

THE EFFECT OF ATTRITION ON VARIANCE AND SAMPLE SIZE IN THE HOUSING COMPONENT OF THE CONSUMER PRICE INDEX

August 2005

William E. Larson

U.S. Bureau of Labor Statistics, 2 Massachusetts Avenue, N.E., Room 3655, Washington, D.C. 20212
Larson_W@bls.gov

Opinions expressed in this paper are those of the author and do not constitute policy of the Bureau of Labor Statistics.

Key Words: Attrition, variance, sample size, housing, rent, REQ, CPI

The Bureau of Labor Statistics (BLS) began to initiate the current housing sample in 1998 after selecting the sample based on weights derived from the 1990 Census. The Bureau has completed a continuous rotation plan that will update both the geographic areas and the housing units on a regular basis. This was projected to start in 2006, but the program did not receive funding for the current fiscal year. If the continuous rotation plan is not put into effect then the current housing sample will be used for the foreseeable future. This paper simulates the future performance of the present housing sample by using past attrition rates to predict the sample size and variance estimates of the housing component of the Consumer Price Index (CPI). The effect of different changes in the methodology of the price relative calculation (PRC) is also examined.

In section one the price index and variance estimators for rent and REQ are described. Section two presents the attrition rates obtained from observing the past performance of the units in the housing sample. In section three discusses the changes in PRC, weighting and replicate assignment that were made in this study, as well as how prices were simulated. Section four presents the variance estimates obtained. Conclusions are given in section five.

1. CPI Shelter Indexes and Price-Change Variances

For a full discussion of the CPI the reader is referred to Chapter 17 of the *BLS Handbook of Methods*. The rent estimates (also called economic rent) used in the CPI are “contract rents”, which include all services, such as utilities, included in the rent payment. The REQ estimates (also called pure rent) are contract rent payments adjusted so that only that the shelter is included.

The CPI includes 31 self-representing PSUs consisting of large Metropolitan Statistical Areas (MSAs), and 56 non-self-representing

PSUs selected to represent medium to small-size MSAs in 4 Census regions and urban non-MSAs in 3 regions.

The 1998 housing sample is a stratified cluster sample that represents housing built before 1990. Census blocks were grouped into segments based on minimum numbers of housing units, and the PSU was divided into 6 geographic strata. Segments were ordered by shelter expenditure within county within a stratum, and chosen by a systematic sample. The weight for a segment (W_s) was assigned based on segment shelter expenditure as a proportion of total PSU shelter expenditure. The shelter weight is divided between rent and REQ. The rent and REQ indexes are estimated separately using a 6-month chained estimator. The PRC aggregates the weighted rents for the units (i) in the Index Area (a) for the current period (t) and for the 6-months previous ($t-6$) and then computes a 6-month price relative:

$$REL_{a,t,t-6} = \frac{\sum_{i \in a} W_s * R_{i,t}}{\sum_{i \in a} W_s * R_{i,t-6}}$$

where R =rent.

The rent and REQ indexes need a 1-month price relative, so the 6th root of the $REL_{a,t,t-6}$ is derived:

$$REL_{a,t,t-1} = \sqrt[6]{REL_{a,t,t-6}}$$

and passed to the Index Estimation system where:

$$IX_{a,t} = IX_{a,t-1} * REL_{a,t,t-1}$$

The variance is calculated by a stratified random groups method, which is estimated at the index area-item stratum level. Upon selection, segments in self-representing PSUs are assigned to one of two or more independent samples called *replicates*. The set of all observations is called the *full sample*. In the non-self-representing areas entire PSUs are assigned to

replicates. Variance is estimated for a k -month price change. The index-area full-sample price change is:

$$PC_{a,f,t,t-k} = \left[\frac{IX_{a,f,t}}{IX_{a,f,t-k}} - 1 \right] * 100$$

where f is full sample. The replicate price change is given as:

