Physical Real Estate : Risk Factors and Investor Behaviour[1]

Preliminary version

March 2001

Fabrice Barthélémy[*]

Michel Baroni[**]

Mahdi Mokrane[***]

Abstract

In this paper we investigate the risk factors associated with real estate investment. We explore a rich database of over 100 000 transactions mainly for residential properties in the Paris area over the 1973 – 1998 period. The main risk factors are identified using a Principal Component Analysis as well as a Stepwise WLS Regression Method. The first method indicates that linear or log-linear combinations of factors such as interest rates, interest rate spreads, equity market returns, rents, unemployment, or even market traded real estate cannot wholly capture physical real estate return risk. The second method indicates it is nevertheless possible to derive a factor model for real estate risk, and that the consistent factors are rents, unemployment, and listed real estate. Comparisons of our factor model index with the IPD index and the Notaires/INSEE square-metre price index, as well as statistical probe of the database, yield interesting implications concerning real estate risk, market participant behaviour, and the nature of the so-called 1990s 'speculative bubble'.

[1] The authors wish to thank the Bureau Van Dijk for graciously providing us with the data base CD-Bien. The usual disclaimer applies.

[*] THEMA, University of Cergy-Pontoise, 33, Bd du Port, 95011, Cergy-Pontoise Cedex, France.
Tel : (33) 1 34 25 62 53. Mail : barthele@u-cergy.fr.

[**] ESSEC Business School, Avenue Bernard Hirsch – B.P. 105, 95021, Cergy-Pontoise Cedex, France.
Tel : (33) 1 34 43 30 92. Mail : baroni@essec.fr

[***] THEMA, University of Cergy-Pontoise, 33, Bd du Port, 95011, Cergy-Pontoise Cedex, France, and ESSEC Business School. Corresponding author : Tel–Fax : 01 39 62 77 45. Mail : mmokrane@wanadoo.fr (corresponding author)

Introduction

Every real estate investor faces an objective difficulty concerning the measurement of real estate investment performance and risk. The reasons explaining this difficulty are numerous : an absence of centralised trading, or even price lists, a low degree of buildings or apartments turnover in investor portfolios, a lack of transparency in transactions, the heterogeneity and indivisibility of real estate properties, and a tradition of confidentiality in the industry.

Does this imply that investors should disregard investing in real estate? This would be a mistake if real estate investments represented a consistent diversification vehicle from a global portfolio management perspective. This mistake would be even greater if real estate provided an efficient inflation hedge.[2]

Real estate markets are relatively often subject to price shocks whose amplitude prove to be very high and welfare decreasing. According to R. Shiller (1998)[3], these shocks are as difficult to explain as those that affect equity or debt markets. One of the main problems in trying to measure real estate volatility in France is that there does not exist satifactory historically long time series for price and rent evolution (see Section 1 of the present paper). A possibility for creating such indices would be to use a methodology based on observed price transactions (Section 2). To achieve this goal we explore a rich transactions database containing information on single-family homes transactions, as well as office, mixed (professional and housing) and commerce properties for the Paris and its surrounding area.

Such indices may prove very useful, for example, in serving as "benchmarks" or in risk measurement. Indeed, real estate performance is today at the heart of investor focus, whome, after a period of withdrawl in the mid nineties, are flowing back, but with the concern of trying to better control for the risk and return of real estate investments. According to F. Savel (1998)[4], delegated real estate management seems to benefit from a promising future, at the condition that this sector's risk be well identified and hedged. Section 3 presents a statistical analysis of our database and highlights interesting specific investor behavior.

Identifying risk factors for real estate enables the investor, be he or she a landowner, a multi-billion dollar institutional investor or a real estate debt holder, to measure and therefore hedge his (her) investment position. This process also permits the debt holder to assess the risk of a real estate loan. A unified approach to real estate risk is the only way to bring market participants to trade risks using real estate derivatives. In this context, we isolate real estate risk factors (Section 4), and construct a factorial model capable of synthetising real estate systematic price dynamics (Section 5). Our synthesis in the form of risk-returns comparisons is presented in Section 6. The last section gathers our concluding remarks.

1. Existing Indices and Risk Measures

1.1 Existing Indices

There are three measures capable of calibrating real estate risk in France (and in particular in the Paris area) today. They correspond to three indices of relatively different nature:

- The square metre index provided by the Chambre des Notaires de Paris and INSEE. This index is computed every six months using a sample of transactions of unoccupied apartments aged five years and over. The index for a given date is the weighted average of transactions prices per square metre.
- The IPD index (International Property Databank) : This index is constructed using property appraisal estimates covering a representative sample of real estate properties owned by majors

[2] For a study of the link between inflation and real estate in the US, see Sirman and Sirmans (1987). For the case of France, see Friggit (1999).

[3] Shiller (1998).

[4] Savel (1998).

insurance and banking institutions in the Paris area. The methodology adopted consists in gathering a representative sample whose value is periodically appraised using expert valuation models. The aggregation of appraised values provides the index's value for a given date. For the case of France, IPD publishes such an index every year[5].

- The index for listed real estate is published each day by Datastream-Financial Times (code RLDEVFR) based on the liquidation prices of the following listed investment trusts : GECINA, INTERBAIL, KLEPIERRE (CIE FONCIERE), LOCINDUS, SEFIMEG, SILIC, SIMCO, SOCIETE FONCIERE LYONNAISE, SOGEPARC, UIF, UNIBAIL.

1.2 Inability of theses Indices to Measure Real Estate Risk

The following figure illustrates the evolution of the above-mentioned indices since 1986. The IPD and Notaires/INSEE indices represent the return on capital for housing, whereas the type of market sector represented by the listed real estate index depends on the type of assets the investment trusts holds.

IPD, Notaires/INSEE and Listed Real Estate Indices (return on capital)

