DISCUSSION PAPER SERIES

No. 7344

REPEAT SALES INDEXES: ESTIMATION WITHOUT ASSUMING THAT ERRORS IN ASSET RETURNS ARE INDEPENDENTLY DISTRIBUTED

Rachel Campbell, Kathryn Graddy and Jonathan Hamilton

INDUSTRIAL ORGANIZATION

Centre for Economic Policy Research

www.cepr.org

REPEAT SALES INDEXES: ESTIMATION WITHOUT ASSUMING THAT ERRORS IN ASSET RETURNS ARE INDEPENDENTLY DISTRIBUTED

Rachel Campbell, Tilburg University and University of Maastricht
Kathryn Graddy, Brandeis University and CEPR
Jonathan Hamilton, University of Florida

ABSTRACT

Repeat Sales Indexes: Estimation Without Assuming that Errors in Asset Returns Are Independently Distributed

This paper proposes an alternative specification for the second stage of the Case-Shiller repeat sales method. This specification is based on serial correlation in the deviations from the mean one-period returns on the underlying individual assets, whereas the original Case-Shiller method assumes that the deviations from mean returns by the underlying individual assets are i.i.d. The methodology proposed in this paper is easy to implement and provides more accurate estimates of the standard errors of returns under serial correlation. The repeat sales methodology is generally used to construct an index of prices or returns for unique, infrequently traded assets such as houses, art, and musical instruments which are likely to be prone to exhibit serial correlation in returns. We demonstrate our methodology on a dataset of art prices and on a dataset of real estate prices from the city of Amsterdam.

Rachel Campbell
Room 1.021B, Faculty of Economics
and Business Administration
PO Box 616
6200 MD Maastricht
THE NETHERLANDS

Email:
R.Campbell@FINANCE.unimaas.nl

Kathryn Graddy
Brandeis University
Department of Economics MS 0221
415 South Street
Waltham, MA 02454
USA

Email: kgraddy@brandeis.edu

Jonathan Hamilton
Department of Economics
University of Florida
Gainesville FL 32611-7140
USA

Email: hamilton@ufl.edu

1. Introduction

The repeat sales methodology has become an important technique to determine price trends and returns for idiosyncratic assets, including real estate, art, and antique musical instruments. The basic principle is to improve upon hedonic regression techniques by using pairs of observations on the same asset. In effect, this is using fixed effects estimation with the usual advantage of controlling for unobserved characteristics of individual assets.

Bailey, Muth, and Nourse [1963] were the first to propose a methodology for repeat sales regressions, simply using ordinary least squares. Case and Shiller [1987] improved on this with a three-stage generalized least squares (GLS) methodology. Under the assumption that deviations from the mean single-period returns by the underlying assets are independently and identically distributed, the variance of returns grows linearly when single-period returns are aggregated over the holding period of an asset. Thus, the errors in the repeat sales regression are heteroskedastic. In order to correct for this heteroskedasticity, one proceeds by first estimating OLS regressions using dummy variables for time periods between sales (or -1 and +1 indicators for first and second sale dates if one is estimating index levels). Then, the squared residuals are regressed against the length of the holding period.[1]   The estimates from the second stage regression are then used to form weights for the third-stage GLS regressions. While both the OLS and GLS regressions provide unbiased estimates of the return coefficients for each period, the

---

[1] The second stage regressions also include a constant. Since the repeat sales approach purges estimates of "house-specific" errors, the constant represents the variance of a "transaction-specific random error" which is independent of the holding period. With this explanation, the constant must be positive.

GLS regressions will be more efficient.[2] Our goal in this paper is to study the implications of non-i.i.d. errors on the specification of the second-stage regression. We show that in many cases, the i.i.d. hypothesis on the individual asset returns is rejected and that efficiency can be easily improved by a more general specification of the second stage. We demonstrate our method using two different datasets: art sold in Amsterdam between 1780 and 2007, and real estate sold along the Herengracht canal in Amsterdam between 1630 and 1972.

There has been considerable research since Case and Shiller on repeat sales techniques. It is well-known that the standard specification of logarithmic returns estimates the geometric mean of property returns. The estimate of the arithmetic mean in a period depends on the geometric mean return and the cross-sectional variance of the per-period return of the individual assets. Thus, any biases in estimates of the variance affect point estimates of the arithmetic mean return. Goetzmann [1992] proposed using the second stage of the Case-Shiller method to estimate the cross-sectional variance. We show that under any assumptions other than i.i.d., the Goetzmann correction provides a biased estimate of the variance in per-period returns, and thus a biased estimate of the per period arithmetic return.

In Section 2, we explore work that has used the Case-Shiller method and some of their findings. In Section 3 we detail the Case-Shiller methodology and look at its effects on our two applications.   Section 4 explores different assumptions regarding the asset return errors.   Section 5 proposes a non-parametric alternative to the standard Case-

---

[2] It is well-known that the last few periods in the sample have only a small fraction of transactions spanning them. The estimated returns in these last periods may have large standard errors, making the GLS efficiency improvements particularly critical.

Shiller method. In Section 6 we discuss further applications and extensions and in Section 7 we conclude our analysis.

2. The Importance of Repeat Sales Indexes

and the Case-Shiller Method

In the current economic environment and the resulting subprime crisis, economists are paying very close attention to estimates of house price movements. Efficient and consistent estimates are important for gauging the state of the economy.

Both the OFHEO (Office of Federal Housing Enterprise Oversight) and the S&P/ Case-Shiller home price indexes use a variation of the Case-Shiller method to calculate their indexes.   While the Case-Shiller method proposes using a linear specification for the second-stage regressions, regressing the squared errors from the first stage on the time between sales (which theoretically results from the i.i.d. assumption on the individual asset returns), the OFHEO approach (see Calhoun [1996] for details) fits a quadratic equation—regressing the squared error on time between sales and the square of time between sales. Abraham and Schauman [1991] discussed fitting a quadratic term in the holding period, but they did not discuss the theoretical implications. Calhoun [1996] states that, in practice, the constant term in the second-stage regression is often negative, which is inconsistent with the Case-Shiller explanation. Calhoun suggests forcing the constant to zero and re-estimating, which is the approach taken by the OFHEO in their regressions.   The OFHEO presents the coefficients from their second stage regression so that users may go from the geometric to the arithmetic mean.

Case and Shiller directly estimate an arithmetic index but still use the standard Case-Shiller correction to correct for heteroskedasticity. The addition of non-linear terms

4

to the Case-Shiller correction, as discussed below, could possibly improve the efficiency of their estimates.

In the literature on estimating returns to art, Goetzmann [1993] and Mei and Moses [2002] use the second stage regression coefficients in the Case-Shiller estimation scheme to estimate the variance of the cross-sectional return and continue to make use of the i.i.d. assumption on the errors in the single-period returns. These papers do not present the second-stage regressions, so readers obtain little evidence on the fit in this regression.

Quigley [1995] proposes estimation of a hybrid model using both repeat sales and hedonic estimates for properties sold once and fits the squared residuals to a quadratic function of elapsed time without a constant. He refers to this as the implication of a random walk, with little further explanation. In a subsequent paper, Hwang and Quigley [2004, p. 165] specify an error distribution "to include mean reversion as well as a random walk" in which they model autoregression in the errors in price levels. This may be limiting and could have an effect on the efficiency of the index estimation. We therefore start by dropping the i.i.d. assumption and deriving the statistical properties of the effect of the holding period on the variance of returns. Unlike Hwang and Quigley, we consider violations of independence in the errors in returns, rather than errors in price levels. It turns out to be relatively straightforward to model the effects of short-term violations of independence (such as first-order moving average processes) on the variances at different holding periods. In most cases, it is not possible to identify the statistical properties of the error process on the individual asset returns from the pattern of residuals.

3. Does the Case-Shiller Method Correct for Heteroskedasticity?

Assume that there are N observations of repeat sales in a data set. Each observation consists of the purchase (buy) date, b_i, the purchase price, B_i, the sale date, s_i, and the sale price, S_i. The purchase and sales dates span the interval from $t = 0$ to $t = T$.[3]

3.1 The Basic Case-Shiller Model

(i.i.d. errors on individual one period asset returns)

Define the length of the holding period as $\tau_i = s_i - b_i$. Let $y_i = \log\left(\dfrac{S_i}{B_i}\right)$ be the log of the compound return on property i. We can write the compound return as the sum of the returns to property i in each period between purchase and sale, or:

$$y_i = \sum_{t=b_i}^{s_i} r_{i,t} \quad \text{where} \quad r_{i,t} \equiv \log\left(\frac{P_{i,t}}{P_{i,t-1}}\right),$$

and $P_{i,v}$ is the price of property i in period v (which is not observed except for $v = s_i$ and b_i). The standard assumption in repeat sales regressions is that $r_{i,t} = \mu_t + \varepsilon_{i,t}$, where $\varepsilon_{i,t}$ is independent and identically normally distributed. Then, we can rewrite the compound return as:

$$y_i = \sum_{t=b_i}^{s_i} \mu_t + \sum_{t=b_i}^{s_i} \varepsilon_{i,t}.$$

Most analyses using repeat sales regressions are primarily interested in the estimated μ_t series, either to uncover the time pattern of property returns or their expectation and variance over time.

