Data quality issues in the property market: some evidence and implications for financial intermediaries

Francesca Battaglia, Claudio Porzio, Gabriele Sampagnaro

EMFI WPs 4 - 2010

Data quality issues in the property market: some evidence and implications for financial intermediaries

Francesca Battaglia[a], Claudio Porzio[a] and Gabriele Sampagnaro[a]*

University of Naples Parthenope, Italy

Abstract

This paper addresses the issue of data quality in the real estate market. In many countries, the returns indices for direct markets are provided by several sources differing in terms of the methodology adopted and index weights. These differences produce a lack of informative standardization, which could negatively affect the ability of market participants to make predictions. Focusing on the Italian real estate market, the aim of this paper is therefore twofold: to investigate the reliability of property data sources, and to assess the impact for financial intermediaries involved in real estate investing. Our results show a significant level of divergence between the data, and considerable implications for those financial institutions dealing with them. These findings conflict with the requirements of an efficient (or at least sub-efficient) market.

Keyword: real estate, data divergence, IRR, LGD, efficient frontier.

Jel Classification: G11, G21, L85, L15

June 2010

Department of Research in Business and Finance at University Parthenope, Via Medina 40, Naples 80133, Italy; e.mail: sampagnaro@uniparthenope.it, phone /fax: +39.0815474851/ +39.0815522313.

Introduction

This paper[1] discusses issues concerning the quality of property data from different sources and resulting implications for market participants. It is divided into two sections. In the first section, we discuss the nature and availability of property data for the Italian market. Our initial exploration of the quality and accessibility of some domestic data documents the presence of many property data sources, each of which uses different methods of data collection. The high number of data sources and their methodological heterogeneity produce excessive data discrepancies, hardly compatible with efficient research and professional investment processes. Using a set of longitudinal aggregated property values, we proceed to estimate the level of uniformity of data using correlation and cointegration analysis.

The second section provides an examination of the potential effects of data non-uniformity on the decision-making process. To this end, we describe three simulations which deal with the implications of the lack of uniformity of data for mortgage loan lenders, real estate investment vehicles, and asset allocation planners.

1. Conceptual framework

In general, the nature of the real estate asset makes a convincing comparison with the other traditional asset classes difficult. Stephan (1999) examines the criteria behind collecting and evaluating real estate data, focusing on the difficulties encountered in compiling them as well as the various limitations of the data themselves. Some factors, such as the heterogeneity of property, the low frequency of transactions, and the high fragmentation (urbanization) of the markets, seriously limit the possibility of estimating real-estate relationships accurately (Sirmans et al., 1998). These features lead the research team to prefer secondary data rather than primary data. While primary data are gathered by researchers (specifically, for the problem at hand), secondary data are collected from other sources: universities, government agencies, real estate appraisals, and market research firms. Secondary data are certainly less expensive and less time consuming than primary data, but they are seldom expressed appropriately for the intended purpose. Moreover, transaction records frequently contain empty data fields, and class definitions seldom fit analysts' needs exactly (Greer and Kolbe, 2006).

Because of the indisputable link between economic forecasting and the reliability of data, many governments have taken on the task of gathering data, and their involvement plays a vital role in providing reliable and valid information for property market performance. In many cases, however, data is often inaccurate or incomplete at the time of collection, rendering adjustment procedures necessary to ensure the validity and reliability of the dataset available (Ge and Harfield, 2007).

A large number of agencies, and limited government involvement in data collection, represent two factors able to reduce greatly the efficiency of the real estate market. In those countries where these characteristics are found, there is also a noticeable lack of coordination and centralization between data-collectors, which negatively affects the possibility to calculate homogeneous estimates. As for implications for the market, the inefficiency of data collection systems undoubtedly affects the operation of financial institutions involved in the real estate sector, including performance

[1]Author's contribution: GS conceived the study, designed and implemented the cointegration analysis, the IRR and mean-variance simulations and wrote the manuscript. FB assembled the input data, conducted the correlation/ratio analysis and carried out the LGD simulation, and helped GS to write paragraph 3.2. CP participated in the design and coordination of the study and gave conceptual advice. All authors have read and approved the final manuscript.

measurement of real estate funds, mortgage loan management by banks, and the identification of the real estate weight in a mixed-asset portfolio of pension or mutual funds.

In general, two types of indices have been proposed as measures for real estate values: appraisal-based and transaction-based indices. Appraisal-based indexes are based on appraisal estimates while transaction-based indices on actual transactions. The low frequency of transactions experienced by the direct real market results in greater reliance on appraisal-based indices.

However, appraisal-based indices have various drawbacks, such as a methodological bias produced by the smoothing effect of the returns, which have been well documented in the literature[2]. The main issues deriving from returns smoothing include the overestimation of real estate in mixed portfolios, because of its unnaturally low correlation with the other asset classes, and methodological problems connected with the availability of real estate indices. These drawbacks give rise to the need for revision of the construction and measurement of real estate appraisal-based indices; indeed, the NCREIF Property Index itself, being an appraisal-based benchmark, is affected by the same kind of issues.

The autocorrelation of appraisal indices return is still under discussion in the literature, and there are contrasting opinions about its level of intensity. The relevance of appraisal smoothing is supported by a broad stream of studies; it is worth mentioning those by Geltner (1989), Geltner (1991), Ross and Zisler (1991), Quan and Quigley (1991), Giacotto and Clapp (1992), Geltner and Goetzmann (2000), Geltner and Ling (2006), Lin and Vandell (2007)[3]; a lower number of contributions attribute a smaller role to the smoothing phenomenon[4].

Finally, a third stream of studies, although more limited, is worth mentioning; this stream denies the existence of any autocorrelation between returns. Of these, the study by Lai and Wang (1998) can be quoted, in which the authors underline how some observations referring to the real estate market are justified by causes different from the serial correlation between real estate returns (i.e. the investor risk premium). In order to neutralize the return smoothing or to reset volatilities artificially lower than the real ones, several methodologies (originating from Geltner's studies) of return unsmoothing have been proposed[5].

When focusing on the real estate fund case, one of the more sensitive activities involved in its management is certainly the planning and monitoring of the expected financial results. From this point of view, the determination of an *interim* internal rate of return (IRR) represents an indisputably useful tool of measurement. The estimation of an expected internal rate of return must take into account multiple factors regarding the size and time of the in-coming and out-going cash flows. Concentrating on the incoming cash flows only, fund performance is typically attributable to the following three categories: a) ground rent; b) property sale values; c) returns from liquid management. Consistent with this consideration, the availability of reliable data for each of these sources of return (ground rent, property sale values, liquidity returns) constitutes an indispensable requisite from which to construct a financial plan endowed with objectivity and, in terms of its usability by third parties (investors, analysts, etc.), an adequate level of transparency. As a consequence, the unreliability of real estate data produces instability in IRR calculation (because it can change depending on the data source), thus hindering the making of appropriate investment choices.

Moving on to the case of the loan mortgage lender, the relationship between data quality and credit management is to some extent observable and measurable. The quality of property data affects the bank's residential mortgage management with regards, for example, to the Loss Given Default

[2] The smoothing effect is related to the tendency of a time series of returns to show a link between the return *t* with the previous *t-1*; smoothed returns leading to the following effects : 1) return autocorrelation; 2) low standard deviation of returns; 3) low correlation with the returns of other asset classes not characterized by smoothing.

[3] For a more detailed review of the appraisal bias literature see Yiu C., Tang B., Chiang Y. and Choy L. (2006).

[4] See Edelstein, R., Quan, D., "How Does Appraisal Smoothing Bias Real Estate Returns Measurement?" Journal of Real Estate Finance and Economic, 2006, 32 ; Webb, R. B., Miles M., and D.Guilkey, Transactions-Driven Commercial Real Estate Returns: The Panacea to Asset allocation Models?, AREUEA Journal, 1992, 20.

