

The University of Hong Kong Real Estate Index Series (HKU-REIS)

Index Construction Method for The University of Hong Kong All Residential Price Index (HKU-ARPI)

and

the following Sub-Regional Residential Price Indices:

The University of Hong Kong Hong Kong Island Residential Price Index (HKU-HRPI)

The University of Hong Kong Kowloon Residential Price Index (HKU-KRPI)

The University of Hong Kong New Territories Residential Price Index (HKU-NRPI)

Professor K W Chau
Chair of Real Estate and Construction
Department of Real Estate & Construction
The University of Hong Kong



VERSITECH LTD.
The University Technology Transfer Company

Version 1.26
September 2007

Please cite this article as:

K W Chau (2006) *Index Construction Method for The University of Hong Kong All Residential Price Index*, Versitech Limited, The University of Hong Kong, Hong Kong.

Copyright

All materials in this document are protected by copyrights, trademarks and/or other intellectual property rights. Versitech owns all rights, title and interest in and to these materials or these material is used with the permission of their owners. No part of this document may be copied, reproduced, republished, uploaded, posted, transmitted or distributed in any way for commercial purposes without the prior written consent of Versitech.

Introduction

The University of Hong Kong All Residential Price Index (HKU-ARPI) is a monthly real estate price index that tracks the changes in the general price level of residential properties in Hong Kong over time. The index is constructed based on actual transactions of completed private residential properties registered with the Hong Kong SAR Government (the Land Registry). The index covers the entire Hong Kong Special Administrative Region and is a weighted average of sub-indices for three sub-regions in Hong Kong, namely The University of Hong Kong Hong Kong Island Residential Price Index (HKU-HRPI), The University of Hong Kong Kowloon Residential Price Index (HKU-KRPI), The University of Hong Kong New Territories Residential Price Index (HKU-NRPI). The weights reflect the market value of the total stock of residential units in each of the three sub-regions. The indices are constant quality indices that are designed to capture only changes in price levels over time. The method of index construction is based on a set of mechanical computation and sampling rules that are transparent. Given the same set of data, a third party could replicate the index values by following the same set of computation and sampling rules. These rules are designed with the following objectives:

- To adjust for variation in the quality of the transacted properties;
- To eliminate transaction records that deviate substantially from market prices such as non-arm's length transactions¹, special deals² and data entry error.
- To minimize the index revision problem³, i.e. to ensure that the published index in any month is robust and not affected by future transactions.

Repeat Sales Real Estate Price Index

The HKU-ARPI and its component sub-regional indices (hereafter the Indices) are constructed based upon a modified repeat sales methodology⁴. The original repeat sales model was first proposed by Bailey, Muth and Nourse⁵ in 1963. Since all properties are unique, constructing a constant quality index based on the transaction prices of different properties is a challenge. The repeat sales model is based on the idea that attributes, and therefore the quality, of a property remain largely the same over time, so that the difference in the transaction price of identical properties during different periods, represents

¹ Examples include; transfer of title as gifts to a third party or sale of properties to family members at a substantially discounted price.

² Some properties may be transacted at extremely low or high prices due to special conditions included in the sales and purchase agreement, such as escape clauses or special financial arrangements.

³ Index revision refers to difference the index values at time t that are computed based on transaction data up to time T and that based on transaction data up to T+1. Index revision is a potential problem that exists for all transaction based real estate price indices. The nature and magnitude of the difference varies across markets. Index revision imposes psychological but not practical issues in using the index for contract settlement purposes, since on certain occasions after settlement, the contracting parties may find that the index value that they had settled on has changed.

⁴ A "repeat sale" can be defined as more than one pair of transactions of the same property within the observation period.

⁵ Bailey, M.J., Muth, R.F. and Nourse, H.O. (1963). "A regression model for real estate price index construction," Journal of the American Statistical Association, 58(304), 933-942.



pure price level changes ⁶. Given a large number of pairs of transactions of the same property over an observation period, we can construct a price index that tracks the general changes in price levels of properties during this period. The following hypothetical example (Table I below) illustrates the idea:

Property A was transacted in 2000 and 2001 at prices of HKD\$10 million and HKD\$11 million respectively. This pair of transactions implies that its price has increased by 10% during 2000-2001. Similarly transactions of Property B imply that its price has increased by 15% during 2001-2002. The price index derived from the transaction prices of property A and property B can therefore be determined. However, during the same period, there are other repeat sales pairs that may imply a different change in price levels. For example, Property C implies that the cumulative growth rate during 2000-2002 is 25% as opposed to a cumulative growth of 27%, implied by Property A and B. The index, being a market wide indicator represents a 'compromise' with growth rates somewhere between those implied by Properties A, B and C. The repeat sales model uses a statistical technique known as regression analysis to find the set of changes in price levels that best fit the changes in price levels implied by all the repeat sales pairs. Appendix I shows a formal derivation of the repeat sales model.