---

[3] Our presentation adopts some of the notation in Goetzmann and Peng [2002]. We use their notation as they work in returns, which is the approach that we have adopted. Case and Shiller originally worked in levels, and then differenced the levels, which resulted in an index being estimated, rather than returns.

In the above case, the error term $\sum_{t=b_i}^{s_i} \varepsilon_{i,t}$ grows linearly with the length of the holding period. Under this assumption, one need not actually use the three-stage procedure since one can simply use $\dfrac{1}{\sqrt{s_i - b_i}}$ as the weights for GLS.

Case and Shiller [1987] assumed that:

$$\log(P_{i,t}) = C_t + H_{i,t} + N_{i,t}$$

where C_t is the value of the index in period t (cumulated returns to time t), $H_{i,t}$ is the value of a random walk process for property i at time t, and $N_{i,t}$ is the "sale-specific random error". This is the Case-Shiller framework with "transaction risk". The above expression is equivalent to writing the price of property i in period T as:

$$P_{iT} = \exp\left(\sum_{1}^{T} \mu_t + \sum_{1}^{T} \varepsilon_{it} + \upsilon_{iT} \right)$$

where $\upsilon_{iT} \neq 0$ only if a transaction occurs in period T. Taking logs and differencing prices at two different transactions, we obtain:

$$\ln\ P_{iT}\ - \ln\ P_{iT-k}\ = \sum_{T-k}^{T} \mu_t + \sum_{T-k}^{T} \varepsilon_{it} + \upsilon_{iT} - \upsilon_{i,T-k}.$$

Note that this expression holds even if other transactions occur between periods T-k and T (unlike if there is additional variation in the *return* in a period with a transaction).

Then let $\kappa_i = \sum_{T-k}^{T} \varepsilon_{it} + \upsilon_{iT} - \upsilon_{i,T-k}$ be the residual for property i in the first-stage regression. Since the transaction errors are independent of the return errors, we can use all our earlier results. We now would have:

$$E\left[\kappa_i^2\right] = E\left(\sum_1^\tau \varepsilon_t\right)^2 + 2\sigma_\upsilon^2$$

where σ_υ^2 is the expectation of $\upsilon_{i,t}^2$, under the assumption that the $\upsilon_{i,t}$ are i.i.d.

(Quigley et al. allow for serially correlated $\upsilon_{i,t}$).

3.2 Applications

We start by looking at how well the Case-Shiller method corrects for heteroskedasticity empirically with two distinct repeat-sales datasets. The first dataset we use comprises 1,468 observations of repeat sales of art sold in Amsterdam between the years 1780 and 2007 and was put together by Rachel Campbell using catalogues at the Rijksmuseum in Amsterdam, analyzed in Campbell & Kraeussl (2008). The second dataset consists of 3,577 repeat sales of houses that were sold along the Herengracht canal in Amsterdam between the years 1630 and 1972. This dataset is analyzed in Eichholtz [1997].

We employ the Case-Shiller method as follows. In the first stage regressions we regress the difference in the logs of the price change on dummy variables $X_{i,j}$. When asset i is purchased in period b_i and sold in period s_i, $X_{i,j}$ take on the value 1 for $b_i < j \le s_i$ and zero otherwise. In the second stage of the regressions, we regress the square of the residuals from the first stage on the holding period and a constant. For the 3rd stage regressions, we construct a weighted least squares regression by dividing the regressand and the regressors from the first stage by the square root of the predicted values from the second stage. For the art dataset, following Goetzmann (1993) we estimate the model using 10-year periods. For the Herengracht dataset, following Eichholtz (1997) we estimate the model using 2-year periods.

8

We present summary results in Table 1 (Appendix Tables 1 and 2 present full results of the first and third stage regressions). In Table 1 we also present the test results from the Koenker-Basset test for heteroskedasticity. In this test, the squared residuals from the regression model $(\hat{u}_i^2)$ are regressed on the squared estimated predicted values of the dependent variable $(\hat{Y}_i^2)$ and a constant: $\hat{u}_i^2 = \alpha_1 + \alpha_2(\hat{Y}_i^2) + v_i$. The null hypothesis is that $\alpha_2 = 0$. If this is not rejected, then one could conclude that there is no heteroskedasticity.

As shown in Table 1, the stage I R^2 for the Amsterdam art dataset is about .69 and about .60 for the Herengracht dataset, which is typical of repeat sales datasets. For the Amsterdam art dataset, we marginally cannot reject that there is no heteroskedasticity, but for the Herengracht dataset, we reject the hypothesis of no heteroskedasticity. The coefficient on time between sales is significant in both datasets. The Stage III R-squareds fall slightly relative to OLS, and again we cannot reject that there is no heteroskedasticity in the art dataset, but we continue to reject the hypothesis of no heteroskedasticity in the Herengracht dataset.

3.3 Evidence of Non-i.i.d. Errors

Table 2 and Table 3 present different specifications of the second stage of the Case-Shiller method. For the Campbell art dataset, the standard Case-Shiller specification (shown in column 1) is dominated using the adjusted R^2 criterion and some other measures by all other specifications that include a constant. For the Herengracht dataset, the standard Case-Shiller specification does not appear to be unequivocally dominated. Nonetheless, according to the Koenker-Basset test, the Case-Shiller method fails to get rid of the heteroskedasticity, as shown in Table 1 above.

Yet more evidence of non-i.i.d. errors is presented in Table 4. If the errors are

i.i.d., the variance of the return error for an individual property grows linearly with time.

Hence, the coefficient on time between sales when estimating two year period returns

should be twice the coefficient on time between sales when estimating one year period

returns. Likewise, the coefficient on time between sales when estimating 5 and 10 year

periods should be respectively five and ten times the coefficient on the time between

sales for one year periods. Looking at the art dataset, this is clearly not the case. Ten

times the coefficient on the coefficient on time between sales when estimating yearly

returns is statistically significantly less than the coefficient on the time between sales

when estimating ten year returns. In contrast, in the Herengracht dataset, ten times the

coefficient on the time between sales for the one year return does not appear to be

statistically significantly different than the coefficient on the time between sales for the

ten year returns.

4. Individual Asset Errors that Are Serially Correlated Across Periods

A possible reason that the Case-Shiller correction may not be working as it should

be is that the return errors are not i.i.d. We now drop the assumption that return errors

are i.i.d. For Goetzmann's [1992] study of the behavior of repeat sales regressions using

stock market data, the i.i.d. assumption seems appropriate. Considerable evidence finds

few deviations from market efficiency. Market liquidity, low trading costs, and the

ability to sell shares short all contribute to rapid transmission of new information into

asset prices. In contrast, for many of the asset classes studied in repeat sales regressions,

these features are not present.[4] Houses and individual works of art have idiosyncratic features, making simple observations of prices of other assets in the class only signals of the "true price" of an asset. Trading costs are also significant (5-6% commissions plus transactions taxes and other costs for houses in the U.S. and a 10%-20% buyer's commissions plus seller's commissions for art sold at auction), and short sales are essentially impossible. Price data are also only available with some lag for houses (the interval between contract date and closing date at a minimum).

Note that the statistical issue is whether the error term on the individual asset returns is correlated between periods. In repeat sales data, only the residuals on the aggregated time periods are observed. We shall see that this prevents us from uncovering much of the fine structure of the time series processes of the error returns.

In what follows, we shall drop the subscript i for the individual property since all calculations are with respect to a single property.

4.1 The MA Process

Suppose instead of i.i.d. errors that the errors follow a first-order moving average (MA(1)) process: $\varepsilon_t = \eta_t + \theta\eta_{t-1}$, $t = 1, \tau$, where $-1 < \theta < 1$. Under this assumption, the expectation of $\left(\sum_{1}^{\tau}\varepsilon_t\right)^2$ is:

$$E\left(\sum_{1}^{\tau}\varepsilon_t\right)^2 = E\left(\sum_{1}^{\tau}\varepsilon_t{}^2\right) + 2\theta E\left(\sum_{1}^{\tau-1}\varepsilon_t\varepsilon_{t+1}\right),$$

---

since all other $\varepsilon_t \varepsilon_{t+k}$ have zero correlation for $k > 1$,

$$E\left(\sum_1^\tau \varepsilon_t\right)^2 = \tau \left[1+\theta^2\right]\sigma_\eta^2 + 2(\tau-1)\theta\sigma_\eta^2 = \tau\left[1+\theta\right]^2\sigma_\eta^2 - 2\theta\sigma_\eta^2.$$

Thus, regressing the square of the residual in the repeat sales regression $\left(\sum_1^\tau e_t\right)^2$ on τ

and a constant should yield:

$$\hat{\alpha} = -2\theta\sigma_\eta^2 \text{ and } \hat{\beta} = \left[1+\theta\right]^2\sigma_\eta^2.$$

Importantly, this provides a different explanation for the constant term than that of Case and Shiller [1987]. Unlike their formulation, $\hat{\alpha} < 0$ is not an anomaly, but arises whenever $\theta > 0$ (unlike first-order autoregressive processes, there is no presumption that θ take a particular sign). Thus, the issue raised in Calhoun [1996] about a negative constant term may be evidence of non-i.i.d. error processes.