$$PC_{a,r,t,t-k} = \left[\frac{IX_{a,r,t}}{IX_{a,r,t-k}} - 1 \right] * 100$$

where r is the replicate. Index-area variance is calculated as:

$$V(PC_{a,t,t-k}) = \frac{1}{R_a(R_a - 1)} \sum_{r=1}^{R_a} (PC_{a,r,t,t-k} - PC_{a,f,t,t-k})^2$$

where R_a = number of replicates in the index area. Area indexes are multiplied by their aggregation weight to form cost weights:

$$CW_{a,t} = IX_{a,t-1} REL_{a,t,t-1} AGGWT_{a,t}$$

These cost weights can then be combined.

$$CW_{A,t} = \sum_{a \in A} CW_{a,t}$$

Full sample price changes can then be calculated:

$$PC_{A,f,t,t-k} = \left(\frac{CW_{A,f,t}}{CW_{A,f,t-k}} - 1 \right) * 100$$

Higher level area replicate price changes are calculated as:

$$PC_{a,r,t,t-k} = 100 * \left(\frac{CW_{A,f,t} - CW_{a,f,t} + CW_{a,r,t}}{CW_{A,f,t-k} - CW_{a,f,t-k} + CW_{a,r,t-k}} - 1 \right)$$

Variance for higher level areas is formed by

$$V(PC_{A,t,t-k}) = \sum_{a \in A} \frac{1}{R_a(R_a - 1)}$$

$$\sum_{r=1}^{R_a} (PC_{a,r,t,t-k} - PC_{A,f,t,t-k})^2$$

2. Attrition Rates

The current housing panel, based on 1990 Census data, consists of housing units chosen from selected segments within the PSUs. A segment consists of one or more 1990 Census blocks. The housing units were initiated for use in the CPI beginning in 1998. Since then there have been yearly augmentations using new construction units, consisting of units selected by Census from lists provided by building permit offices in the PSUs. The sample is divided into six panels, with each panel re-priced every six months.

There is attrition in the sample, as units convert from renter-occupied to owner-occupied, and as respondents drop out of the survey. This paper looks at how attrition may affect future variance estimates in the Rent and REQ estimates of the CPI. Variances are estimated for the six panels from March 2004 to August 2004, using first the full sample, and then using 20 samples that have been reduced at estimated attrition rates through 2015.

The housing sample from March 1998 to August 2004 was examined to see if inscope attrition rates vary by different classification variables. A unit is deemed inscope if it is initiated into the sample and continues to qualify for repricings, whether or not an actual quote is obtained. If a quote is obtained and is eligible for inclusion in the PRC, then it is deemed usable. The proposed classification variables were as follows:

- number of times the unit had been re-priced
- original units vs. new construction units
- percent renter using 2000 Census block groups
- region
- decade the structure was built

The resulting attrition rates were applied to the 20 samples in an attempt to project what the sample might look like over the next eleven years.

Number of repricings. In terms of attrition, there is a large difference between the first repricing and subsequent repricings. It seems that once we get past the first repricing, there is a

much greater chance that the unit will continue in the sample, so all original unit attrition rates have been calculated after eliminating all first repricing attrition data. New construction units will be treated differently according to whether they are first or subsequent repricings.

Original units vs. new construction. Here the rate of attrition does vary depending on whether the unit is an original unit or new construction. So this classification variable was included in our analysis.

Percent renter. Each housing unit was mapped to a 2000 Census block group. For original units, different attrition rates were calculated according to percent renter. Groupings were formed according to differences found within regions. No differences were found in attrition for this variable in new construction units.

Region. There were differences found between Region 1, Region 2, and Regions 3 and 4 grouped together.

Decade structure built. There were no observable differences in attrition rates based on decade built. This classification was not used.

