SPARSE DATA AND COMMERCIAL PROPERTY PRICE INDEXES

MICK SILVER AND BRIAN GRAF[1]
International Monetary Fund, Washington DC, USA[2]
Statistics Department
Corresponding author's email: msilver@imf.org

Both appraisals of the value of properties and transaction prices are used as the underlying database used for the measurement of commercial property price indexes (CPPIs). The former is based on judgment and may smooth and lag transaction prices. The latter is based on actual transactions and may have sample selectivity bias and limited sample sizes for heterogeneous properties. In principle, actual transaction prices are preferred, but their inherent problem of sparse data on heterogeneous properties limits their use in practice. We examine such issues using counts information on granular CPPIs for the United States (US).

---Draft version not to be cited without author's permission---

JEL Classification Numbers: C43, E30, E31, R31.

Keywords: Commercial Property Price Indexes; Commercial Property Inflation; Appraisal Indexes; Transaction Indexes; Weighted Least Squares.

1. Introduction

In principle commercial property price indexes (CPPIs) should be based on actual transactions. In practice, commercials properties are heterogeneous and transactions irregular thus complicating comparisons of average transaction prices for a fixed-quality bundle of properties over time—the type of methodology that would be applied for a consumer price index. Even where matched (repeat) transactions are used, the sample of properties sold more than once in the period of the index can be very limiting and its implicit selection procedure biased. The problem of sparse transactions on heterogeneous properties is worse when measurement really matters, as we go into and during recessions.[3]

---

[1] Acknowledgements are due to Real Capital Analytics (RCA) for the supply of the data used in this study and subsequent technical advice on the features of the data. We note our due recognition of caveats as to the confidentiality of the data. Acknowledgements for helpful comments are given to Jeff Fisher (RCA), David Geltner (MIT), and Xingwei Hu (IMF).

[2] The views expressed herein are those of the author and should not be attributed to the IMF, its Executive Board, or its management.

[3] Similar problems arise with residential property price indexes (RPPIs) (Silver, 2012) though even during recessions the number of transactions remains reasonable for estimation purposes and the heterogeneity is not as marked as commercial property.

This shortcoming of CPPIs being based on sparse data has limited their provision and use. Instead of using actual transactions-based CPPIs, appraisal-based indexes using valuations are more commonly used. However, as outlined in Silver (2013), they too suffer from limitations: they are subjective and not every property is regularly or properly reappraised every period, see Kanutin (2013). There is evidence of a dampening or smoothing of market price volatility and a tendency of appraisal indexes to lag market price indexes. Users have an established preference for transaction-based indexes. Kanutin (2013, page 4).

Accurate measure of commercial property price indexes (CPPIs) are recognized as important for economic analysis, monetary policy, financial stability and prudential supervision. The International Monetary Fund (IMF)/Financial Stability Board (FSB) G-20 Data Gaps Initiative (DGI) includes CPPIs in its 20 recommendations on data gaps (see Heath, 2013). Yet their measurement is highly problematic.

We use, for the United States (US) from 2000Q4–2012Q4, granular CPPI series disaggregated by type of transaction (apartment, industrial, office, and retail) in specified geographical areas with accompanying series on the "counts"—number of transactions— used each quarter to estimate the index. We use a more efficient estimator due to the availability of information on counts, the source of our problem of sparse data.

The features of appraisal and transaction-based indexes are outlined in Silver (2013). Section 2 outlines methodological issues that take account of sparse data in estimating commercial property price inflation; section 3 provides empirical results using US data; and conclusions are drawn in section 4.

2. CPPIs and sparse data: method

The analysis below is based on "granular" transaction-based CPPI series provided by Real Capital Analytics (RCA CPPI).[4] It extends from 2000:Q4 to 2012:Q4 and focuses on relatively high-value transactions; RCA data from 2000 covered transactions of over $5 million but in 2005 extended this to transactions over $2.5 million.

Attention in this paper will first be focused on an aggregate CPPI for apartments only derived from sub-indexes for 34 metropolitan areas, and second, on an aggregate CPPI derived from sub-indexes for four types of properties: apartments, industrial, office, and retail.

Importantly, to gain insights into the problem of sparse data, RCA have provided us with confidential data on the "counts"—number of transactions—underlying each of their granular indexes in each period. Our interest here is whether by incorporating this counts data as weights in a WLS estimator, the measurement problems associated with sparse data can be ameliorated.