Under the assumption that the ε_t follow an MA(1) process, we can identify point estimates of θ and σ_η^2 from $\hat{\alpha}$ and $\hat{\beta}$.

We can extend this approach to higher-order MA processes. For an MA(2) process:

$$\varepsilon_t = \eta_t + \theta\eta_{t-1} + \gamma\eta_{t-2}.$$

Then $E\left(\sum_1^\tau \varepsilon_t\right)^2 = \tau\left[1+\theta^2+\gamma^2+2\theta+2\theta\gamma+2\gamma\right]\sigma_\eta^2 - 2\sigma_\eta^2\left[\theta+\theta\gamma+\gamma\right]$.

Similar calculations reveal that all MA(k) processes will have an intercept term and a constant multiplying τ, but no terms multiplying higher powers of τ. The slope term will be positive, but the intercept may be positive or negative, depending on the specific

parameters of the process. Note that, for any MA process other than a first-order one, we

cannot identify the parameters since we observe only a slope and intercept.

4.2 The AR Process

The other classical form of time dependence is an autoregressive process.

Suppose that ε_t, $t = 1, \tau$ follows an AR(1) process:

$$\varepsilon_t = \eta_t + \rho\varepsilon_{t-1}, \mathrm{t} = 1, \ \tau.$$

Economic processes usually exhibit positive autocorrelation, so we generally assume that

$0 \leq \rho < 1$, although in principle it is only required that $|\rho| < 1$ for stationary processes.

Using the fact that that $E\left[\varepsilon_t\varepsilon_{t-k}\right] = \rho^k \dfrac{\sigma_\eta^2}{1-\rho^2}$, we find that:

$$E\left(\sum_1^\tau \varepsilon_t\right)^2 = \tau\sigma_\varepsilon^2 + 2\sigma_\varepsilon^2\left[\tau\left(\frac{\rho-\rho^\tau}{1-\rho}\right) - \frac{\rho-\rho^\tau(\tau-1)}{1-\rho} - \rho\left[\frac{\rho-\rho^{\tau-1}}{1-\rho}^2\right]\right].$$

The calculations for the above result are shown in Appendix A.   As τ grows, for $\rho > 0$,

this expression increases at an increasing rate and asymptotes to an increasing straight

line. For $\rho < 0$, it increases at a decreasing rate.[5] Thus, only negative first-order serial

correlation is consistent with a positive coefficient on time between sales and a negative

coefficient on its square, which is a common finding in the literature. Given that negative

autocorrelation is not common in economic data, it seems unlikely that an AR(1) error

process explains the commonly observed pattern. Higher-order AR processes will also

results in $E\left[\sum_1^\tau \varepsilon_t\right]^2$ being a nonlinear function of the holding period.

---

[5] Positive autocorrelation in return errors does not result in a mean-reverting process, in
contrast to Hwang and Quigley's description of autocorrelation in price level errors.

One could fit a structural nonlinear regression using τ to $E\left(\sum_1^\tau \varepsilon_t\right)^2$. As with the

MA(1) process, one would only identify parameters of the stochastic process conditional

on the assumption about the order of the AR process.

4.3 The ARMA Process

Another possibility is a mixed autoregressive and moving average process. An

ARMA(p, q) process has p[th]-order autoregression and q[th]-order moving average. For p =

q = 1, an ARMA(1, 1) process can be written as:

$$\varepsilon_t = \phi\varepsilon_{t-1} + \eta_t + \theta\eta_{t-1}, \text{ t} = 1, \tau.$$

The expected values of the variances and covariances of errors equal:

$$E\left[\varepsilon_t^2\right] = \sigma_\eta^2\left[\frac{1+\theta^2+2\phi\theta}{1-\phi^2}\right], \; E\left[\varepsilon_t\varepsilon_{t-1}\right] = \phi\sigma_\varepsilon^2 + \theta\sigma_\eta^2 = \frac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2}\sigma_\eta^2,$$

$$\text{and } E\left[\varepsilon_t\varepsilon_{t-k}\right] = \phi^k\hat{\beta} \text{ where } \hat{\beta} = \frac{\sigma_\eta^2}{\phi}\frac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2} \text{ for k} \geq 2.$$

Using this fact, we find that

$$E\left(\sum_{t=1}^\tau \varepsilon_t\right)^2 = \tau\left(\frac{1+\theta^2+2\phi\theta}{1-\phi^2}\right)\sigma_\eta^2 + 2\left(\tau-1\right)\left(\frac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2}\right)\sigma_\eta^2$$

$$+2\sigma_\eta^2\frac{\phi^\tau - (\tau-1)\phi^2 + \tau - 2\,\phi}{1-\phi^{\,2}}\left(\frac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2}\right)$$

$$= \tau\sigma_\eta^2\left[\left(\frac{1+\theta^2+2\phi\theta}{1-\phi^2}\right)+2\left(\frac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2}\right)\right] -2\sigma_\eta^2\left(\frac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2}\right)$$

$$+2\sigma_\eta^2\frac{\phi^\tau - (\tau-1)\phi^2 + \tau - 2\,\phi}{1-\phi^{\,2}}\left(\frac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2}\right).$$

See the Appendix for details of the calculations. As with the MA process, there is an intercept (which is easily negative), and a constant coefficient on the time horizon. As with the AR process, there is also a term which decays exponentially, so one could estimate a polynomial in τ and then find the parameters to match up.

For $\phi > 0$ (the "normal" case), the expectation of the square of the sum of the residuals increases with the length of the holding period. For $\phi > |\theta|$, it increases as an increasing rate, while for $0 < \phi < |\theta|$, it increases as at decreasing rate. This last possibility is consistent with finding a positive coefficient on the linear term and a negative coefficient on the quadratic term in regressing the squared residual on the holding period length. It also seems to be the "minimal" assumption on the return error process to generate such concavity. As with AR processes, higher-order ARMA processes will continue to result in $E\left(\sum_{t=1}^{\tau} \varepsilon_t \right)^2$ being a nonlinear function of the holding period.

The conclusion of the above exercise is that different data generating processes result in different specifications for the second stage Case-Shiller regressions. However, the aggregated data that is commonly available does not provide enough information to specify the underlying data generating process. Hence, we need a more general specification for the second stage regressions.

5. A More Efficient Methodology

A better outcome would likely result if we did not have to rely on the i.i.d assumption for the second stage regression. Hence, we propose a non-parametric second stage, where the squared residuals from the first stage are regressed on dummy variables

15

for each holding duration. Hence, in the second stage, we propose regressing $\hat{u}_i^2$ on a matrix which has a row of dummy variables for each asset in the sample. The dummy variable Z_{ij} takes on the value 1 if $s_j - b_j$ equals j-i and zero otherwise. As before, in the third stage we divide the regressors and the regressand by the square root of the predicted value from the second stage. Table 5 presents the regression results from the new second and third stage regressions along with the Koenker-Basset test for heteroskedasticity in the third stage. The coefficients and error terms are presented in the last three columns of Appendix Tables 1 and 2.

As is evident from both Table 5 and the Appendix Tables, our non-parametric approach is a significant improvement, both over OLS and over standard Case-Shiller. On average, the non-parametric approach has decreased the errors on the coefficients by about 12% for the Campbell dataset, and by about 7% for Herengracht dataset.   As shown in Table 5, the fit has also improved, going from an OLS R-squared of .6886 to a non-parametric R-squared of .6956 for the Campbell dataset, and going from an OLS R-squared of .6037 in the Herengracht dataset to a non-parametric R-squared of .6589.

For the Campbell art dataset, the mean nominal return calculated from the coefficients in Appendix Table 1 over the period 1780-2007 was 3.13%. The mean nominal return calculated from the coefficients in Appendix Table 2 for the Herengracht data was just slightly over 1% over the period 1630 to 1972.   If we compare the two Dutch assets for the overlapping time period of 1780 to 1970, we find a 3.41% nominal return for art, and a 1.19% return for the Herengracht real estate, demonstrating that the returns to art sold in Amsterdam heartily outperformed the Herengracht canal real estate market during that time period!

6. Further Applications and Extensions

6.1 Real Estate Indexes

Calhoun's [1996] discussion of the second stage regression was part of a technical description of the Office of Federal Enterprise Oversight's (OFHEO) methodology in constructing price indexes, a methodology they are still currently using. The OFHEO restricts the constant term to be zero, but includes a quadratic term in the second stage. Table 6 presents their published results for 3rd quarter 2007. Although the significance levels are not published by the OFHEO, the OFHEO stated that for both terms, significance resulted in most regions and states. The pattern of a positive coefficient on the linear term and a negative coefficient on the quadratic term is present throughout – but note that pattern resulted in our data also, when the constant was restricted to be zero.

