

Robust Desmoothed Real Estate Returns

Jean-Christophe Delfim*

Martin Hoesli**

29 February 2020

*University of Geneva
Geneva School of Economics and
Management
40 boulevard du Pont-d'Arve
CH-1211 Geneva 4
Switzerland

Email: jean-christophe.delfim@unige.ch

Phone: +41 22 379 9264

**University of Geneva
Geneva School of Economics and
Management
40 boulevard du Pont-d'Arve
CH-1211 Geneva 4
Switzerland

University of Aberdeen Business School,
U.K., and

Kedge Business School, France

Email: martin.hoesli@unige.ch

Phone: +41 22 379 8122

Abstract

This research starts from the observation that common desmoothing models are likely to generate some extreme returns that will distort risk measurement and hence can lead to investment decisions that are suboptimal relative to those that would be made if a transaction-based index were available. Thus, we propose to improve the desmoothing models by incorporating a robust filter into the procedure. We report that in addition to properly treating for smoothing, the method prevents the occurrence of extreme values. As shown with U.S. data, our method leads to desmoothed series whose characteristics are akin to those of transaction-based indices.

Keywords: Desmoothing models; Robust filter; Appraisal-based index; Private real estate; Unlevered REITs.

JEL codes: C32, C61, G10, R33

Introduction

Smoothing is a well-known issue encountered when relying upon appraisal-based indices that are often the only source of information on returns for private investments such as real estate. It originates from two main sources. First, the temporal aggregation bias which appears because assets are not appraised at a single point in time but rather over a given window (Geltner 1993a). Second, the anchoring bias which relates to the behavior of appraisers who when assessing the current value of a property rely heavily on the value they estimated for the previous period, while making some adjustment to account for changes in the economic environment (Geltner 1989, Diaz and Wolverton 1998). Smoothing leads to return series with high autocorrelation levels, which in turn implies that standard deviations are shrunk toward zero, while higher moments of the return distribution may also be distorted. This could potentially have serious consequences in various applications such as mixed-asset portfolio optimization with the share of real estate being overstated because of its perceived low level of risk. Thus, correcting appraisal-based series is crucial for a meaningful risk assessment.

Models designed to deal with the smoothing bias have existed for quite some time. The most common desmoothing model explicitly represents the appraiser behavior mentioned above (Quan and Quigley 1989, Geltner 1993b, Fisher and Geltner 2000). Despite the fact that this classical type of desmoothing model treats smoothing by correcting the high autocorrelation and the dampened volatility estimates, one however often observes extreme values generated through the desmoothing process when changes in the underlying appraisal-based return series are large. These extreme outlying values potentially lead to a wrong estimation of real estate risk and can alter the return distribution features with respect to what they should be for the unobserved true market returns. More sophisticated specifications of the aforementioned desmoothing model, such as the regime-switching model developed by Lizieri, Satchell and Wongwachara (2012), are found to be able to limit extreme desmoothed values. However, some outliers still remain even with such models.

The critical issue of extreme values generated by desmoothing models is emphasized in the example below. Figure 1 compares the total returns of an uncorrected appraisal-based index (NPI), a transaction-based index (TBI), and a desmoothed version of the appraisal-based index¹. The corresponding summary statistics are reported in Table 1, in addition to the resulting optimal allocation to real estate in a mixed-asset portfolio². We observe first that the NPI underestimates volatility, with its standard deviation of 2.07% being half the 4.32% observed for the TBI. The expected shortfall at 5% (ES 5%) and the maximum drawdown (MaxDD) are underestimated as well. Second, the desmoothed NPI usually displays a pattern close to the one of the TBI, but we identify extreme unrealistic returns around the early 1990s and the 2008 global financial crisis (GFC). Therefore, risk measured either by volatility, ES 5%, or MaxDD is overestimated. In addition, the desmoothed series cannot correct the low skewness measured on the original series and strongly amplifies its kurtosis.