[5] See, for example, D. Geltner and W. Goetzmann (2000), D. Galtzaff and D. Geltner (2000), Bond, Hwang and Marcato (2006).

(LGD) estimation. In lending, Loss Given Default is the magnitude of likely loss on exposure and is expressed as a percentage of this.It is assumed that LGD may be explained largely by loan characteristics, the nature and the expected value of the underlying property, as well as variables measuring the default, foreclosure and settlement process (see Qi and Yang, 2009). The role of property values as an explanatory variable to predict LGD entails the use of reliable data sources. Despite extensive literature about the requirements for an accurate property appraisal and for high quality control from a management viewpoint, few papers focus on the link between the quality of property data sources and LGD predictions. According to the Italian version of the regulation for prudential supervision of banks, the exposures secured by real estate property have to satisfy certain conditions, including an adequate monitoring of property value used as collateral (Banca d'Italia, 2006)[6]. In an attempt to clarify the relationship between the quality of property data and mortgage loan management, we carry out some simulations out in order to submit the regular calculation of LGD to a variety of scenarios related to the set of available time series. The simulation results show the existence of a precise functional relationship between the value of the collateral and the data sources from a regulatory (capital requirements) and management (pricing policy) point of view.

Finally, we turn to the issue of the effect of data divergence on investment portfolio composition in a mean-variance framework. The role of real estate in diversifying mixed-asset portfolios has been well recognized in the literature (*inter alia*, Seiler et al, 1999). The majority of these studies focus on the effects of including real estate investment in a mixed-asset portfolio, revealing results consistent with a diversification benefit. This benefit is typically explained by some attributes of real estate investment, such as low correlation with other traditional asset classes, its suitability for inflation-hedging, and its high level of risk-adjusted performance, etc. (Hudson-Wilson, Fabozzi, Gordon, 2003).

After including a sub-set of comparable property indices (provided by different data sources) in a set of *n* financial asset classes, we launch a series of portfolio optimizations in order to identify the sensitivity of efficient frontier curvature to the property data source employed. In this case too, the results show how the non-uniformity of data constitutes a significant issue in ensuring correctness and validity in investment choices.

2. Study of domestic data harmonization

2.1 Data description

The set of data is composed of 21 time series, provided by 5 different data sources and containing the historical values of property indices for two geographical areas: Italy (10 out of 21) and Milan (11 out of 21). Each data-provider offers coverage of all or part of the traditional market segments: residential, commercial, office, and industrial[7]. The time interval of the data varies from a minimum of 5 to a maximum of 42 years as shown in Table 1.

A preliminary analysis of the data reveals the presence of a different frequency of time observations, since in some cases the index values are monthly while in others they are 6-monthly or

[6] "Accordingly: i) the value of the property shall be verified at least once every three years for residential property and once every year for commercial real estate, or more frequently where the market is subject to significant changes in conditions. Statistical methods may also be used to monitor the value of the property and to identify property that requires verification; ii) where the verifications under point i) reveal a material decline in the value of the property, a valuation shall be made by an independent valuer, based on a value that shall not exceed the market value", Banca d'Italia, 2006, pp. 21-22.

[7] The privacy disclaimers of some sources of data do not authorise the use of data for external research. For this reason, the historical series available are identified by code (data source 1, data source 2, etc.). As a guarantee of the truthfulness of the results, the authors are prepared to reveal their data sources upon private request (eieffe@uniparthenope.it).

yearly. To ensure the uniformity of the comparisons between data, we standardize the data frequency on a common quarterly basis using linear interpolation.

The adoption of a linear interpolation raises legitimate questions about the significance of a comparison between manipulated time series rather than raw time series. However, an interpolation generates a smoothing out of values and, in general, this contributes to the blunting of outliers rather than to their amplification. In other words, while the use of data interpolation certainly affects the correctness of the results, its most likely effect would be an underestimation of data dissimilarity which, from a prudential point of view, represents a more acceptable effect than an overestimation.

< INSERT TABLE 1 >

For each region, the data represents average prices over time of housing, commercial, office and industrial properties. Because our raw data for index values is expressed in different units of measurement, we proceed to standardize the time series families in order to allow a straight comparison between them. After converting the data into index numbers (with the base value equal to 100) we use a log transformation to stabilize the variance of series and then we estimate the first log difference. The choice of first log difference rather than log levels is explained in a series of graphs (omitted for brevity): while all the time series of log levels show an overall positive trend as a reflection of the domestic market upturn of the last decade, the scatter plot of the first log difference shows various contrasting movements between data related to the same property category but provided by different sources. This preliminary evidence reveals a substantial discrepancy in the rates of change of the indices, reflecting the lack of homogeneity in data collection methods.

The data were gathered adopting both transaction-based and appraisal-based methods, but we are unable to identify the prevalent approach used by each data source due to limited transparency and the incompleteness of the methodology descriptions available. While one of the five time-series families, named Source#5, certainly follows a transaction-based approach to gathering data, the other sources' data-collection methods of the other data sources appear indistinguishable.

2.2 Methodology

To investigate the discrepancies between data, we adopt a three-step analysis consisting of 1) a dissimilarity test, 2) correlation and 3) cointegration analysis. Thus, first of all in order to detect dissimilarities between the data, we implement a simple test based on the ratio between property index values. For each pair of comparable time series, we calculate the value of a ratio R_{XY}. The ratio R_{XY} is the simple average of the quotients between the values of two comparable time series (X and Y) with a time interval length of *m*:

$$R_{XY} = \frac{1}{m}\sum_{i=1}^{m}\frac{X_i}{Y_i} \qquad (1)$$

More precisely, two time series are comparable if they are related to a common time interval and to a homogeneous class of property. The interpretation of the ratio is straightforward: the closer the ratio gets to one, the more the two series analyzed will be statistically equal; conversely, the further the ratio moves away from one, the less homogeneous the series will be. To assess the significance of the relationship between the two series, we test the null hypothesis H_0: ratio = 1 by using the F statistic.

The second step is the calculation of the correlation matrix of property indices both for log-levels and for first-differences. The aim of the correlation analysis is to validate the previous graph (omitted) according to which the log-level time series appears to be characterized by a quasi-similar trend, while the change rate (first log difference) is not. The lack of a single data-gathering approach, and the consequent low level of standardization of information, make a low positive correlation coefficient probable, while the presence of a negative value would be considered unexpected and symptomatic of a more significant phenomenon of data divergence.

The cointegration analysis represents our third step towards reaching a definitive assessment on the issue of data uncertainty. The lack of homogeneity potentially observed in the two previous steps does not appear to be final since it does not take into consideration the possibility that despite the divergence between the returns in the interval observed, two or more series can show a long-term equilibrium relationship. For this purpose, one can proceed to verify the existence of a common trend between the time series, whose presence would moderate the opinion expressed about the dissimilarity between sources of property data, and the consequent inefficiency of market information processes.

Generally speaking, two variables are cointegrated if they have a common stochastic trend, that is, if they move together for a long period of time despite the trend not always being (visually) observable. More formally, two variables that are stationary in their first differences but non-stationary in their levels, are said to be cointegrated if there is a stationary linear combination between them. In order for the two historical series to be considered as cointegrated, it is necessary for both to be i) integrated by the same n level, and ii) their linear combination (i.e. *cointegration relationship)* to be integrated by a level less than n. The general relationship from which the identification of a cointegration phenomenon proceeds is:

$$y_t = \beta_0 + \beta_1 x_t + \xi_t \tag{2}$$

The model illustrated by equation (2) represents the so-called cointegration regression, and can be interpreted as the stochastic representation of the relationship that connects the variables to each other (it is also worth mentioning that $y_t = \beta_1 x_t$). The error term ξ_t is representative of the deviations from the equilibrium relationship. To test for cointegration it is therefore necessary to investigate the stationarity of the error term; in the case of stationarity of the residuals ξ_t there is cointegration between *X* and *Y*. When there is cointegration between the variables, and therefore a long-term relationship between them, assuming certain conditions, it would be possible to establish an ECM (*Error Correction Mechanism*) able to estimate the velocity of convergence of the dependent variable (Y) with the equilibrium relationship corresponding to each variation of the independent variable (X).

