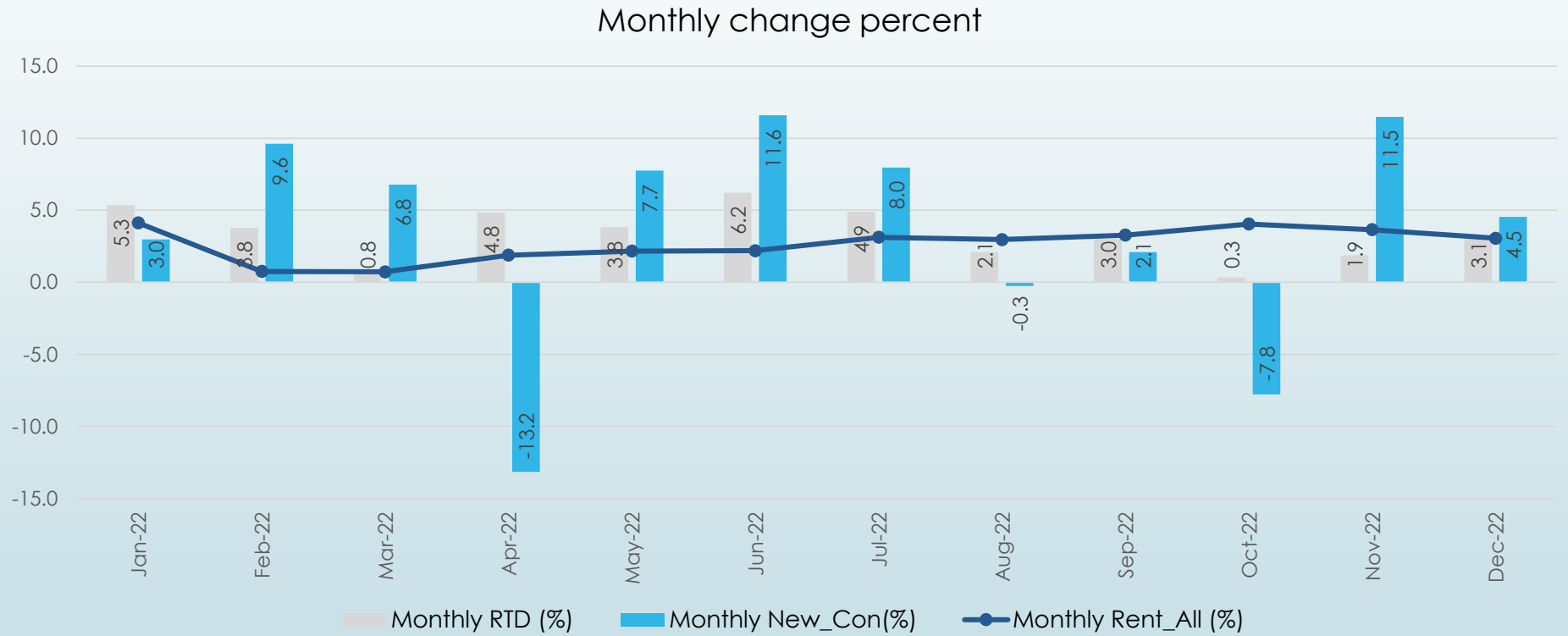| variables	Coeff	variables	Coeff	variables	Coeff	variables	Coeff	variables	Coeff
(Intercept)	14.46	factor(Province)4	-0.78	factor(Province)16	-1.05	factor(Province)28	-0.93	factor(Month)13	0.03
Log_Area	0.71	factor(Province)5	-0.47	factor(Province)17	-0.56	factor(Province)29	-1.11	factor(Month)14	0.12
factor(Age2)2	-0.12	factor(Province)6	-1.12	factor(Province)18	-0.76	factor(Province)30	-0.80	factor(Month)15	0.17
factor(Age2)3	-0.19	factor(Province)7	-1.25	factor(Province)19	-0.82	factor(Province)31	-1.32	factor(Month)16	0.20
factor(Age2)4	-0.21	factor(Province)9	-1.16	factor(Province)20	-1.01	factor(Province)32	-1.26	factor(Month)17	0.20
factor(Age2)5	-0.25	factor(Province)10	-1.40	factor(Province)21	-0.71	factor(Skeleton)1	0.09	factor(Month)18	0.23
factor(EstateType)2	-0.32	factor(Province)11	-0.81	factor(Province)22	-1.10	factor(Skeleton)2	0.22	factor(Month)19	0.26
factor(EstateType)3	-0.30	factor(Province)12	-1.48	factor(Province)23	-1.09	factor(Skeleton)3	0.15	factor(Month)20	0.30
factor(Province)1	-0.88	factor(Province)13	-0.90	factor(Province)24	-0.83	factor(Skeleton)4	0.00	factor(Month)21	0.36
factor(Province)2	-1.05	factor(Province)14	-0.89	factor(Province)26	-0.83	factor(Month)11	0.10		
factor(Province)3	-1.36	factor(Province)15	-1.36	factor(Province)27	-0.79	factor(Month)12	0.10		

➥ Residual standard error: 0.4393 on 671,086 degrees of freedom

➥ Multiple R-squared: 0.5709,   Adjusted R-squared: 0.5709

➥ F-statistic: 1.717e+04 on 52 and 671,086 DF, p-value: < 2.2e-16

Result

Register-base index vs Survey-base indices, Jan2021= 100

Rent indices

Legend: RTD — Rent-New Contract — Rent-All Contract — RPPI

Monthly change percent

Monthly change percent

Conclusion

- we try to using register-base data for indexation
- Challenges:
  - we couldn't built index at province level.
  - Monthly change percent not sensible.
  - It is out of mind that RPPI index lower than RTD or R-NC index.

- We look forward to learning from the NSOs who use hedonic method.

Thank you!

Mostafa Farrokhfal

Price indices office

Statistical Center of Iran

m.farokhfal@gmail.com