As the average return of real estate is about the same whatever the series we consider, the features mentioned above have an impact on the Sharpe ratio and hence on asset allocation. Indeed, the NPI overestimates the Sharpe ratio, while the desmoothed series underestimates it, compared with the TBI one. Then, the optimal allocation to real estate is very large (73.2%) when relying on the NPI and only 21.2% with the desmoothed series, instead of the 38.2% with the transaction-based series. This simple example demonstrates that treating extreme values

¹ This preliminary analysis is performed on U.S. quarterly data from 1984Q2 to 2018Q3. For real estate, we use the appraisal-based NCREIF NPI index, a transaction-based index (TBI), and a desmoothed version of the NPI. The TBI series we use is a spliced series of the former hedonic MIT TBI (MTBI) index until 2010Q4 and of the current NCREIF TBI (NTBI), which uses a sale price to appraisal ratio method (SPAR). We refer to this spliced series as TBI. Motivations about why we use both the MTBI and the NTBI series are further detailed in the Appendix A, where we conclude that the NTBI is comparable to the MTBI since 1997Q2. The desmoothing parameter α we use for this example is equal to 0.8, which is a simple common choice for quarterly data. Also, the NPI index is adjusted in terms of property type mix in order to match the composition of the TBI index.

² The Sharpe ratio is reported in annualized terms and is computed taking the returns in excess of the three-month T-Bill rate. For the portfolio allocation results, stocks and bonds are considered along with real estate through the U.S. MSCI and the Citigroup U.S. 7-10 years Treasury bonds indexes, respectively. The optimal allocation is defined as the one maximizing the Sharpe ratio.

generated through the desmoothing process is at least as important as desmoothing appraisal-based series per se for a proper assessment of investment risk and any related decision making.

[Figure 1 here]

[Table 1 here]

In this paper, we propose to address the issue of extreme values generated through the desmoothing process by applying robust statistical techniques to desmoothed series. Indeed, robust techniques have been devised to correct unlikely extreme observations (Huber 1972). In particular a whole stream of the robust literature has developed filters for time series (Fried, Einbeck and Gather 2007), which would allow to treat outliers in desmoothed series in a simple and convenient manner. Hence, we proceed as follows. First, we desmooth appraisal-based series using the classical model and its regime-switching version. We then apply a robust filter to the desmoothed series and we compare these series with the transaction-based series to identify which of them better matches the true market return series. In addition, both desmoothing models and their robust versions are compared with the following alternative methods: (1) the Getmansky, Lo and Makarov (2004) model that was originally developed for the hedge funds industry and is often applied to private equity return series as well as its robust version; (2) a multivariate desmoothing model inspired by the Pedersen, Page and He (2014) and Cho, Hwang and Lee (2014) models; and (3) a method relying on indices of unlevered REIT returns (Fisher, Geltner and Webb 1994)³.

The paper contributes to the literature in the following ways. First, we compare various indices that can be used as a proxy for real estate performance measured by transaction-based benchmarks. Second, we suggest relying on robust statistical techniques for improving series obtained with desmoothing models. For this purpose, we design a flexible robust time series filter. Third, in addition to assessing the performance of various methods by comparing the

³ Robust versions of the latter two methods are not needed as they rely on observed rather than appraised data.

return characteristics with those of their transaction-based counterparts, we formally analyze similarities of and dissimilarities between series by applying a wide range of statistical tests. These tests not only focus on the central tendency and dispersion of return distributions, but also on higher distribution moments, on linkages with the reference transaction-based series, and the propensity to generate extreme desmoothed values.

We report the following main results. First, we show that all of the desmoothing methods we consider are able to correct at least some of the drawbacks of appraisal-based indices. However, various characteristics of the series we obtain with these methods are not always in line with those of transaction-based series. For instance, the multivariate desmoothing model produces returns that still exhibit very high autocorrelation figures. Also, the Getmansky, Lo and Makarov (2004) model tends to provide series that still underestimate the transaction-based series volatility, while the usual desmoothing model, and to a lesser extent its regime-switching version, leads to overestimated volatility, kurtosis, and skewness negativity due to extreme values generated through the desmoothing process. Second, once the robust filter is applied to the desmoothed series, most of the outliers are correctly identified and treated. Indeed this filter proves particularly useful in dealing with unrealistically large returns generated by the desmoothing process and yields series whose characteristics are much closer to those of the corresponding transaction-based series, allowing for more precise estimation of real estate risk. Third, we report that unlevered REIT series have distribution features generally in line with the benchmark. Finally, we identify as the best methods the robust regime-switching model, the robust classical model, and to a lesser extent the unlevered REIT method.