The considerations expressed above thus make it necessary to consider the issue of cointegration by means of an investigation into the level of stationarity present in the residuals from the cointegration regression (1). The stationarity of the residuals can be assessed in two ways: graphically or descriptively. A graphic assessment involves the observation of both the residuals (*Y-axis*) plotted against time (X-axis), and the residuals at time *t* (*Y-axis*) plotted against those at time *t-1*. If in either of these cases a stationary dynamic is evident, it is possible to plausibly hypothesize the presence of cointegration. The use of a statistical-descriptive method for identifying the presence of cointegration requires, on the other hand, that the regression of cointegration be submitted to a series of tests. The first is the Durbin Watson test (*DW test*) which, to summarize briefly, estimates the presence of autocorrelation between the residuals; a value of the DW statistic near to zero is indicative of the presence of autocorrelation in the residuals and, therefore, of their non-stationarity. Vice versa, a higher value of the DW statistic is indicative of stationarity in the residuals (the absence of autocorrelation) and, therefore, the presence of cointegration. A second investigative test (consistent with the approach adopted by Engle and Granger, 1987) consists in putting the residuals from cointegration regression through the ADF test (Augmented-Dickey-Fuller); when the residuals are stationary, there will be cointegration, while otherwise it is not possible to establish the presence of an equilibrium relationship in the long-term between the two variables (the variables are not cointegrated)[8]. A third investigation method for testing for the presence of the cointegration phenomenon is represented by the Phillips-Perron test (PP test); its mechanism can be approximated to that of the ADF test in so far as it offers the advantage of offering greater robustness to the heteroschedasticity of the error terms and does not require the choice of a lag optimal for the lags in its base regression.

2.3 Empirical results

The main results of the three-step process are reported in Table 2 (dissimilarity ratio, R_{XY}), Table 3 (correlation analysis), and Tables 4-5 (ADF-t and cointegration analysis) respectively.

< INSERT TABLE 2 >

We extended the R_{XY} ratio analysis (step 1) to the 21 couples of comparable time series in terms of time interval and class of property extracted from the original dataset. Achieving results with values close to 1 would be indicative of a convergence between data, while the results reported in Table 2 are consistent with a preliminary indication of the lack of harmonization between data.

The average value of R_{XY} for all the 21 cases is 1.196, with a standard deviation equal to 0.509 (max = 2.049, min= 0.037). If we distinguish between the two geographical areas, we note a slight increase in divergence for the indices relating to Italy: in this case the R_{XY} average value is 1.297 (with a standard deviation of 0.718, approx. 55%) while for the indices related to Milan, the R_{XY} average value is 1.134 (with a standard deviation of 0.345, approx. 30%).

The results of the correlation analysis seem to confirm the hypothesis of discrepancy in the data, especially for those relating to the national indices rather than a single urban area (Milan). The correlation matrix shown in Table 3 reveals a wide range of correlation coefficients, most of which are statistically significant.

< INSERT TABLE 3 >

With regard to the national indices, the range of correlations is $-0.698 \leq \rho \leq 0.721$ (with a standard deviation equal to 0.568), while for the city of Milan, we observe a smaller interval of $-0.2349 \leq \rho \leq 0.7484$ (st. dev.= 0.374). The presence of some negative signs in the correlation matrix is surprising since it reveals an outcome more compatible with a comparison between (different) asset classes rather than within a (similar) asset class. Although we cannot exclude some bias in the data, as we shall discuss shortly, these findings clearly demonstrate the existence of significant data divergence and raise some legitimate doubts about the informational efficiency of the domestic real estate market and the accuracy of information provided on it.

[8] Other investigation methods also exist for verifying the phenomenon of cointegration, including: the restricted vector autoregression test, RVAR; augmented restriction vector autoregression, ARVAR; the unrestricted vector autoregression test, UVAR; and the augmented restriction autoregression vector, AUVAR. For a review see Engle and Granger, 1991.

In both cases (ratio and correlation analysis), the incongruity of data-base systems appears stronger for the national index data (11 out of 21 indices). This result could be explained by the adoption of an advanced data collection procedure in the urban area of Milan (provided by the local board of trade) not yet widespread in the rest of the market.

Our interest in this issue invites a further level of investigation aimed at investigating the existence of a long-term relationship able to moderate or refine the assessment of dishomogeneity shown above. To this end, we proceed to run a cointegration test for the time series of national indices.

< INSERT TABLE 4 >

In order to enhance the statistical significance of the results, we select the time-series pairs with an adequate time interval, excluding from the cointegration analysis any data with a time-interval of less than ten years. Imposing this selection criterion, we obtain three pairs of time series provided by two data property sources (Source #1-Italy and Source #2-Italy), covering the residential, commercial and office sectors respectively. Each time series pair is then submitted to an ADF test (Augmented Dickey Fuller) to estimate its order of integration (see Table 4). The resulting cointegration outcomes are reported in summarized form in Figure 1, while the detailed presentation of the residual test results of the cointegration regression are given in Table 5.

< INSERT TABLE 5 >

< INSERT FIGURE 1 >

The results are consistent with the absence of cointegration for each of the cases analyzed, revealing non-negligible independence among data structures. Moreover, data show that the absence of cointegration is observed both for the historical series of absolute values (logarithmic levels) as well as for the returns series (first differences). Consequently, the lack of cointegration is an obstacle to achieving an ECM in order to approximate the convergence velocity to an equilibrium relationship between variables.

To summarise, the lack of a long-term relationship between data gathered from distinct sources, as well as the aforementioned characteristic of nonconformity, lead us to believe that the real-estate information systems are not at all adequate. However, some possible explanations of the empirical findings need to be explored. For example, the poor traceability of the valuation dates for the real estate portfolio to which the indices are linked impedes the correct synchronization of time series which, incidentally, may render the results of a comparison between information sources inefficient or implausible. Further, in spite of the traditional asset class markets (i.e. stocks and bonds), where the indices reflect the performance of a group of a similar type, real estate indices are based on portfolios that differ in terms of urban location and other hedonic variables (green areas, proximity to transport infrastructures, carparks, etc.).

3. Data quality and property management: implications for market participants

The accuracy of real estate data is a topic of interest to many market participants. In this section we develop three simulations to show how the level of data homogeneity impacts on some of the operations carried out by financial institutions involved in real estate investment management. The simulations refer respectively to: 1) the impact of disharmonized data on IRR calculation of real estate funds; 2) the sensitivity of LGD (loss given default) values to the pricing of the collateral for bank mortgage management processes; 3) the relationship between the weights of an optimal mixed-asset portfolio and the source of property data, in a mean-variance optimization framework.

3.1. The impact of time series heterogeneity on IRR funds: a simulation

To assess the impact of data divergence on the valuation of a real estate fund, we conduct a sequence of back-tests for the internal rate of return (IRR) calculation, adopting a different real estate data source for each iteration. As is well known, the IRR calculation of a real estate investment is a function of three parameters: 1) rental cash flow, 2) cash management, and 3) the end value of properties. Assuming a real estate fund with an extremely simplified structure of assets, we design a procedure of IRR backtesting consisting in an IRR sensitivity analysis, setting different values for the third of the above parameters (the end value of the properties in the portfolio), keeping the other two constant. The back-testing procedure is iterated n times, where n represents the number of sub-periods selected and related to the different property end values.