Table I: Illustration of the repeat sale index

	2000	2001	2002
Property A	\$10M	\$11M	
Property B		\$20M	\$23M
Property C	\$10M		\$12.5M
Index (AB)	100	110	127
Growth (AB)		10%	15%
Index (C)	100		125
Growth (C)		12%	12%
Growth (ABC)		11%	13%
Index (ABC)	100	111	126

⁶ Another commonly used method is a hedonic based property price index. This method controls quality constant by including all the quality variables (pricing influencing property attributes such as floor level, age, size etc) in the hedonic price model. This method is however not suitable for Hong Kong since much of the price influencing information is lacking. Even if we had data on all property attributes, the hedonic based property index construction method also suffers from a lack of prior knowledge about the functional form of the hedonic price model.



Data source

Property transaction data is obtained in electronic form from a data provider EPRC. In Hong Kong, all property transactions have to be registered with the Land Registry. All registered transactions have been recorded in electronic form since July 1991. The EPRC has purchased all electronic transaction data from the Land Registry and has added to the database other attributes that are not in the Land Registry database such as gross floor area, net floor area and district codes etc. We use the following information from the EPRC database to construct the Indices:

- Property types (e.g. Residential, Office, Industrial, and Commercial)
- Property addresses (including district code, street name, building name, floor level, and flat number)
- Occupation Permit dates (on which the building is completed)
- Transaction prices (in millions of Hong Kong dollars)
- Transaction dates (on which the Sale and Purchase Agreement is signed)
- Transaction instrument (transaction type): Sale and Purchase agreement, Assignment, Sub-sale.

We extract from EPRC the records that match the following criteria:

- Property type: Residential
- Transaction instrument: Sale and Purchase Agreement
- Transaction price: larger than zero but smaller than \$100 million
- Transaction date: after building completion

The first two criteria are trivial. The third criterion is needed to exclude non-arm's length transactions and bulk (or multi-asset) purchases. The last criterion excludes all pre-sales transactions. We also remove records that are duplicate or clearly erroneous. The resulting dataset contains approximately 1.5 million transactions of private residential properties taking place between 1st July 1991 and 31st July 2006.

Address Matching

The next step is to identify repeat sales pairs from the dataset. Two transaction records are said to be a repeat sales pair if both have exactly the same address. Since the EPRC data does not always have a consistent format for addresses (e.g. "Garden" is sometimes expressed as "Gdn" and "A, B" as "A&B"), manual checks are undertaken to standardize the format of all addresses. The records with the same address are then matched into pairs and sorted to form a repeat sales dataset. This is shown in Figure 1 below. After address matching, there are approximately 430,000 pairs of repeat transactions of completed private residential properties from the period 1st July 1991 to 31st July 2006.

Figure 1: Example of creation of repeat sales database from the EPRC records.

Records in the EPRC database			
	Property address	Sale price (\$m)	Sale date
...			
Record 10	Flat A, 10/F, AA Building, AA Street, Kowloon	3.5	1/3/1999
...			
Record 25	Flat A, 10/F, AA Building, AA Street, Kowloon	2.8	26/11/2004
...			

↓

Repeat sales dataset					
	Property address	1 st sale price	2 nd sale price	1 st sale date	2 nd sale date
...					
Pair 5	Flat A, 10/F, AA Building, AA Street, Kowloon	3.5	2.8	1/3/1999	26/11/2004
...					

Regional Segregation

Since properties in different locations may follow different price trends, different sub-indices need to be estimated for different locations. The repeat sales dataset is therefore segregated into three sub-regional datasets, namely Hong Kong Island (HKI), Kowloon (KL) and New Territories (NT)⁷. The repeat sales model is applied to each region to estimate the corresponding sub-regional indices, i.e. HKU-HRPI, HKU-KRPI and HKU-NRPI. The territory-wide residential price index (i.e. HKU-ARPI) is then compiled as a weighted average of the three regional sub-indices. The weighting scheme is discussed later under the section "Weighted Average". After regional segregation, there are approximately 140,000, 100,000, and 190,000 pairs of repeat sales transactions in HKI, KL, and NT, respectively.