Table 1. Attrition Rates for Inscope Units

<i>New Construction Units</i>		
First repricing		-0.027
Two or more repricings		-0.009
<i>Original Units</i>		
REGION	PERCENT RENTER	SIX-MONTH ATTRITION RATE
1	>=80%	-0.011
	40%-80%	-0.019
	<40%	-0.026
2	>=80%	-0.007
	40%-80%	-0.014
	10%-40%	-0.020
	<10%	-0.030
3 and 4	>=80%	-0.007
	40%-80%	-0.016
	20% to 40%	-0.030
	10% to 20%	-0.043
	<10%	-0.050

The attrition rates were applied to all units in the housing sample from March 2004 through August 2004. Each unit was assigned a random number from U(0,1). If the random number was greater than 1 plus the unit's attrition, then the unit was dropped from the sample. This

procedure was repeated 23 times, simulating the sample from the second half of 2004 through the first half of 2015. (Note: I am calling the months of March through August the first half of the year, and August through the following February the second half of the year, since new construction is introduced starting in March of every year.) When original units were dropped, they were not reintroduced back into subsequent samples. But when new construction units were dropped, they were put in a pool to be reintroduced back in the next year in order to simulate normal new construction augmentation. New construction units initiated each year average about 190 units. In our simulations, the number of dropped new construction units never reached higher than 130, so all of the dropped new construction units were effectively reintroduced each year. When they were reintroduced, their attrition rate was set to the first repricing rate.

In addition to determining the attrition rate among units that are inscope, we also looked at whether there was a pattern of attrition in the usability of a unit, that is, whether a rental price could be obtained that could be used in the PRC. We found that there is a marked drop in usability over time among units that are inscope. The same factors of new construction, region and percent renter were examined for significant differences in usability attrition. Table 2 shows the usability attrition rates.

Table 2. Attrition Rates for Usable Units

<i>New Construction Units</i>		
First repricing		-0.061
Two or more repricings		-0.014
<i>Original Units</i>		
REGION	PERCENT RENTER	SIX MONTH ATTRITION RATE
1	>=80%	-0.033
	40%-80%	-0.050
	<40%	-0.055
2	>=80%	-0.019
	40%-80%	-0.031
	10%-40%	-0.044
	<10%	-0.050
3 and 4	>=80%	-0.016
	40%-80%	-0.033
	20% to 40%	-0.050
	10% to 20%	-0.069
	<10%	-0.082

Chart 1 shows the result of applying the obtained inscope and usable attrition rates to the housing sample projected through 2015. The number of inscope units falls from over 36,000 to an estimated 23,700. More importantly, the number of units with usable quotes falls by over one-half, from 28,400 to an estimated 12,600. It is important to note that these estimates were compiled using data from the first six years of this housing sample. As the length of time that units remain in the sample increases there could be an increase in “response fatigue” that may lead to increases in attrition. Since this has not been factored into the attrition rates used here, it is probably best to regard the obtained results as conservative estimates of the decline in the housing sample.

3. Changes in the PRC, Weighting and Replicate Assignment and Price Simulation

These simulations were done using two changes to the PRC that have been approved but not yet incorporated into the production of the housing component of the CPI. The first is that only collected rents will be used for imputations. The second is that there will be a minimum of five units, if possible, to be used for imputations. Both changes have been tested and have been shown to give greater stability to the PRC results. The other change is that all units have been reweighted to reflect the 2000 Census. The current sample uses weights from the 1990 Census. Reweighting to the 2000 Census is part of the proposed continuous rotation plan, but no alternative reweighting is currently scheduled. There are two issues involved in reweighting the current sample. The first, and probably most important, is that we have more up-to-date information and the sample will be more representative of the country if we use that information. However, the second consideration is that the current sample was drawn according to probabilities determined by the weights from the 1990 Census. A reassignment of weights to the current sample can give some problematic results. In the current weighting the upper range for a unit’s REQ weight is about 11,000. In reweighting the current sample, several units have weights that over 100,000. This can have an inordinate effect on variance if a highly-weighted unit happens to have an extreme price relative. However, given that these simulations attempt to project estimates for 2015, it is likely the arguments for using more current weights will prevail over arguments for using weights

that will be 25 years old. Hence the sample was reweighted to reflect the 2000 Census in this study. This reweighting also necessitated new replicate assignment. An entire segment is assigned to a replicate. The segment composition changed; in the current scheme it is based on 1990 Census blocks. In the proposed scheme it is based on 2000 Census block groups. In certain cases there is a one-to-many correspondence between 1990 segments and 2000 segments, so a new replicate assignment scheme was necessary.