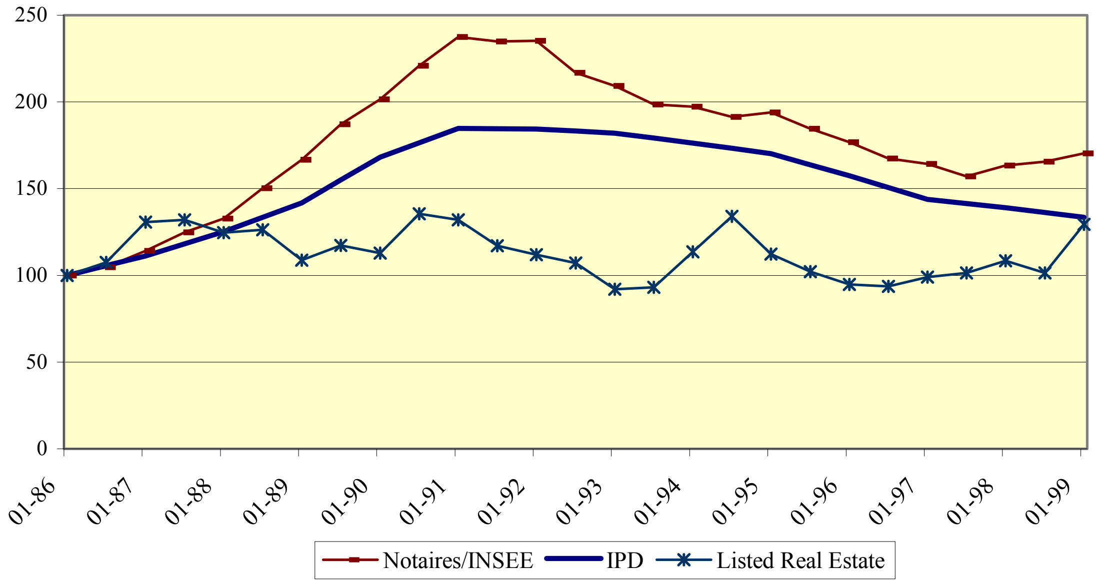

Figure 1

The main drawback of the Notaires/INSEE index is that it is based on a weighted average of prices for heterogeneous real estate properties. This implies that, even if the general market trend may indeed be correctly captured, the volatility measurement by the index is biased.

For the case of the IPD index, its main drawback is linked to the fact that it is not based on real transactions prices, but rather on appraised values. This induces a certain degree of inertia in the index that takes its source in:

- the fact that the appraisal experts periodically valuing a property generally tend to adjust preceding valuations, conditioning valuations period to period.
- the time interval separating transactions information and the processing of this information in the form of an appraised value. This bias is an increasing function of the index's periodicity (note for example that the IPD index is monthly in Great-Britain and annual in France).

---

[5] For more details on the IPD index as well as other related information surf the following web site www.ipdindex.co.uk, http://www.propertymall.com/ipdindex/ipfrisk.html.

These elements have a non negligible impact when one attempts to measure real estate risk, since they both lead to an underestimation of volatility (inertia leads to a lower index volatility).

Listed real estate naturally contains real estate risk, but the total risk measured is simultaneously affected by two other sources of risk. One is linked to the relatively high liquidity (daily listing) of investment trusts, which is not representative of physical real estate whose markets are relatively illiquid in comparison. The second is linked to the role that these trusts play in global portfolio diversification, which implies that demand for these securities is more directly linked to capital markets fluctuations. This risk is for example illustrated by the "kink" in the index during the year 1987, the same year securities markets began to experience rapid growth and developed liquid and organised derivatives compartments (MATIF, MONEP etc…).

One may then suggest that listed real estate may have been turned down in favour of then more attractive securities such as stocks and bonds for which risk management and hedging became easier thanks to the development of sophisticated equity and FX derivatives. If one believes that listed real estate is perceived by capital market participants as an alternative investment with respect to other capital markets securities, then it may not be surprising that listed real estate is much more highly sensitive to capital markets risk factors than physical real estate, that may have its own factorial risk representation. For all these reasons, the former may thus not satisfactorily represent the latter's risk structure. One of the aims of this paper is to shed light on this question.

Based on the above analysis, one may temporarily conclude that it is indeed very difficult to obtain consistent and reliable results into the risk-return measurement of physical real estate based on existing market representations, in particular when historical time series are short. What's more, as stressed by Hoesli (1993) [6], the evidence suggests a clear difference in terms of the relationship between inflation and listed real estate on the one hand and physical real estate on the other hand, which confirms this risk measurement difficulty. The real estate investor therefore is in search of a transactions-based index, and therefore a listing of transactions prices. One such database exists for the case of Paris and the neighbouring "départements" : the CD-Bien database. Note that this base contains a very high proportion of housing transactions (more than 80%).

2 The Database

To analyse real estate returns, we constructed a database containing not only real estate returns but also corresponding returns from economic or financial variables as well as hedonic[7] information for each transaction. The latter are extracted for the CD-Bien database who lists all real estate transactions written in front of a notary for the Paris and near surrounding area (Hauts-de-Seine, Seine Saint-Denis, and Val de Marne). From this database, we extracted 121 327 transactions for which we had the information on both the initial price and date (post 1st of January 1973) at which the properties had been bought as well as the price and date for the following resale[8]. These "complete" transactions represent 22,2% of the total number of transactions.

Having at our disposal both the initial price of the property P_1 as well as its resale price P_2, we may calculate more than 120 000 returns $R = P_2/P_1$. Using the dates for both transactions, T_1 and T_2, we may annualise these returns in the following way $R_{annual} = (P_2/P_1)^{365/(T2-T1)}$.

To every observation in the database, the following hedonic characteristics can be associated :

- *Holding duration (duration)* : *duration* is simply the difference $T_2 - T_1$ expressed in days.

[6] See Hoesli (1993).

[7] In this paper hedonic is to be taken in the loose sense of transactions characteristics.

[8] We have to note an important feature concerning the database's structure: we only observe those transactions whose second transaction has taken place after 1990. We will come back to this point and point out where it may be cause of concern in the course of the analysis presented below.

- *Holding horizon (horiz)* : to facilitate statistical processing and interpretation, variable *duration* has been translated[9] into a discrete variable (*horiz*).
- *Type of purpose (purpose)* : variable *purpose* may take any one of four values depending on the type of purpose declared : 1 for family housing, 2 for commerce, 3 for a mixed-purpose (housing and professional), and 4 for offices (commercial or professional).
- *Catgeo* is a variable that classifies transactions according to a semi-geographical partition based on the square-metre average price of 1997. Its value for a given transaction depends on P_2/(number of m^2) : it will take a value of 1 if the price is lower than 10 000 FF, 2 for a price falling between 10 and 15 000 FF, 3 if the price is comprised between 15 and 20 000 FF, and finally 4 for a transaction price greater than 20 000 FF.

The quest for explanatory factors of real estate capital returns necessitates the selection of indicators that one a priori believes to have some form of explanatory power. Ten factors were selected based on two criteria. The first asked of potential factors to have a clear economic interpretation and presupposed links with real estate markets. The second was that it was necessary for the times series to run back as far as possible in time. The data we explored in Datastream ran back, for France, as far as the first of January 1973.