The modulation of end values follows a mechanism defined as follows: given the i_{th} (i=1..n) subperiod of m years, and given the availability of property index data provided by the j_{th} source (j=1..h), the i_{th} property end value is set as equal to the (hypothetical) initial value of the property compounded at m annual yields intrinsic to the corresponding time series interval. Following this mechanism, we select six five-year subperiods (n=6, and m=5) and three commercial property indices related to the city of Milan and provided by three different data-sources (h=3). The six subperiods started from 1998:1 and each is separated from the previous one by a year; thus we obtain the following sequence of subperiods: 1st) 1998:1-2002:12; 2nd) 1999:1-2003:12; 3rd) 2000:1-2004:12; 4th) 2001:1-2005:12; 5th) 2002:1-2006:12; 6th) 2003:1-2007:12.

We then assume a five-year investment in a real estate fund invested in only two properties (A and B) whose financial characteristics are illustrated in the upper part of Table 6. With these established conditions, for each of the three data sources selected, we calculate six property portfolio end values and, consequently, six IRR values (keeping the other cash flow constant). The results are shown in Table 6, where some sensitivity measures are used.

< INSERT TABLE 6>

The last row and last column of Table 6 show the standard deviation of the IRR "within" subperiods and "between" data-sources respectively. While analysis of the "within subperiod" indicator is not so important for our aim, an inspection of the results of the second indicator appears indispensable. By looking at the results of "between data-source" standard deviation, it becomes clear how the choice of data sources may affect the evaluation of the IRR in each subperiod; this influence is also significant in some cases, and varies between 2.5% and 25.6%.

These findings are consistent with the previous remarks about the existence of scarcely negligible data divergence for the Italian real estate market. In general, the results of this IRR simulation confirm how important it is to have access to comprehensive, reliable and timely evidence of property transactions in order to make informed predictions, and how this represents an issue of great concern to both market participants and policymakers who rely on price signals for decision-making (Lum, 2004).

3.2. Reliability of real estate time series and LGD evaluation

Recent turbulence in international financial systems originating in the mortgage markets highlights the close relationship between developments in real estate prices and the soundness of the financial sector (Koetter and Poghosya, 2008). The exposure of a bank to the real estate market has implications for the sensitivity of property-collateralized loan value to housing price changes. We address this issue through a simulated Loss Given Default calculation. According to the risk management terminology adopted in the Basel II framework, Loss Given Default (LGD) denotes the fraction of exposure that will not be recovered following default. The purpose of the simulation described here is to analyze how the real estate data source affects LGD prediction for a property-secured loan. In general, the regulatory formula for a collateralized loan can be approximated as follows:

$$LGD = 1 - \frac{\sum_{t=1}^{n} \frac{E(RV)_t}{(1+i)^t} - \sum_{t=1}^{n} \frac{Exp_t}{(1+i)^t}}{EAD} \qquad (3)$$

The term *E(RV)* denotes the expected recovery value of the collateral (property or properties). While cash recoveries are easy to evaluate, the evaluation of non-cash recovery, such as the repossession of properties, is complex, and can be tackled on an ad-hoc basis using reliable property market information. *Exp* indicates post-default administrative expenses and can be split into direct costs (court costs, attorney and bailiff fees, cost of appraisal, etc.), and indirect costs (i.e. the operating costs of the lender's recovery department). The discount rate *i* can be inferred by following an appropriate risk-based hypothesis, while *t* is the time at which the lender obtains recoveries or pays the costs. Finally, EAD is the exposure at the time of default, i.e. the sum of the flows relating to the position outstanding, and discounted at the date of default.

For the purpose of our analysis, we use a simplified back-testing technique, which analyzes the sensitivity of LGD predictions to the different expected recovery values estimated using a set of properly-selected property data sources. In particular, we select six time-series Milan housing prices from four sources of data split into two groups: three indices of new housing prices and three indices of previously-occupied housing prices. We then assume default by the debtor three years from the starting date (1997:11) on a secured loan repayment, the structure of which is illustrated in the upper part of Table 7. By setting a predetermined purchase price (housing market value at t_0), we simulate six probable property recovery values at time t_{rec} subsequent to the time of the debtor's default t_{def} (t_{def} >t_{rec}). Each recovery value is calculated as the property market value at t_0 compounded at the annual growth rate extracted over the period t_0-t_{def} in one of the six time series employed.

The results of the LGD-data sensitivity relationship are shown in Table 7, where different LGD values are given. The confirmation of LGD sensitivity to the property data source is consistent with the results which emerged from the previous analytical test (see Section 2), and raise some issues about the risk management processes of lenders (principally banks).

< INSERT TABLE 7>

The IRB method (internal rating method) introduced by the Basel Committee on Banking Supervision, requires banks, under certain conditions, to carry out stress tests on their portfolio of loans in order to measure overall exposure to credit risk and, consequently, to adapt their capital adequacy. The available data on which the stress tests are based are typically obtained from both internal and external sources. Internal data consist of loan information, default outcomes and internal payment records. Data obtained externally instead consist of accounts, external payment records and property data. This confirms the fact that banks are interested in using accurate estate databases, since there is a correlation (albeit slight) between the quality of external data sources and the assessment of portfolio risk (and therefore a prudent capital requirement). Moreover, this is a further argument for the involvement of financial authorities, such as central banks, in the collection and dissemination of reliable real estate data.

3.3 Data divergence, portfolio optimization and investment choices

Finally, we come to the last issue discussed in this paper: the relationship between data property divergence and the quality of investment choices. The basic idea is to select a set of asset class indices, including domestic real estate, and to create a sequence of portfolio optimizations, varying the property data at each iteration. By changing the property index at each optimization, we analyze the sensitivity of portfolio weights to the data source switch, measuring the consequent implications for the investment choices with an appropriate variable (DARaP, see below).

The tenet of portfolio theory is diversification within a mix of asset classes with an appropriate risk-return profile and a low correlation, to mitigate risk to the whole portfolio. In spite of its limitations (Chopra and Ziemba, 1999), the Markowitz Mean-Variance approach is widely used and represents the most suitable model for achieving optimal portfolio selection.

Using the classical principles of efficient frontier construction, we select 5 asset classes and estimate their expected returns, as well as their covariance matrix. The set of asset classes is made up of equity, bond, and real estate indices listed as follows: 1) S&P500 (US stock market); 2) MIBTEL (Italian stock market); 3) MTS BTP 10Y (long-term domestic government bonds); 4) domestic risk free-rate (MTS BOT); 5) a property index selected from the available set (Section 2.1 and Table 1).

The estimation of efficient frontier input represents an issue widely discussed in the literature. However, the merely descriptive purpose of this paragraph leads us to choose a simplified approach rather than more refined models (e.g. the Black and Litterman model, the Bayes-Stein approach, etc.). Thus, the expected returns are expressed as the annualized average of historical quarterly returns; the historical approach is then extended to the estimation of the covariance matrix.

From the available set of property indices, we recognize three triplets of comparable time series belonging to three data sources and related to both geographical area and the three main real estate segments (residential, commercial, office). For each triplet (i.e. for each data source) we can potentially proceed to the construction of three efficient frontiers, by changing the property index at each optimization iterate. However, to improve the significance of optimization outcomes, we exclude from the subset of (nine) series those with less than ten years of data. Imposing this criterion, we identify two time series triplets (six series), provided by two different sources and related respectively to the residential, commercial and office segments of the city of Milan. We then carry out six portfolio optimizations (one for each time series), obtaining three pairs of comparable efficient frontiers as shown in Figure 2. The expected returns and risks are estimated on an annual basis and are equal to the historical average and the standard deviation for the period 1997:6 to 2008:6 respectively[9].

To determine the sensitivity of portfolio composition to each data-change, we use a proxy of return/risk ratio for each frontier, which we call DARaP (Decile Average Risk adjusted Performance). In detail, the mean DARaP variable may be explained as follows: for each efficient frontier it represents the average value of the return-to-risk ratio of ten "decile portfolio", where this term describes the portfolio with a risk equal to a decile of the volatility interval (max σ - min σ). Formally, we write:

$$DARaP = \frac{1}{10}\sum_{i=1}^{10}\frac{R_i}{\sigma_i} \tag{4}$$

where with Ri and σi we denote the return and risk (σi) respectively of the optimal portfolio corresponding to the ith decile of the volatility interval of the frontier.