After the weights were recalculated for the segments and the segments were reassigned, new price quotes were calculated for each successive version of the housing sample. Normalized rent relatives for usable quotes from September 2003 through August 2004 were used to simulate prices. The rent relatives were examined on the basis of relevant categories to determine the assignment scheme. First of all, the relatives for rent-controlled units were found to be significantly different from non rent-controlled units. Within rent-controlled units there was no difference between regions. There was a difference between Region 1 and Regions 2, 3, and 4 in the non rent-controlled units. For each reduced sample the normalized rent relative was randomly sorted within the significant rent-control and regional variables and assigned to units marked as usable. The assigned relative was multiplied by the previous period’s economic and pure rent to form the new rents. This was repeated for all of the 22 reduced versions in all 20 simulations.

4. Variance Estimates

The repeatedly reduced samples were run in the PRC to obtain price relatives, and variances were estimated according to the methods described in Section 1. The results are shown in Chart 2. The 12-month price change variance estimate for rent increased from 0.037 obtained from the actual housing sample to an average of 0.065 projected for the reduced sample in 2015. The variance estimates for the 20 simulations in 2015 ranged from 0.034 to 0.192. There was an increase in the variance estimates for REQ, from 0.038 to an average of 0.087. The variance estimates for the 20 simulations in 2015 ranged from 0.042 to 0.220. Chart 3 shows the monthly 12-month price change variance estimate for REQ for each of the 20 simulations, and Chart 4 shows the Rent results. There are marked spikes in variance at certain points. These points were

examined and it turned out in each case that a randomly assigned price relative that was extremely high or low was matched to a unit that carried an extremely high weight. For example, there is a spike in variance for one of the simulations in Chart 3 that starts at the projection for November 2006. Chart 5 is the estimated Region 1 variance for REQ, and it shows a more exaggerated spike. A closer examination shows that this spike is entirely due to one housing unit in PSU A111 that had a randomly assigned rent relative of 5.88 and a base weight of over 69,000.

5. Conclusions

By using past attrition rates the current housing sample is projected to lose over half of the units that provide usable quotes over the next ten years. It is important to note that these estimates were compiled using data from the first six years of this housing sample. As the length of time that units remain in the sample increases there could be a increase in “response fatigue” that may lead to increases in attrition. Since this has not been factored into the attrition rates used here, it is probably best to regard the obtained results as conservative estimates of the decline in the housing sample.

Likewise, the method used to assign rent prices to individual units was conservative. Only the price relatives from September 2003 through August 2004 were used. Chart 6 shows 1-month price change estimates from 1998 through August 2004. The price relatives used are represented by the broken lines at the left of the chart. This is one of the more stable periods of the graph. Other time periods, like October 1999 through November 2001 are more volatile. It is reasonable to assume that in looking ahead to 2015 there will be similar periods of volatility.

The 12-month price change variance estimates for 2015 in Chart 2 show that almost all of the simulations increase in variance, and a few show a marked increase. Budgetary considerations will determine whether the present housing sample is kept in place or whether the new continuous rotation plan will replace it.

