Boundary effects and repeat-sales

Arnaud Simon
Université Paris-Dauphine
CEREG
Place du Maréchal de Lattre-de-Tassigny
75775 Paris Cedex 16
Arnaud.simon@dauphine.fr
Boundary effects and repeat-sales

Abstract

We study in this article the consequences of the double censoring (right and left) applied on every sample of repeat-sales before calculating a repeat-sales index. We show that, independently of the specificities of each dataset, there exist some intrinsic features for this kind of index that directly depend on the censoring mechanism. Working with the standard deviation of the estimator and the well known bias of the Cass, Shiller model we bring to fore a time structure for these two quantities. Their behaviour is different near the edges of the interval compared to the middle, and this phenomenon is generally more pronounced for the right side than for the left side. A sensitivity analysis is developed using a “neutral sample” in order to study the way this U-shape evolve in relation with the fundamental parameters. With this analysis we also try to determine the optimal length of the estimation interval, the consequences of the data dispersion and the impact of the liquidity levels in the market (number of goods sold at each date and resale speed).

Key words: repeat-sales, boundary effects, sensitivity analysis, optimal parameters.
1. Introduction

The two seminal articles for the repeat-sales technique are Bailey, Muth and Nourse (1963), in an homoscedastic situation, and Case, Shiller (1987) for the heteroscedastic context. Since these two papers the repeat-sales approach has become one of the most popular index because of its quality and its flexibility. It used not only for residential but also for commercial real estate, cf. Gatzlaff, Geltner (1998). One can also refer to Chau et al. (2005) for a recent example of a multisectorial application of the RSI (repeat-sales index) and to Baroni et al. (2004) for the French context. In this article we are going to study the time structure of the errors and how the reliability of the index evolves. For any index technique the errors is generally not one single piece. We can divide the errors in two elements: the first component depends one the methodology employed and the second one depends on the specificities of each dataset. Our goal in this paper consists in studying the first one, although we will give some elements on the second one when we will study the liquidity shocks. The main characteristic of the repeat-sales technique is its double censoring mechanism. Indeed, when we work with an estimation interval \([0,T]\), the repeat-sales with a purchase before 0 or a resale after \(T\) are not considered during the estimation process. As we will see, this feature has some important consequences on the time-structure of the errors: we will generally observe an asymmetric U-shape. With a sensitivity analysis we will try to determine which are the main intrinsic factors that drive this structure. We will study the impacts of the noises, the length of the estimation interval, the level of liquidity in the market and the resale speed. As a by-product we could give some clue about the optimal choice of parameters to estimate a RSI. For instance, the number of transactions collected for the first date seems to be a important driver of the global reliability of the index. The rest of this paper is organised as follows. Section 2 presents the repeat-sales index theory. If we want to study the structural
determinants of the errors we have to work with a “neutral sample” in order to remove the impact of the specificities of a real dataset; Section 3 presents the neutral sample. Section 4 is devoted to the sensitivity analysis. The last section will conclude.

2. The estimation of the RSI

In this paragraph we summarize briefly the theoretical framework established in Simon(2007).

2.1. The classical estimation of the repeat-sales index

In the repeat-sales approach, the price of a property k at time t is decomposed in three parts:

\[ \ln(p_{k,t}) = \ln(\text{Index}_t) + G_{k,t} + N_{k,t} \]

\( \text{Index}_t \) is the true index value, \( G_{k,t} \) is a Gaussian random walk representing the asset’s own trend and \( N_{k,t} \) is a white noise associated to the market imperfections. If we denote \( \text{Rate} = (\text{rate}_0, \text{rate}_1, \ldots, \text{rate}_{t-1})' \) the vector of the continuous rates for each elementary time interval \([t,t+1]\), we have \( \text{Index}_t = \exp(\text{rate}_0 + \text{rate}_1 + \ldots + \text{rate}_{t-1}) \), or equivalently \( \text{rate}_t = \ln(\text{Index}_{t+1}/\text{Index}_t) \). For a repeat-sale we can write at the purchase time \( t_i \) : \( \ln(p_{k,i}) = \ln(\text{Index}_i) + G_{k,i} + N_{k,i} \) and at the resale time \( t_j \) : \( \ln(p_{k,j}) = \ln(\text{Index}_j) + G_{k,j} + N_{k,j} \). Thus, subtracting, we get

\[ \ln(p_{k,j}/p_{k,i}) = \ln(\text{Index}_j/\text{Index}_i) + (G_{k,j} - G_{k,i}) + (N_{k,j} - N_{k,i}) \]

The return rate realised for the property \( k \) is equal to the index return rate during the same period, plus the random walk and the white noise increments. Each repeat-sale gives a relation of that nature, under a matrix form we have \( Y = D*\text{LIndex} + \varepsilon \). Here, \( Y \) is the column vector of the log return rates realised in the estimation dataset and \( \text{LIndex} = (\ln(\text{Index}_1), \ldots, \ln(\text{Index}_T))' \). \( \varepsilon \) is the error term and \( D \)
is a non singular matrix. Moreover, if we notice that there exists an invertible matrix $A$, such that $L \text{Index} = A \text{Rate}$, we can also write $Y = (DA) (A^{-1} L \text{Index}) + \varepsilon = (DA) \text{Rate} + \varepsilon$. In the estimation process, the true values Index and Rate will be replaced with their estimators, respectively denoted $\text{Ind} = (\text{Ind}_1, \ldots, \text{Ind}_T)'$ and $\text{R} = (r_0, r_1, \ldots, r_{T-1})'$. The usual estimation of $Y = D^* L \text{Index} + \varepsilon$ or $Y = (DA) \text{Rate} + \varepsilon$ is carried out in three steps because of the heteroscedasticity of $\varepsilon$. Indeed, the specification of the error term leads to the relation $\text{Var}(\varepsilon_k) = 2\sigma_N^2 + \sigma_G^2(j-i)$ in which the values $\sigma_N$ and $\sigma_G$ are the volatilities associated to $G_{k,t}$ and $N_{k,t}$, and $j-i$ is the holding period for the $k^{th}$ repeat-sale. The first step consists in running an OLS that produces a residuals series. These residuals are then regressed on a constant and on the length of the holding periods to estimate $\sigma_N$, $\sigma_G$ and the variance-covariance matrix of $\varepsilon$, denoted $\Sigma$. Finally, the last step is an application of the generalised least squares procedure with the estimated matrix $\Sigma$. In Simon (2007) we establish that this traditional approach is equivalent to an algorithmic decomposition (cf. Figure 1), where the informational framework becomes explicit. The next paragraphs present briefly the mechanism of this algorithm.

2.2. Notations, basic concepts and decomposition of the RSI

2.2.1. Time of noise equality

The variance of the residual $\varepsilon_k$ measures the quality of the approximation $\ln(p_{k,j}/p_{k,i}) \approx \ln(\text{Ind}_j/\text{Ind}_i)$ for the $k^{th}$ repeat-sale. This quantity $2\sigma_N^2 + \sigma_G^2(j-i)$ can be interpreted as a noise

---

1 D is a matrix extracted from another matrix $D'$; the first column has been removed to avoid a singularity in the estimation process. The number of lines of $D'$ is equal to the total number of the repeat-sales in the dataset and its $T+1$ columns correspond to the different possible times for the trades. In each line -1 indicates the purchase date, +1 the resale date and the rest is completed with zeros.

2 A is a triangular matrix whose values are equal to 1 on the diagonal and under it, 0 elsewhere.

3 The basic rules of linear algebra indicate that the matrix $DA$ gets as many lines as the number of repeat sales in the sample, and that the columns correspond to the elementary time intervals. In each line of $DA$, if the purchase occurs at $t_i$ and the resale at $t_j$, we have $0 \ldots 0 1(t_i) 1 \ldots 1(t_{j-1}) 0 \ldots 0$. Therefore, the relation $Y = (DA) \text{Rate} + \varepsilon$ simply means that $\ln(\text{return}) = \text{rate}_{i+1} + \ldots + \text{rate}_{j-1} + \varepsilon$

4 $\Sigma$ is a diagonal matrix with a dimension equal to the size of the repeat-sales sample.
measure for each data. As a repeat-sale is composed of two transactions (a purchase and a resale), the first noise source $N_{k,t}$ appears twice with $2\sigma_N^2$. The contribution of the second source $G_{k,t}$ depends on the time elapsed between these two transactions: $\sigma_G^2(j-i)$. Consequently, as time goes by, the above approximation becomes less and less reliable. For the future developments it is useful to modify slightly the expression of the total noise, factorising by $\sigma_G^2$: $2\sigma_N^2 + \sigma_G^2(j-i) = \sigma_G^2[(2\sigma_N^2/\sigma_G^2)+(j-i)] = \sigma_G^2[\Theta+(j-i)]$. What does $\Theta = 2\sigma_N^2/\sigma_G^2$ represent? The first noise source provides a constant intensity ($2\sigma_N^2$) whereas the size of the second is time-varying ($\sigma_G^2(j-i)$). For a short holding period the first one is louder than the second. But as the former is constant and the latter is increasing regularly with the length of the holding period, there exists a duration where the two sources will reach the same levels. Then, the Gaussian noise $G_{k,t}$ will exceed the white noise. This time is the solution of the equation: $2\sigma_N^2 = \sigma_G^2 \times \text{time} \Leftrightarrow \text{time} = 2\sigma_N^2/\sigma_G^2 = \Theta$. For that reason, we will call $\Theta$ the “time of noise equality”. Below in the formula, the function $G(x) = x/(x+\Theta)$ will sometimes appear. For an holding period $j-i$ we have $G(j-i) = (j-i)/(\Theta+(j-i)) = \sigma_G^2(j-i)/[2\sigma_N^2+\sigma_G^2(j-i)]$. $G(j-i)$ is the proportion of the time-varying noise in the total noise; these numbers will be used subsequently as a system of weights.