Pooled ordinary least squares (OLS) and weighted least squares (WLS) fixed cross-section and time-period estimators were used to derive estimates of commercial property price inflation for apartments. Consider a pooled estimator for national property price inflation for apartments derived from the i=1,…34 US metropolitan areas over the period 2001Q1 to 2012Q12:

[4] See https://www.rcanalytics.com/Public/rca_cppi.aspx; for details on methodology see: presentation by David Geltner under "MIT/CRE Historical Development of CPPI" at http://web.mit.edu/cre/research/credl/ and Geltner and Pollakowski (2007).

$$dCPPI_{it} = \ln\left(\frac{CPPI_{it}}{CPPI_{it-1}}\right) = \sum_i \delta_i FEmetro_{it} + \sum_t \lambda_t FEtime_{it} + \nu_{it} \qquad (1)$$

where $dCPPI_{it}$ is commercial property price inflation for property type *i* in period *t*; $FEmetro_{it}$ and $FEtime_{it}$ are respective fixed-effects for metro areas and quarterly time periods with associated coefficients δ_i and λ_t; and ν_{it} is an error term with the usual desirable properties. The model is estimated using the least squares dummy variable (LSDV) method (Wooldridge, 2002) by both OLS and WLS estimators; the weights for WLS are $\sqrt{count_{it}}$ where $count_{it}$ are the number of observed price transactions for metro area *i* in period *t*. The assumption is that $V(\nu_{it}) = \sigma^2 / count_{it}$; as count increases, the variance decreases. Estimates of λ_t form the basis of the estimates of the property price inflation series in Figure 1 (and in the next section's Figure 2).[5] *Taken as a whole series and paying due regard to the quality of the counts data*, the WLS more efficient estimates can be argued to better reflect changes in commercial property price inflation.[6]

First, the OLS series relies heavily on a homoskedasticity assumption, one that is questionable given the variability in the counts data, both across metro areas and, more particularly, over time. OLS gives less precisely measured observations more influence than they should have and more precisely measured ones too little influence. WLS using, in this context, counts information assigns a weight to each observation that reflects the uncertainty of the measurement and thus improves the efficiency of the parameter estimates.

Second, this focus on the efficiency of the estimator is in line with the literature on "errors in measurement" of the dependant variable, the commercial property price index—Hausman (2001) and Wooldridge (2002). This literature shows such measurement errors result in OLS parameter estimates that are unbiased, but inefficient, with reduced precision and associated lower *t*-statistics $\bar{R}^2$.

Third, estimates of price changes from small samples are more likely to contain extreme values compared with those from larger samples and, with a skewed distribution, have a directional bias. Extreme values in regression-based frameworks have undue leverage effects on the parameter estimates (Hausman, 2001). WLS gives les emphasis to such extreme values.

3. CPPIs and sparse data: empirical results

The series considered in section A are quarter-on-quarter price inflation from 2000:Q1 to 2012:Q4 for apartments broken down by 34 metropolitan areas, and in section B, by four types of properties

[5] We follow Kennedy (1981) and use as the estimate of the proportionate impact of the period *t* time dummy, the consistent (and almost unbiased) approximation: $\left[\exp\left(\hat{\lambda}^t\right) / \exp(V(\hat{\lambda}^t / 2))\right] - 1$ where $\hat{\lambda}^t$ is the OLS estimator of λ_c in equation (1) above and $V(\hat{\lambda}^t)$ its estimated variance, see also Giles (2011).

[6] The above framework cannot be applied to an individual isolated bilateral quarter-on-quarter inflation estimate, RCA repeat sales estimates are, however, based on the extended sample of transactions data and do not utilize isolated bilateral comparisons.

(apartment, industrial, office, and retail); these in turn constitute $CPPI_{it}$ on the left-hand-side of equation (1) above. [7] The former is used to estimate CPPI for apartments and the latter for overall commercial property price inflation. In both sections we consider RCA CPPI series estimated using pooled OLS and, making use of the counts series, using pooled WLS.

A natural first step is to identify whether there is any association between the number of observed price quotes in any period constituting the index[8] and commercial property price inflation. We would expect a positive relationship; the collapse during this current recession in commercial property price inflation to be associated with a fall in transaction counts and, similarly, movement out of the recession with increasing inflation and increased counts. Table 1 shows there is a consistent positive association between inflation and counts for each of office, retail, industrial and apartment properties and their respective counts for the period 2001Q1 to 2012Q12 and within pre-recession and recession sub-periods; the association is statistically significant at a 5 percent level for each type of property in the period of the recession and for the aggregate of all properties in each period considered.