The indices selected to serve as factors were thus constructed with base 100 at the start of 1973. They are the following : *long term rate (LtR), short term rate (StR), consumer price index (Consum), MSCI*[10] *equity market index (Equity), listed real estate (ListRE), rents (Rent), demographic index (Demog), unemployment (Unemp), savings as a percentage of disposable income (Saving), and yield spread (Spread).*

For each transaction, we computed the corresponding return of every potential factor for the period separating date T_1 from T_2. The whole set of returns was then transformed into semester returns[11]. The resulting variables are thus the following: R_S, $Equity_S$, $Consum_S$, $Rent_S$, LtR_S, StR_S, $Demog_S$, $ListRE_S$, $Unemp_S$, $Saving_S$, $Spread_S$.

3 Statistical Analysis of the Database : The Broad View

3.1 Extreme Value Elimination

Let us begin by pointing out at the presence of a significant number of extreme values for variable R_S, the presence of whom may severely bias statistical estimation. We have therefore begun by eliminating transactions whose semester return exceeded 10 (or 1000%). One may note that all of these excluded transactions correspond either to very short holding durations, for which one of the two prices (at date 1 or date 2) is clearly not a market price, or to transactions conducted using a symbolic 1 FF price.

The following table illustrates the major differences between transactions based real estate returns (variable R_S) and those of other variables namely on the grounds of the extent or dispersion (standard-

---

9 The partitioning of variable *duration* is arbitrary and corresponds to a number of observations concerning market participants and their presumed investment horizons. We will come back to this point to motivate it. Variable *horiz* will take value 1 if *duration* is lower than 3 years, 2 if it is comprised between 3 and 5 years, 3 if *duration* is between 5 and 10 years, 4 if it is comprised between 10 and 15 years, and finally 5 if *duration* exceeds 15 years.

10 We used the Morgan Stanley Capital International (MSCI) Index for France, which runs farther back in time than the CAC40 Index .

11 The choice of the semester as a unit period is linked to the fact that, as we will see in the next sections, this time step suits best our factorial model of index construction Hence, for every transaction *i*, characterised by dates T_1 and T_2, we construct a series of variables $F_{js}(i)$, using the following formula: $F_{js}(i) = (F_j(T_2)/F_j(T_1))^{182.5/(T2-T1)}$, where $F_j(T)$ represents the level of variable (time series) F_j (the MSCI index for example) for any date *T*.

deviation), of the asymmetric feature or skewness (numerous returns above the average), of the presence of persistent extreme values (excess kurtosis[12]).

	Minimum	*Maximum*	*Mean*	*Standard deviation*	*Skewness*	*Excess Kurtosis*
RS	**0,00**	**9,77**	**1,032**	**0,196**	**22,93**	**734,40**
Equit	0,36	2,44	1,054	0,052	4,14	85,53
Consum	0,97	1,04	1,014	0,007	1,06	0,71
Rent	1,00	1,04	1,022	0,010	-0,52	-0,67
LtR	1,02	1,05	1,041	0,006	-0,16	-0,29
StR	1,02	1,06	1,038	0,009	-0,97	0,21
Demog	1,00	1,01	1,002	0,001	-0,95	1,07
ListRE	0,53	1,81	1,009	0,047	1,30	27,57
Unemp	0,90	1,12	1,015	0,015	-0,50	4,01
Saving	0,97	1,04	1,003	0,011	0,14	1,20
Spread	0,12	2,72	1,027	0,073	2,44	31,58

Table 1

These results suggest pursuing the analysis by taking the natural logarithm of returns of all the above variables, will be the resulting variable names being prefixed by an *Ln*. Remark that this method is rather classical in finance when analysing return distributions. The histogram for variable R_S, for values comprised between 0,7 and 1,3 is the following.

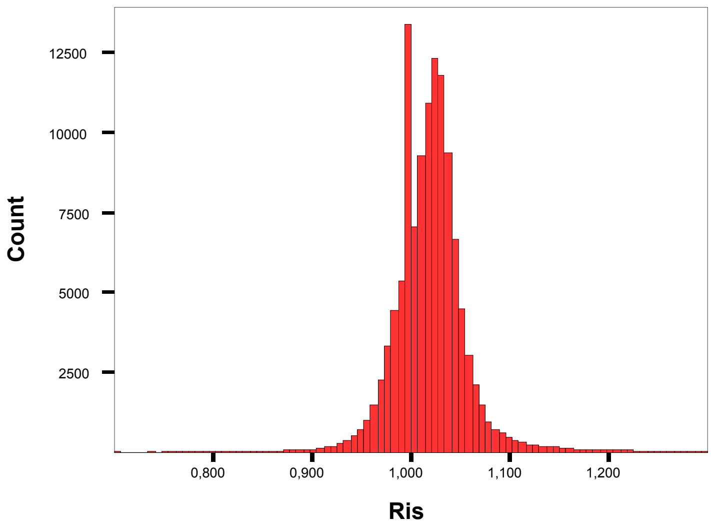

Figure 2

3.2 « No-Loss » Behaviour and Real Estate Crisis

The above figure points to an anomaly corresponding to a large number of transactions at or very near a return of 1, or equivalently a rate of return of zero. Taking a close-up of the 11 491 observations whose returns lie between 0,9975 and 1,0025, yields the following results:

[12] Skewness measures the tendency of a distribution to give more weights to values greater than (positive skewness) or lower than (negative skewness) the mean. Kurtosis measures the degree of flatness of a distribution. A normal distribution has a kurtosis of 3 or equivalently an excess kurtosis of 0. An positive (negative) excess kurtosis indicates a distribution with fat (thin) tails with respect to the normal distribution, which indicates the presence (absence) of extreme values.

	Minimum	*Maximum*	*Mean*
Rs	0,998	1,002	1,000
Date 1	01/03/73	01/11/98	**01/07/92**
Date 2	01/11/91	01/12/98	**01/10/96**
Price1	14 000	130 000 000	806 865
Price2	14 000	131 000 000	**807 243**
Duration	28	9 345	**1 535,91**
Purpose	1	4	**1,03**
Horiz	1	5	**1,96**
CatGeo	1	4	1,92

Table 2

This portion of returns represent nearly 9% of our observations, and correspond to the histogram's peak. Remark that all these transactions exhibit a resell date T_2 post November 1991, that is after the beginning of the real estate crisis in France. Observing the *Average* column of this table helps characterise these transactions. The average acquisition date of these properties is July 1992, whereas the average resale date is October 1996. The type of property is quasi-exclusively housing.

If one considers the Notaires/INSEE index during the same period, its level dropped by 10,5%. One may conclude that an important number of property sellers had a "No-loss" behaviour during this crisis period, in which they refused to sell their property at a price lower than the one they paid for it. A more thorough analysis of variables P_2 and *CatGeo* indicates that these properties had a resale price not exceeding 1 000 000 FF and located in areas with an average square-foot price below 15 000 FF/ m^2.