< INSERT FIGURE 2>

In general, if we calculate the DARaP for two efficient frontiers (A and B) differing in terms of the source of data of one (or more) asset classes, a proxy of the sensitivity of portfolio composition to the data change would be shown by the value $\Delta DAR3_{A,B}$, where:

$$\Delta DARaP = \left|\frac{\max\{DARaP_A, DARaP_B\} - \min\{DARaP_A, DARaP_B\}}{\min\{DARaP_A, DARaP_B\}}\right| \tag{5}$$

The $\Delta DARaP_{A,B}$ variable captures the geometric translation of the efficient frontier when a data change occurs. Thus, a high (low) value of $\Delta DARaP_{A,B}$ is consistent with discrepancies (convergence) between sources of data. The results of efficient frontier comparison are summarized in Table 8, where rows indicate the data source of the property index inserted in the portfolio optimization and the $\Delta DARaP_{A,B}$ values, while the columns are indicative of the category of real estate indices.

< INSERT TABLE 8>

The ΔDARaP value is between 29.22% and 46.27%, revealing a significant change in portfolio weights due to the substitution of the property data source. These findings are consistent with those of the previous simulations, and suggest much caution is needed in the selection of the property benchmark to include in portfolio optimization tests, especially for those which are mean-variance based. The most serious practical limitations of the mean-variance approach are, in fact, the ambiguity and instability of portfolios. Small changes in input assumption often lead to large changes in the composition of optimized portfolios (Michaud 1998). Therefore, optimal weights will change

[9] For reasons of brevity, the covariance matrix and expected return are not shown here, but are available from the author on request.

significantly over time as a direct result of making estimation errors (Kallberg and Ziemba, 1984 and Adler, 1987). Thus, in the case of a high level of divergence between property indices (e.g. the office sector in Table 8), and to impede the amplification of estimation errors, it would be appropriate to adopt a procedure at least able to mitigate the discrepancy of the data (i.e. the calculation of average index values).

4. Conclusions.

Data quality plays a vital role in providing reliable and valid information for property market performance. Its relationship with the assessment of financial stability and monetary policy is much debated among academics and policymakers alike. The complexity of the market itself and differences in its functioning impede the adoption of standardized data collection procedures in different countries. Thus, gathering reliable and comparable data on property markets has proved very difficult (Zhu, 2005). Furthermore, it is not uncommon to find markets where multiple, very different data collection methods coexist.

By focusing on the Italian real estate market, we have discussed the reliability of domestic property data sources, taking into account variables such as the frequency of collection, data-gathering methodology, and the area covered. Furthermore, we have conducted three simulations in order to measure the impact of data divergence for, respectively, real estate investment vehicles, loan mortgage lenders, and the asset allocation of optimized portfolios. Our results show a poor level of homogeneity between data both for national time series and for urban data time series. These findings raise the issue of how important it is to have quick access to comprehensive and reliable evidence of property transactions in order to make informed predictions, and how this represents a critical question for both policymakers and market participants who rely on price signals for decision-making. Looking forward, there is the need for action aimed at improving the quality of property data and enhancing the comparability of across-data sources.

References,

Adler, M., Global Asset Allocation: Some Uneasy Questions, *Investment Management Review*, September/October, 1987, 13-18.

Banca d'Italia, "New regulations for the prudential supervision of banks", circular n. 263 of 27 December 2006.

Bank for International Settlements, Real estate indicators and financial stability, Bis Paper, 2005, Vol. 21.

Biasin, M., L'investimento Immobiliare, Il Mulino Bologna: 2005.

Bond, S., Hwang, S. and MARCATO, G., Evaluating Unsmoothing Procedures for Appraisal Data, *ARES 2006, NUS Symposium 2006*.

Cacciamani, C., (a cura di), Il Rischio Immobiliare: una Soluzione di Rating dell'Investimento Immobiliare, EGEA, 2003

Carretta, A., Fiordelisi, F., Mattarocci, G., New Drivers of Performance in a Changing Financial World, Palgrave MacMilliam: HoundMills, 2008.

Cesarini, F., (a cura di), Banca e Finanza Immobili*are,* Bancaria Editrice: Rome, 2003.

Cheng P., and Roulac, E., Measuring the Effectiveness of Geographical Diversification, *Journal of Real Estate Portfolio Management*, 2007, 13, 29-44.

Chopra, V. K.. and Ziemba, W. T., The effects of errors in means, variances, and covariances on optimal portfolio choice, *Journal of Portfolio Management*, 1993, 20, 6–11.

Engle, R. F. and Granger, C., Cointegration and error correction: representation, estimation and testing, *Econometrica*, 1987, 55, 251-76.

Engle, R F. and Yoo, B. S., Forecasting and testing in co-integrated systems, *Journal of Econometrics*, 1987, 35, 143-159.

Engle, R.F. and Granger, C.W.J., *Long Run Economic Relationships*, Oxford University Press, 1991.

Fisher J., Geltner D. and Webb, R., Value Indices of Commercial Real Estate: A Comparison of Index Construction Methods, *Journal of Real Estate Finance and Economics*, 1994, 9, 137-164.

Gatzlaff, D. and Geltner, D., A Transaction-Based Index of Commercial Property and its Comparison ot the NCREIF Index, *Real Estate Finance*, 1998, 15, 7-22.

Ge, X. and Harfield, T., The Quality of Data and Data Availability for Property Research, The 13th Pacific Rim Real Estate Society Conference, Perth, Australia, January 2007.

Geltner, D., Bias in Appraisal-Based Returns, *AREUEA*, 1989, 17, 338-352.

Geltner, D., Smoothing in Appraisal-Based Returns, *Journal of Real Estate Finance and Economics*, 1991, 4, 327-345.

Geltner, D., and Goetzmann, W., Two Decades of Commercial Property Returns: A Repeated Measure Regression-Based Version of the NCREIF Index, *Journal of Real Estate Finance and Economics*, 2000, 21, 5-21.

Geltner, D., and Ling, D., Indices for Investment Benchmarking and Return Performance Analysis in Private Real Estate, *International Real Estate Review,* 2007, 10, 17-36.

Giaccotto, C. and J. Clapp, Appraisal-Based Real Estate Returns under Alternative Market Regimes, *AREUEA Journal*, 1992, 20, 1-24.

Giannotti, C. and Mattarocci, G., Real Estate Selection and Portfolio Construction model: Data Analysis from the Italian Market, in Carretta, A., Fiordelisi, F., Mattarocci G., (eds). New drivers of performance in a changing financial world, Palgrave MacMilliam: HoundMills, 2008.

Giliberto, M., A Note on the Use of Appraisal Data in Indexes of Performance Measure, *Journal of the American Real Estate and Urban Economics Association*, 1988, 77–83.

Greer, G. E., and Kolbe, P. T., *Investment analysis for real estate decisions*, 6th ed., Chicago: Deaborn Real estate Education, 2006.

Hudson-Wilson, S., Fabozzi, F J. and Gordon, J.N., Why Real Estate? - An Expanding Role for Institutional Investors, Journal of Portfolio Management (Special Real Estate Issue), 2003.

Hung, K., Onayev Z. and Tu, C., Time-Varying Diversification Effect of Real Estate in Institutional Portfolios: When Alternative Assets Are Considered, *Journal of Real Estate Portfolio Management*, 2008, 14, 241-261.

Jorion, P., Portfolio optimization with constraints on tracking error, *Financial Analysts Journal*, 2003, 59, 70-82.

Kallberg, J.G., and Ziemba, W. T., Mis-Specification in Portfolio Selection Problems, in Bamberg, G. & Spremann, A. (eds.). Risk and Capital, Springer-Verlag: New York, 1984.

Koetter, M., and Poghosyan, T., Real estate markets and bank distress, Discussion Paper Series 2: Banking and Financial Studies, Deutsche Bundesbank, Research Centre,2008, 18.