Table 1, Correlation coefficients: count and commercial property price inflation:

	All	Office	Retail	Industrial	Apartment
Whole period: 2001Q1: 2012Q4	0.298**	0.215	0.286*	0.312**	0.239
Pre-recession: 2000Q4: 2007Q4	0.538***	0.702***	0.007	0.721***	0.168
Recession: 2008Q1: 2012Q4	0.612***	0.487**	0.601***	0.578***	0.622***

***,**,* denotes statistically signifcicant for two-tailed test at 1, 5, and 10 percent levels respectively.

a. Disaggregation of apartment price inflation by 34 metropolitan areas

The estimation of commercial property price inflation in this section is for a single property type, that is apartments, using granular apartment indexes for 34 metropolitan areas. Such inflation measures, if derived for all four property types, can then be explicitly weighted together to compile an aggregate series of commercial property price inflation covering the four property types. The resulting index would benefit from flexibility in applying explicit weights rather than the implicit equal weights in the regression-based aggregates by RCA and above. Details of the regression results are available from the authors.

Figure 1 shows there is some apparent convergence between WLS estimated inflation and OLS/RCA inflation for US apartments going into the recession, coming out, and thereafter. However, (i) WLS estimated inflation for apartments in the first half of the decade and lead up to the recession was substantively higher than estimates by OLS and RCA, and (ii) the amplitude of the trough around 2009Q1 measured by WLS was markedly smaller. Less precisely measured index values, due to smaller sample sizes, are more likely to have extreme values with relatively high leverage; WLS accordingly gave the extreme values in the downturn less weight. [9] This downgrading has served here

[7] The null hypothesis of individual unit roots for this pooled data set was rejected when tested.
[8] For the RCA CPPI data the count in a given quarter is the number of transactions in that period for which there are earlier transaction-pairs on the same property. It may be argued that if there is more than two such transactions, the earlier data contribute to the effective sample.

[9] The recessionary trough is characterized by negative skewness in regional apartment property inflation while positive skewness is the norm outside of this period.

to reduce the extreme falls, for some metropolitan areas, in inflation leading to WLS estimates exceeding those estimated by OLS. Commercial property price inflation for apartments in the US was understated in the first-half of the decade and lead up to the recession and its fall in the initial phase of the recession overstated.

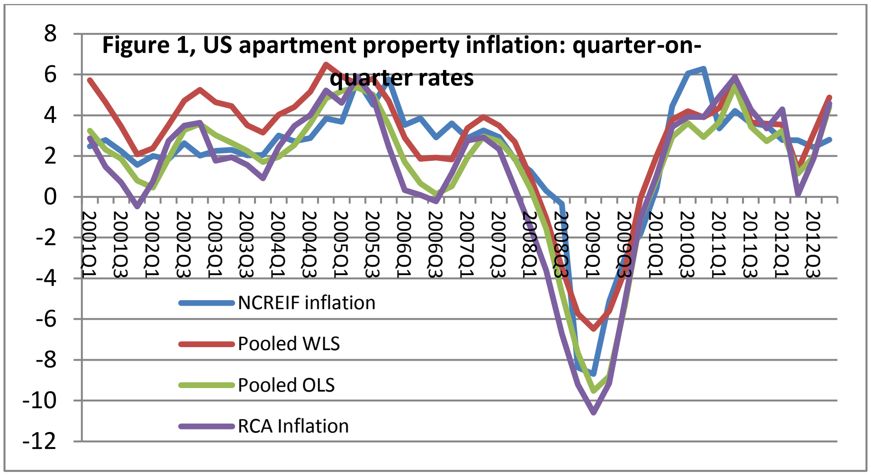

Figure 1, US apartment property inflation: quarter-on-quarter rates

b. Disaggregation by four property types: apartment, industrial, office, and retail

Figure 2 show overall commercial property price inflation estimated from a pooled quarterly series over 2001Q1 to 2012Q12 of the four property type in a model that includes fixed-time and cross-property type effects, as outlined above, both without using the counts series, an OLS pooled estimator, and with using the counts series as weights in a WLS pooled estimator. Also included for comparison are the RCA's overall commercial property and the NCREIF's valuation-based commercial property inflation series. It should be noted that the series by RCA are the result of a repeat sales regression made up from the sample of repeat purchases.

Figure 2 shows the pooled OLS estimates to closely follow the RCA (regression-based) estimates, as would be expected. The WLS and OLS estimates are very close, only noticeable departing during 2010. Aggregation at this higher level of counts is not affected by the differential relative counts of the underlying transactions for each series. None of the four series have extreme movements associated with a relative paucity in the numbers of observations. Were it the case, the WLS estimator would give them less emphasis in determining the series, compared with OLS.