Our analysis suggests that not only should the constant be included in the regressions, but that a non-parametric second stage regression would improve their analysis. If the constant term is negative, it is likely to be indicative of an ARMA process, given the quadratic term.

The above analysis also suggests that a non-parametric second stage could sensibly be used in the construction of housing price indexes to allow for the possibility that the errors follow an AR process or ARMA process. The non-parametric second stage could lead to more precise estimates of the index.[6]

6.2 Art Indexes

Repeat sales regressions are used in numerous articles that estimate price indexes from art auctions including Pesando [1993, 2006], Goetzmann [1993] and Mei

---

[6] Andrew Leventis [2007] has a nice discussion of the differences between the OFHEO and S&P/Case-Shiller House Price Indexes.

and Moses [2002]. Ashenfelter and Graddy [2003, 2006] provide a survey. Most of the articles on art auctions do not provide the results from the second stage regressions and Pesando [1993] does not use the Case-Shiller method but uses OLS. As art is an infrequently traded asset with many of the same properties as musical instruments and housing, it is very likely that many of these studies could benefit from the non-parametric specification in the second stage of the repeat sales regressions.

6.3. Implications of Serial Dependence

It is well-known that the logarithmic specification of the dependent variable results in a geometric mean across assets for each time period of the index. Goetzmann [1992] suggested a way to calculate the arithmetic mean, using the second stage of the Case-Shiller regressions. He suggested that the coefficient on the time between sales should be used as an estimate of the cross section variance to give the following formula for the arithmetic mean:

$$\mu^a \cong \exp\left(\mu^g + \frac{\sigma^2}{2}\right) - 1$$

where μ^a and μ^g are the arithmetic and geometric means and σ^2 is the cross-section variance. Goetzmann [1993] and Mei and Moses [2002] utilized this for art price indexes. This correction becomes problematic in the case of non-i.i.d. errors.

Without a specific assumption on the errors in the returns of the individual assets, the single period return variance in an asset cannot be identified from the second stage of the Case-Shiller regression results.[7] Calhoun [1996] proposes using:

---

[7] The S&P/Case-Shiller® Home price index directly estimate an arithmetic index in order to circumvent this problem.

$$\sigma_t^2 = At + Bt^2$$

(where A and B are the linear and quadratic coefficients from the second stage—with no constant) as the variance in the geometric to arithmetic correction formula (in index form). There is a problem with this simple adaptation of the Goetzmann approach once the second stage includes more than a simple linear term: the estimated variance for any property becomes a function of the holding period. Even using the variance per period ($A + Bt$) depends on the holding period. Any estimate of the arithmetic return must really depend on the planned holding period if the return errors are not i.i.d.

The geometric index has a simpler specification, and is intuitively appealing to many in the finance literature because of the continuously compounding interpretation of the log returns. Depending on the purpose of the regression, one can simply recognize that this is a geometric mean across assets, or a lower bound on the arithmetic mean. While this is an unintuitive interpretation, for many standard assets such as stocks, the geometric and the arithmetic mean can be directly computed. Thus, one could compare a geometric index of a standard asset directly with the geometric index of the alternative asset.

Some of the basic ideas in our paper have been utilized in other contexts. For example, studying the effects of holding periods on return variances is one standard technique to identify mean reversion in asset returns.[8] Furthermore, variance-ratio tests are commonly used to test for non-i.i.d. returns. We are, in effect, using a form of the variance-ratio test to test for non-i.i.d. deviations from the mean one-period returns by the

---

[8] See, for example, Poterba and Summers [1988] and Lo and MacKinlay [1988]. That approach generally uses a nonparametric framework with respect to the error process.

underlying assets. As suggested in other contexts, nonlinearities with respect to the holding period in the second-stage regressions may diagnose longer-horizon violations of independence as well.

Variance ratio tests calculate the variance of returns over different holding periods. If returns are i.i.d., then the variance of annual returns should be 12 times the variance of monthly returns. Likewise, if the deviations from mean returns for the underlying assets are i.i.d., than the variance in the deviations in annual returns should be 12 times the monthly variance. This would precisely match the Case-Shiller modeling of the second-stage regression on repeat sales without the "house-specific error". The fact that a polynomial specification is indicated in the Campbell art data reveal that deviations from mean returns for the individual assets violate serial independence. Abraham and Schauman [1991] report that the variance in the errors in returns for a housing data set is maximized at a holding period of twenty to thirty years, so they find concavity in the relationship as well. An essential difference between mean reversion and the processes modeled in Section 4 is that mean reversion can operate over a much longer time horizon than conventional AR and MA processes—it is effectively a more general version of serial dependence.

7. Conclusions and Issues for Further Study

Nonlinearities with respect to the holding period in estimated variances of property returns may be indicative of a failure of the assumption that errors in property returns are statistically independent over time. For assets where the usual conditions which induce market efficiency are not present, this should come as little surprise. More

20

efficient estimates can result if researchers take account of these nonlinearities in determining GLS weights for the final stage. Researchers should also take account of the failure of independence in using the calculated index values to interpret market returns.

In further work, we plan to extend our exploration of the implications of serial dependence in the deviations from mean returns on the standard errors of the return coefficients. Through Monte Carlo simulation, we can estimate true standard errors under a known non-i.i.d. process and compare it to the estimates of returns using the conventional procedures and our non-parametric method. Because the estimated coefficients for the last few periods depend on only a small number of transactions, biases in computing standard errors may matter most here.

References

Abraham, J., and W. Schauman, 1991, New Evidence on Home Prices from Freddie Mac
    Repeat Sales, *AREUEA Journal* 19: 333-352.

Bailey, M., R. Muth, and H. Nourse, 1963, A Regression Method for Real Estate Price
    Index Construction, *Journal of the American Statistical Association* 58: 933-942.

Calhoun, C., 1996, OFHEO House Price Indexes: HPI Technical Description, mimeo,
    Office of Federal Housing Enterprise Oversight.

Campbell, R., and R. Kraeussl, 2009, A Long Run Amsterdam Art Index, Working paper,
    Maastricht University.

Case, K., and R. Shiller, 1987, Prices of Single Family Homes Since 1970: New Indexes
    for Four Cities, Cowles Foundation Discussion Paper No. 851, Yale University.

Eichholz, P.M.A., (1997), A Long Run House Price Index: The Herengracht Index, 1628-
    1973. *Real Estate Economics*, 25: 175-192.

Goetzmann, W., 1992, The Accuracy of Real Estate Indices: Repeat Sale Estimators,
    Journal of Real Estate Finance and Economics 5:5-53.

Goetzmann, W., 1993, Accounting for Taste: Art and Financial Markets over Three
    Centuries, *American Economic Review*, 83: 1370-1376.

Goetzmann, W., and L. Peng, 2002, The Bias of the RSR Estimator and the Accuracy of
    Some Alternatives, *Real Estate Economics* 30: 13-39.

Graddy, K., and P. Margolis, 2007, Fiddling with Value: Violins as an Investment?,
    CEPR Discussion Paper 6583.

Hwang, M., and J. Quigley, 2004, Selectivity, Quality Adjustment and Mean Reversion
    in the Measurement of House Values, *Journal of Real Estate Finance and
    Economics* 28: 161-178.

Leventis, A., 2007, A Note on the Differences between the OFHEO and S&P/Case-
    Shiller House Price Indexes, OFHEO working paper.

Lo, A., and A. C. MacKinlay, 1988, Stock Market Prices Do Not Follow Random
    Walks: Evidence from a Simple Specification Test, *Review of Financial Studies*
    1: 41-66.

Mei, J and M. Moses, 2002, Art as an Investment and the Underperformance of
    Masterpieces", American Economic Review 92: 1656-1668.

Poterba, J., and L. Summers, 1988, Mean Reversion in Stock Prices: Evidence and Implications, *Journal of Financial Economics* 22: 27-59.

Quigley, J., 1995, A Simple Hybrid Model for Estimating Real Estate Price Indexes, *Journal of Housing Economics* 4: 1-12.

Roll, R., 1983, On Computing Mean Returns and the Small Firm Premium, *Journal of Financial Economics* 12: 371-386.

Shiller, R., 2007, Historic Turning Points in Real Estate, Cowles Foundation Discussion Paper No. 1610, Yale University (http://ssrn.com/abstract=991107).