The type of property described above exhibits a relatively high resistance to price drops even in the context of a rough downside market. One may also interpret this behaviour as being one of owners willing to sell only at the condition that they do not loose in terms capital, or accepting to sell with no gain rather than waiting for a hypothetical market upturn. This behaviour also illustrates the relatively dominant weight of the seller over the acquirer for this type of property. For the analysis in the next sections, we exclude these transactions based on the assumptions that they correspond to very particular types of market participant behaviour during a peculiar period.

3.3 Investor Behaviour During the Crisis

The second conclusion worth mentioning here concerns the market behaviour of "retired" individuals during the 1987-1997 period. Thanks to the precision of the database information concerning the status active/retired of the acquirer and seller, we were able to see that the retired population massively entered the physical real estate market during the end of the 1980s, early 1990s. Indeed, during this period, more than 25% of the market entrants were retired individuals. What's more, it is the retired who marched out of this same market during the period 1996-1997, probably to profit from the bullish capital markets France experienced at that time.

We may also note that only 7% of market participants with a an investment horizon lower than three and a half years are property traders. Hence, it appears that the market participants that have the possibility of profiting from arbitrage opportunities from the capital markets are not the legal entities, but the individual entities, and more precisely retired individuals. We will come back to this point when presenting our results on factorial models and their link with listed real estate.

4 Common Risk Factors

4.1 Preliminary Analysis : Horizon Threshold and Data Selection

On the basis of the above analysis (descriptive statistics and out-of-market data elimination) we may now proceed with a Principal Component Analysis (PCA[13]) whose goal is to isolate the common risk factors associated with real estate returns.

A first result of our exploration stresses the absence of a clear linear relationship between physical real estate returns and the macroeconomic and financial variables selected[14]. A PCA analysis on hedonic variables shows the relative importance of variable *horizon* and more specifically of the atypical behaviour of returns associated with short holding durations (less than three years). This result is confirmed and refined by the study of the error terms in regressions of variable LnR_S on the selected variables. Indeed, it seems suitable to treat the short-horizon transactions separately. The main partition appears at horizon 3,5 years. The characteristics of return distributions for short holding periods (less than 3,5 years) are markedly different from those of longer holding periods. Note that this result is consistent with the findings of the previous section, where we stressed "No-loss" and "retired" behaviours which were both associated to short horizon.

It seems therefore reasonable to assume that transactions whose *duration* is less than 3,5 years behave according to a model that cannot be correctly captured by combining our variables. For longer *durations* the prospects of elaborating a consistent and reliable model seems more encouraging. For these transactions we have proceeded using a similar systematic analysis of error terms, leading to a further screening of outliers and/or influential data.

Data selection and screening is a delicate and rather cumbersome exercise[15] since one has to both eliminate unsuitable or outlying data and at the same time be careful not to discard valuable information. This often implies a careful examination of a large number of observations on a case to case basis: sale condition, price comparisons, property inhabited or not at the time of sale, transaction parties' status, symbolic francs transaction…). This screening is not that penalising for those property purposes for which we had relatively few observations originally, i.e. mixed (952 observations before and 632 after screening), commerce (1 500 before and 1 199 after screening), and office (572 before and 435 after screening). For the housing purpose we had more than 100 000 transactions at our disposal, of which we extracted a random sample of 7 000 observations. After screening, this led to a total of 5 609 observations for the housing purpose. We believe this provides us with a suitable amount of relevant information to derive real estate factors.

4.1.1 Results Per Property Purpose for LnR_S

Before moving further, let us make sure the screened data imply a reasonably normal distribution for variable LnR_S. Table 3 shows that for all property purposes, semester log-returns distributions exhibit a significant positive excess kurtosis, which means that the distributions tails are still relatively thicker than normal. However, these statistics are much less alarming than before data screening (see Table 2). Skewness to the right or to the left (for mixed purpose) seems very reasonable, being quite close to zero. Finally, the examination of kurtosis and skewness suggests differentiating the analysis with respect to property purpose. This will be further confirmed later on in the analysis.

---

[13] PCA is a data analysis method used in a number of management disciplines (marketing, finance, insurance…). Consider a table containing individuals (lines) and many variables measured on each individual (columns), PCA aims at creating a restricted number of linear variable combinations (called factors) capable of summarising the essence of information contained in the initial table.

[14] More detailed results, concerning "outlier detection" in particular, are presented in Barthélémy, Baroni and Mokrane (2001).

[15] For a more thorough discussion on outlier detection see the seminal paper of Cook (1977), and the more recent book by Atkinson (1985).

Purpose	*Number of observations.*	*Minimum*	*Maximum*	*Mean*	*Standard deviation*	*Skewness*	*Excess Kurtosis*
Housing	5 609	-0,126	0,168	**0,020**	0,028	**0,128**	**2,072**
Mixed	632	-0,073	0,128	**0,026**	0,031	**-0,395**	**0,691**
Commerce	1 199	-0,075	0,143	**0,024**	0,032	**0,017**	**0,581**
Office	435	-0,060	0,130	**0,025**	0,028	**0,091**	**0.616**

Table 3

4.1.2 Bivariate analysis on the temporal structure per type of purpose

The following figure represents, for the Housing purpose, and for variables *date1*, *date2*, *duration* and LnR_S, a set of univariate sorting (frequency distributions on the diagonal) as well as cross-sorting (bivariate) of selected variables. For example, the lower left-hand side graph indicates the distribution of values taken by variable LnR_S as a function (sorted by) of *date1*.

For graph readability, we have randomly selected a set of 1 000 transactions for Housing (similar results obtain but are not shown here for Commerce, and for the sets of available transactions for Mixed and Office purposes). Note that this does not imply any loss of generality of the qualitative and quantitative insights these graphs enable us to draw.

Transactions dates, returns and market volume

Whereas the negative linear relationship between variables *date1* and *duration* is obvious, it is worth underlining the absence of any linear relationship between *date2* and *duration*. Indeed, we believe the very small positive linear relationship between variables *date2* and *duration* is due to the database's own structure : we only observe those transactions whose second transaction has taken place after 1990.

Variable *date1's* distribution illustrates, to a certain extent, the real estate market's activity. The graph's observation leads us to a time partition into three broad phases : a relatively low but increasing volume of home acquisitions from 1973 to the mid-1980s, a strong increase in market activity (acquisitions) from 1985 to 1990, followed by a brutal decline in buys coinciding with the real estate bust of 1991, with the mention that the slow acquisitions activity has lasted till the end of 1998.

By comparing real estate returns with the market volume as measured by *date1*, we are able to relate LnR_S to the three distinct phases identified above : until the mid 1980s, real estate returns are relatively stable both in level and variability. During the 1985-1990 phase, returns decreased steeply with time (*date1*) and their variability is distinctly higher than during the preceding phase. Finally, during the post-1990, returns relate positively to *date1*, but one observes an even higher variability.