Outlier Removal

There are a small number of outliers⁸ in the sample. They are likely to be a result of

- 1) data entry error in the original EPRC database,
- 2) non-arm's length transactions⁹, and
- 3) special deals¹⁰

⁷ A finer classification of districts is possible. But since the aggregate market value (i.e. the weight) of properties is not available for all districts, we leave this application for future development.

⁸ Outliers can be defined as a repeat sales pair that deviates significantly from the market price trend.

⁹ See footnote 1

¹⁰ See footnote 2



These need to be removed from the dataset before computation of the Indices. Since a priori information to identify outliers is unavailable, they are determined iteratively by examining the residuals of a preliminary regression of the repeat sales model¹¹. An observation is regarded as an outlier if the magnitude of its residual (in absolute terms) is larger than one¹². These outliers constitute less than 1% of observations in the repeat sales dataset.

Removing Transactions with Very Short Holding Period

A practical problem of transaction-based constant quality property indices is revision (i.e. previous index values change as new transactions become available.) For example, revision occurs when index values derived from data up to time T , $\{P_{1t}, \dots, P_{Tt}\} / I_{Tt}$, are different from index value derived data up to time $T+1$, $\{P_{1t}, \dots, P_{T+1t}\} / I_{T+1t}$. For the repeat sales model, revision is inevitable as all index values are re-estimated when more transaction data is available over time. Some studies have further found that revisions are mostly downward. That is, index values are found to be lower than previously estimated index values when more transaction data becomes available. Such examples are studies by Clapham, et al (2006)¹³ and Clapp and Giaccotto (1999)¹⁴ using European and US data respectively. We found a similar pattern in Hong Kong when the indices are estimated using the same conventional repeat sales approach¹⁵.

Since substantial revisions are undesirable for derivatives transactions, it is necessary to minimize revisions as much as possible. Consistent with Clapp and Giaccotto¹⁶, we found that in Hong Kong, transactions with very short holding periods, or flips (i.e. properties bought and sold in quick succession), are the major source of revisions. When these 'flips' are removed, revisions are substantially reduced and the index does not show any systematic revision. Sensitivity analysis further suggests that removing data with holding periods of less than four months produces the best results. The data sample used for computing the property price indices are therefore restricted to transactions with holding periods of at least four months. The removal of the repeat sales pairs with holding periods of less than four months results in a further reduction of sample size by approximately 15%. In addition, removing flips, also reduces the possibility of market price manipulation by speculators towards settlement dates.

¹¹ The model finds a set of index values that best fit the observed pairs of repeat sales transactions and is expressed formally as equation (6) in the Appendix. Also see section on "Regression" below.

¹² Since the repeat sales index is a "compromised average" that best fits the observed repeat sales pairs, most observations (repeat sales pairs) deviate from the index at different degrees. The "residual" of each observation reflects the signed deviation of the change in price level of the subject property from the index (a positive value indicates a higher than average price growth while a negative value indicates a lower than average price growth). A large residual (in absolute terms) suggests that the change in the price level implied by the repeat sales pair differs substantially from market average. The magnitude of residual in excess of one means that the price of the property has grown more / less than the market average by 100 percentage points.

¹³ Clapham, E., Englund, P., Quigley, J.M. and Redfearn, C.L. (2006). "Revisiting the past and settling the score: index revision for house price derivatives," *Real Estate Economics*, 34(2), 275-302.

¹⁴ Clapp, J. and Giaccotto, C. (1999). "Revisions in repeat-sales indexes: here today, gone tomorrow?" *Real Estate Economics*, 27(1), 79-104.

¹⁵ Downward bias also exists in Hong Kong using Case and Shiller's (1989) weighted least squares method of computing the repeat sales index.

¹⁶ *ibid.*



Rolling Window

The EPRC data does not include information on changes in property attributes (e.g. major renovation, change in internal layouts and environmental changes) and hence these quality changes cannot be explicitly accounted for in the repeat sales model. This is one of the problems of the repeat sales index. However, this problem is much less serious in Hong Kong for several reasons. First, most residential units in Hong Kong are apartment units and substantial remodeling is prohibited by building regulations. Second, land prices in Hong Kong are high: a relatively high proportion of the market value of the property is attributable to the land and a relatively low proportion is attributed to the building structure. The effect of renovation on property value is small compared with the case of a detached house in a sub-urban area where the land price is relative cheap. Third, the holding period for the repeat sales pairs in Hong Kong is relatively short (average holding period is less than 3 years). The probability of a substantial change in price influencing attributes is therefore small. Fourth, the private residential market in Hong Kong is actively transacted. The large volume of repeat sales data a permits us to ignore transaction with long holding periods (prone to change in price influencing attributes) but still retain a sufficiently large volume of data for estimation of the repeat sales index.