2.2.2. Quantity of information delivered by a repeat-sale

The theoretical reformulation developed in this article brings the concept of information at crucial place. As $\Theta+(j-i)$ is a noise measure, its inverse can be interpreted as an information measure. Indeed, if the noise is growing, that is if the approximation $\text{Ln}(p_{k,j}/p_{k,i}) \approx \text{Ln}(\text{Ind}_j/\text{Ind}_i)$ is becoming less reliable, the inverse of $\Theta+(j-i)$ is decreasing. Consequently,
is a direct measure\(^5\) (for a repeat-sale with a purchase at \(t_i\) and a resale at \(t_j\)) of the quality of the approximation or, equivalently, of the quantity of information delivered. In the estimation process, the small weights for the long holding periods make these observations less contributive to the index values.

### 2.2.3. Subsets and algorithmic decomposition of the RSI

The set of the repeat-sales with a purchase at \(t_i\) and a resale at \(t_j\) will be denoted by \(C(i,j)\). For a time interval \([t',t]\), we will say that an observation is relevant if its holding period includes \([t',t]\); that is if the purchase is at \(t_i \leq t'\) and the resale at \(t_j \geq t\). This sub-sample will be denoted \(S_{t',t}\). For an elementary time-interval \([t,t+1]\), we will also used the simplified notation \(S_{t,t+1} = S_{t,t+1}\). If we organize the dataset in an triangular upper table, the sub-set \(S_{t',t}\) will correspond to the cells indicated in Table 1. From the optimization problem associated to the general least squares procedure, we demonstrated in Simon (2007) that the repeat-sales index estimation could be realised using the algorithmic decomposition presented in Figure 1. The left-hand side is related to the informational concepts (for example the matrix \(H\)), whereas the right-hand side is associated to the price measures (for example the mean of the mean rates \(\rho_t\)). The final values of the index come from the confrontation of these two parts.

### 2.3. The real distribution and its informational equivalent

The time is discretized from 0 to \(T\) (the present), and divided in \(T\) sub-intervals.

\[
\begin{array}{cccccccccc}
0 & 1 & 2 & \ldots & t & t+1 & \ldots & T-1 & T
\end{array}
\]

\(^5\) These measures are relative ones. The matter is their relative sizes and not the absolute levels. They can be defined up to a constant in order to standardize the measures.
We assume that the transactions occur only at these moments, and not between two dates (the step can be for example a month or a quarter, depending on the data quality). Each observation gives a time couple \((t_i; t_j)\) with \(0 \leq t_i < t_j \leq T\); thus we have \(\frac{1}{2}T(T+1)\) possibilities for the holding periods. The number of elements in \(C(i,j)\) is \(n_{i,j}\), and we denote \(N = \sum_{i,j} n_{i,j}\) the total number of the repeat-sales in the dataset. Table 2a is a representation of the real distribution of the \(\{n_{i,j}\}\). As each element of \(C(i,j)\) provides a quantity of information equal to \((\Theta+(j-i))^{-1}\), the total informational contribution of the \(n_{i,j}\) observations of \(C(i,j)\) is \(n_{i,j} (\Theta+(j-i))^{-1}\) = \(n_{i,j} / (\Theta + (j-i)) = L_{i,j}\). Therefore, from the real distribution \(\{n_{i,j}\}\) we get the informational distribution \(\{L_{i,j}\}\) (cf. Table 2b), just dividing its elements by \(\Theta+(j-i)\). The total quantity of information embedded in a dataset is: \(I = \sum_{i,j} L_{i,j}\).

### 2.4. Averages for the noise proportions, the periods and the frequencies

The number of repeat-sales in \(\text{Spl}\) is \(n = \sum_{i \leq t < j} n_{i,j}\). For an element of \(C(i,j)\), the length of the holding period is \(j - i\). Using the function \(G\), we can define the G-mean\(^6\) \(\zeta_t\) of these lengths in \(\text{Spl}\) by \(\sum_{i \leq t < j} \sum_k G(j-i) = n G(\zeta_t)\). The first sum enumerates all the classes \(C(i,j)\) that belong to \(\text{Spl}\), the second all the elements in each of these classes. Moreover, as \(G(j-i)\) measures the proportion of the time varying-noise \(G_{k,t}\) in the total noise for a repeat-sales of \(C(i,j)\), the quantity \(G(\zeta_t)\) can also be interpreted as the mean proportion of this Gaussian noise in the global one, for the whole sub-sample \(\text{Spl}\). In the same spirit, we define the arithmetic average \(F_t\) of the holding frequencies \(1/(j - i)\), weighted by the \(G(j-i)\), in \(\text{Spl}\): \(F_t = (n G(\zeta_t))^{-1}\)

---

\(^6\) We recall here that the concept of average is a very general one. If a function \(G\) is strictly increasing or decreasing the G-mean of the numbers \(\{x_1, x_2, \ldots, x_n\}\), weighted by the \(\{\alpha_1, \alpha_2, \ldots, \alpha_n\}\), is the number \(X\) such that: \(\alpha G(X) = \alpha_1 G(x_1) + \alpha_2 G(x_2) + \ldots + \alpha_n G(x_n)\) with \(\alpha = \sum_{i=1}^{n} \alpha_i\). An arithmetic mean corresponds to \(G(x) = x\), a geometric one to \(G(x) = \ln(x)\) and the harmonic average to \(G(x) = 1/x\)
\[
\sum_{i \leq t < j} \sum_{k'} G(j-i)^n (1/(j-i)) = I^t / (n^t G(\zeta_t)). \]

Its inverse \( \tau^t = (F^t)^{-1} \) is then the harmonic average\(^7\) of the holding periods \(j-i\), weighted by the \(G(j-i)\), in \(\text{Spl}^t\). If at first sight the two averages \(\zeta^t\) and \(\tau^t\) can appear as two different concepts, in fact it is nothing of the sort. We always have, for each sub-sample \(\text{Spl}^t\), \(\zeta^t = \tau^t\).

2.5. Two matrixes

The matrix \(\eta\) is a diagonal one, its \(T\) diagonal coefficients are: \(n^0G(\zeta^0), \ldots, n^{T-1}G(\zeta^{T-1})\). The \(i^{th}\) element gives the number of repeat-sales relevant for \([i,i+1]\), multiplied by \(G(\zeta^i)\). Now we are working with the interval \([t',t+1]\). A repeat-sale provides information on it if the purchase is at \(t'\) or before and if the resale takes place at \(t+1\) or after. The quantity of information relevant for \([t',t+1]\) is thus \(I_{[t',t+1]} = \sum_{i \leq t'} \sum_{t < j} L_{i,j}\). For an interval \([t,t+1]\) we simply denote \(I^t\) for \(I_{[t,t+1]}\). As exemplified in Table 3, \(I^t_{[t',t+1]}\) can be calculated buy-side with the partial sums \(B_0^t, B_1^t, \ldots, B_{t'}^t\) or sell-side with \(S_0^{t'}, S_1^{t'}, \ldots, S_{t'}^{t}\). But we always have \(I^t_{[t',t+1]} = B_0^t + \ldots + B_{t'}^t = S_0^{t'} + \ldots + S_{t'}^{t+1}\). For all the intervals included in \([0,T]\) we get this way the related quantities of information. These values are arranged in a symmetric matrix \(\hat{I}\).

\[
\hat{I} = \begin{pmatrix}
I_{[0,1]} & I_{[0,2]} & I_{[0,3]} & \ldots & I_{[0,T]} \\
I_{[0,2]} & I_{[1,2]} & I_{[1,3]} & \ldots & I_{[1,T]} \\
I_{[0,3]} & I_{[1,3]} & I_{[2,3]} & \ldots & I_{[2,T]} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
I_{[0,T]} & I_{[1,T]} & I_{[2,T]} & \ldots & I_{[T-1,T]}
\end{pmatrix}
\]

\(^7\) We have \((n^t G(\zeta^t)) / \tau^t = \sum_{i \leq t < j} \sum_{k'} G(j-i) * (1/(j-i)) = I^t\)
2.6. The mean prices

Within each repeat-sales class \((i,j)\), we calculate the geometric and equally weighted averages of the purchase prices: 
\[
h_p^{(i,j)} = (\Pi_{k'} p_{k',i})^{1/n_{i,j}},
\]
and of the resale prices: 
\[
h_f^{(i,j)} = (\Pi_{k'} p_{k',j})^{1/n_{i,j}}.
\]
For an elementary time-interval \([t, t+1]\), the relevant classes \(\zeta(i,j)\) are the ones that satisfy to the inequalities \(i \leq t < j\). With these classes, we calculate the geometric average \(H_p(t)\) of the \(h_p^{(i,j)}\), weighted by their corresponding \(L_{i,j}\) (here, the total mass of the weights is \(I^t = \sum_{i \leq j} L_{i,j}\)):
\[
H_p(t) = \left( \Pi_{i \leq t < j} (h_p^{(i,j)})^{L_{i,j}} \right)^{1/I^t} = \left( \Pi_{i \leq t < j} (\Pi_{k'} p_{k',i})^{1/(\Theta + (j-i))} \right)^{1/I^t}.
\]
As indicated in the second part, \(H_p(t)\) is also the geometric mean of the purchase prices, weighted by their informational contribution \(1/(\Theta +(j-i))\), for the investors who owned real estate during at least \([t, t+1]\). Similarly, we also define the mean resale price \(H_f(t)\):
\[
H_f(t) = \left( \Pi_{i \leq t < j} (h_f^{(i,j)})^{L_{i,j}} \right)^{1/I^t} = \left( \Pi_{i \leq t < j} (\Pi_{k'} p_{k',j})^{1/(\Theta + (j-i))} \right)^{1/I^t}.
\]
As we can see, \(H_p(t)\) can be interpreted as a mean purchase price weighted by the informational activity, buy-side, of the market. The interpretation is the same for \(H_f(t)\) with the informational activity of the market, sell-side.