The remainder of the paper is structured as follows. After the literature review in the next section, we present the robust filter we design in the methodology section. The data and results sections follow, while a final section contains some concluding remarks.

Literature Review

As discussed by Geltner, MacGregor and Schwann (2003), the smoothing issue has been approached using various methods. In the context of mixed-asset portfolio construction, early studies recognizing the issue of real estate volatility being dampened due to smoothing proposed simple correction rules. For example, Firstenberg, Ross and Zisler (1988) take real estate volatility to be equal to that of stocks. Also from a portfolio allocation perspective, Blundell and Ward (1987) propose a simple desmoothing model relying on two elements. First, they assume that the private real estate market is efficient and hence that the true unobserved return series follows a random walk. Second, they model the way appraisers rely on both past appraised values and current market information in order to estimate the current value of a property. This second element is referred to as the appraisal rule, or transfer function, that is presented in detail in Appendix B. Combining both elements yields the desmoothing parameter that must be applied to treat the appraised series for smoothing and get an estimate of the true unobserved return series. Quan and Quigley (1989) also describe how smoothing arises due to appraiser's behavior as suggested by the aforementioned transfer function. Quan and Quigley (1991) simplify the previous model and augment it with a buyers and sellers model. They consider that appraisers extract relevant signals from the whole market as well as from each particular property.

The Geltner (1989, 1991) model relies on the Blundell and Ward (1987) transfer function and on assumptions regarding the true real estate volatility. Among the volatility values assumed in the literature, Fisher, Geltner and Webb (1994) propose to take half of the stock market volatility, while Hoesli, Lekander and Witkiewicz (2004) suggest it should be equal to the average of the volatilities of stocks and bonds. Alternatively, Geltner, Miller, Clayton and Eichholtz (2007), for example, propose to determine the optimal desmoothing parameter

according to the time span it should take for the appraised values to converge toward the market price⁴.

Initial desmoothing models usually rely on a random walk describing the generating process of the unobserved returns because they assume that the private real estate market is efficient. This implies that the autocorrelation in the true returns series must be zero. Nevertheless, as this assumption is likely to be unrealistic, Geltner (1993b) as well as Barkham and Geltner (1994) relax it, which makes it possible to represent the return generating process differently. As suggested, a possible way is to model it as an autoregressive process of order one, an AR(1) process. This class of models is referred to as reverse engineering. Fisher and Geltner (2000) further develop the model and apply it successfully to the NCREIF Property Index (NPI). Cho, Kawaguchi and Shilling (2003), improving the Fisher, Geltner and Webb (1994) model, also relax the random walk assumption and maintain that this helps achieving more realistic results. More sophisticated processes than an AR(1) can also be used. For example, with ARFIMA(p, d, q) models, Bond and Hwang (2007) document lower levels of smoothing than the ones reported in previous studies, while Bond, Hwang and Marcato (2012) report that the smoothing level is lower at the property level than at the aggregate level.

The possibility of having a time-varying desmoothing parameter is recognized by Quan and Quigley (1989, 1991). Matysiak and Wang (1995) also suggest that the magnitude of smoothing should depend on market conditions. Clayton, Geltner and Hamilton (2001) conclude that smoothing is the greatest when the market is the most illiquid; however, the range in which the time-varying desmoothing parameter value evolves is close to the value of the fixed desmoothing parameter. Chaplin (1997) considers up to six different regimes with a specific desmoothing parameter for each. The parameter value is supposed to be the largest