To reduce potential bias due to omission of unmeasured quality changes, Case and Shiller¹⁷ propose to down-weight repeat pairs (i.e. lower their influence on the repeat sales index) with significantly longer holding periods because the probability of changes in quality (attributes) increases as holding period increases. This solution, however, is not suitable for Hong Kong¹⁸. In view of the large sample size in Hong Kong, we simply discard transactions with long holding periods (longer than 10 years). In addition, we believe that transactions that took place a long time ago are of little value in estimating the recent change in price levels. We therefore further exclude remote observations whose first sale took place more than 10 years ago. This sampling restriction results in a 10-year rolling estimation of the repeat sales model. That is, as time goes by, new transactions will be added and old transactions removed from the sample data set.

Regression

The time coefficient a_t in the following repeat sales model is estimated using repeat sales pairs in each sub-region (with the above sampling restrictions imposed),

$$\ln \frac{P_{i,t_2}}{P_{i,t_1}} = \sum_{t=1}^T a_t \cdot D_{i,t} + \ln \epsilon_i$$

¹⁷ Case, K.E. and Shiller, R.J. (1989). "The efficiency of the market for single-family homes," American Economic Review, 79(1), 125-137.

¹⁸ Case and Shiller's weighted least squares approach down weights repeat sales pairs with long holding periods and therefore increases the weights of repeat sales pairs with short holding periods. This approach may aggravate the index revision problem as we have found that that in Hong Kong, repeat sales pairs with very short holding periods have caused index revision.



where the dependent variable is the natural logarithm of the ratio of the second to first sales prices of each repeat sales pair. The independent variable (Di,T) is a series of time indicators which equal -1 (when first sale took place), +1 (when second sale took place) or 0 (any other time). These time indicators take “month” as the unit of measurement so as to construct a monthly price index. The time coefficient at determines the index value, which is estimated using regression analysis techniques¹⁹. The resultant index is found by converting the time coefficients to index values. All indices are scaled to have the same value of 100 in January 2000, which is chosen as the common base period (see Appendix I for detail).

Weighted Average

The residential price indices for the three sub-regional areas are aggregated to form the market-wide residential price index using the following weighted average formula:

$$HKU - ARPI_t = \left(w_H \frac{HKU - HRPI_t}{HKU - HRPI_{10}} + w_K \frac{HKU - KRPI_t}{HKU - KRPI_{10}} + w_N \frac{HKU - NRPI_t}{HKU - NRPI_{10}} \right) \times HKU - ARPI_{10}$$

Where w_H , w_K and w_N represent the weightings for Hong Kong Island, Kowloon and the New Territories respectively. To ensure that the HKU-ARPI tracks capital returns of a fully diversified portfolio of residential properties in Hong Kong, the weights should reflect the relative size of the market value of the stock of residential properties in each sub-region. We have used total ratable values of completed residential units in each sub-region as proxies for market value weights. The Rating and Valuation Department have the most comprehensive records of the stock of completed residential units in Hong Kong. Those ratable values are closely correlated with market values, particularly at the aggregate level. The latest ratable values (2005) are used as weights (rounded off to the nearest percentage point) in computing the HKU-ARPI. Table 2 below shows the sub-regional index weights based on ratable values in 2005. In the future, the weights will be updated when new ratable values become available, usually on an annual basis. The weights will be updated in the first month of the year that follows the year of publication of the new ratable values. Since the weights only change slowly from year to year, the updates should have little immediate effect on the index.

Table 2: Percentage ratable value of all completed residential units (2006)

	Weights
Hong Kong Island	38.5%
Kowloon	26.5%
New Territories	35.0%
Total	100%

Source: Rating and Valuation Department, HKSAR Government

¹⁹ A technique that minimizes the sum of the squares of the deviations of the observations from the estimated index trend.