Appendix A: Calculations for the AR Process

Using the fact that that $E\left[\varepsilon_t \varepsilon_{t-k}\right] = \rho^k \dfrac{\sigma_\eta^2}{1-\rho^2}$., we find that

$$\left(\sum_1^\tau \varepsilon_t\right)^2 = \sum_1^\tau \varepsilon_t^2 + 2\sum_1^{\tau-1}\varepsilon_{t+1}\varepsilon_t + 2\sum_1^{\tau-2}\varepsilon_{t+2}\varepsilon_t + \ldots + 2\sum_1^{\tau-(\tau-1)}\varepsilon_{t+\tau-1}\varepsilon_t .$$

Thus, $E\left(\sum_1^\tau \varepsilon_t\right)^2 = \tau\sigma_\varepsilon^2 + 2\rho(\tau-1)\sigma_\varepsilon^2 + 2\rho^2(\tau-2)\sigma_\varepsilon^2 + \ldots + 2\rho^{\tau-1}\sigma_\varepsilon^2$

$$= \tau\sigma_\varepsilon^2 + 2\sigma_\varepsilon^2\left[\rho(\tau-1) + 2\rho^2(\tau-2) + \ldots + 2\rho^{\tau-1}\right]$$

where $\left[\rho(\tau-1) + 2\rho^2(\tau-2) + \ldots + \rho^{\tau-1}\right] = \sum_1^{\tau-1}\rho^k(\tau-k) = \tau\sum_1^{\tau-1}\rho^k - \sum_1^{\tau-1}\rho^k k$.

The first term in this expression equals $\tau\left(\dfrac{\rho-\rho^\tau}{1-\rho}\right)$, using

$$Z = \rho + \rho^2 + \ldots + \rho^{\tau-1} \quad \text{and} \quad \rho Z = \rho^2 + \ldots + \rho^\tau ,$$

while the second term equals $-\left[\dfrac{\rho - \rho^\tau(\tau-1) + \rho\left[\dfrac{\rho-\rho^{\tau-1}}{1-\rho}\right]}{1-\rho}\right]$, using

$$Y = \rho + 2\rho^2 + (\tau-1)\rho^{\tau-1} \quad \text{and} \quad \rho Y = \rho^2 + 2\rho^3 + (\tau-1)\rho^\tau + \tau\rho^\tau , \text{ and}$$

$$Y - \rho Y = \rho - \rho^\tau(\tau-1) + \rho^2 + \rho^3 + \ldots + \rho^{\tau-1}$$

$$= \rho - \rho^N(N-1) + \rho\left(\dfrac{\rho-\rho^{N-1}}{1-\rho}\right).$$

Hence, $E\left(\sum_1^\tau \varepsilon_t\right)^2 = \tau\sigma_\varepsilon^2 + 2\sigma_\varepsilon^2\left[\tau\left(\dfrac{\rho-\rho^\tau}{1-\rho}\right) - \dfrac{\rho-\rho^\tau(\tau-1)}{1-\rho} - \rho\left[\dfrac{\rho-\rho^{\tau-1}}{1-\rho^2}\right]\right].$

Calculations for the ARMA process

For p = q = 1, an ARMA(1, 1) process can be written as:

$$\varepsilon_t = \phi\varepsilon_{t-1} + \eta_t + \theta\eta_{t-1}, \text{ t} = 1, \tau.$$

The expected values of the variances and covariances of errors equal:

$$E\ \varepsilon_t^2 = \sigma_\eta^2\left[\frac{1+\theta^2+2\phi\theta}{1-\phi^2}\right], \ E\ \varepsilon_t\varepsilon_{t-1} = \phi\sigma_\varepsilon^2 + \theta\sigma_\eta^2 = \frac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2}\sigma_\eta^2,$$

$$\text{and } E\ \varepsilon_t\varepsilon_{t-k} = \phi^k\hat\beta \text{ where } \hat\beta = \frac{\sigma_\eta^2}{\phi}\frac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2} \text{ for k} \geq 2.$$

Then, we can write the expected value of the square of the sum of the residuals as:

$$E\left(\sum_{t=1}^{\tau}\varepsilon_t\right)^2 = E\left(\sum_{t=1}^{\tau}\varepsilon_t^2\right) + 2E\left(\sum_{t=1}^{\tau-1}\varepsilon_t\varepsilon_{t+1}\right) + 2E\left(\sum_{t=1}^{\tau-2}\varepsilon_t\varepsilon_{t+2}\right) + ... + 2E\left(\sum_{t=1}^{\tau-\tau-1}\varepsilon_t\varepsilon_{t+\tau-1}\right)$$

$$= E\left(\sum_{t=1}^{\tau}\varepsilon_t^2\right) + 2E\left(\sum_{t=1}^{\tau-1}\varepsilon_t\varepsilon_{t+1}\right) + 2\left\{\sum_{J=1}^{\tau-2}JE\ \varepsilon_t\varepsilon_{t+\tau-J}\right\}.$$

Taking the expectations, we obtain:

$$E\left(\sum_{t=1}^{\tau}\varepsilon_t\right)^2 = \tau\sigma_\varepsilon^2 + 2\ \tau - 1\ \frac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2}\sigma_\eta^2 + 2\hat\beta\phi^\tau\sum_{J=1}^{\tau-2}J\phi^{-J}.$$

For the last term, let $K = \tau - 2$ and $\omega = \frac{1}{\phi}$. Then let $Z = \sum_{J=1}^{K}J\omega^J$.

Now $Z = \omega + 2\omega^2 + 3\omega^3 + ... + K\omega^K$ and $\omega Z = \omega^2 + 2\omega^3 + ... + K\omega^{K+1}$, so

$$Z - \omega Z = \omega + \omega^2 + \omega^3 + ... + \omega^K - K\omega^{K+1}.$$

Let $Y = \omega + \omega^2 + \omega^3 + ... + \omega^K$. Then $\omega Y = \omega^2 + \omega^3 + ... + \omega^{K+1}$, so $Y - \omega Y = \omega - \omega^{K+1}$, and

$Y = \dfrac{\omega - \omega^{K+1}}{1-\omega}$. Substituting this into the earlier formula,

25

$$Z - \omega Z = \frac{\omega - \omega^{K+1}}{1-\omega} - K\omega^{K+1} = \frac{\omega - (K+1)\omega^{K+1} + K\omega^{K+2}}{1-\omega}.$$

Hence, $Z = \dfrac{\omega - (K+1)\omega^{K+1} + K\omega^{K+2}}{(1-\omega)^2}$

Thus, we have $E\left(\displaystyle\sum_{t=1}^{\tau}\varepsilon_t\right)^2 = \tau\sigma_\varepsilon^2 + 2(\tau-1)\left(\dfrac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2}\right)\sigma_\eta^2 + 2\hat{\beta}\phi^\tau\sigma_\eta^2.$

Substituting, $E\left(\displaystyle\sum_{t=1}^{\tau}\varepsilon_t\right)^2 = \tau\left(\dfrac{1+\theta^2+2\phi\theta}{1-\phi^2}\right)\sigma_\eta^2 + 2(\tau-1)\left(\dfrac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2}\right)\sigma_\eta^2$

$$+2\sigma_\eta^2\,\frac{\phi^\tau-(\tau-1)\phi^2+(\tau-2)\phi}{(1-\phi)^2}\left(\frac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2}\right)$$

$$= \tau\sigma_\eta^2\left[\left(\frac{1+\theta^2+2\phi\theta}{1-\phi^2}\right)+2\left(\frac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2}\right)\right] -2\sigma_\eta^2\left(\frac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2}\right)$$

$$+2\sigma_\eta^2\,\frac{\phi^\tau-(\tau-1)\phi^2+(\tau-2)\phi}{(1-\phi)^2}\left(\frac{\theta+\phi+\theta^2\phi+\theta\phi^2}{1-\phi^2}\right).$$

Table 1

OLS and Case and Shiller Results

	Campbell Dataset	Herengracht Dataset
Number of Transaction Pairs	1468	3577
Stage I, R^2	0.68	0.60
Stage I, Koenker Basset α_2	0.04	0.06
	(1.80)	(4.73)
Stage II, R^2	0.01	0.00
Stage II, Constant	2.21	0.16
	(7.34)	(15.31)
Stage II, Coefficient	0.20	0.00
	(3.36)	(3.71)
Stage III, R^2	0.64	0.60
Stage III, Koenker Basset α_2	-0.02	0.06
	-(0.59)	(4.62)

T-statistics are noted in parentheses.