⁴ For this purpose, they suggest using the following formula: $T = \frac{1}{1-\alpha} - 1$. Assuming it takes one year for appraised values to converge toward market price, α must be equal to 0.5 using yearly data and equal to 0.8 using quarterly data.

when changes in rents and capitalization rates are the lowest. However, in the latter study, the desmoothing parameter values are not estimated but rather they are set a priori with respect to various assumptions. This limitation is removed in the regime-switching model proposed by Lizieri, Satchell and Wongwachara (2012) allowing for optimal parameter estimation. This model replaces the autoregressive desmoothing process in the Geltner (1993b) model by a threshold autoregressive process (TAR). By also allowing for a threshold autoregressive return generating process these authors achieve the so-called TAR-TAR desmoothing model.

Treating the smoothing bias is not specific to real estate but is also of importance for any private market for which mainly appraisal-based series are available. Various approaches have been proposed for other asset classes. In the hedge fund literature, Getmansky, Lo and Makarov (2004) investigate the case of illiquidity and smoothing and consider that the current observed smoothed return is a weighted average of previous unobserved true returns with several lags. Starting from a similar approach than Geltner (1991), they however make various assumptions regarding the true return generating process and the smoothing profile. Their model is applied by Cumming, Hass and Schweizer (2013) to private equity series. Also in the private equity literature, Shepard and Liu (2014) develop a model relying on Bayesian techniques in which a single parameter links smoothed and desmoothed returns. The smoothing parameter is estimated at the same time as parameters modeling the structure of desmoothed returns.

The studies discussed above describe models using mainly information from the series to be desmoothed itself. However, several researchers have proposed relying on additional information to explain the smoothing phenomenon and hence treating appraisal-based return series (Geltner 1989, Wang 2006). For example, Ling, Naranjo and Nimalendran (2000) use a latent variable model to estimate the desmoothed return series while Fu (2003) uses REIT returns as additional information in the Quan and Quigley (1991) model.

Cho, Hwang and Lee (2014) apply a conditional desmoothing model relying on fundamentals, such as economic and financial series, to estimate the smoothing level in a time-varying way. They report that smoothing increased gradually since the mid-1990s while it was inexistent before. Smoothing is positively driven by credit spread and negatively by investor sentiment, short-term interest rates, inflation, and the term spread. It is also higher during periods of uncertainty. Pedersen, Page and He (2014), who rely on the work by Scholes and Williams (1977) as well as by Dimson (1979), consider appraisal-based returns as a moving average of current and past unobserved true returns depending on exogenous risk factors. It results that the observed smoothed returns are linear combinations of weighted lagged risk factors. The estimated parameters of this model provide the impacts of current realizations of each risk factor on the current true returns. Then, parameter estimates and risk factors are combined in order to build the unobserved true return series.

Finally, rather than attempting to desmooth appraisal-based return series, another stream of the literature proposes to use adjusted series of REIT returns. This line of thought, initiated by Fisher, Geltner and Webb (1994), takes advantage of the information transmission channel existing between the public and the private real estate markets as reported for example by Hoesli, Oikarinen and Serrano (2015) and Ling and Naranjo (2015).

As discussed above, a variety of desmoothing models and associated methods are available. In this context, Marcato and Key (2007) emphasize that correctly calibrating the desmoothing model is even more important than the choice of the desmoothing model itself. In this research, we propose improvements to the desmoothed series based on the observation that commonly used models tend to generate extreme and implausible returns when changes in the original appraisal-based series are relatively important. This notably leads to an overestimation of risk and misleading values for the higher moments of the return distributions. In order to overcome this issue, we suggest including robust statistical methods in the desmoothing process. Robust

statistics consists of identifying unlikely outlying data and reducing, or even annihilating, their influence in the estimation process. Regarding the case of robust techniques for time series several methods exist, such as repeated median (Siegel 1982), least median of squares regression (Rousseeuw 1984), deep regression (Rousseeuw and Hubert 1999), as well as weighted repeated median and associated methods (Fried 2004, Fried, Einbeck and Gather 2007).