Knight, J. R., Sirmans, C.F., Gelfand, A.E. and Ghosh, S. K., Analyzing real estate data problems using the Gibbs sampler, *Real Estate Economics*, 1998, 26, 469-492.

Lai, T. and K. Wang, Appraisal Smoothing: The Other Side of the Story, *Real Estate Economics*, 1998, 26, 511-535.

Lin, Z., and Vandell K., Illiquidity and Pricing Biases in the Real Estate Markets, Real Estate Economics, 2007, 35, 291-330.

Lum, S. K., Property price indices in the Commonwealth – Construction methodologies and problems, *Journal of Property Investment & Finance*, 2004, 22, 25-54

Markowitz, H. M., Portfolio Selection", *Journal of Finance*, 1952, 7, 77–91.

Michaud, R., Efficient Asset Management, New York: Oxford University Press 1998. First published by Harvard Business School Press.

Phillips, P. C. B. and Perron, P., Testing for a Unit Root in Time Series Regression, *Biometrica*, 1988, 75, 335-346.

Phillips, P. and Ouliaris, S., Asymptotic Properties of Residual Based Tests for Cointegration, *Econometrica*, 1990, 58, 165-93.

Qi, M. and Yang, X., Loss Given Default of High Loan-to-Value Residential Mortgages, *Journal of Banking & Finance*, 2009, 33, 788-799.

Quan, D. and Quigley, J., Price Formation and the Appraisal Function in Real Estate, *Journal of Real Estate Finance and Economics*, 1991, 4, 127-146.

Ross, S. and Zisler R., Risk and Return in Real Estate, *Journal of Real Estate Finance and*

Economics, 1991, 4, 175-190.

Seiler, M.J, Webb, J.R. and Meyer, F.C.N., Diversification Issues in Real Estate Investment, *Journal of Real Estate Literature* , 1999, 7, 163-179.

Stephan, V.A., Obtaining real estate data: criteria, difficulties and limitations" BIS Papers chapters, in: Bank for International Settlements (ed.), *Real estate indicators and financial stability*, 2005, 21, 63-69.

Yiu, C.Y., Tang, B.S., Chiang, Y.H. and Choy, L.H.T., Alternative Theories of Appraisal Bias, *Journal of Real Estate Literature*, 2006, 14, 321-344

Zhu, H.B., The importance of property markets for monetary policy and financial stability, in: Bank for International Settlements (ed.), Real estate indicators and financial stability, 2005, 21, 9-29.

Table 1.
Data Description: time intervals and geographical markets

Real estate market: Italy					
Real estate category	Source#1	Source #2	Source #3	Source #4	Source #5
Residential	1988-2008	1997-2008	2002-2007	n. a.	n. a.
Commercial	1988-2008	1997-2008	n. a.	n. a.	n. a.
Office	1988-2008	1997-2008	2002-2007	n. a.	n. a.
Industrial	n. a.	1997-2008	2002-2007	n. a.	n. a.
Real Estate market: Milan					
Real estate category	Series#1	Series#2	Series#3	Series#4	Series#5
Residential	1965-2008	1993-2007	n. a.	2001-2008	1995-2007
Commercial	1965-2008	1993-2007	n. a.	2001-2008	2001-2007
Office	n. a.	1993-2007	n. a.	n. a.	1997-2007
Industrial	n. a.	1993-2007	n. a.	n. a.	n. a.

Table 2
Dissimilarity ratio analysis

1st Panel – Dissimilarity ratio between national time series

Ratio denominator	***S.rce# 1 retail***	***S.rce #1 comm.***	***S.rce #1 office***	***S.rce #2 retail***	***S.rce #2 comm.***	***S.rce #2 office***	***S.rce #2 ind.***	***S.rce #3 retail***	***S.rce #3 office***	***S.rce #3 ind.***
	Ratio numerator									
Sources#1_ret	1									
Source#1_com		1								
Source#1_off			1							
Source#2_ret	0.601**			1						
Source#2_com		1.1504			1					
Source#2_off			1.113			1				
Source#2_ind							1			
Source#3_ret	2.048**			1.887				1		
Source#3_off			0.036**			2.021**			1	
Source#3_ind							1.5131			1

2nd Panel – Dissimilarity ratio analysis between Milan time series

Ratio denominator	***S.rce# 1 retail***	***S.rce #1 comm.***	***S.rce #2 retail***	***S.rce #2 comm.***	***S.rce #2 office***	***S.rce#4 retail***	***S.rce#4 comm.***	***S.rce #4 office***	***S.rce #5 retail***	***S.rce #5 comm.***
	Ratio numerator									
Source#1_ret	1									
Source#1_com		1								
Source#2_ret	1.162		1							
Source#2_com		0.739		1						
Source#2_off					1					
Source#4_ret	1.374		0.9598			1				
Source#4_com		0.9768		1.234			1			
Source#4_off					1.371			1		
Source#5_ret	1.108		0.879			0.848			1	
Source#5_com		2.074		1.145			0.875			1

*Note: ** statistically significant at 95% (H_0: ratio=1; H_A: ratio ≠ 1); grey cells indicate time series pairs which are not comparable for analysis purposes.*

Table 3
Correlation matrix

1st Panel – Correlation coefficients between national time series

Market: Italy	***Source#1 retail***	***Source#1 Office***	***Source#1 comm.***	***Source#2 retail***	***Source#2 office***	***Source#2 comm.***	***Source#2 industrial***	***Source#3 Retail***	***Source#3 office***	***Source#3 industrial.***
Source#1_ret	1									
Source#1_off	0.9171**	1								
Source#1_com	0.8765**	0.8431**	1							
Source#2_Iret	**0.2700**	0.0401	0.041	1						
Source#2_off	0.6366**	**0.6233****	0.4855**	0.2661	1					
Source#2_com	0.246	0.0832	**-0.0867**	0.2164	0.484**	1				
Source#2_ind	0.1749	0.2056	0.0725	0.0217	0.4206**	0.1338	1			
Source#3_ret	**-0.6722****	-0.6622**	-0.695**	**-0.6982****	-0.4458**	-0.469**	0.5829**	1		
Source#3_off	0.3126	**0.5508****	0.2913	0.6186**	**0.7212****	-0.0119	-0.2218	-0.51	1	
Source#3_ind	-0.4242	-0.5042**	-0.4134	-0.6812**	-0.2679	-0.2123	**0.4401**	0.7571**	-0.5459**	1

2nd Panel – Correlation coefficients between time series for Milan city

Market: Milan	***Source#1 retail***	***Source#1 comm.***	***Source#2 retail***	***Source#2 comm.***	***Source#2 office***	***Source#2 ind.***	***Source#4 retail***	***Source#4 comm.***	***Source#4 Office***	***Source#5 retail***	***Source#5 comm.***
Source#1_ret	1										
Source#1_com	0.4659**	1									
Source#2_ret	**0.4528****	0.4393**	1								
Source#2_com	0.466**	**0.3524****	0.6909**	1							
Source#2_off	0.4739**	0.5137**	0.9337**	0.6828**	1						
Source#2_ind	0.4866**	0.4134**	0.7826**	0.8327**	0.8088**	1					
Source#4_ret	**0.5837****	0.5212**	**0.8937****	0.711**	0.8328**	0.7526**	1				
Source#4_com	0.418**	**0.6421****	0.7546**	**0.4125****	0.7694**	0.5299**	0.8317**	1			
Source#4_off	0.3932**	0.6909**	0.7137**	0.3271**	**0.7484****	0.3955**	0.74**	0.9788**	1		
Source#5_ret	**-0.2349**	0.2592	**0.2641**	0.6578**	0.1799	0.3489**	**0.4976****	0.3628**	0.3082	1	
Source#5_com	0.154	**-0.0188**	-0.1962	**-0.1623**	-0.0788	0.1561	-0.2921	**-0.2098**	-0.1752	-0.2066	1

Note: ** statistically significant at 95% (H_0: $\rho=0$; H_A: $\rho \neq 0$). Bold numbers indicate the correlation of two comparable time series. The time intervals for each correlation coefficient are shown in Table 1.