Heteroscedasticity

The analysis of the figure's last line (corresponding to variable LnR_S) exhibits strong heteroscedasticity for physical real estate log returns as a function of *duration.* Indeed, the shorter the holding period, the higher the log-return variability. We believe that part of this variability takes its source in the three-phase structure linked with *date1* indicated above. This holding period-linked heteroscedasticity will have to be corrected when specifying our factorial model for real estate physical returns.

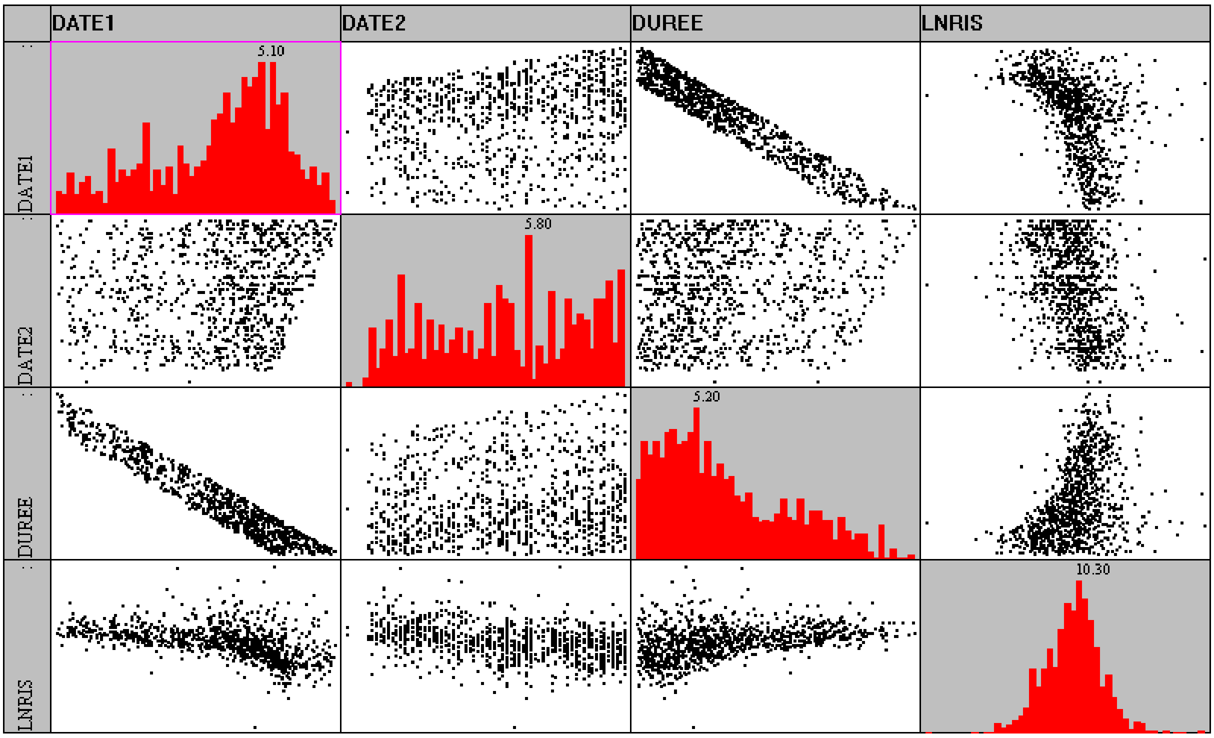

Figure 3

4.1.3 How Many Real Estate Risk Factors?

Let us now turn to the search for systematic relationships between real estate returns and macroeconomic, financial and hedonic variables. PCA indicates that only four factors (linear combinations of 12 variables) are sufficient to capture nearly 85% of total variance of our dataset. The first factor is able to explain roughly 51% of total variance, followed by Factor 2 (15,9%), Factor 3 (10,3%), and finally Factor 4 (7,7%). The following table indicates linear correlations between variables and factors:

Variables	***Factor 1***	***Factor 2***	***Factor 3***	***Factor 4***
LnR_S	0,551	0,152	-0,388	0,084
Horiz	**0,825**	-0,281	0,056	-0,027
$LnConsum_S$	**0,920**	-0,010	0,260	-0,013
$LnRent_S$	**0,906**	0,389	-0,047	-0,004
$LnLtR_S$	**0,970**	0,144	0,043	-0,007
$LnCtR_S$	**0,858**	0,473	0,007	0,002
$LnDemog_S$	**0,750**	0,235	-0,369	0,023
$LnListRE_S$	**0,752**	-0,551	0,099	-0,023
$LnSaving_S$	**-0,716**	0,522	-0,318	0,011
$LnEquity_S$	0,381	**-0,854**	-0,106	-0,035
$LnUnemp_S$	0,388	0,454	**0,719**	0,016
CatGeo	0,047	-0,079	0,002	**0,992**
$LnSpread_S$	-0,593	-0,041	**0,577**	0,060

Table 4

4.1.4 Interpretations

The first factor comprises interest rate –linked (long and short term rates) as well as economic variables (consumer price, rents, demographic index, and to a lesser extent savings rate and listed real estate). Note that all of these variables are linked to investment horizon and hence to holding duration. The second factor reflects the volatility of equity market returns, whereas the third factor captures business cycles by combining unemployment and interest rate spreads. Finally, the fourth factor sorts returns using the geographical variable (average price per square metre).

Introducing real estate returns into the analysis does not significantly change the results (coefficients). Real estate returns react positively to Factor 1 shocks (an increase in long term rates being the clearest example), negatively to Factor 2 shocks (MSCI French Equity index). Factor 3 opposes physical real estate variations to those of cyclical variables such as unemployment and interest rate spreads.

Variable LnR_S is essentially captured by Factors 1 and 3 (their combined explanatory power roughly 45%[16]). We might add that since the first four factors explain less than 48% of real estate returns, this level constitutes somewhat the maximum R^2 any regression model of these returns on linear combinations of explanatory variables might attain. Remark that Factor 4 is not correlated with real estate return, which, to some, might come as a surprise. Indeed, this means that physical real estate returns are not linked to properties' average square metre prices.

Finally, note the very different factor projections of physical real estate and listed real estate (apart from Factor 1 coefficients, the other three have opposite signs). What's more, listed real estate variance is almost entirely captured by the fist two factors (loosely speaking interest rate and equity market returns explain more than 87% of variance). This comes as a confirmation of the elements presented in the previous sections, and reinforces the conclusion that physical and listed real estate consistently behave differently, and are thus subject to different risk factor structures.