2.7. The mean of the mean rates

For a given repeat-sales \(k'\) in \(\zeta(i,j)\), with a purchase price \(p_{k',i}\) and a resale price \(p_{k',j}\), the mean continuous rate realised on its holding period \(j-i\) is 
\[
r_{k'}^{(i,j)} = \frac{\ln(p_{k',j}/p_{k',i})}{j-i}.
\]
In the subset \(S_{pl}^t\), we calculate the arithmetic mean of these mean rates \(r_{k'}^{(i,j)}\), weighted\(^8\) by the \(G(j-i)\): 
\[
\rho_t = \sum_{i \leq t < j} L_{i,j} \cdot (\Pi_{k'} p_{k',j})^{1/(\Theta + (j-i))}.
\]

\(^8\) The total mass of these weights is \(n'G(\zeta)\)
This value is a measure of the mean profitability of the investment for the people who owned real estate during \([t, t+1]\), independently of the length of their holding period. The weights in this average depend on the informational contribution of each data. We demonstrated in Simon (2007) that we can also write \(\rho_t\) in a simpler way with the following formula: 

\[
\rho_t = \left( \frac{1}{\tau^t} \right) \ast (\ln H_f(t) - \ln H_p(t)).
\]

For Spl\(^t\), this relation is actually the aggregated equivalent of 

\[
\rho_{k,j}^{(i,j)} = \ln(p_{k,j} / p_{k,i}) / (j-i),
\]

with the harmonic mean of the holding periods \(\tau^t\), the mean purchase price \(H_p(t)\) and the mean resale price \(H_f(t)\). All these averages are weighted by the informational activity of the market. We denote the vector of these mean rates 

\[
P = (\rho_0, \rho_1, \ldots, \rho_{T-1}).
\]

### 2.8. The index and the relation \(\hat{I} R = \eta P\)

The estimation of the RSI can now be realised just solving the equation: 

\[
\hat{I} R = \eta P \Leftrightarrow R = (\hat{I}^{-1} \eta) P.
\]

The unknown is the vector \(R = (r_0, r_1, \ldots, r_{T-1})\) of the monoperiodic growth rates of the index. The three others components of this equation (\(\hat{I}\), \(\eta\) and \(P\)) are calculated directly from the dataset. The main advantages of this formalism are its interpretability and its flexibility: the matrix \(\hat{I}\) gives us the informational structure of the dataset, the matrix \(\eta\) counts the relevant repeat-sales for each time interval \([t, t+1]\) and the vector \(P\) indicates the levels of profitability of the investment for the people who owned real estate at the different dates.
3. **The exponential benchmark**

3.1. Hypotheses

A repeat-sales sample is never an intuitive object because of the left censoring at 0 and the right censoring at T. These cuts generate a lower data density near the edges, whereas the quantity of information is higher in the middle of the interval. In this situation, what does a uniform distribution of repeat-sales look like? The answer is probably not unique. The benchmark sample we study in this article is based on the two following assumptions:

1) *The quantities of goods traded on the market at each date are constant and denoted K.*

2) *The length of the holding period is deterministic. The survival curve for each cohort is an exponential with a parameter $\lambda > 0$ (the same for all cohorts): $S(t) = K e^{-\lambda t}$.  
   
   This last hypothesis means that the resale intensity is constant. Such a choice is of course unrealistic because it implies that the percentage of sold houses in the next year is not influenced by the length of the holding period. In a probabilistic context, if we introduce the hazard rate\(^9\) which measures the instantaneous probability of resale: $\lambda(t) = (1/\Delta t) \cdot \text{Prob}(\text{resale} > t+\Delta t | \text{resale} \geq t)$, it is a well-known fact that the choice of an exponential distribution is equivalent to the choice of a constant hazard rate. In the real world things are of course different. For the standard owner (cf. Figure 2) we can reasonably think that the hazard rate is first low (quick resales are scarce). In a second time, it increases progressively up to a stationary level, potentially modified by the economic context (residential time). Then, as time goes by, the possibility of moving due to the retirement, or even the death of the householder, would bring the hazard rate to a higher level (ageing). Nevertheless, the aim of

---

\(^9\) $\lambda(t)$ is a classical concept in the survival models, cf. Kalbfleisch, Prentice (2002). It appears for example in the econometrical studies for the prepayment and the default options embedded in the mortgages, cf. Deng, Quigley, Van Order (2000).
our benchmark does not consist in describing exactly the reality. We just try to modelize a basic, uniform and deterministic behaviour in order to have some comparative elements. Under these assumptions we study below the various building blocks that appear in the RSI decomposition.

3.2. The real and the informational distributions

3.2.1. Notations

We denote $\alpha = e^{-\lambda}$ the instantaneous survival rate, that is the proportion of goods that are not sold during one period of time. As $\lambda \approx 0^+$ we have $\alpha \approx 1 - \lambda$. After $k$ periods of time, the disappearance rate is $d(k) = 1 - \alpha^k$. Thus, the mean linear disappearance rate per time unit is $\lambda_{lin} = d(k)/k$. These concepts are illustrated with Figure 3. $K$ represents the number of goods traded on the market at each date. In the next paragraph the quantity $K' = K (1-\alpha)/\alpha$ will sometimes appear. If we use the approximation $\alpha \approx 1 - \lambda$, we get $K' = K (1 - e^{-\lambda})/e^{-\lambda} = K (e^\lambda - 1) \approx K\lambda \approx K(1-\alpha)$. $K'$ is approximately equal to the number of resales observed in the next period within a sample of $K$ initial elements.

3.2.2. The real distribution

We consider the $K$ goods purchased at time $i$ on the market. The number of goods that are still alive at time $i+1$ is $K\alpha$, at time $i+2$ we have $K\alpha^2$, ..., and at time $j$ it remains $K\alpha^j$. From this we get $n_{i,i+1} = K - K\alpha = K (1-\alpha) = K (1 - \alpha)/\alpha$, $n_{i,i+2} = K\alpha - K\alpha^2 = K\alpha (1-\alpha) = K\alpha^2$ and more generally $n_{ij} = K\alpha^{j-i} - K\alpha^{j-i} = K\alpha^{j-i}$. The real distribution for the benchmark sample is illustrated in Table 4a. The sums $b_i$ (on the lines) give the number of repeat-sales with a purchase at $i$, whereas the sums $s_j$ (on the columns) give the number of repeat-sales

---

10 We have $\alpha = S(t+1)/S(t) = S(t'+1)/S(t')$ for all $t$ and $t'$
11 Here the “death” means a resale
with a resale at j. As we have geometric progressions, we get very easily the closed formulas for these two quantities. Moreover, we can establish that the total number of repeat-sales observed with this benchmark sample is: \( N = KT \left[ 1 - \alpha \frac{\lambda_{\text{lin}}(T)}{(1-\alpha)} \right] \approx KT \left[ 1 - \frac{\lambda_{\text{lin}}(T)}{\lambda} \right]. \)

Among the K goods traded on the market at each date, some of them will not give a repeat-sales observation because the resale will occur after T. During the time interval \([0,T]\), as K goods are purchased at each date \(t (t = 0,\ldots,T-1)\) we have a potential number of repeat-sales that is equal to KT. If we compare KT and the expression that we got for N, we can establish directly that the proportion of the missing goods is simply equal to \( \pi(T) = \frac{\lambda_{\text{lin}}(T)}{\alpha} \approx \frac{\lambda_{\text{lin}}(T)}{\lambda}. \) (cf. appendix A for the details).