Table 4
ADF Unit Root Test (time interval: 1997:6 to 2008:6)

Time Series	R.E. category	Levels	*p*-value	1st /2nd Difference	*p*-value	Order of integration
Source#1	*Residential*	-1.290 (4) [-3.536]	0.8904	-4.562 (4) [-3.544]	0.0012	I(2)
Source#2	*Residential*	-1.364 (4) [-3.536]	0.8712	-6.577 (2) [-3.536]	0.0000	I(2)
Source#1	*Commercial*	-2.178 (1) [-3.524]	0.5021	-4.081 (1) [-3.540]	0.0067	I(1)
Source#2	*Commercial*	-2.565 (1) [-3.524]	0.2963	-3.741 (1) [-3.528]	0.0198	I(1)
Source#1	*Office*	-2.734 (1) [-3.524]	0.2222	-4.863 (4) [-3.540]	0.0004	I(1)
Source#2	*Office*	-3.150 (1) [-3.524]	0.0948	-4.385 (4) [-3.536]	0.0023	I(1)

Note: Augmented Dickey Fuller (ADF) test to check the stationarity of a series under the null hypothesis that series is non-stationary. We present a model with trend and constant. The ADF statistics are obtained from:

$$\Delta x_t = a_0 + b_0 u_{t-1} + \sum_{j=1}^{p} c_{0j} \Delta x_{t-j} + \varepsilon_t$$

where Δ is the difference operator, a_0, b_0 and c_0 are the coefficients to be estimated, x is the variable whose time series are examined and w is the white-noise error term. Values in parentheses show the lag length of the ADF test. Values in square brackets indicate 5% critical value adopted from MacKinnon (1991). Details of the ADF regression (trend and constant)are not included to save space but are available on request.

Table 5
Cointegration analysis between two real estate data sources

Real estate category		*Residential*		*Commercial*		*Office*	
dep variable		log (Source#2)	Δlog(Source#2)	log (Source#2)	Δlog(Source#2)	log (Source#2)	Δlog(Source#2)
ind. variable		log(Source#1)	Δlog(Source#1)	log(Source#1)	Δlog(Source#1)	log(Source#1)	Δlog(Source#1)
				Cointegration regression			
β (t value) [p –value]		1.59 (28.47) [0.000]	0.634 (1.83) [0.074]	0.818 (29.88) [0.000]	-0.116 (-0.57) [0.571]	0.916 (56.32) [0.000]	0.745 (5.23) [0.372]
R^2 (adjR^2)		0.948 (0.947)	0.0739 (0.0724)	0.953 (0.952)	0.075 (-0.0156)	0.986 (0.986)	0.388 (0.374)
	Test statistic			***Residual- based test***			
CRDW[a]	DW	0.052 (*1.03*)	0.495 1.03	0.104 1.03	0.647 1.03	0.1261 1.03	0.878 1.03
ADF[b]	ADF-t	-0.744 . (lag 1) (*-3.5136*)	-2.545 (lag 1) (*-3.5136*)	-2.946 (lag 1) (*-3.5136*)	-2.908 (lag 4) (*-4.7690*)	-3.264 (lag 3) (*-4.3993*)	-3.195 (lag 1) (*-3.5136*)
PP [c]	Z_p	-2.33 (*-19.42*)	-18.2 (*-19.34*)	-5.03 (*-19.42*)	-16.1 (*-19.34*)	-5.91 (*-19.42*)	-18.1 (*-19.34*)
	Z_t	-1.56 (*-3.52*)	-3.47 (*-3.52*)	-1.53 (*-3.52*)	-3.13 (*-3.52*)	-1.72 (*-3.52*)	-3.306 (*-3.52*)

[a]The critical values of the cointegrating regression Durbin-Watson test are reported in Engle and Yoo (1987).
[b]The critical values for the ADF test are from MacKinnon (1991). The lag length was chosen according to the Schwartz criterion.
[c] The critical values of the Phillips-Perron test are taken from Philips and Ouliaris (1990).
The numbers in italics in parentheses are critical values.

Figure 1.

Summary of cointegration analysis results

So urc #1 - Italy		**Source #2-Italy**	
	Residential	*Offices*	*Commercial*
Residential	not		
Office		not	
Commercial			not

*Note: *the absence of cointegration is observed for both the levels and first differences.*

The outcomes of the residual-based test are reported in Table 5.

Table 6
Simulating IRR calculation: main results.

Assumptions:

Portfolio composition	*Date of investment*	*Date of liquidation*	*Initial Price*	*Annual Rental*	*Costs*
Property A	t_0	t_5	100	1	0
Property B	t_0	t_5	200	2	0

End values of the fund

Sub- period	Jan/1998-Dec/2002	Jan/1999-Dec/2003	Jan/2000-Dec/2004	Jan/2001-Dec/2005	Jan/2002-Dec/2006	Jan/2003-Dec/2007	SDWSP*
Data Source #1	413.3	409.1	409.3	414.2	399.2	388.9	2.41%
Data Source #2	433.3	691.7	565.2	453.1	433.3	339.8	25.47%
Data Source #3	416.6	444.4	413.3	389.9	364.5	345.8	9.19%
*SDDS***	2.5%	29.9%	19.2%	7.6%	8.6%	7.5%	

Internal Rate of Return (IRR) of the fund

Sub period	Jan/1998-Dec/2002	Jan/1999-Dec/2003	Jan/2000-Dec/2004	Jan/2001-Dec/2005	Jan/2002-Dec/2006	Jan/2003-Dec/2007	SDWSP
Data Source #1	18.1%	17.9%	17.9%	18.1%	17.5%	17.0%	2.49%
Data Source #2	19.0%	28.3%	24.1%	19.8%	19.0%	14.6%	22.87%
Data Source #3	18.2%	19.4%	18.1%	17.0%	15.8%	14.9%	9.67%
*SDBDS***	2.5%	25.6%	17.6%	7.6%	9.0%	8.3%	

**SDWSP: standard deviation within sub-periods*
***SDBDS: standard deviation between data sources*
SDWSP and SDBDS are expressed as percentages of the IRR average value

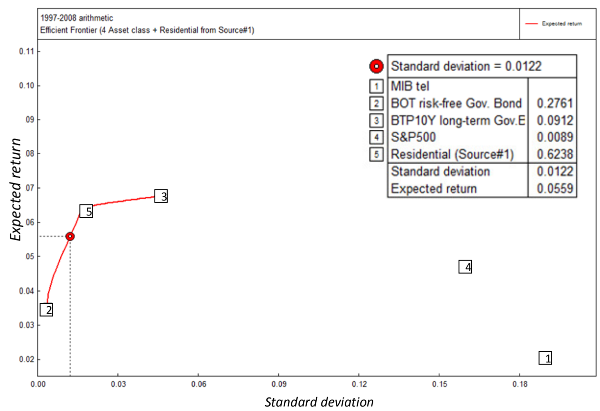
Table 7
Simulation of Loss Given Default prediction

Simulation assumptions			
Loan amount	€ 600,000	Date of default	t_{def}= t_4 (2001)
Predetermined property purchase price (€)	€ 300,000	Date of recovery*	t_5 (2002)
Loan start date	t_0 (1997:11)	Exp*	1.8% of outstanding
Date of first installment payment	t_1 (1998:1)	Discount rate	3%
Annual installment (€)	€ 44,149.05	EAD**(€)	€ 598,709.77