5 Towards a Factor Model of Physical Real Estate

A factor model enables to "synthetize" a real estate index capable of representing systematic risk per property purpose. The selected methodology is that of a multi-factor model resulting from a linear regression on the selected explanatory variables presented above. We will now analyse, thanks to a "stepwise" regression method, the relationship between physical real estate, macroeconomic, financial, and idiosyncratic factors. For each transaction i, the estimator for $LnRi_S$ is thus:

$$LnRi_S(i) = \alpha + \sum_{j=1}^{k} \beta_j \, LnF_{jS}(i) + \varepsilon(i) \qquad (1)$$

where α is the regression's constant estimator, $F_{jS}(i)$ represents the semester return for variable F_j during period $[T_1, T_2]$ separating the two sales of property i (see footnote n°11), β_j is the loading of the j^{th} variable LnF_{jS}, and finally $\varepsilon(i)$ is a normally distributed random variable with mean 0 and constant variance σ^2. It represents the error term for transaction i. We assume away any correlation between any two transactions' error terms.

Ordinary Least Squares (OLS) are used to estimate the above model's parameters. The following table summarises selected models as well as the corresponding coefficient estimates. The value in parentheses represent the Student-t estimates for each coefficient[17].

Purpose	***Constant***	***$LnRent_s$***	***$LnUnempl_s$***	***$LnListRE_s$***	***R^2 (%)***
Housing	-0.0278 (-22.6)	2.38 (42.1)	-0.75 (-18.5)	0.091 (4.9)	31.4
Mixed	-0.0193	2.40	-1.11	0.200	28.2

[16] The explanatory power of a factor with respect to a variable's variance (the R^2 obtained by regressing the variable on the factor) is obtained by taking the squared value of the corresponding coefficient. Hence, Factor 1 explains $(0,551)^2$ = 30,4% of real estate returns, whereas Factor 2 explains $(-0,388)^2$ = 15,1%, with a combined explanatory power of 45,5%.

[17] Note here that all coefficients are significant to the 99.9% confidence level.

	(-4.4)	(11.9)	(-8.5)	(3.2)	
Commerce	-0.0194 (-5.8)	1.95 (13.6)	-0.41 (-3.7)	-	14.3
Office	-0.0330 (-6.1)	2.57 (11.0)	-0.73 (-4.5)	-	24.1

Table 5

5.1 Residuals Analysis : Heteroscedasticity

The above proposed modelling assumes that the variance associated to each purpose does not depend on observation (*i*). If these assumptions are not validated, the model must be amended so that its specification corresponds to our data structure. Inside each purpose class, the White test[18] clearly indicates heteroscedasticity. To study its correct nature (the search for variables that are at the source of heteroscedasticity), we use the Goldfeld-Quant (GQ) test.

We begin by ordering regression residuals as functions of variables, and then study whether residual variance is constant across classes. Several variables may be candidate sources for heteroscedasticity : the ones contained in the table above, which were identified by the White test, as well as temporal variables such as *duration*, *date1*, and *date2*. Table 6 below indicates the *p*-values for different varaibles and the four purposes.

	Housing	**Mixed**	**Commerce**	**Office**
duration	e-175	e-22	e-28	e-8
date1	e-137	e-18	e-21	e-6
date2	e-26	e-6	e-7	e-4
$\boldsymbol{LnUnempl_s}$	0.001	0.62	0.88	0.85
$\boldsymbol{LnListRE_s}$	e-8	e-13	e-24	e-5
$\boldsymbol{LnRent_s}$	e-15	0.002	e-4	0.04

Table 6

By gathering these results with those of the bivariate analysis (see 4.1.2), the variable at the source of heteroscedasticity is clearly *duration*. The study of the graph mapping the residuals as a function of *duration* logically suggests an inverse relationship between the error term's variance for observation *i* and the total holding period for the asset as measured by *duration*.

The new model[19] selected is thus a modified version of (1) in which $\varepsilon(i)$ is now considered to be a random variable that follows a normal distribution with zero mean and variance equal to $\sigma^2/duration(i)$. This model can then be estimated using Weighted Least Squares (WLS).

In modifying the model's specification to take into account this variance's modelling, we are able to eliminate the sources of heteroscedasticity that appeared in Table 6. Indeed, the White test does not detect any other source of heteroscedasticity based on our WLS estimates. As for the GQ tests, all *p*-values

[18] The tests presented thereafter are described in detail in Greene (1997).

[19] It may be possible to try and find α in the specification $\sigma^2(i) = \sigma^2/duration(i)^{\alpha}$ using the maximum likelihood method, but this is rendered very difficult by the fact that we do not know the residuals distribution function (normality is rejected here).

become high except for variable *date2*, and only for the case of the housing purpose. Our conclusion is that the test results in Table 6 for the case of variables $LnUnempl_s$, $LnListRE_s$, and $LnRent_s$ were significant only due to their correlation with *duration*.

This new specification for the factorial model illustrates the importance of variable *duration* ; by being able to include it in the variance but not in the level (different levels of returns depending on horizon and different factorial relations) enables us to construct a synthetic index for mid-term and long-term physical real estate[20].

5.2 Selected WLS Factor Models for Real Estate Risk

The following table summarises selected models as well as the corresponding WLS coefficient estimates[21].

Purpose	*Constant*	$LnRent_s$	$LnUnempl_s$	$LnListRE_s$	R^2 *(%)*
Housing	-0.0314 (-23.9)	2.50 (40.46)	-0.72 (-17.74)	0.065 (3.47)	45.3
Mixed	-0.0213 (-4.28)	2.37 (10.26)	-0.92 (-6.67)	0.166 (2.53)	40.8
Commerce	-0.0227 (-5.8)	2.12 (13.6)	-0.52 (-3.7)	-	25.5
Office	-0.0391 (-5.69)	2.84 (9.0)	-0.81 (-4.58)	-	31.1

Table 7

A first comment on these results is that the variables selected are similar for the four types of property purposes, with the exception of variable *ListRE (listed Real Estate)* which comes significantly into action for housing and mixed (professional and housing) purposes. The message delivered by this result is that, even if the levels of return demanded by investors differ according to the property purpose, the sources of systematic risk are the same for housing/mixed and commerce/office, noting however that variables $LnRent_s$ *(rents)* and $LnUnempl_s$ *(level of unemployment)* come into action whatever the property purpose.

It may appear at first glance quite surprising that listed real estate intervenes in property purposes such as housing and not in office or commerce purposes. However, this result is to be paralleled with the one put to light concerning the retired individuals during the real estate crisis period (see section 3.3). Indeed, we had stressed an arbitrage-type behaviour between physical real estate and capital markets on the part, not of professional real estate investors, but of individual buyers. This arbitrage is therefore to be found in the returns demanded by these market participants in the form of the listed real estate factor for risk, as illustrated in Table 7.