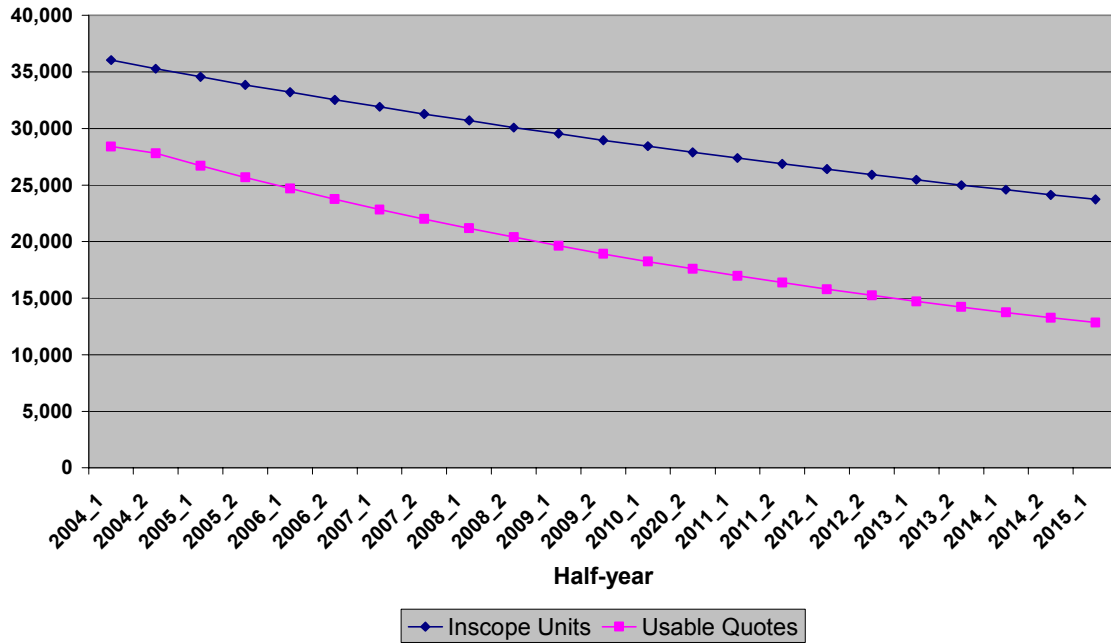
6. Acknowledgements

I would like to thank David Johnson, William Johnson, Robert Poole, and Frank Ptacek for their assistance with this paper and for their insights and questions..

7. References

Chapter 17, *BLS Handbook of Methods*, on the Internet at http://www.bls.gov/opub/hom/homch17_a.htm.

Chart 1. Number of Inscope Units and Units with Usable Quotes 2004 (Actual) through 2015 (Projected) Averaged over 20 Simulations



**Chart 2. All U.S. Rent and REQ
Mean 12-month Price Change Variance Estimate Comparison
2004 (Actual) vs. 2015 (20 Simulations)**