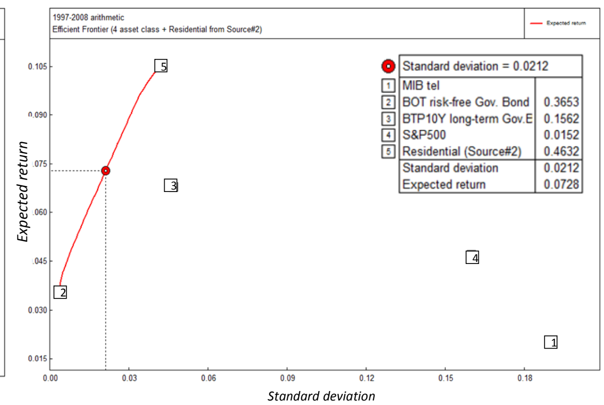
Previously- occupied housing prices (Milan)		*New housing prices (Milan)*	
Time series	Estimated LGD	Time series	Estimated LGD
Source#1	42.00%	*Source#1*	38.48%
Source#2	35.98%	*Source#3*	36.45%
Source#3	31.33%	*Source#4*	33.30%

Notes: ** Exp indicates the default administrative expenses; its value (1.8%) is consistent with the results of a central bank survey (see Supervisory Bulletin of the Bank of Italy n.12, 2001).*
*** Assuming default occurred in 2001, the exposure at default, EAD, is equal to the sum of the annual installment for that year and the one following, discounted at the default time. The LGD values assume property repossession one year after the default.*

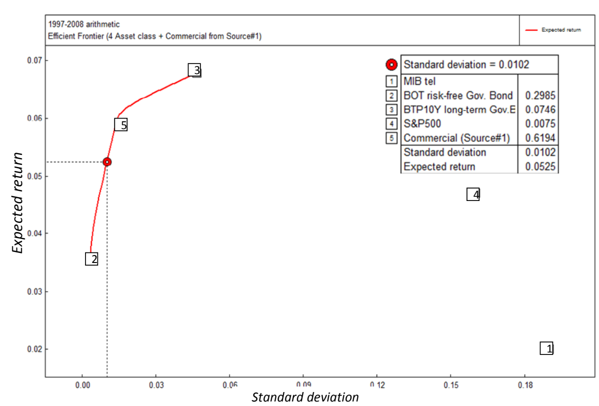
Figure 2
Set of comparable efficient frontiers (input is historical values, 1997:6 – 2008:6)



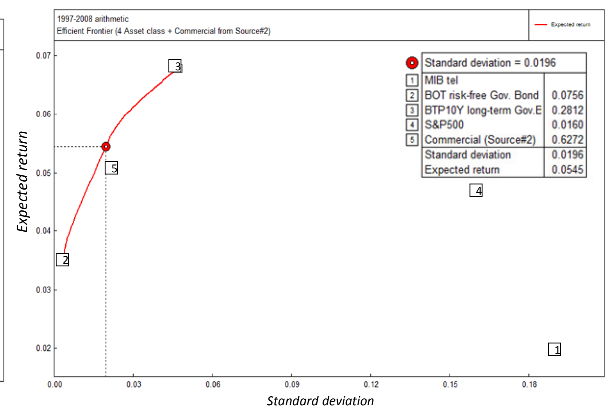
Real Estate Class: Residential Index (Source #1)



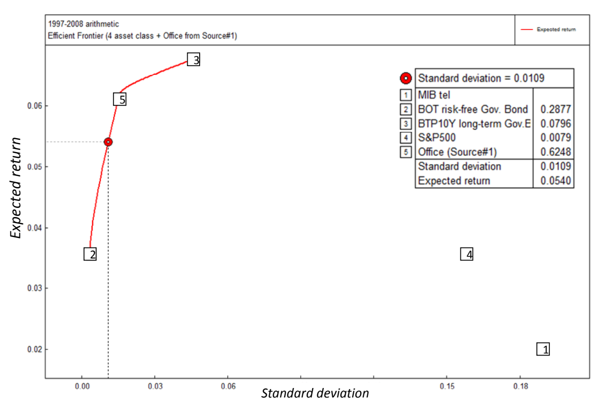
Real Estate Class: Residential Index (Source #2)



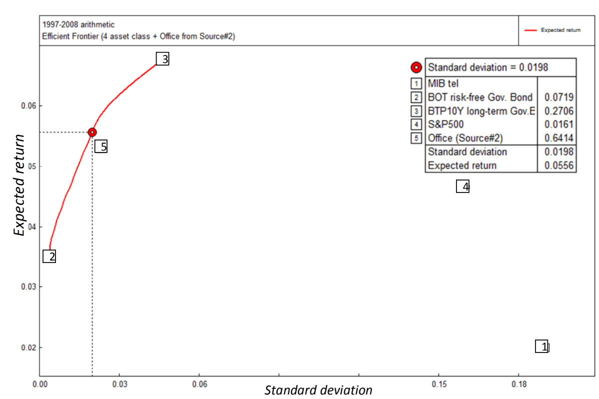
Real Estate Class: Commercial Index (Source #1)



Real Estate Class: Commercial Index (source #2)



Real Estate Class: Office Index (Source #1)



Real Estate Class: Office Index (Source #2)

Table 8
Map of *Decile Average Risk adjusted Performance* (DARaP) values

	Real estate category		
Data	*Residential*	*Commercial*	*Office*
Source #1 (A)	3.668	4.727	5.349
Source #2 (B)	4.747	3.527	3.657
$\Delta DARaP_{A,B}$ *(%)*	29.22	34.02	46.27

For each efficient frontier DARaP denotes the average value of the return to risk ratio of ten "*decile portfolio*", i.e. a portfolio with a risk equal to one decile of the frontier volatility interval (max σ - min σ).

Appendix.
Data Summary

#	Data source	market	Real estate category	Original frequency of obs.	Starting date	Ending date	Nominal /Real
ITALY							
1	Source#1	Italy	Residential	6-monthly	1988, 2nd quarter	2008, 2nd quarter	Real
2	Source#1	Italy	Commercial	6-monthly	1988, 2nd quarter	2008, 2nd quarter	Real
3	Source#1	Italy	Office	6-monthly	1988, 2nd quarter	2008, 2nd quarter	Real
4	Source#2	Italy	Residential	monthly	1997, 1st quarter	2008, 3rd quarter	Real
5	Source#2	Italy	Commercial	monthly	1997, 1st quarter	2008, 3rd quarter	Real
6	Source#2	Italy	Office	monthly	1997, 1st quarter	2008, 3rd quarter	Real
7	Source#2	Italy	Industrial	monthly	1997, 1st quarter	2008, 3rd quarter	Real
8	Source#3	Italy	Residential	yearly	2002, 1st quarter	2007, 2nd quarter	Real
9	Source#3	Italy	Office	yearly	2002, 1st quarter	2007, 2nd quarter	Real
10	Source#3	Italy	Industrial	yearly	2002, 1st quarter	2007, 2nd quarter	Real
MILAN							
11	Source#1	Milan	Residential	6-monthly	1965, 2nd quarter	2008, 2nd quarter	Real
12	Source#1	Milan	Commercial	6-monthly	1965, 2nd quarter	2008, 2nd quarter	Real
13	Source#2	Milan	Residential	Yearly	1993, 1st quarter	2007, 4th quarter	Real
14	Source#2	Milan	Commercial	Yearly	1993, 1st quarter	2007, 4th quarter	Real
15	Source#2	Milan	Office	Yearly	1993, 1st quarter	2007, 4th quarter	Real
16	Source#2	Milan	Industrial	yearly	1993, 1st quarter	2007, 4th quarter	Real
17	Source#4	Milan	Residential	6-monthly	1stsem 1998	1stsem 2008	Real
18	Source#4	Milan	Commercial	6-monthly	2001, 1st quarter	2008, 2nd quarter	Real
19	Source#5	Milan	Residential	6-monthly	1995, 4th quarter	2007, 4th quarter	Real
20	Source#5	Milan	Commercial	6-monthly	1997, 2nd quarter	2007, 4th quarter	Real
21	Source#5	Milan	Office	6-monthly	1997, 2nd quarter	2007, 4th quarter	Real

www.adeimf.it