### 3.2.3. The informational distribution

The \( n_{ij} \) elements of \( C(i,j) \) provide a quantity of information \(^{12}\) \( L_{ij} = n_{ij} / (\Theta + (j-i)) \). Table 4b illustrates this distribution. The informational sums on the lines \( B_i \) and on the columns \( S_j \) depends on the values \( u_n = \frac{\alpha}{(\Theta + 1)} + \frac{\alpha^2}{(\Theta + 2)} + \frac{\alpha^3}{(\Theta + 3)} + \ldots + \frac{\alpha^n}{(\Theta + n)} \). This number represents the quantity of information provided by a sample with \( K = 1/\lambda \) repeat-sales, initiated at time \( t \), and observed during \( n \) periods of time. We denote \( \ell \) the limit of \( (u_n) \); if we approximate \( \Theta \) with the closer integer we have \( \ell = -[\ln (1-\alpha) + \alpha + \alpha^2 / 2 + \ldots + \alpha^{\Theta} / \Theta] / \alpha^{\Theta}. \)

The total quantity of information provided by these \( N \) repeat-sales is equal to: \( I = K' \left[ (T + \Theta + 1) u_T - T \pi(T) \right] \), cf. appendix A. If we would not have any limit of observation on the right of the interval \([0,T]\), the repeat-sales initiated between 0 and T-1 would provide a quantity of information equal to \( K'T\ell \). We establish in appendix A that the proportion of the missing information \( \mu = (K'T\ell - I) / K'T\ell \), that is the informational equivalent of \( \pi \), is equal to \( \mu = \left[ \ell + \pi(T) - (T + \Theta + 1) \left( u_T / T \right) \right] / \ell. \)

\(^{12}\) The parameter \( \Phi \) is fixed exogenously. It may refer to the value obtained for a specific and real dataset.
3.3. Informational matrix and other quantities

3.3.1. \( n^t \) and \( n^{[t',t+1]} \)

\[ n^t = \sum_{i \leq t < j} n_{i,j} \] is the number of repeat-sales relevant for \([t,t+1]\) and \( n^{[t',t+1]} = \sum_{i \leq t' < j} n_{i,j} \)
gives the number of repeat-sales relevant for \([t',t+1]\). With the classical formula for the geometric progression we establish immediately that

\[ n^t = \left( \frac{K}{1 - \alpha} \right) \frac{d(T - t)}{d(t + 1)} \]

and

\[ n^{[t',t+1]} = \left( \frac{K}{1 - \alpha} \right) \left( 1 - \frac{d(t - t')}{d(T - t)} \right) \frac{d(T - t)}{d(t' + 1)} \]

3.3.2. \( I^t \) and \( I^{[t',t+1]} \)

We introduce for \(-1 \leq m \leq n\) the notation

\[ U(m,n) = u_n - u_m = \frac{\alpha^{m+1}}{(\Theta + m + 1)} + \ldots + \frac{\alpha^n}{(\Theta + n)} \]

with the conventions \( u_0 = 0 \) and \( u_1 = -1/\Theta \). Moreover we denote for \( 0 \leq t' \leq t < T \):

\[ \Upsilon(t',t,T) = (1 + \Theta/2)U(t,T) - [(t+\Theta/2)U(t-1,t) - (t'-\Theta/2)U(t'-1,t) - (t'-\Theta/2)U(t'-1,T-t'-1)] + (T+\Theta/2)U(T-t'-1,T) \]

With these notations we establish in the appendixes B and C that the formulas for \( I^t \) and \( I^{[t',t+1]} \) are similar. More precisely we have:

\[ I^{[t',t+1]} = \left( \frac{1 - \alpha}{\alpha} \right) \left[ n^{[t',t+1]} + K \Upsilon(t',t,T) \right] \]

\[ I^t = I^{[t',t+1]} = \left( \frac{1 - \alpha}{\alpha} \right) \left[ n^t + K \Upsilon(t,t,T) \right] \]

We can now directly compute the informational matrix \( \hat{I} \) with these two expressions.

4. Boundary effects and sensitivity

4.1. The problem

The aim of this section consists in analyzing the non-uniform behavior of the RSI across time. Indeed, as we will see several of its characteristics have a time structure; the behavior near the
edges is generally different from the one observed in the middle of the interval. We are going to
study this point for two specific aspects:

- The standard error\textsuperscript{13} of the estimator Ind\textsubscript{t} (actually, we will work with the quantity
  100*σ(Ind\textsubscript{t})/Ind\textsubscript{ext}, that gives the standard error as a percentage of the true value)
- The multiplicative bias of the index\textsuperscript{14} (in percentages)

When we measure these two quantities the results of course depends of each dataset. But what
we claim here is that, even if we take a “neutral sample”, we can observe some features that
do not depend on the nature of the observations but rather than on the intrinsic characteristics
of the model. The exponential benchmark developed in the section above will be our “neutral
sample”. Its real distribution is known as soon as we know K and λ. In order to apply the
various informational formulas, we also need to choose the values of T, θ and σ\textsubscript{G}. We
introduce here a reference situation\textsuperscript{15} that corresponds to: T = 30, σ\textsubscript{G}\textsuperscript{2} = 0,001, θ = 10, K = 200, λ = 0.1. In the next paragraphs we are going to study the time structure for the bias and
the variance of the estimators modifying each parameter separately for our reference.

\textbf{4.2. Time structure and sensitivity to the error terms σ\textsubscript{G} and θ}

In the Figures 4a and 4b the reference corresponds to σ\textsubscript{G}\textsuperscript{2} = 0,001. The two curves (error and
bias) present a U-shape. The values on the left of the interval are a little bit higher than the
ones in the middle but the edge effect is more pronounced for the right side. For the

\textsuperscript{13} We recall here that Ind\textsubscript{t} ~ \mathcal{CN} ( L\textsuperscript{ext}; diag(τ(L\text{Ind})) ), where \mathcal{V}(L\text{Ind}) = σ\textsubscript{G}\textsuperscript{2} A \bar{A}' A. A is a triangular
matrix whose values are equal to 1 on the principal diagonal and under it, 0 elsewhere (cf. Simon(2007)). Thus
we have: σ(Ind\textsubscript{t})/Ind\textsubscript{ext} = \exp(2*L\text{Ind\textsubscript{t}} + \mathcal{V}(L\text{Ind}))\textsuperscript{0.5} \left( \exp(\mathcal{V}(L\text{Ind})) - 1 \right)\textsuperscript{0.5} / \text{Index\textsubscript{t}}

\textsuperscript{14} It is a well-known fact that the RSI is slightly biased (Goetzmann 1992; Goetzmann and Peng 2002). With our
notations the multiplicative bias is equal to: \exp( \frac{1}{2}\mathcal{V}(L\text{Ind}) ). We multiply this quantity by 100 to get a
percentage. (cf. Simon 2007 for the details).

\textsuperscript{15} T = 30 : it corresponds to an index estimated on thirty years annually, or fifteen years on a six month basis.
The level chosen for the values σ\textsubscript{G}\textsuperscript{2} and θ are quite standard if we refer for instance to the article Case, Shiller
(1987). This value of λ corresponds to a mean holding period of ten years, and with this choice of K = 200 we
have approximately N = 4100 repeat-sales in the sample.
sensitivity analysis we choose to let the parameter $\sigma_G^2$ varies between 0,0001 and 0,0025. The first extreme case means that the random-walk is small, that is we are close to the Bailey, Muth and Nourse (BMN) model. In this situation the boundary effects are small; they could even be neglected. In the second extreme case, the idiosyncratic components modelled by the random walk, are strong. As we can see the impact of $\sigma_G$ on the errors is important (it doubles for $\sigma_G^2 = 0,0025$ compared to the reference). The heterogeneity of the price movements in a city has some consequences for the quality of the index. In the Case-Shiller context the U-shape becomes stronger when the heterogeneity increases; this phenomenon is very clear for the bias. Regarding the parameter $\theta = 2\sigma_N^2/\sigma_G^2$ it allows studying the impact of the white noise $\sigma_N^2$ (we let $\sigma_G^2$ at its base level 0,001). When $\sigma_N^2$ varies its impact has the same magnitude than the one generated by a variation of the random-walk (Figures 5a and 5b). If the white noise is stronger the bias and the errors evolve in the same direction (upward). Globally we also find the U-shape but a small difference appears when $\theta$ (that is $\sigma_N$) is close to zero. This choice of parameter means that the white noise can be neglected and that the error term in the regression is mainly directed by the idiosyncratic component. If the Case, Shiller situation is in the middle, if the BMN context is a pole, then $\theta \approx 0^+$ corresponds to the second polar situation. Here, it seems that the left boundary-effect is reversed. For instance for $\theta = 1$ the errors and the bias are smaller for the beginning of the interval. The pivot is at $\theta = 5$. In the article of Case, Shiller (1987) it approximately corresponds to the city of San Francisco (cf. Table 6). For the right side we do not find the same phenomenon for the small values of $\theta$.

4.3. Time structure and sensitivity to the length of the estimation interval $T$

Figures 6a and 6b present the result when the time horizon is varying from $T = 5$ to $T = 50$. Here also we have a time-structure. For the short intervals ( $T = 5$ and $T = 10$) we find an
additional error and an additional bias. From $T = 15$ the first values seem to reach their asymptotic levels. Building a repeat-sales index on a too short interval is not always reliable; as we can see the bias can double in this kind of situation. The U-shape is always there, with a greater magnitude for the right side for all the classical situations ($T \geq 15$). We could have thought that the only problem with respect to the length of the time interval would have appeared for the small intervals; however it is not completely true. Indeed, for the long interval the errors and the bias for the more recent dates increase slightly. This phenomenon appears after $T = 40$. Consequently, our results suggest that there exist an optimal length for the estimation interval (at least under the errors and the bias point of view). For our base situation this optimal value is approximately $T = 30$.