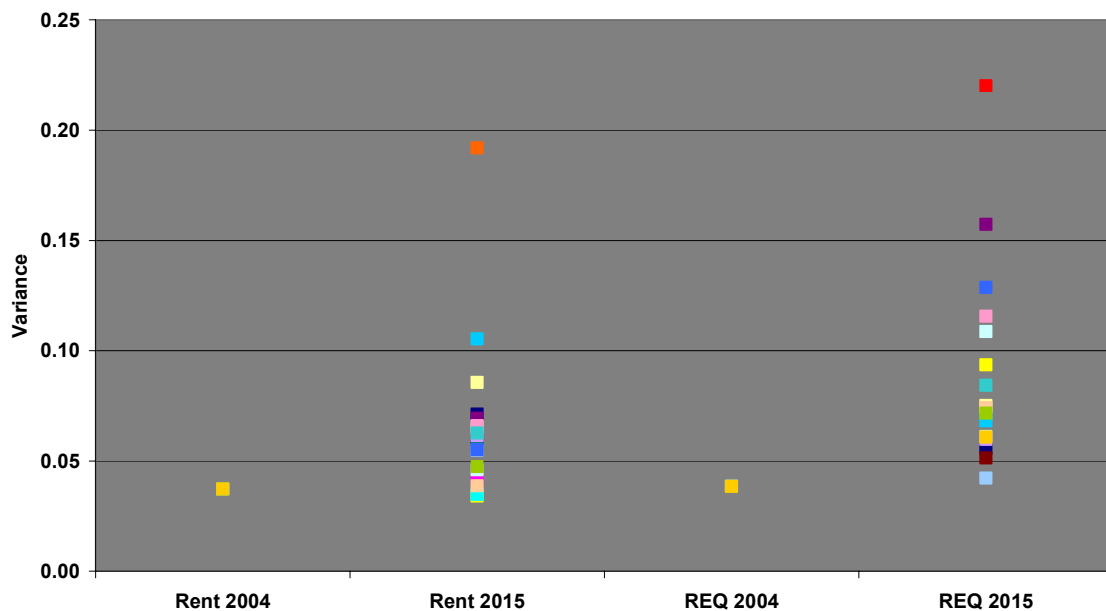


Chart 3. All U.S. REQ
12-month Price Change Variance Estimates
20 Simulations, March 2004 through August 2015

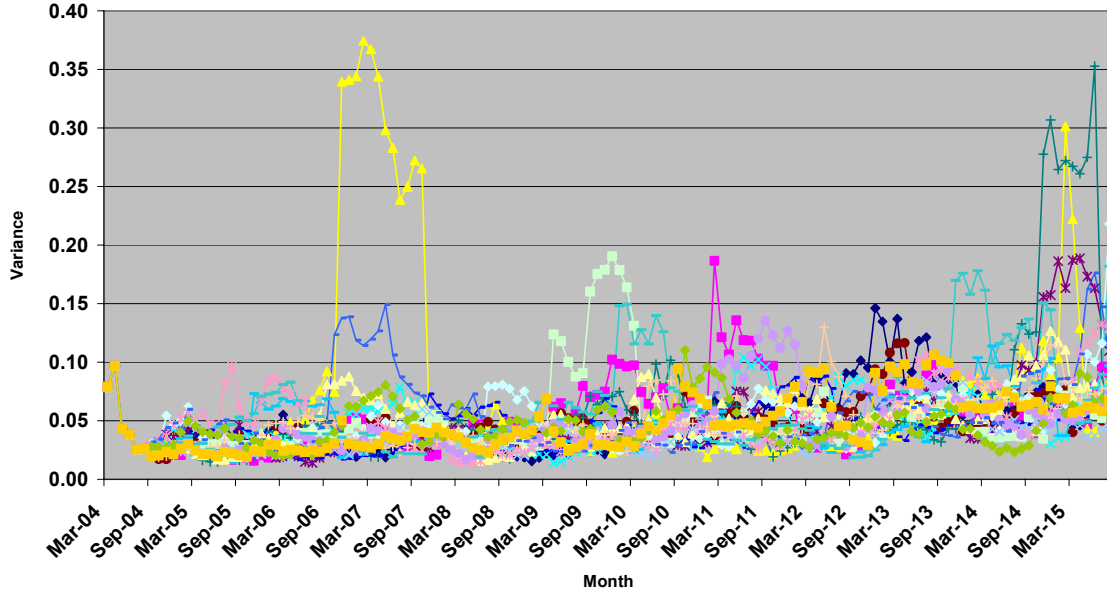


Chart 4. All U.S. Rent
12-month Price Change Variance Estimates
20 Simulations, March 2004 through August 2015