Table 2
Second Stage Regression Results: Campbell Art Dataset

	1		2		3		4		5		6		7	
	coef	t-stat	coef	t-stat	coef	t-stat	coef	t-stat	coeff	t-stat	coef	t-stat	coef	t-stat
TBS	0.20	(3.36)	0.49	(3.20)	0.96	(2.59)					0.52	(12.72)	0.99	(11.99)
TBS^2			-0.02	(0.01)	-0.09	(0.05)							-0.04	-(6.50)
TBS^3					0.00	(1.40)								
ln(TBS)							1.05	(4.08)						
Duration Dummies									22					
Cons	2.21	(7.34)	1.63	(3.93)	1.01	(1.67)	1.92	(5.94)	1.87	(0.39)				
F-Stat	11.26		7.74		5.81		16.65		1.81		161.74		104.28	
Prob>F	0.000		0.001		0.001		0.000		0.014		0.000		0.000	
Adj R^2	0.007		0.009		0.010		0.011		0.011		*		*	
AIC	10197		10195		10195		10192		10210		10248		10208	
BIC	10208		10211		10216		10202		10327		10243		10219	
Obs	1468		1468		1468		1468		1468		1468		1468	

Table 3
Second Stage Regression Results: Herengracht Dataset

	1		2		3		4		5		6		7	
	coef	t-stat	coef	t-stat	coef	t-stat	coef	t-stat	coeff	t-stat	coef	t-stat	coef	t-stat
TBS	0.002	(3.710)	0.002	(2.410)	0.003	(1.520)					0.006	(19.710)	0.011	(18.330)
TBS^2			0.000	-(0.980)	0.000	-(0.530)							0.000	-(10.070)
TBS^3					0.000	(0.290)								
ln(TBS)							0.024	(3.620)						
Duration Dummies									105.000					
Cons	0.157	(15.310)	0.149	(11.410)	0.146	(0.016)	0.125	(0.018)	0.274	(0.630)				
F-Stat	13.75		7.36		4.93		13.09		1.33		388.42		250.34	
Prob>F	0.000		0.001		0.002		0.000		0.016		0.000		0.000	
Adj R^2	0.004		0.004		0.003		0.003		0.010		*		*	
AIC	4197		4198		4200		4198		4279		4422		4325	
BIC	4210		4217		4225		4210		4934		4429		4337	
Obs	3577		3577		3577		3577		3577		3577		3577	

Table 4

Second-Stage results with Varying Time Periods

Cambell Art Dataset

	10-year	5-year	2-year	1-year
TBS	0.209	0.074	0.014	0.000
	(3.435)	(2.700)	(1.356)	(0.084)
Cons	2.182	2.236	2.265	2.290
	(7.201)	(8.208)	(9.099)	(9.865)
Obs	1463	1463	1463	1453
R^2	0.008	0.005	0.001	0.000

Herengracht Dataset

	10-year	5-year	2-year	1-year
TBS	0.008	0.004	0.002	0.001
	(3.508)	(3.870)	(3.694)	(3.409)
Cons	0.183	0.163	0.152	0.144
	(14.434)	(13.389)	(13.354)	(14.041)
Obs	3199	3199	3199	3199
R^2	0.209	0.074	0.014	0.000

T-statistics are noted in parentheses.

30

Table 5
OLS and Non-parametric Results

	Campbell dataset	Herengracht dataset
Number of Transaction Pairs	1468	3577
Stage I, R^2	0.69	0.60
Stage I, Koenker Basset α_2	0.04	0.06
	(1.80)	(4.73)
Stage II, R^2	0.03	0.04
Stage II: $F(21,1446)$; $F(105,3471)$	1.81	1.33
Stage III, R^2	0.70	0.66
Stage III, Koenker Basset α_2	-0.01	0.00
	-(0.61)	(0.57)

The F-statistic tests for joint significance of the
dummy variables. T-statistics are noted in brackets.

Table 6
OFHEO Second Stage Regression Statistics for their QUARTERLY Housing Index : 3rd Quarter 2007
(Constant is Restricted to be Zero)

	Linear term	Quadratic Term		Linear term	Quadratic Term
East North Central	0.001680477	-0.00000299	Maine	0.002206272	-0.00001042
East South Central	0.001413392	-0.0000018	Michigan	0.001802758	-0.00000883
Middle Atlantic	0.002074802	-0.00000055	Minnesota	0.001778828	-0.00000748
Mountain	0.002409786	-0.00001399	Missouri	0.001510312	-0.00000427
New England	0.002154035	-0.00001018	Mississippi	0.001711552	-0.00000806
Pacific	0.002429178	-0.0000141	Montana	0.001928588	-0.00000922
South Atlantic	0.002187866	-0.00000785	North Carolina	0.001565033	-0.00000252
West North Central	0.00177686	-0.00000562	North Dakota	0.001039276	-0.00000135
West South Central	0.001778648	-0.00000617	Nebraska	0.001253644	-0.00000316
Alaska	0.001670519	-0.00001364	New Hampshire	0.002027072	-0.0000166
Alabama	0.001518371	-0.00000204	New Jersey	0.002004723	-0.00001082
Arkansas	0.001386812	-0.00000118	New Mexico	0.00149599	-0.00000543
Arizona	0.001633332	-0.00000734	Nevada	0.001229675	-0.00000624
California	0.001746513	-0.0000077	New York	0.002274004	0.00000173
Colorado	0.001809468	-0.00000946	Ohio	0.001425313	-0.00000283
Connecticut	0.001763343	-0.00000839	Oklahoma	0.001735064	-0.00001022
District of Columbia	0.002680667	-0.00001472	Oregon	0.0018648	-0.00000801
Delaware	0.001384419	-0.00000705	Pennsylvania	0.001574089	0.00000112
Florida	0.001908697	-0.00000436	Rhode Island	0.001733826	-0.00001077
Georgia	0.00148583	0.00000037	South Carolina	0.001729218	-0.000002
Hawaii	0.002267208	-0.00001294	South Dakota	0.001414512	-0.00000234
Iowa	0.001448536	-0.00000564	Tennessee	0.001334526	-0.00000102
Idaho	0.001885403	-0.00001133	Texas	0.001765041	-0.00000488
Illinois	0.001269782	0.00000665	Utah	0.001433671	-0.00000553
Indiana	0.001651597	-0.00000569	Virginia	0.001594024	-0.00000623
Kansas	0.001308308	-0.00000343	Vermont	0.001695607	-0.0000107
Kentucky	0.001301519	-0.00000324	Washington	0.001739829	-0.00000457
Louisiana	0.00161069	-0.00000721	Wisconsin	0.001578844	-0.0000063
Massachusetts	0.001960005	-0.00001126	West Virginia	0.00218767	-0.00001027
Maryland	0.001480009	-0.00000671	Wyoming	0.001830342	-0.00000946

Appendix Table I
Estimated Coefficients and Errors for Campbell dataset

period	coef	OLS std error	t-stat	coef	Case and Shiller std error	t-stat	coef	Non-parametric Case and Shiller std error	t-stat
1780	0.752	0.882	0.85	0.375	0.949	0.40	0.927	0.662	1.4
1790	-0.069	0.676	-0.10	-0.255	0.735	-0.35	-0.140	0.525	-0.27
1800	0.381	0.703	0.54	0.024	0.771	0.03	-0.230	0.583	-0.39
1810	-0.655	0.792	-0.83	-0.223	0.880	-0.25	0.142	0.682	0.21
1820	0.724	0.922	0.79	1.135	1.061	1.07	1.010	0.766	1.32
1830	-0.051	0.791	-0.06	-0.296	0.835	-0.35	-0.443	0.639	-0.69
1840	-0.469	0.991	-0.47	-0.748	1.018	-0.74	-0.562	0.964	-0.58
1850	0.215	1.142	0.19	0.353	1.249	0.28	0.272	1.118	0.24
1860	1.286	1.057	1.22	1.231	1.276	0.96	0.470	0.829	0.57
1870	-1.201	0.888	-1.35	-1.122	1.084	-1.04	-0.201	0.678	-0.3
1880	1.329	0.587	2.26	1.297	0.668	1.94	1.055	0.494	2.13
1890	-1.111	0.485	-2.29	-1.008	0.523	-1.93	-0.825	0.480	-1.72
1900	1.218	0.421	2.89	1.090	0.461	2.37	1.014	0.463	2.19
1910	0.798	0.350	2.28	0.800	0.380	2.10	0.808	0.380	2.13
1920	0.005	0.243	0.02	0.016	0.258	0.06	0.004	0.256	0.02
1930	-0.679	0.242	-2.80	-0.662	0.253	-2.62	-0.647	0.253	-2.56
1940	1.461	0.259	5.64	1.455	0.265	5.50	1.514	0.270	5.61
1950	-0.460	0.245	-1.87	-0.452	0.244	-1.85	-0.475	0.253	-1.88
1960	1.431	0.215	6.66	1.457	0.209	6.96	1.453	0.219	6.64
1970	1.271	0.162	7.86	1.255	0.154	8.12	1.234	0.155	7.96
1980	0.462	0.099	4.66	0.446	0.092	4.82	0.424	0.086	4.91
1990	0.617	0.088	7.04	0.624	0.082	7.57	0.633	0.076	8.31
2000*	-0.456	0.103	-4.45	-0.445	0.099	-4.49	-0.438	0.095	-4.59
Average std error		0.537			0.589			0.475	
R^2		0.684			0.641			0.691	
Obs		1468			1468			1468	