4.4. Time structure and sensitivity to the size of the sample: $K$ and $\lambda$

In this section we study the time-structure in relation to the size of the sample. With the neutral exponential dataset the number of repeat-sales is a function of two parameters: $K$ and $\lambda$. The first one is a measure of the volume traded globally on the market at each date and the second one is associated to the length of the holding period. If $K$ increases, or $\lambda$ decreases, the number of repeat-sales is higher. Table 7a and 7b give the size of the samples when each parameter is modified separately from the base. The results for the errors and the bias are illustrated with Figures 7a, 7b for $K$, and Figures 8a, 8b for $\lambda$. What is the minimal volume of transactions required to build a reliable RSI? With our base situation ($T = 30$, $\sigma_G^2 = 0.001$, $\theta = 10$, $\lambda = 0.1$) it seems that $K = 50$ is acceptable. With this choice the standard-error of the estimators is approximately 2.5% and the bias is small. The associated sample size is $N = 1025$. If we divide $N$ by the number of dates we get: $1025/30 \approx 34$. The repeat-sales index is

16 We recall here that only a part of the $N = KT$ transactions provide observable repeat-sales; the missing part is given by the formula $\pi(T) = \lambda_{\text{ind}}(T)\alpha/(1-\alpha) = \lambda_{\text{ind}}(T)/\lambda$. (cf. 3.2.2).
acceptable if we have on average 34 repeat-sales data for each date, or equivalently if we work on a market in which we can observe 50 simple transactions at each date. This low level indicates that the RSI can be reasonably computed for a very narrow market, contrary to hedonic index that requires a bigger dataset. For \( K = 10 \) and \( N = 205 \), the limit is reached; the errors become important (6%) and the bias also. If we want to a greater accuracy (for instance 1% for the errors) we would need approximately \( K = 300 \), that is \( N \approx 5000 \) pairs. As we can see with all these simulations the bias stays at a very acceptable level, the main criterion for the analysis is actually the size of the errors. Concerning the time-structure, the results do not depend on \( K \); the U-shape stays unchanged whatever be the value of this parameter. We are now going to examine the impact of \( \lambda \). As the resale decision is exponentially distributed, the mean holding period is equal to \( 1/\lambda \) (10 units of time for the base). \( \lambda \) is a measure of the turnover rate and it seems that there exists an optimal level for this value. In Figures 8a and 8b the errors and the bias hit a low threshold for \( \lambda \) between 0.1 and 0.2; that is for a mean holding period between 5 and 10 units of time. For the base, if the turnover rate becomes lower, it deteriorates the quality of the index. On the other hand, a sample with very short holding periods is not really an advantage. Indeed, as we can see for \( \lambda = 0.5 \), the errors surprisingly increases. This type of goods, called “flips” by Clapp and Giaccotto (1999), produces higher errors and higher bias. Moreover the time structure changes, the U-shape disappears and it is replaced with an increasing line. The edge effect becomes small on the left and much more important on the right side. This result confirms once more that the flips are a very specific category. Here we demonstrated that this result holds not only because of a specific price dynamic as in Clapp and Giaccotto (1999), but also because of a structural issue related to the econometrical properties of the RSI.
4.5. Liquidity shocks

In the previous paragraph the liquidity shock is global; it acts on the full range on the interval. We are now going to consider time-localized shocks. With the exponential benchmark we assumed that at each date $K$ goods were negotiated on the market ($t = 0,\ldots, T-1$). In order to deepen the analysis we distinguish these quantities denoting $K_0,\ldots,K_{T-1}$ the volumes for these dates. Similarly, the resale intensity at $t$, that is the instantaneous resale rate for the goods still alive after $t-1$, is denoted $\lambda_t$ (for $t = 1,\ldots,T$). We assume that the same rate $\lambda_t$ is applied to all the goods alive after $t – 1$, whatever be their initial purchase date. Globally, the $K_t$ and the $\lambda_t$ will stay at the same levels than for the reference situation. A liquidity shock just consists in a variation of a few $K_t$ or a few $\lambda_t$. A change of $K_t$ means that the volume traded on the market at $t$ is smaller or greater ($K_t = 100$ or $K_t = 400$, the level of the reference is $K = 200$). A variation of $\lambda_t$ means a change in the length of the holding period: a decrease of $\lambda_t$ is associated to a lengthening of the period and reciprocally (the shocks correspond to $\lambda_t = 0.2$ and $\lambda_t = 0.05$, the level of reference being $\lambda_t = 0.1$). We consider three kinds of shocks: on a single date, on a small sub-interval in the middle of the global interval, on the right or left half-interval. For each situation, we calculate the percentages of variation for the biases and for the errors between the exponential benchmark and the shocked sample. Results are presented in the Figures 9, 10, 11, 12, 13.

4.5.1. Shock on a single date

The Figures 9 give the results when the shock (upward or downward) concerns $K_0, K_{15}$ and $K_{29}$. For all the dates between $t = 1$ and $t = 29$ the results are very intuitive and similar to the

---

17 Indeed, we could imagine some more sophisticated shocks. For instance the resale intensity could depend on the cohort, that is the purchase date $i : \lambda_{t,i}$
ones obtained for $K_{15}$ and $K_{29}$; we just observe higher biases and higher errors when $K_t$ decreases (and inversely when it increases the biases and the errors become smaller) for the date $t$. This effect is more pronounced for the first dates ($t = 1, 2, 3$), then it decreases progressively: the impact is at the lowest level for $t = 29$. However, there is a notable exception for $K_0$. The volume negotiated on the market at the first date seems to be the main factor. Indeed, if we observe some moderate variations for $K_t$ when $t > 0$, the things change drastically for $t = 0$. Not only the impact on the errors and on the bias at $t = 0$ is more important but this type of shock also affects the subsequent dates, in a very significant proportion. This strong boundary effect on the left side should invite to choose very carefully the first date when estimating an index. A bad choice, that is a date with a low level of transaction, could easily increase the bias and the error by 100% on the whole interval. The Figures 10 present the variations when the shock concerns $\lambda_1$, $\lambda_{14}$ and $\lambda_{30}$. The effect is globally higher when it concerns the most recent dates. But we do not observe the same kind of boundary effect than the one we have for $K_0$; the effect simply stays localized on the shocked date.

4.5.2. Shock on a sub-interval

Figures 11 present the effects of two shocks for four joint dates in the middle of the estimation interval. These shocks are only downward; they reduce the size of the sample. When the change affects the liquidity levels on the market $K_t$, the errors and the bias for the previous date are not modified. The effect is maximal for the shocked dates, but this perturbation also deteriorates the quality of the index for the rest of the interval. If we now study a shock on the resale rate $\lambda_t$, it also appears at its maximal level for the modified dates. But surprisingly the effect is quite null after the shock, or even slightly negative: improvement of the index
quality… For the dates before the shock, we also obtain a quite counter-intuitive result: a shock on $\lambda_t$ that takes place after $t = 12$ deteriorates the quality of the index before $t = 12$…

4.5.3. Shock on a semi-interval

The Figures 12 and 13 correspond to various shocks on a semi-interval, separately with $K_t$ and $\lambda_t$, and in a sense that decreases the size of the dataset ($K_t$ lower, $\lambda_t$ higher). The first case is quite specific; it is mainly a consequence of the shock on $K_0$. The effect is maximal on the shocked period but it also acts after. We can also note that the variations on the left side are exactly at the same level. With the second case ($K_t = 100$ for $t > 15$) we simply observe a variation on the shocked part. For the third and the fourth case (shock on the length of the holding period), the effect is maximal for the modified dates but surprisingly we also observe some modifications before and after...

5. Conclusion

The reliability of the RSI is generally an asymmetric U-shape; the estimation is more accurate in the middle of the interval compared with the edges and the left side usually provides a better estimation compared with the right side. We studied the impact of various parameters that do not depend on the specificities of each dataset but are rather some intrinsic features of the repeat-sales model. We found that in a BMN context the boundary effect is smaller, or even null. At the opposite side of the range, that is when we just have a random walk without the white noise for the residuals, we found an increasing shape: the left side provides the better estimators. Between these two poles, the classical Case-Shiller model generates an
asymmetric U-shape. When we looked at the length of the estimation interval we found that a short interval creates higher errors and an increasing framework. Then, an optimal level appears with the U-shape. Surprisingly, the large intervals are not the better ones. With the liquidity level and the resale speed we found that high levels of liquidity and quick resales produce the best situations. With a notable exception when the resale intensity is very high (flips). Indeed, in this situation the framework evolves from the asymmetric U-shape to an increasing curve and the reliability of the index globally becomes lower. This result speaks for a distinct treatment of the flips. With the sensitivity analysis developed in this paper we can define what an optimal situation is: small noises ($\sigma_G$ and $\sigma_N$), length of the interval around 30, at least 50 transactions observed at each dates, a first date with the highest possible level of transaction. Some parameters matter more than others. The most important are the liquidity levels and the resale speeds. Then, we have the length of the interval, and finally the noises.
References

Bailey, Muth, and Nourse. 1963. A Regression Method for Real Estate Price Index Construction. *Journal of the American Statistical Association* 58.

Baroni, Barthélémy, Mokrane. 2004. “Physical real estate : A Paris repeat sales residential index”. *ESSEC Working paper* DR 04007, ESSEC Research Center, ESSEC Business School

Case and Shiller. 1987. Prices of Single Family Homes since 1970: New Indexes for Four Cities. *New England Economic Review* September/October: 45-56.

Chau, Wong, Yiu, Leung. 2005. "Real estate price indices in Hong-Kong" *Journal of real estate literature* 13(3) : 337-356

Clapp, Giaccotto. 1999. “Revisions in repeat-sales price indexes: Here today, gone tomorrow?” *Real estate economics* 27(1) : 79-104

Deng, Quigley, Van Order. 2000. “Mortgage terminations, heterogeneity and the exercise of mortgage options”. *Econometrica* 68(2) : 275-308

Gatzlaff, Geltner. 1998. "A repeat-sales transaction-based index of commercial property" *A study for the real estate research institute*

Goetzmann. 1992. The Accuracy of Real Estate Indices: Repeat Sales Estimators. *Journal of Real Estate Finance and Economics* 5: 5-53.
Goetzmann, Peng. 2002. The Bias of the RSR Estimator and the Accuracy of Some Alternatives. *Real Estate Economics* 30(1): 13-39.