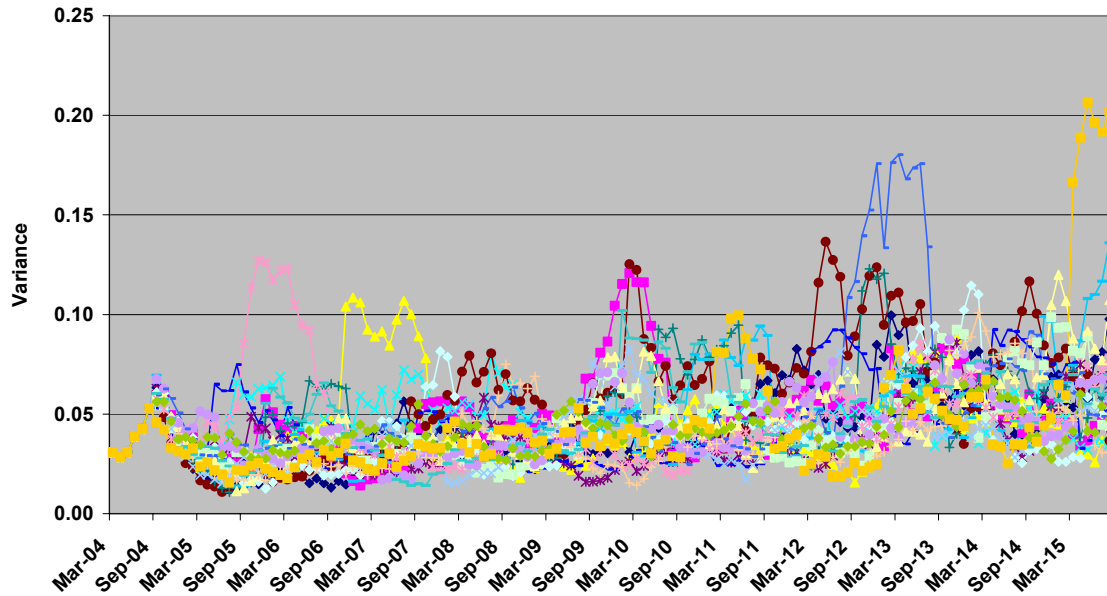


Chart 5. Region 1 REQ
12-month Price Change Variance Estimates
20 Simulations, March 2004 through August 2015

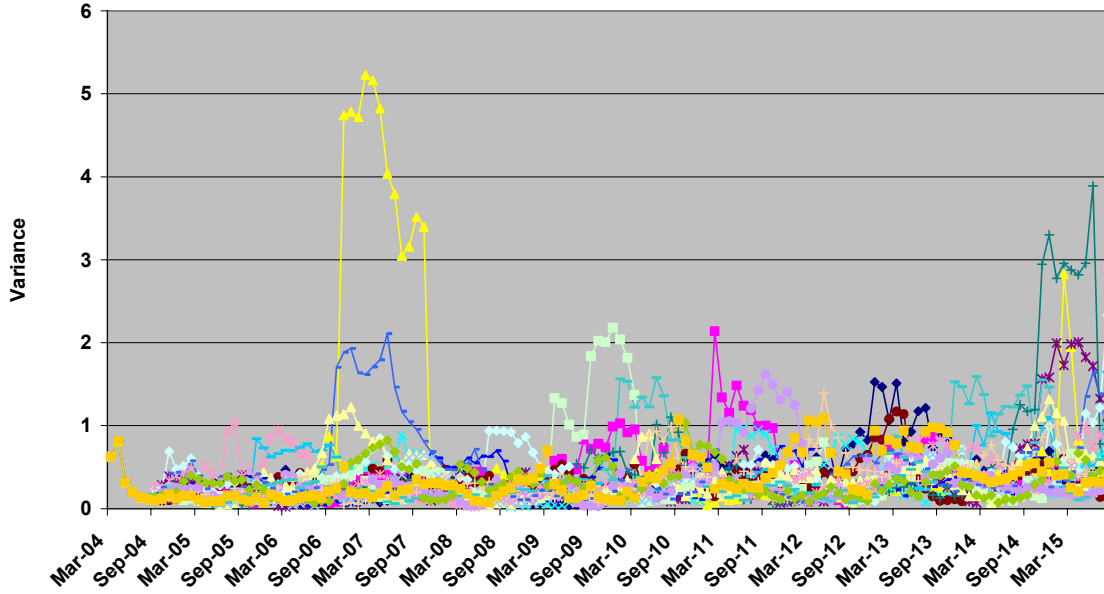


Chart 6. All U.S. Rent and REQ
1 Month Price Change
1998 to August 2004