5.3 Visualisation of indices

The indices constructed using the above factor models are represented in the following figure, which also contains the listed real estate index to facilitate comparisons.

The factorial indices for the four purposes are quite similar (see Figure 3) and move significantly away from the listed real estate index starting from the beginning of 1982. One may however remark that notwithstanding the July 1990-July 1994 period corresponding to the real estate crisis, the physical real

---

[20] Recall that in Section 4.1 we had shown that it was necessary not to include transactions whose *durations* were less than 3.5 years because of their very different correlation structure with our then potential risk factors.

[21] Note here again that all coefficients are significant to the 99.9% confidence level.

estate changes in trend seem to be portended by the listed index, in the sense that the trend kinks seem to be 'anticipated' by a whole year or even more by the listed index.

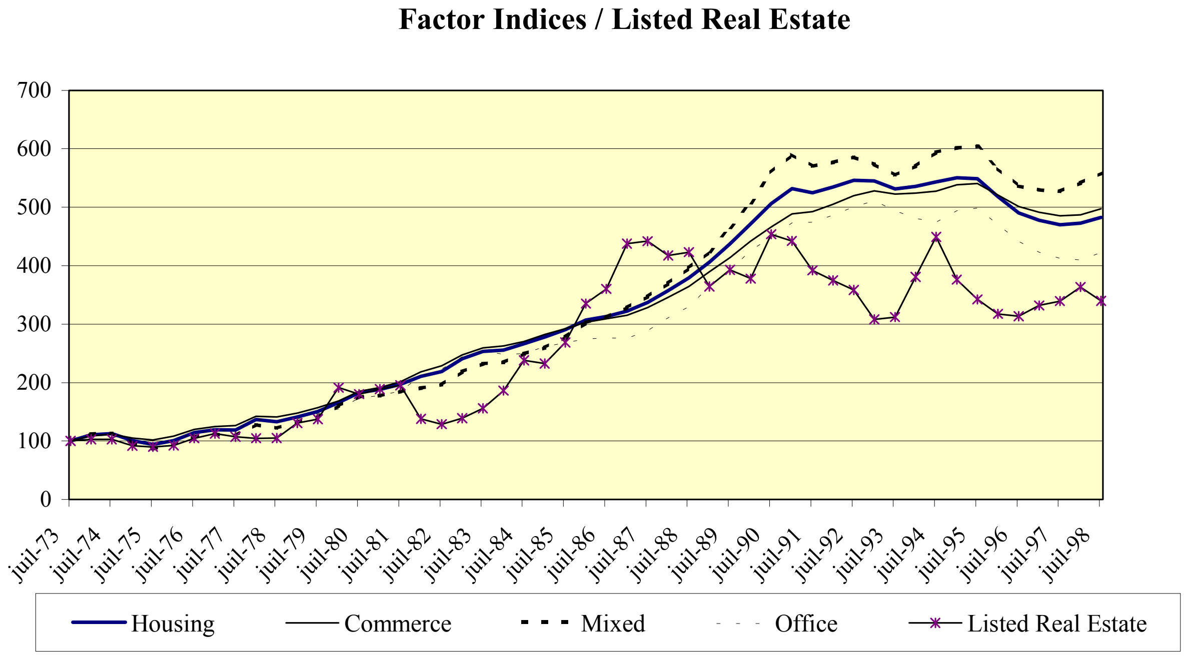

Figure 4

Representing the Housing Factor Index, the Notaires/INSEE index and the factor variables (rents, unemployment and listed real estate), yields the following figure.

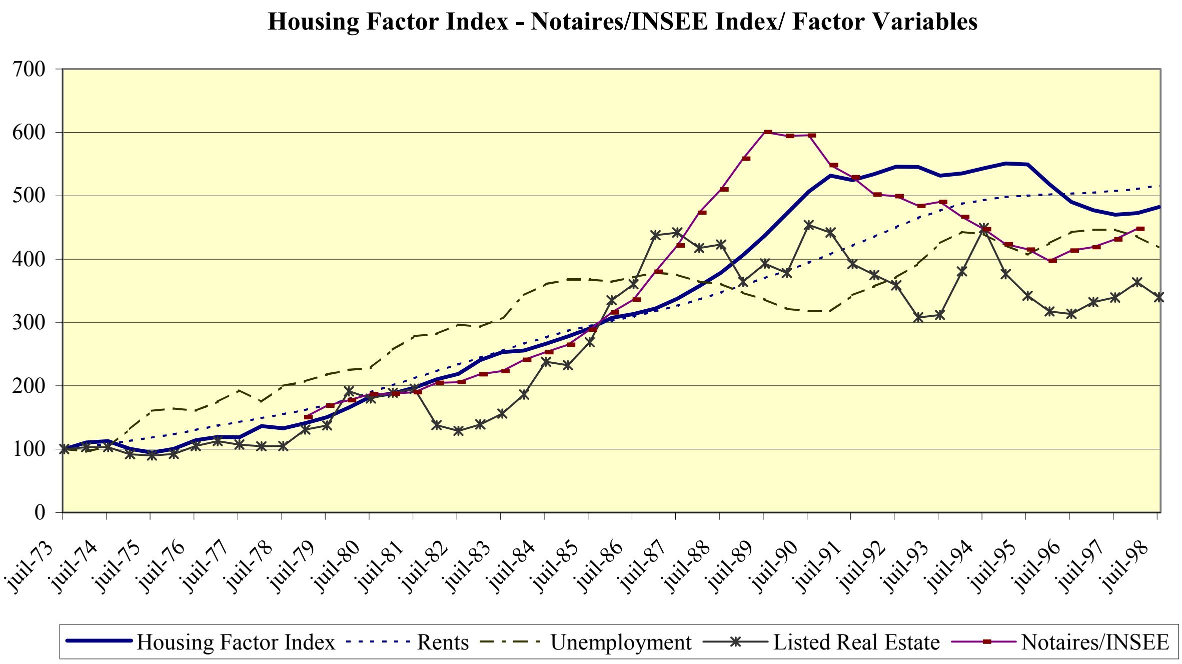

Figure 5

It seems interesting to note that the Housing Factor Index's trend is driven by the Rents Index whereas its departures from its central trend are linked to the Listed Real Estate and Unemployment Indices. Whenever the latter drop, the Factor Index climbs and conversely. This feature is particularly notable for the period 1987-1990, when real estate price experienced fast growth. Note that the Notaires/INSEE Index depicted in a rather exaggerated manner this fast growth. An explanation of this feature may be that the

Notaires/INSEE index includes all types of transactions, and does not control for those with very short holding periods (*durations*) which we have analysed separately since they corresponded to significantly different investment behaviours. Therefore, our factor model may be seen as representing the fundamental real estate market evolution. Note finally that the Notaires/INSEE index moves away from the Factorial Index only during the 1987-1994 period which is precisely the time when the so called "real estate bubble" formed and deflated.