Kalbfleisch, Prentice. 2002. “The Statistical Analysis of Failure Time Data” *Wiley-Interscience*

Simon. 2007. “Information and repeat-sales”. Working paper, available on SSRN
**Appendix A**: Formulas for the real distribution

\[ b_i = K' \alpha + K' \alpha^2 + \ldots + K' \alpha^{T-i} = K' \alpha (1 - \alpha^{T-i})/(1 - \alpha) = K (1 - \alpha^{T-i}) = K \, d(T-i) \]

\[ s_j = K' \alpha + K' \alpha^2 + \ldots + K' \alpha^j = K' \alpha (1 - \alpha^j)/(1 - \alpha) = K (1 - \alpha^j) = K \, d(j) \]

\[ N = b_0 + \ldots + b_{T-1} = K \, (d(1) + \ldots + d(T)) = K \, (T - (\alpha + \ldots + \alpha^T)) = K \, (T - \alpha(1-\alpha^T)/(1-\alpha)) \]

\[ N = KT \, (1 - \lambda_{\text{lin}}(T)/\lambda) \approx KT \, [1 - \lambda_{\text{lin}}(T)/\lambda] \]

\[ \pi = (KT - N) / KT = \lambda_{\text{lin}}(T) \, \alpha/(1-\alpha) \approx \lambda_{\text{lin}}(T)/\lambda \]

\[ I = K' \sum_{k=1,...,T} [(T - k + 1) / (\Theta + k)] \, \alpha^k \quad \text{(sum on the diagonals)} \]

\[ = K' (T + \Theta + 1) \sum_{k=1,...,T} \alpha^k / (\Theta + k) - K' \sum_{k=1,...,T} \alpha^k \]

\[ = K' (T + \Theta + 1) \, u_T - K' (\alpha/(1-\alpha)) \ast (1 - \alpha^T) \]

\[ = K' \left[ (T + \Theta + 1) \, u_T - T \pi(T) \right] \]

\[ \mu = (K'T \ell - I) / K'T \ell \]

\[ = \left[ T \ell - (T + \Theta + 1) \, u_T + T \pi(T) \right] / T \ell \]

\[ = \left[ \ell + \pi(T) - (T + \Theta + 1) \left( u_T / T \right) \right] / \ell \]
Appendix B: Formula for $I^{[t',t+1]}$ with $t' < t$

In Figure 5a, $I^{[t',t+1]}$ corresponds to the sum of the $L_{i,j}$ within the area D. We know that the sum of the $L_{i,j}$ on the area $A + B + C + D$ is equal to $I = K' \left( (T + \Theta + 1) u_T - T \pi(T) \right)$. As the areas C, A + C, B + C are triangular they are similar to the total area I. Consequently we get very easily the expressions for the sums of the $L_{i,j}$ on these areas just adapting the formula for I. Now, if we notice that $D = I - (A + C) - (B + C) + C$, we get for $I^{[t',t+1]}$:

$$I^{[t',t+1]} = K' \left( (T+\Theta+1) u_T - (t+\Theta+1) u_t - (T-t'+\Theta) u_{T-t'-1} + (t-t'+\Theta) u_{t-t'-1} \right) = K' \left[ (1+\Theta/2)(u_T - u_t) - [(t+\Theta/2)( u_t - u_{t,t'-1}) - (t'-\Theta/2)( u_{T-t'-1} - u_{t,t'-1})] + (T+\Theta/2)( u_T - u_{T,t'-1}) \right]$$

If we introduce the notation $U(m,n) = u_n - u_m = a^{m+1}/(\Theta+m+1) + \ldots + a^n/(\Theta+n)$, for $0 \leq m \leq n$ with the convention $u_0 = 0$, the square brackets becomes:

$$(1+\Theta/2) U(t,T) - \left[ (t+\Theta/2) U(t-t'-1,t) - (t'-\Theta/2) U(t-t'-1,T-t'-1) \right] + (T+\Theta/2) U(T-t'-1,T)$$

For $0 \leq t' < t < T$ we denote this expression $\mathcal{U}(t', t, T)$. Thus, the formula for $I^{[t',t+1]}$ is:

$$I^{[t',t+1]} = (1-\alpha)/\alpha \left[ n^{[t',t+1]} + K \mathcal{U}(t', t, T) \right]$$
Appendix C: Formula for $I^t = f^{t,t+1}$

The formula established in appendix B is valid if $t' = t - 1$ but not for $t' = t$. In that case the demonstration is a little bit different because the area C does not exist and the two areas A and B are separated by a column, as illustrated in Table 5b. However, we can use the same kind of technique just writing that $D = I - A - B$.

\[
I^t = K' \left( (T+\Theta+1)u_T - \left(\alpha / (1 - \alpha)\right) (1 - \alpha^T) \right) \\
- K' \left( (t+\Theta+1)u_t - \left(\alpha / (1 - \alpha)\right) (1 - \alpha^t) \right) \\
- K' \left( (T-t+\Theta)u_{T-t-1} - \left(\alpha / (1 - \alpha)\right) (1 - \alpha^{T-t-1}) \right)
\]

The second parts give:

\[
(-K) \left[-\alpha^T + \alpha^t + \alpha^{T-t-1} - 1\right] = K \left[ 1 - \alpha^t + \alpha^1 (1 - \alpha^{T-t} - \alpha^{t+1} + \alpha^{T+1}) \right]
\]

\[
= K \left[ -(1-\alpha)^{\alpha} + \alpha^{t} d(T-t) d(t+1) \right] = -K' + (1-\alpha)^{\alpha} \ n^t
\]

We choose to let the quantity $(1-\alpha)^{\alpha} n^t$ on the right side whereas we integrate $(-K')$ in the left side. It gives:

\[
K' \left[ (T+\Theta+1)u_T - (t+\Theta+1)u_t - (T-t+\Theta)u_{T-t-1} - 1 \right]
\]

\[
= K' \left[ (T+\Theta+2+\Theta+2+1)u_T - (t+\Theta+2+\Theta+2+1)u_t - (T-t+\Theta+2+\Theta+2)u_{T-t-1} - 1 \right]
\]

\[
= K' \left[ (1+\Theta/2) (u_T - u_t) - (t+\Theta/2)u_{T-t-1} + 1 \right] + (T+\Theta/2) ( u_T - u_{T-t-1} )
\]

\[
= K' \left[ (1+\Theta/2)U(t,T) - \left(\alpha / (1 - \alpha)\right) (1 - \alpha^{T-t} - \alpha^{t+1} + \alpha^{T+1}) \right] + (T+\Theta/2)U(T-t-1,T)
\]

If we denote $u_{-1} = -1/\Theta$, we have:

\[
K' \left[ (1+\Theta/2)U(t,T) - \left(\alpha / (1 - \alpha)\right) (1 - \alpha^{T-t} - \alpha^{t+1} + \alpha^{T+1}) \right] + (T+\Theta/2)U(T-t-1,T)
\]

That is $K' \mathcal{U}(t, t, T)$, if we extend the definition for $\mathcal{U}(t', t, T)$ assuming that we can have $t = t'$. Therefore, the formula for $I^t$ is: $I^t = (1-\alpha)^{\alpha} \left[ n^t + K' \mathcal{U}(t, t, T) \right]$. 
Table 1: Relevant repeat-sales for the time interval \([t', t]\)

| Purchase date | 0 | … | t' | … | t | … | T |
|---------------|---|---|----|---|---|---|---|
| 0             |   |   |    |   |   |   |   |
| t'            |   |   |    |   |   |   |   |
| t             |   |   |    |   |   |   |   |
| T             |   |   |    |   |   |   |   |
### Table 2a: Real distribution for the repeat-sales sample

|     | 0        | 1        | 2        | 3        | … | t | t+1 | … | T–2 | T–1 | T   |
|-----|----------|----------|----------|----------|---|---|-----|---|-----|-----|-----|
| 0   | n₀,₁     | n₀,₂     | n₀,₃     | n₀,t     | n₀,t+₁ | n₀,T-2 | n₀,T-1 | n₀,T |
| 1   | n₁,₂     | n₁,₃     | n₁,t     | n₁,t+₁   | n₁,T-2 | n₁,T-1 | n₁,T   |
| 2   | n₂,₂     | n₂,₃     | n₂,t     | n₂,t+₁   | n₂,T-2 | n₂,T-1 | n₂,T   |
| 3   | n₃,₂     | n₃,₃     | n₃,t     | n₃,t+₁   | n₃,T-2 | n₃,T-1 | n₃,T   |
| t   |           |          |          |          | n₉,t+₁ | n₉,T-2 | n₉,T-1 | n₉,T |
| t+1 |           |          |          |          |        | n₉,t+₁ | n₉,T-1 | n₉,T |
| T–2 |           |          |          |          |        |        |        | n₉,t+₁ |
| T–1 |           |          |          |          |        |        |        |        | n₉,T |
| T   |           |          |          |          |        |        |        |        |        | n₉,T |

Vertical axis: purchase date  Horizontal axis: resale date

### Table 2b: Informational distribution for the repeat-sales sample

|     | 0        | 1        | 2        | 3        | … | t | t+1 | … | T–2 | T–1 | T   |
|-----|----------|----------|----------|----------|---|---|-----|---|-----|-----|-----|
| 0   | L₀,₁     | L₀,₂     | L₀,₃     | L₀,t     | L₀,t+₁ | L₀,T-2 | L₀,T-1 | L₀,T |
| 1   | L₁,₂     | L₁,₃     | L₁,t     | L₁,t+₁   | L₁,T-2 | L₁,T-1 | L₁,T   |
| 2   | L₂,₂     | L₂,₃     | L₂,t     | L₂,t+₁   | L₂,T-2 | L₂,T-1 | L₂,T   |
| 3   | L₃,₂     | L₃,₃     | L₃,t     | L₃,t+₁   | L₃,T-2 | L₃,T-1 | L₃,T   |
| t   |           |          |          |          | L₉,t+₁ | L₉,T-2 | L₉,T-1 | L₉,T |
| t+1 |           |          |          |          |        | L₉,t+₁ | L₉,T-1 | L₉,T |
| T–2 |           |          |          |          |        |        |        | L₉,t+₁ |
| T–1 |           |          |          |          |        |        |        |        | L₉,T |
| T   |           |          |          |          |        |        |        |        |        | L₉,T |