Estimated Coefficients and Errors for Herengracht Dataset

| period | OLS | | | Case and Shiller | | | Non-parametric Case and Shiller | | |
	coef	std error	t-stat	coef	std error	t-stat	coef	std error	t-stat
1630	0.390	0.376	1.04	0.504	0.418	1.21	0.354	0.435	0.82
1632	0.457	0.328	1.39	0.509	0.344	1.48	0.683	0.277	2.47
1634	-0.573	0.366	-1.57	-0.620	0.389	-1.59	-0.777	0.316	-2.46
1636	-0.002	0.317	-0.01	0.045	0.340	0.13	0.167	0.310	0.54
1638	0.152	0.220	0.69	0.081	0.236	0.34	-0.049	0.218	-0.23
1640	0.309	0.190	1.62	0.359	0.201	1.78	0.351	0.158	2.22
1642	0.269	0.188	1.43	0.294	0.194	1.51	0.367	0.146	2.50
1644	-0.084	0.169	-0.50	-0.135	0.173	-0.78	-0.167	0.161	-1.04
1646	0.100	0.159	0.63	0.125	0.159	0.79	0.133	0.155	0.86
1648	-0.092	0.195	-0.47	-0.089	0.197	-0.45	-0.038	0.193	-0.20
1650	0.099	0.180	0.55	0.109	0.181	0.60	0.049	0.178	0.28
1652	-0.165	0.209	-0.79	-0.146	0.210	-0.70	-0.138	0.212	-0.65
1654	-0.030	0.224	-0.13	-0.036	0.226	-0.16	-0.038	0.225	-0.17
1656	0.268	0.185	1.45	0.264	0.191	1.38	0.211	0.185	1.14
1658	0.151	0.203	0.74	0.111	0.204	0.54	0.112	0.202	0.55
1660	0.102	0.176	0.58	0.104	0.176	0.59	0.165	0.165	1.00
1662	-0.168	0.154	-1.09	-0.154	0.166	-0.93	-0.209	0.133	-1.57
1664	0.007	0.337	0.02	0.058	0.380	0.15	-0.099	0.209	-0.47
1666	-0.360	0.372	-0.97	-0.345	0.413	-0.83	-0.268	0.260	-1.03
1668	0.110	0.242	0.46	0.060	0.252	0.24	0.186	0.233	0.80
1670	0.126	0.182	0.69	0.139	0.184	0.75	0.133	0.177	0.75
1672	-0.213	0.182	-1.17	-0.229	0.186	-1.24	-0.226	0.168	-1.35
1674	-0.149	0.176	-0.84	-0.158	0.182	-0.87	-0.124	0.152	-0.82
1676	-0.472	0.229	-2.06	-0.415	0.246	-1.68	-0.577	0.201	-2.87
1678	0.323	0.236	1.37	0.302	0.250	1.21	0.428	0.219	1.96
1680	0.347	0.159	2.18	0.316	0.161	1.96	0.286	0.155	1.85
1682	-0.422	0.137	-3.08	-0.381	0.141	-2.71	-0.371	0.133	-2.78
1684	0.053	0.128	0.41	0.048	0.131	0.37	0.124	0.122	1.01
1686	0.257	0.137	1.87	0.234	0.139	1.69	0.219	0.116	1.90
1688	-0.065	0.174	-0.37	-0.071	0.180	-0.39	-0.136	0.151	-0.90
1690	0.138	0.174	0.80	0.118	0.180	0.65	0.168	0.161	1.05
1692	-0.195	0.143	-1.36	-0.188	0.143	-1.31	-0.236	0.137	-1.72
1694	0.164	0.157	1.04	0.173	0.156	1.11	0.242	0.149	1.63
1696	-0.024	0.177	-0.14	-0.014	0.181	-0.08	-0.148	0.164	-0.90
1698	-0.027	0.164	-0.17	-0.027	0.171	-0.16	0.068	0.150	0.45
1700	0.111	0.130	0.85	0.121	0.134	0.90	0.098	0.122	0.80
1702	-0.064	0.121	-0.53	-0.060	0.124	-0.49	-0.013	0.117	-0.11
1704	0.022	0.139	0.16	0.011	0.143	0.08	-0.070	0.126	-0.55
1706	-0.107	0.154	-0.70	-0.104	0.159	-0.66	-0.087	0.149	-0.58

1708	0.034	0.139	0.24	0.030	0.143	0.21	0.026	0.143	0.18
1710	0.003	0.118	0.03	0.006	0.119	0.05	0.033	0.116	0.28
1712	0.110	0.114	0.97	0.103	0.114	0.90	0.130	0.102	1.28
1714	-0.078	0.110	-0.71	-0.083	0.113	-0.73	-0.133	0.080	-1.67
1716	0.097	0.099	0.98	0.107	0.100	1.07	0.091	0.077	1.17
1718	-0.098	0.107	-0.92	-0.097	0.108	-0.90	-0.105	0.082	-1.28
1720	0.242	0.117	2.07	0.225	0.119	1.90	0.293	0.095	3.08
1722	0.119	0.109	1.10	0.134	0.110	1.21	0.093	0.105	0.88
1724	0.188	0.105	1.79	0.200	0.108	1.85	0.249	0.099	2.52
1726	-0.054	0.109	-0.50	-0.058	0.112	-0.52	-0.076	0.103	-0.74
1728	-0.079	0.101	-0.78	-0.079	0.101	-0.77	-0.100	0.095	-1.05
1730	0.133	0.106	1.26	0.126	0.107	1.18	0.146	0.101	1.44
1732	0.045	0.114	0.40	0.043	0.113	0.38	0.069	0.111	0.62
1734	-0.013	0.111	-0.12	-0.012	0.110	-0.11	-0.081	0.106	-0.76
1736	0.095	0.126	0.75	0.082	0.126	0.65	0.124	0.116	1.06
1738	-0.180	0.137	-1.31	-0.157	0.140	-1.12	-0.163	0.133	-1.23
1740	0.135	0.116	1.16	0.128	0.120	1.06	0.120	0.116	1.03
1742	-0.536	0.148	-3.63	-0.556	0.147	-3.77	-0.543	0.143	-3.79
1744	0.193	0.148	1.30	0.213	0.148	1.44	0.213	0.143	1.49
1746	0.035	0.128	0.28	0.059	0.130	0.46	-0.004	0.123	-0.03
1748	0.016	0.138	0.12	-0.037	0.139	-0.27	0.001	0.130	0.01
1750	-0.132	0.111	-1.19	-0.108	0.111	-0.97	-0.113	0.104	-1.09
1752	0.303	0.119	2.54	0.282	0.122	2.32	0.288	0.116	2.48
1754	-0.080	0.120	-0.66	-0.042	0.123	-0.34	-0.039	0.119	-0.32
1756	-0.115	0.110	-1.05	-0.132	0.110	-1.20	-0.129	0.109	-1.18
1758	-0.042	0.105	-0.40	-0.038	0.105	-0.36	-0.026	0.102	-0.26
1760	0.001	0.104	0.01	-0.001	0.104	-0.01	-0.001	0.095	-0.01
1762	0.044	0.103	0.42	0.047	0.103	0.46	0.028	0.096	0.29
1764	0.132	0.090	1.46	0.131	0.091	1.44	0.067	0.072	0.92
1766	0.046	0.100	0.46	0.047	0.101	0.46	0.116	0.081	1.42
1768	0.071	0.095	0.75	0.079	0.095	0.83	0.070	0.089	0.78
1770	-0.015	0.095	-0.16	-0.019	0.096	-0.20	-0.006	0.092	-0.07
1772	-0.015	0.095	-0.16	-0.023	0.095	-0.24	-0.025	0.094	-0.27
1774	-0.119	0.106	-1.12	-0.111	0.107	-1.03	-0.131	0.100	-1.30
1776	0.067	0.101	0.67	0.066	0.102	0.64	0.092	0.093	1.00
1778	0.204	0.095	2.15	0.182	0.095	1.92	0.160	0.091	1.76
1780	-0.241	0.100	-2.40	-0.210	0.101	-2.09	-0.153	0.099	-1.55
1782	0.199	0.104	1.91	0.185	0.105	1.77	0.148	0.103	1.44
1784	0.024	0.099	0.24	0.020	0.100	0.20	-0.006	0.095	-0.06
1786	-0.101	0.100	-1.01	-0.093	0.100	-0.93	-0.095	0.095	-1.01
1788	-0.220	0.107	-2.05	-0.199	0.108	-1.84	-0.148	0.104	-1.43
1790	0.048	0.104	0.46	0.040	0.105	0.38	-0.011	0.105	-0.10
1792	0.065	0.099	0.66	0.047	0.101	0.46	0.079	0.099	0.80
1794	-0.045	0.096	-0.47	-0.046	0.098	-0.47	-0.074	0.094	-0.79
1796	-0.278	0.106	-2.62	-0.298	0.109	-2.74	-0.288	0.102	-2.83