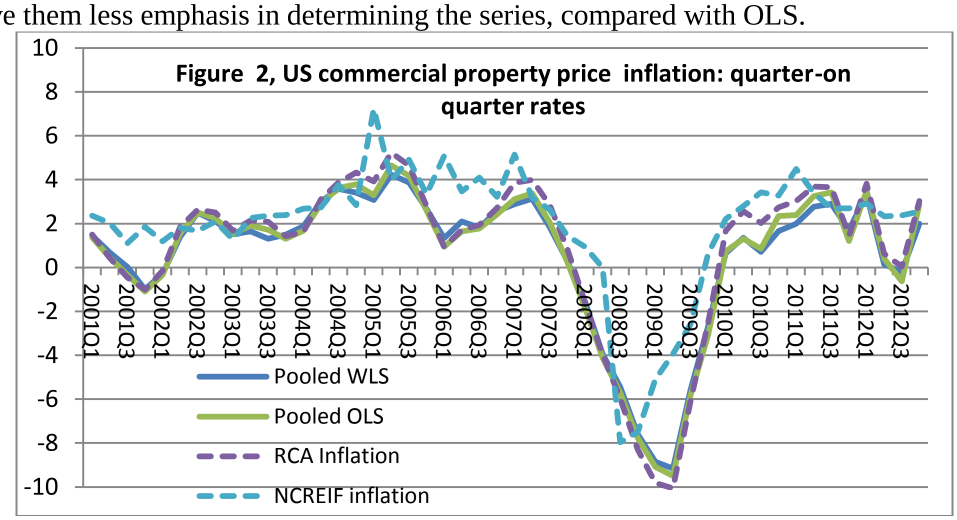

Figure 2, US commercial property price inflation: quarter-on quarter rates

4. Conclusions

CPPIs based on transaction data, although generally considered to be desirable, have been plagued by concerns over their derivation from sparse data. Organizations with a need for CPPI macroeconomic series, such as the European Central Bank (ECB), acknowledge their user-preference for transaction-based series, but have turned to using valuation-based series with their incumbent problems, because of sparse data problems associated with CPPIs. We consider methods of improving transaction data first, in this paper, by explicitly making use of information on "counts," the number of transactions each period, to improve the efficiency of the regression-based estimates.

We demonstrate that when compiling series from disaggregate regional series using regression-based estimators, incorporating counts data by means of a WLS estimator markedly changes, and given its increased efficiency, improves, estimated inflation. We show that, for example, the bottom of the recessionary trough in 2009Q1 to be about 3 percentage points less than would otherwise be considered. We demonstrate that at higher levels of aggregation, by property type into national inflation measures, the gains from using counts data in this manner are minimal. Thus a proposed strategy arising out of this work would be to use counts data to derive aggregate measure from regional data for each property type, and then explicitly weight, say using relative values of the stock of properties, the resulting indexes to form a national commercial property price index.

Further research into improving estimates of commercial property price inflation is underway to first, explore inclusion of spatial autocorrelation of the residuals into the modeling and second, identify means by which valuation-based data may be used to improve CPPI estimates.

References

Geltner, David and Henry Pollakowski (2007), A Set of Indexes for Trading Commercial Real Estate Based on the Real Capital Analytics Transaction Prices Database, MIT Center for Real Estate, *Working Paper*, Release 2.

Giles, David E., (2011), Interpreting Dummy Variables in Semi-logarithmic Regression Models: Exact Distributional Results, University of Victoria, Department of Economics, *Econometrics Working Paper* EWP1101, January.

Hausman, J. (2001). Mismeasured Variables in Econometric Analysis, Problems from the Right and Problems from the Left. *Journal of Economic Perspectives* 15, 4, 57-67.

Heath, Robert (2013), Why are the G-20 Data Gaps Initiative and the SDDS Plus Relevant for Financial Stability Analysis? *IMF Working Paper Series*, WP/13/6.

Kanutin, Andrew (2013), ECB Progress Towards a European Commercial Property Price Index. Paper presented at the 13th Ottawa Group Meeting held from 1–3 May, 2013, Copenhagen, Denmark. Available at: http://www.dst.dk/da/Sites/ottawa-group/agenda.aspx.

Kennedy, Peter, E., (1981), Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations, *American Economic Review*, 71, 801.

Silver, Mick (2012), Why house price indexes differ: measurement and analysis, *IMF Working Paper* WP/12/125.

Silver, Mick (2013), Understanding Commercial Property Price Indexes, *World Economics*, 14, 3,

Wooldridge, Jeffrey M. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, Massachusetts and London, England: The MIT Press.