Vertical axis: purchase date  Horizontal axis: resale date
Table 3: Relevant repeat-sales for \([t',t+1]\) and quantity of information associated

|       | 0          | ... | \(t'\) | ... | \(t\) | \(t+1\) | \(T\) | Sum  |
|-------|------------|-----|--------|-----|-------|--------|------|------|
| 0     |            |     |        |     | \(L_{0,t'}\) |         |      | \(B_0^t\) |
| \(t'\)|            |     |        |     | \(L_{t',t} \) | \(L_{t',t+1}\) | \(L_{t',T}\) | \(B_{t'}^t\) |
| \(t\) |            |     |        |     | \(L_{t,t+1}\) | \(L_{t',T}\) | \(L_{t',T}\) | \(L_{l',T}\) |
| \(T\) |            |     |        |     | \(L_{0,T}\) | \(L_{0,T} \) | \(L_{0,T}\) | \(L_{0,T}\) |

\[ B_0^t = L_{0,t+1} + \ldots + L_{0,T} \quad B_{t'}^t = L_{t',t+1} + \ldots + L_{t',T} \quad \text{sum on the lines (buy-side)} \]

\[ S_{t+1}^t = L_{0,t+1} + \ldots + L_{t',t+1} \quad S_T^t = L_{0,T} + \ldots + L_{t',T} \quad \text{sum on the columns (sell-side)} \]

\[ I_{[t',t+1]} = B_0^t + \ldots + B_{t'}^t = S_T^t + \ldots + S_{t+1}^t \]
Table 4a: The real distribution for the benchmark sample

|   | 0          | 1       | 2       | ... | t       | t + 1   | ... | T   | b_i |
|---|------------|---------|---------|-----|---------|---------|-----|-----|-----|
| 0 | K'α       | K'α^2   | K'α^{t-1} | ... | K'α^{t+1} | K'α^t   | ... | K'α^{T-1} | K (1 – α^t) |
| 1 | K'α       | K'α^{t-1} | K'α^t   | ... | K'α^{T-1} | K (1 – α^{T-1}) |
| 2 | K'α^{t-2} | K'α^{t-1} | K'α^{T-2} | ... | K (1 – α^{T-2}) |
|   | ...       | ...     | ...     | ... | ...     | ...     | ... | ... | ... |
| t | K'α       | K'α^{T-1} | K (1 – α^{T-1}) |
| t + 1 | ... | ...     | ...     | ... | ...     | ...     | ... | ... | ... |
| T – 1 | ... | ...     | ...     | ... | ...     | ...     | ... | ... | ... |
| T | ... | ...     | ...     | ... | ...     | ...     | ... | ... | ... |
| s_j | K(1–α) | K(1–α^t) | K(1–α^{t+1}) | ... | K(1–α^T) | N |

Table 4b: The informational distribution for the benchmark sample

|   | 0          | 1       | 2       | ... | t       | t + 1   | ... | T   | B_i |
|---|------------|---------|---------|-----|---------|---------|-----|-----|-----|
| 0 | K'/((θ+1) | K'/((θ+2) | K'/((θ+t) | ... | K'/((θ+t+1) | K' (θ+T) | K' u_T |
| 1 | K'/((θ+1) | K'/((θ+t-1) | K'/((θ+t) | ... | K'/((θ+T-1) | K' u_{T-1} |
| 2 | K'/((θ+t-2) | K'/((θ+t-1) | K'/((θ+T-2) | ... | K' u_{T-2} |
|   | ...       | ...     | ...     | ... | ...     | ...     | ... | ... | ... |
| t | K'/((θ+1) | K'/((θ+T-t) | K' u_{T-t} |
| t + 1 | ... | ...     | ...     | ... | ...     | ...     | ... | ... | ... |
| T – 1 | ... | ...     | ...     | ... | ...     | ...     | ... | ... | ... |
| T | ... | ...     | ...     | ... | ...     | ...     | ... | ... | ... |
| S_j | K' u_1 | K' u_2 | K' u_1 | ... | K' u_{T-1} | K' u_T | I |

\[ u_n = \frac{\alpha}{(\Theta + 1)} + \frac{\alpha^2}{(\Theta + 2)} + \frac{\alpha^3}{(\Theta + 3)} + ... + \frac{\alpha^n}{(\Theta + n)} \]
Table 5a: relevant information for \([t', t+1], t' < t\)

|       | 0 | 1 | ... | t' | t'+1 | ... | ... | t | t+1 | ... | T |
|-------|---|---|-----|----|------|-----|-----|---|-----|-----|---|
| 0     |   |   |     |    |      |     |     |   |     |     |   |
| 1     |   |   |     |    |      |     |     |   |     |     |   |
| t'    |   |   |     |    |      |     |     |   |     |     |   |
| t'+1  |   |   |     |    |      |     |     |   |     |     |   |
| t     |   |   |     |    |      |     |     |   |     |     |   |
| t+1   |   |   |     |    |      |     |     |   |     |     |   |
| T     |   |   |     |    |      |     |     |   |     |     |   |

Table 5b: relevant information for \([t, t+1]\)

|       | 0 | 1 | ... | t | t+1 | ... | T |
|-------|---|---|-----|---|-----|-----|---|
| 0     |   |   |     |   |     |     |   |
| 1     |   | A |     |   |     |     | D |
| t     |   |   |     | t |     |     |   |
| t+1   |   |   |     |   |     |     |   |
| T     |   |   |     |   |     |     |   |

Diagram A: Relevant information for interval \([t', t+1], t' < t\).
Diagram B: Relevant information for interval \([t, t+1]\).
Table 6: Times of noise equality in Case, Shiller (1987)

| City     | Atlanta | Chicago | Dallas | San Francisco |
|----------|---------|---------|--------|---------------|
| $\Theta$ | 12.89   | 9.11    | 6.77   | 4.20          |
Table 7a: Size of the samples (Figure 7a and 7b)

| K  | 10  | 50  | 100 | 200 | 500 | 1000 | 2000 | 5000 | 10000 |
|----|-----|-----|-----|-----|-----|------|------|------|-------|
| N  | 205 | 1025| 2050| 4100| 10249| 20498| 40996| 102489| 204979 |

Table 7b: Size of the samples (Figure 8a and 8b)

| λ  | 0,01 | 0,02 | 0,05 | 0,1  | 0,2  | 0,5  |
|----|------|------|------|------|------|------|
| N  | 816  | 1488 | 2893 | 4100 | 5002 | 5600 |
Figure 1: Algorithmic decomposition of the repeat-sales index

1. Sample of estimation

2. Distributions of the purchase and the resale prices $h_{p}^{(i,j)}$ and $h_{f}^{(i,j)}$

3. Real distribution $\{n_{i,j}\}$

4. Informational distribution $\{L_{i,j}\}$

5. Mean of the holding periods $\tau^{t}$

6. Time of noise equality $\Theta$

7. Mean purchase and resale prices for $[t,t+1]$ $H_{p}(t)$ and $H_{f}(t)$

8. Mean of the mean rates $\rho_{t}$

9. Matrixes $\hat{I}$ and $\eta$

10. Index $\hat{I} \ R = \eta \ P$
Legend of the Figure 1

1. \( n_{i,j} \): Number of the repeat-sales with a purchase at \( t_i \) and a resale at \( t_j \), organized in an upper triangular table.

2. Estimation of the volatilities \( \sigma_N \) and \( \sigma_G \) for the white noise and the random-walk (step 1 and 2 of the Case-Shiller procedure). The time of noise equality is \( \Theta = 2\sigma_N^2 / \sigma_G^2 \).

3. \( L_{i,j} = n_{i,j} / (\Theta + j - i) \): Quantity of information delivered by the \( n_{i,j} \) repeat-sales of \( C(i,j) \). These numbers are also organized in an upper triangular table.

4. We get the matrix \( \hat{\mathbf{I}} \) from the informational distribution of the \( \{ L_{i,j} \} \) summing for each time interval \([t',t+1]\) the relevant \( L_{i,j} \), that is the ones with an holding period that includes \([t',t+1]\). The diagonal elements of the diagonal matrix \( \mathbf{I} \) are equal to the sums (lines or columns indifferently) of the components of the matrix \( \hat{\mathbf{I}} \).

5. Dividing the diagonal elements of \( \hat{\mathbf{I}} \) by the diagonal elements of \( \mathbf{I} \) we obtain directly the mean holding periods \( \tau^i \).