6 Risk and Return

This section extends the analysis to the nature of the return and risk (standard-deviation) characteristics of physical real estate. Assuming that real estate log-returns (LnR_S), are indeed normally distributed, enables us to consider that our factorial indices follow standard geometric brownian motion dynamics. The estimated model for R_S is hence a diffusion process whose instantaneous expected mean is $\exp(\mu)$ where : $\mu = m + \frac{1}{2}\sigma^2$, and variance is σ^2 (*m,* resp. σ, being the historical average, resp. standard-deviation, of log-returns). The following table presents the characteristics of mean returns and a measure for risk (standard-deviation of rates of return) per property purpose:

Annualised moments (%)	***Return : $\mu = m + \frac{1}{2}\sigma^2$***		***Volatility : σ***	
	Empirical Measure[22]	***Factor Model***	***Empirical Measure***	***Factor Model***
Housing	*6,01*	*6,54*	*4,02*	*6,99*
Mixed	*7,31*	*7,26*	*4,42*	*8,66*
Commerce	*7,05*	*6,56*	*4,49*	*5,27*
Office	*6,99*	*6,06*	*3,90*	*7,74*

Table 8

Note that empirical returns are approximately the same as those computed using the Factor Models. However empirical volatilities are very significantly lower than those of the factor model. This is probably due to the returns' heteroscedasticity that is not corrected for in the empirical measure.

Also, note the apparent homogeneity of returns, and the relatively low Commerce volatility. By construction, a given Factor Index volatility is linked to the volatilities of selected explanatory variables. This does not prevent Factor volatility from being greater than empirical volatility. The transcription of theses results in terms of Factor rates of nominal return and risk is given in Table 7.

Annualised moments (%)	***Rate of Return (R)***	***Risk (Standard deviation)***
Housing	*6,76*	*7,47*
Mixed	*7,52*	*9,33*
Commerce	*6,78*	*5,63*
Office	*6,25*	*8,24*

Table 9

[22] The empirical measures are simply the mean and standard deviation as measured directly from the dataset observations.

Finally, comparing the above results to inflation behaviour during the 1973-1998 period confirms the traditional view that real estate is a hedge for inflation risk. Indeed, average inflation was 6,4% annually which is very close, while being slightly lower than average physical real estate returns. What's more, note in Table 10 the very strong correlation between physical real estate multi-factor indices and inflation during the 1973-1998 period. This feature is present to a lower extent if one uses the INSEE/Notaires Index, and is quite absent if one were to use the IPD Index.

Correlation Coefficient	***Housing Index***	***Mixed Index***	***Commerce Index***	***Office Index***	***INSEE/Notaires Index***	***IPD Index***
Price Index	0,955	0,943	0,966	0,949	0,815	0,504

Table 10

Conclusion

Our study has enabled us to answer a number of questions concerning real estate risk :

- Are we capable of identifying real estate risk factors? Yes.
- Can we distinguish between systematic risk and idiosyncratic risk in real estate? Expressed differently, amongst the risk factors identified, can we point to those that investors really value? Yes.
- What is the maximum number of systematic factors? Two or three.
- Is it necessary to proceed to separate analysises depending on property purposes? Yes.
- Can a thourough analysis of transactions data put to light certain specific market participant behaviour? Yes.
- Has there been a real estate speculative bubble during the late 80s-early 90s? Yes and no depending on the investment horizon.

In this paper we have tried to exhibit systematic sources of risk on the basis of real estate transactions data. We have started by studying the particularities of the semester capital return variable. Interesting behaviours appeared, and in particular one which opposes market participants whose investment horizon is rather short and are keen on arbitrage opportunities (with capital markets for instance), to those with a much longer investment horizon who await a return linked to more fundamental factors such as rents and unemployment.

We then explored the possibility of deriving a factor model of real estate returns, capable of capturing the market's fundamental movements. Our methodology, backed by both Principal Component Analysis and "stepwise" WLS regression techniques, was aimed at bringing forward consistent explanatory variables.

The results are globally encouraging. Firstly, they provide intuition and explanations as to the role played by such variables as rents, listed real estate, and unemployment in return formation. Next, they indicate that there exist combinations of a priori explanatory variables capable of representing in a satisfactory manner systematic real estate returns for different market sectors.

What's more, the results obtained point to the geographical variable as being of little use in explaining real estate risk or return. This might seem surprising to certain categories of real estate professionals. It seems here that location may indeed play a role in the determination of the *transaction price level* but not in real estate *returns or rates of return*.

Our study, grounded in observed transactions prices and dates, has the merit of exposing risk factors, but is not totally operational from the point of view of the real estate investor. These are two reasons to this. The first is that the specific dynamics of risk are neglected since they are considered diversifiable. For such a huge and fragmented market, diversification is not always possible of feasible, in particular when specific risk represents nearly half total risk as it seems to be the case here. The second is linked to the predictability of the factor model which is not necessarily an easy exercise. This question is also linked to the question of whether such a factor model could prove useful in elaborating hedging strategies in real estate investment management.

In sum, the real estate investor is in need of an index representing, in a precise manner, the physical market's movements. Such an index has to be grounded in a method that is both able to capture and separate the systematic as well as the specific market dynamic and that would not be too sensitive to the choice of explanatory variables. Our focus is now turned to this precise direction.

REFERENCES

Atkinson, A.C. (1985), *Plots, Transformations and Regression, UK* : Clarendon Press.

Barthélémy F., M. Baroni and M. Mokrane (2001), « Dynamique des risques en immobilier physique », mimeo

Cook, R.D. (1977), ''Detection of Influential Observations in Linear Regression'', *Technometrics*, 19, 15-18,

Friggit J., (1999), "Quels outils financiers pour une meilleure gestion des risques immobiliers", Ministère de l'Equipement, du Logement, des Transports et du Tourisme, Conseil Général des ponts et chaussées. www.equipement.gouv.fr, February.

Greene, W.H (1997), *Econometric Analysis*, Prentice Hall ed.

Hoesli M. (1993), *Investissement immobilier et diversification de portefeuille*, Paris, Economica.

Savel, F., (1998), « Gestion intermédiée d'actifs immobiliers, un secteur économique qui se structure », *Lettre économique de la CDC*, N°102, September.

Shiller, R. (1998), *Macro Markets, Creating Institutions for Managing Society's Largest Risks*, Clarendon Lectures in Economics, Oxford University Press.

Sirmans, G. S. and C. F. Sirmans, (1987) « The Historical Perspective of Real Estate Returns », *Journal of Portfolio Management*, spring, p. 22.