Index Update

As new transaction data arrives each month, the above steps are repeated. The index values published earlier will not be updated regardless of any revision. The average magnitude of revision within a 2 year period is less than 0.5 percentage point. The new index value at the latest period will be chained to the published index based on the percentage in the last period. For example, given information up to time T , the published index values are $\{\bar{P}_1, \dots, \bar{P}_T\}$. At $T+1$, the estimated index values are $\{P_1, \dots, P_T, P_{T+1}\}$. With all previously published index values retained, the new index value is calculated by:

$$\bar{P}_{T+1} = \frac{P_{T+1}}{P_T} \times \bar{P}_T \quad (10)$$

Appendix I - The Repeat Sales Model.

Definitions

Let r_t be the single period return of a portfolio of property assets from time $t-1$ to t ($t=0, \dots, T$) and P_t be the portfolio price at time t . By definition, the gross return, $1+r_t$, is the ratio of portfolio prices in two consecutive periods:

$$\frac{P_1}{P_0} = 1 + r_1, \quad \frac{P_2}{P_1} = 1 + r_2, \quad \dots, \quad \frac{P_T}{P_{T-1}} = 1 + r_T$$

The return defined here captures price changes before the deduction of any transaction cost (e.g. brokerage fees, legal fees, and stamp duty). Rental income is explicitly excluded because reliable rental data is generally unavailable.

For multiple periods, the price ratio from time 0 to t can be written as the product of a series of gross returns:

$$\frac{P_t}{P_0} = (1 + r_1)(1 + r_2) \dots (1 + r_t)$$

Or simply

$$\frac{P_t}{P_0} = 1 + cr_t \quad (1)$$

where cr_t is the cumulative return of the property portfolio from time 0 to t , with $cr_0=0$

It follows from (1) that for any two periods, t_1 and t_2 (where $T \geq t_2 > t_1 \geq 0$), the gross cumulative return is:

$$\frac{P_{t_2}}{P_{t_1}} = \frac{1 + cr_{t_2}}{1 + cr_{t_1}} \quad (2)$$

Subtracting one from both sides of the equation gives the cumulative return from t_1 to t_2 . As shown later, the expression in (2) proves very useful in understanding the repeat sales model in which P_{t_1} and P_{t_2} are regarded as the prices of repeat transactions of a property asset.

Market Portfolio Assumption

Assuming that the property portfolio is representative of the property market, the cumulative return in (2) can be interpreted as the *general change in price levels of the property market* (simply called the *property market return*) from t_1 to t_2 , which is exactly what a property price index is attempting to measure. The rationale is similar to the Hang Seng Index⁷ which measures Hong Kong's stock market returns by tracking the prices of a representative portfolio of stocks over time.

The above method works well for the stock market because the frequent trading of the shares of each constituent stock constantly reveals the price of the index portfolio in each period. The property market, however, is characterized by illiquidity and heterogeneity. Normally, only a subset of assets in the property portfolio is traded in each period. Without price information on the idle properties, the value of the property portfolio is therefore unknown. This means that the property market return cannot be computed directly from portfolio prices as for Hang Seng Index, but has to be *inferred from the price of a subset of transacted properties in the portfolio*. In fact, due to the lack of complete price information, valuation or appraisal has been common practice in the property market, making inferences from “comparables” through expert opinions on, for instance, quality adjustments and discount rates. But for the construction of price indices for trading purposes, this expert-based approach may not be appropriate because it is prone to manipulation and smoothing. Instead, systematic or rule-based approaches, notably statistical methods, are preferable.

Statistical Model

An intuitive way to infer the portfolio return statistically is to use the “average” returns of transacted properties as an estimate of the property market return. For example, given n individual properties' gross cumulative returns from time 0 to t (i.e. $1 + cr_{it} = P_{it}/P_{i0}$, with $i=1, \dots, n$), the corresponding property market return can be estimated by their geometric mean:

$$\hat{cr}_t = \left[\prod_{i=1}^n (1 + cr_{i,t}) \right]^{\frac{1}{n}} - 1 \quad (3)$$

Hereinafter variables (e.g. r , cr and P) with a subscript i refer to a single property asset, else, they represent the property market portfolio.