6. For each repeat-sales class \( C(i,j) \), the geometric averages of the purchase prices \( h_p^{(i,j)} \), and the resale prices \( h_r^{(i,j)} \) are:

\[
 h_p^{(i,j)} = \left( \prod_{k'} p_{k',i} \right)^{1/n_{i,j}} \\
 h_r^{(i,j)} = \left( \prod_{k'} p_{k',j} \right)^{1/n_{i,j}}
\]

7. For the subset of the people who owned real estate during \([t,t+1]\), that is \( S_{\text{pl}} \), the mean purchase price \( H_p(t) \) (the mean resale price \( H_r(t) \)) is the geometric average of the \( h_p^{(i,j)} \) (respectively the \( h_r^{(i,j)} \)), weighted by the \( L_{i,j} \), for all the relevant repeat-sales classes:

\[
 H_p(t) = \left( \prod_{1 \leq t < j} \left( h_p^{(i,j)} \right)^{L_{i,j}} \right)^{1/1^t} \\
 H_r(t) = \left( \prod_{1 \leq t < j} \left( h_r^{(i,j)} \right)^{L_{i,j}} \right)^{1/1^t}
\]

The mean of the mean rates \( \rho_t \) realised by the people who owned real estate during \([t,t+1]\), can be calculated as a return rate with the fictitious prices \( H_p(t) \) for the purchase \( H_r(t) \) for the resale, and the fictitious holding period \( \tau^i \)

\[
 \rho_t = \left( 1 / \tau^i \right) * \ln \left[ H_r(t) / H_p(t) \right]
\]

8. The vector \( \mathbf{R} \) of the monoperiodic growth rates of the index is the solution of the equation:

\[
 \hat{\mathbf{I}} \mathbf{R} = \mathbf{I} \mathbf{P} \Leftrightarrow \mathbf{R} = ( \hat{\mathbf{I}}^{-1} \mathbf{I} ) \mathbf{P}
\]

\( \mathbf{P} \) is the vector \((\rho_0, \rho_1, \ldots, \rho_{T-1})\).
Figure 2: standard hazard rate

Phase 1: Quick resales are scarce
Phase 2: Residential time
Phase 3: Ageing
Figure 3: Notations

\[ t \rightarrow e^{-\lambda t} \]

Slope = \( \lambda_{\text{lin}}(k) \)
Figure 4a and 4b: Sensitivity to $\sigma_G^2$

**Sensitivity of the standard error to $\sigma_G^2$**

| Time | $\sigma_G^2 = 0.0001$ | $\sigma_G^2 = 0.0005$ | $\sigma_G^2 = 0.001$ | $\sigma_G^2 = 0.0015$ | $\sigma_G^2 = 0.002$ | $\sigma_G^2 = 0.0025$ |
|------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 1    | 0.005                  | 0.005                  | 0.005                  | 0.005                  | 0.005                  | 0.005                  |
| 2    | 0.01                   | 0.01                   | 0.01                   | 0.01                   | 0.01                   | 0.01                   |
| 3    | 0.015                  | 0.015                  | 0.015                  | 0.015                  | 0.015                  | 0.015                  |
| 4    | 0.02                   | 0.02                   | 0.02                   | 0.02                   | 0.02                   | 0.02                   |
| 5    | 0.025                  | 0.025                  | 0.025                  | 0.025                  | 0.025                  | 0.025                  |
| 6    | 0.03                   | 0.03                   | 0.03                   | 0.03                   | 0.03                   | 0.03                   |

**Sensitivity of the bias to $\sigma_G^2$**

| Time | $\sigma_G^2 = 0.0001$ | $\sigma_G^2 = 0.0005$ | $\sigma_G^2 = 0.001$ | $\sigma_G^2 = 0.0015$ | $\sigma_G^2 = 0.002$ | $\sigma_G^2 = 0.0025$ |
|------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 1    | 0.005                  | 0.005                  | 0.005                  | 0.005                  | 0.005                  | 0.005                  |
| 2    | 0.01                   | 0.01                   | 0.01                   | 0.01                   | 0.01                   | 0.01                   |
| 3    | 0.015                  | 0.015                  | 0.015                  | 0.015                  | 0.015                  | 0.015                  |
| 4    | 0.02                   | 0.02                   | 0.02                   | 0.02                   | 0.02                   | 0.02                   |
| 5    | 0.025                  | 0.025                  | 0.025                  | 0.025                  | 0.025                  | 0.025                  |
| 6    | 0.03                   | 0.03                   | 0.03                   | 0.03                   | 0.03                   | 0.03                   |
Figure 5a and 5b: Sensitivity to θ

**Sensitivity of the standard error to Theta**

- Standard error %
- Time

**Sensitivity of the bias to Theta**

- Bias %
- Time
Figure 6a and 6b: Sensitivity to T

**Sensitivity of the standard error to T**

- Time (x-axis): 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49
- Standard error % (y-axis): 0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8

- Lines for different T values:
  - T = 5
  - T = 10
  - T = 15
  - T = 20
  - T = 30
  - T = 40
  - T = 50

**Sensitivity of the bias to T**

- Time (x-axis): 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49
- Bias % (y-axis): 0, 0.002, 0.004, 0.006, 0.008, 0.01, 0.012, 0.014, 0.016

- Lines for different T values:
  - T = 5
  - T = 10
  - T = 15
  - T = 20
  - T = 30
  - T = 40
  - T = 50
Figure 7a and 7b: Sensitivity to K

**Sensitivity of the standard error to K**

| Time | K = 10 | K = 50 | K = 100 | K = 200 | K = 500 | K = 1000 | K = 2000 | K = 5000 | K = 10000 |
|------|--------|--------|---------|---------|---------|----------|---------|---------|-----------|
|                   | 0.00   | 0.00   | 0.00    | 0.00    | 0.00    | 0.00     | 0.00    | 0.00    | 0.00      |
| 1     | 0.01   | 0.02   | 0.03    | 0.04    | 0.05    | 0.06     | 0.07    | 0.08    | 0.09      |
| 2     | 0.02   | 0.04   | 0.06    | 0.08    | 0.10    | 0.12     | 0.14    | 0.16    | 0.18      |
| 3     | 0.03   | 0.06   | 0.09    | 0.12    | 0.15    | 0.18     | 0.21    | 0.24    | 0.27      |
| 4     | 0.04   | 0.08   | 0.12    | 0.16    | 0.20    | 0.24     | 0.29    | 0.36    | 0.42      |
| 5     | 0.05   | 0.10   | 0.15    | 0.20    | 0.25    | 0.30     | 0.36    | 0.45    | 0.55      |
| 6     | 0.06   | 0.12   | 0.18    | 0.25    | 0.31    | 0.38     | 0.46    | 0.58    | 0.72      |
| 7     | 0.07   | 0.14   | 0.21    | 0.29    | 0.37    | 0.47     | 0.60    | 0.80    | 1.00      |
| 8     | 0.08   | 0.16   | 0.24    | 0.34    | 0.45    | 0.59     | 0.85    | 1.15    | 1.50      |
| 9     | 0.09   | 0.18   | 0.27    | 0.39    | 0.51    | 0.75     | 1.10    | 1.60    | 2.10      |
| 10    | 0.10   | 0.20   | 0.30    | 0.42    | 0.55    | 0.85     | 1.25    | 1.80    | 2.50      |

**Sensitivity of the bias to K**

| Time | K = 10 | K = 50 | K = 100 | K = 200 | K = 500 | K = 1000 | K = 2000 | K = 5000 | K = 10000 |
|------|--------|--------|---------|---------|---------|----------|---------|---------|-----------|
|                   | 0.00   | 0.00   | 0.00    | 0.00    | 0.00    | 0.00     | 0.00    | 0.00    | 0.00      |
| 1     | 0.01   | 0.02   | 0.03    | 0.04    | 0.05    | 0.06     | 0.07    | 0.08    | 0.09      |
| 2     | 0.02   | 0.04   | 0.06    | 0.08    | 0.10    | 0.12     | 0.14    | 0.16    | 0.18      |
| 3     | 0.03   | 0.06   | 0.09    | 0.12    | 0.15    | 0.18     | 0.21    | 0.24    | 0.27      |
| 4     | 0.04   | 0.08   | 0.12    | 0.16    | 0.20    | 0.24     | 0.29    | 0.36    | 0.42      |
| 5     | 0.05   | 0.10   | 0.15    | 0.20    | 0.25    | 0.30     | 0.36    | 0.45    | 0.55      |
| 6     | 0.06   | 0.12   | 0.18    | 0.25    | 0.31    | 0.38     | 0.46    | 0.58    | 0.72      |
| 7     | 0.07   | 0.14   | 0.21    | 0.29    | 0.37    | 0.47     | 0.60    | 0.80    | 1.00      |
| 8     | 0.08   | 0.16   | 0.24    | 0.34    | 0.45    | 0.59     | 0.85    | 1.15    | 1.50      |
| 9     | 0.09   | 0.18   | 0.30    | 0.42    | 0.55    | 0.85     | 1.25    | 1.80    | 2.10      |
| 10    | 0.10   | 0.20   | 0.30    | 0.42    | 0.55    | 0.85     | 1.25    | 1.80    | 2.50      |
Figure 8a and 8b: Sensitivity to $\lambda$

Sensitivity of the standard error to $\lambda$

Sensitivity of the bias to $\lambda$
Figures 9: liquidity shocks on a single date for the volumes $K_t$ ($t = 0, 15, 29$)
Figures 10: liquidity shocks on a single date for the resale intensities $\lambda_t$ ($t = 0, 15, 29$)
Figures 11: liquidity shocks on a sub-interval

\[ K_{13} = K_{14} = K_{15} = K_{16} = 100 \]

\[ \lambda_{12} = \lambda_{13} = \lambda_{14} = \lambda_{15} = 0.05 \]
Figures 12: Liquidity shocks on semi-interval for $K_t$
Figures 13: liquidity shocks on a semi-interval for $\lambda_t$.