1798	-0.134	0.109	-1.23	-0.112	0.112	-0.99	-0.107	0.105	-1.03
1800	-0.106	0.098	-1.08	-0.094	0.100	-0.94	-0.115	0.095	-1.21
1802	0.118	0.083	1.41	0.105	0.083	1.26	0.108	0.082	1.32
1804	0.015	0.103	0.15	0.019	0.103	0.19	0.007	0.096	0.07
1806	-0.094	0.125	-0.75	-0.088	0.126	-0.70	-0.068	0.117	-0.59
1808	0.227	0.108	2.10	0.223	0.110	2.04	0.224	0.102	2.20
1810	-0.210	0.099	-2.13	-0.228	0.100	-2.27	-0.219	0.093	-2.36
1812	-0.356	0.182	-1.95	-0.339	0.175	-1.94	-0.340	0.166	-2.04
1814	-0.196	0.190	-1.03	-0.203	0.182	-1.11	-0.247	0.176	-1.41
1816	0.150	0.135	1.12	0.148	0.136	1.09	0.166	0.133	1.25
1818	0.065	0.133	0.49	0.074	0.133	0.55	0.072	0.125	0.58
1820	0.139	0.128	1.08	0.127	0.129	0.99	0.151	0.120	1.26
1822	-0.097	0.127	-0.77	-0.072	0.128	-0.56	-0.064	0.123	-0.52
1824	0.203	0.130	1.56	0.165	0.132	1.25	0.162	0.129	1.26
1826	-0.126	0.159	-0.79	-0.109	0.158	-0.69	-0.124	0.155	-0.80
1828	-0.192	0.153	-1.26	-0.182	0.151	-1.20	-0.197	0.148	-1.33
1830	0.002	0.136	0.02	-0.022	0.134	-0.16	0.016	0.129	0.12
1832	0.012	0.146	0.08	0.032	0.145	0.22	0.011	0.136	0.08
1834	-0.021	0.126	-0.17	-0.021	0.126	-0.17	0.002	0.121	0.02
1836	0.215	0.119	1.80	0.207	0.118	1.75	0.204	0.117	1.75
1838	0.128	0.105	1.22	0.129	0.105	1.23	0.126	0.103	1.22
1840	0.048	0.080	0.60	0.062	0.081	0.77	0.057	0.078	0.74
1842	-0.052	0.072	-0.72	-0.056	0.072	-0.77	-0.064	0.068	-0.94
1844	-0.031	0.074	-0.42	-0.032	0.074	-0.43	-0.046	0.070	-0.66
1846	0.012	0.080	0.15	0.006	0.079	0.08	0.050	0.076	0.66
1848	-0.019	0.083	-0.23	-0.027	0.082	-0.33	-0.048	0.078	-0.61
1850	-0.032	0.087	-0.37	-0.021	0.086	-0.24	-0.036	0.083	-0.44
1852	0.084	0.088	0.95	0.086	0.088	0.97	0.092	0.085	1.08
1854	0.174	0.078	2.22	0.176	0.078	2.25	0.177	0.073	2.41
1856	-0.048	0.075	-0.63	-0.061	0.075	-0.81	-0.062	0.069	-0.89
1858	-0.079	0.076	-1.04	-0.074	0.076	-0.97	-0.071	0.069	-1.03
1860	0.053	0.073	0.73	0.055	0.073	0.75	0.055	0.066	0.83
1862	0.185	0.071	2.59	0.188	0.071	2.65	0.178	0.068	2.63
1864	0.042	0.072	0.59	0.049	0.071	0.69	0.052	0.068	0.76
1866	0.042	0.072	0.59	0.037	0.071	0.52	0.030	0.067	0.45
1868	-0.011	0.073	-0.15	-0.004	0.073	-0.05	-0.005	0.070	-0.07
1870	0.000	0.069	0.00	-0.012	0.068	-0.17	0.004	0.067	0.06
1872	0.115	0.063	1.82	0.110	0.062	1.78	0.101	0.060	1.67
1874	0.209	0.061	3.45	0.211	0.059	3.56	0.218	0.055	3.99
1876	0.019	0.062	0.31	0.027	0.060	0.46	0.028	0.055	0.51
1878	0.075	0.062	1.22	0.074	0.060	1.24	0.073	0.058	1.27
1880	0.042	0.060	0.71	0.041	0.058	0.71	0.044	0.055	0.80
1882	0.017	0.065	0.26	0.023	0.064	0.36	0.016	0.063	0.25
1884	-0.128	0.070	-1.83	-0.132	0.069	-1.91	-0.111	0.068	-1.64
1886	0.001	0.076	0.01	0.001	0.075	0.02	-0.009	0.073	-0.12
1888	-0.088	0.078	-1.13	-0.096	0.076	-1.26	-0.106	0.074	-1.44

1890	-0.057	0.074	-0.77	-0.053	0.071	-0.74	-0.072	0.069	-1.05
1892	0.090	0.086	1.04	0.088	0.084	1.05	0.101	0.081	1.26
1894	-0.156	0.084	-1.87	-0.156	0.081	-1.92	-0.156	0.079	-1.98
1896	-0.028	0.069	-0.40	-0.025	0.067	-0.38	-0.009	0.060	-0.14
1898	0.094	0.071	1.31	0.101	0.070	1.45	0.100	0.064	1.57
1900	0.064	0.068	0.94	0.058	0.066	0.87	0.052	0.065	0.80
1902	0.063	0.063	1.00	0.058	0.061	0.94	0.054	0.060	0.91
1904	-0.073	0.066	-1.12	-0.066	0.064	-1.04	-0.063	0.063	-1.01
1906	-0.050	0.072	-0.70	-0.050	0.070	-0.72	-0.069	0.068	-1.01
1908	-0.012	0.074	-0.16	-0.008	0.072	-0.11	0.004	0.070	0.06
1910	-0.012	0.077	-0.15	-0.020	0.075	-0.27	-0.043	0.070	-0.61
1912	0.129	0.080	1.62	0.142	0.077	1.83	0.156	0.073	2.14
1914	-0.054	0.080	-0.68	-0.068	0.077	-0.88	-0.064	0.075	-0.85
1916	0.253	0.073	3.45	0.251	0.070	3.56	0.243	0.069	3.52
1918	0.283	0.053	5.32	0.294	0.052	5.71	0.304	0.051	6.00
1920	0.363	0.052	7.03	0.361	0.050	7.24	0.359	0.046	7.74
1922	-0.356	0.063	-5.66	-0.361	0.061	-5.96	-0.366	0.057	-6.45
1924	0.068	0.072	0.94	0.075	0.069	1.08	0.050	0.068	0.73
1926	-0.253	0.072	-3.50	-0.250	0.070	-3.59	-0.214	0.068	-3.13
1928	0.081	0.077	1.04	0.082	0.074	1.11	0.072	0.071	1.01
1930	-0.015	0.089	-0.17	-0.020	0.086	-0.23	-0.016	0.082	-0.20
1932	-0.242	0.106	-2.28	-0.240	0.103	-2.33	-0.256	0.100	-2.56
1934	-0.038	0.111	-0.34	-0.053	0.108	-0.49	-0.080	0.105	-0.76
1936	-0.132	0.099	-1.33	-0.146	0.096	-1.51	-0.162	0.094	-1.72
1938	-0.092	0.108	-0.85	-0.069	0.104	-0.66	-0.016	0.102	-0.16
1940	0.201	0.104	1.94	0.204	0.099	2.05	0.208	0.097	2.15
1942	0.119	0.076	1.56	0.122	0.073	1.68	0.129	0.071	1.81
1944	0.084	0.108	0.78	0.089	0.105	0.85	0.051	0.100	0.52
1946	0.117	0.125	0.93	0.111	0.121	0.91	0.126	0.115	1.09
1948	0.213	0.105	2.02	0.218	0.101	2.15	0.231	0.098	2.37
1950	0.037	0.095	0.39	0.045	0.091	0.49	0.048	0.087	0.55
1952	0.027	0.099	0.27	0.036	0.096	0.37	0.005	0.092	0.05
1954	0.134	0.093	1.45	0.108	0.090	1.20	0.152	0.087	1.76
1956	0.285	0.086	3.32	0.296	0.083	3.59	0.262	0.079	3.30
1958	0.112	0.094	1.20	0.123	0.090	1.36	0.133	0.086	1.54
1960	0.374	0.094	3.99	0.365	0.090	4.06	0.361	0.087	4.15
1962	0.033	0.091	0.37	0.004	0.089	0.04	-0.021	0.086	-0.25
1964	0.397	0.094	4.22	0.415	0.092	4.51	0.424	0.089	4.75
1966	-0.366	0.103	-3.56	-0.364	0.100	-3.63	-0.334	0.098	-3.41
1968	0.107	0.105	1.03	0.119	0.102	1.16	0.083	0.097	0.85
1970	0.056	0.098	0.57	0.049	0.096	0.51	0.061	0.091	0.67
1972	0.185	0.098	1.88	0.206	0.096	2.16	0.255	0.094	2.71
Average std error	0.122			0.123			0.114		
R^2	0.604			0.597			0.659		
Obs	3577			3577			3577		