Equation (3) provides a simple case that illustrates how market returns can be estimated from the returns of individual assets statistically. In principle, with many individual asset returns across different periods, the “average” cumulative returns in (3) can be computed for each period to form a price index. However, this averaging approach is not the most efficient as it does not fully utilize the information embedded in returns of varying holding periods. For instance, in a three-period case ($t=0,1,2$), some returns are observed from period 0 to 1 ($cr_{i,01}$), some from period 1 to 2 ($cr_{i,12}$), and the remaining



from period 0 to 2 ($cr_{i,02}$). To estimate the market return from period 0 to 1, $cr_{i,01}$, (3) only uses individual returns from period 0 to 1, $cr_{i,01}$, but ignores the information implied from the returns of other periods, period 0-2 $cr_{i,02}$ and period 1-2 $cr_{i,12}$ ²⁰. As shown by Bailey, *et al.* (1963), a regression-based approach is more efficient in utilizing information embedded in returns of varying holding periods.

To make best use of information on repeat sales pairs with varying holding periods and returns to estimate a property price index, a formal regression based statistical model is employed. Suppose there are N properties, each of which is transacted twice between time 0 and T . Let t_1 be the time of first sale and t_2 the time of second sale. The properties are neither restricted to have the same holding period nor the same transaction date as in (3). Assume, for any property i ($i=1, \dots, N$), the cumulative return between its first and second sales is made up of two multiplicative, uncorrelated components: the property market return (see (2)) and the idiosyncratic return, ϵ_i . In other words,

$$\frac{P_{i,t_2}}{P_{i,t_1}} = \frac{1 + cr_{t_2}}{1 + cr_{t_1}} \times \epsilon_i \quad (4)$$

If the repeat transactions span the whole period from time 0 to T , (4) can be expressed more generally in terms of a series of gross cumulative returns:

$$\frac{P_{i,t_2}}{P_{i,t_1}} = (1 + cr_1)^{D_{i,1}} (1 + cr_2)^{D_{i,2}} \dots (1 + cr_T)^{D_{i,T}} \epsilon_i \quad (5)$$

Where $D_{i,t}$ is a time indicator which equals -1 if t is equal to t_1 , $+1$ if t_2 , and 0 otherwise. Taking the natural logarithm of both sides and making the assumption that $\ln \epsilon_i$ is a random noise with zero mean and constant variance, (5) becomes a linear regression model which can be estimated using the Ordinary Least Squares method²¹:

$$\ln \frac{P_{i,t_2}}{P_{i,t_1}} = \sum_{t=1}^T \ln(1 + cr_t) \cdot D_{i,t} + \ln \epsilon_i \quad (6)$$

Equation (6) is the repeat sales model developed by Bailey, *et al.* (1963). The essence of (6) is the repeat measurement of prices of the *same* property, thus its name “repeat sales”. By using repeat sales transactions to calculate price changes over time, the quality difference between two sales should be minimized, if not eliminated. If some property attributes are known to have changed between sales, it is also possible to account for them by incorporating them into the repeat sales model.²² However, we are not going to discuss this in detail here, because the dataset herein does not encapsulate information on quality changes.

²⁰ Indeed, this also implies that previous estimates would be revised as new information arrives. This might be undesirable if an index is constructed for trading purposes. The index revision problem will be dealt with later by imposing sample restrictions.

²¹ See footnote 19.

²² An exception to this is linear depreciation due to exact multicollinearity (Bailey, *et al.* 1963).



Index Estimation

Given N pairs of transaction prices and dates, the unknown parameter in (6), $\ln(1+cr_t)$, can be estimated. We denote the estimate of $\ln(1+cr_t)$ as α_t , with $\alpha_0=0$. The estimated cumulative market return from time 0 to t is thus²³:

$$\hat{c}r_t = e^{\alpha_t} - 1 \quad (7)$$

To calculate the Indices, one simply needs to set an arbitrary index value at a base period, (for example 100 in January 2000) P_B (where $T \geq B \geq 0$), and calculate the index values for other periods, P_t , by :

$$P_t = P_B e^{\alpha_t - \alpha_B} \quad (8)$$

²³ Strictly speaking, the equality does not hold because $E[\ln(1+cr_t)] \neq \ln[E(1+cr_t)]$ in general. But if the idiosyncratic return, ϵ_t , is assumed to be log-normally distributed, Kennedy (1981)* show that $r_t = e^{\alpha_t - V(\alpha_t)/2} - 1$, where $V(\alpha_t)$ is the variance of α_t . In most cases (including ours), when the sample size is large, $V(\alpha_t)$ becomes negligible. Therefore, the equality still holds approximately.

* Kennedy, P.E. (1981). "Estimation with correctly interpreted dummy variables in semi-logarithmic equations," American Economic Review, 71(4), 801.