Methodology

Desmoothing Models and Related Methods

We desmooth the appraisal-based return series with in turn the Geltner (1993b) autoregressive model (AR hereafter), the Lizieri, Satchell and Wongwachara (2012) regime-switching model (TAR-TAR hereafter), the Getmansky, Lo and Makarov (2004) model (GLM hereafter), and a multivariate desmoothing model (MtV hereafter). Returns of an unlevered REIT index are also used for comparison purposes. The models are presented in Appendix B.

The Robust Time Series Filter

The common AR desmoothing procedure often generates extreme values when the underlying series displays relatively large changes. This is sometimes also the case even with the TAR-TAR model despite the additional flexibility brought by its specification. Treatment of extreme, outlying, observations can be addressed by means of robust statistical techniques. We propose to apply a robust time series filter to desmoothed series in order to fix the issue.

The filter we propose is inspired by Fried (2004) and Fried, Einbeck and Gather (2007). It relies on time-varying central and dispersion distribution parameters with a specific rule for outlier detection and correction⁵. We want a rule that gradually shrinks outlying values

⁵ Rules proposed by Fried (2004) are trimming, downsizing, and winsorization. Trimming implies that observations lying over a given threshold distance from the central value are set to be equal to the central value. Downsizing consists of setting to a closer distance to the center observations being over a given threshold usually defined as a multiple of the dispersion parameter value. Finally, winsorization is similar to downsizing but the new value of outliers is the detection threshold itself.

depending on how far from the center they are. We decide to rely on a loss function designed for robust estimation (Huber 1981) and we select the Tukey biweight (or bisquare) function. Such a function lowers weights of observations whose distance from the center of the distribution is unlikely. The more unlikely the observation, the more shrunk toward the central value of the distribution it is. The value is set equal to the central value for very unlikely observations that trespass a given threshold, the so-called tuning constant, c_w . In order to gauge if an observation is likely with respect to the whole observed distribution, the function requires information on the central parameter of the distribution, μ , and on its dispersion parameter, σ . In the usual static framework, the former can be taken as the average or the median, and the latter as the standard deviation or median absolute deviation.

As we work in a dynamic framework, several adjustments must be made to estimate time-varying Tukey function parameters. Fried (2004) models the time series to be filtered as a succession of regressions on consecutive observation intervals, where the trend, i.e., the central parameter, is estimated by Least Median of Squares (LMS) or Repeated Medians (RM), while the volatility, i.e., the dispersion parameter, is estimated by the Length of the Shortest Half (LSH), the Median Absolute Deviation (MAD) or one of its alternatives, the Rousseeuw and Croux's Q_n and S_n (Rousseeuw and Croux 1993).

We adopt a different approach and consider that the trend and the volatility processes can be expressed in the manner of an ARMA(1, 1)-GARCH(1, 1) model⁶. More precisely, in our case, for the volatility model we rely on an Exponentially Weighted Moving Average model (EWMA_v hereafter) which can be expressed as a specific case of the IGARCH(1, 1) model, as reviewed in Bollerslev (2010)⁷. Regarding the mean model used to estimate the time-varying

⁶ Filtered time series must be stationary. Stationarity is tested by the augmented Dickey-Fuller, Elliot-Rothenberg-Stock, Phillips-Perron, and Zivot-Andrews unit root tests. If stationarity is rejected by a majority of tests, series differences are taken for computing robust weights. Then, those weights are applied to the original series. All return series in this study are found to be stationary.

⁷ The IGARCH model is written as $\hat{\sigma}_t^2 = \omega + \alpha x_{t-1}^2 + \beta \hat{\sigma}_{t-1}^2$ and the EWMA_v model imposes $\omega = 0$, $\beta = \lambda \in [0, 1]$, and $\alpha = 1 - \beta$. Despite the restrictions it involves, the EWMA_v model has the advantage of being

central tendency parameter we follow the same reasoning and apply an EWMA for mean (EWMA_m hereafter) as for example shown in Cox (1961) and Lucas and Saccucci (1990). This way of estimating a series time-varying trend is also related to Holt filters (Holt 2004). In the same way as ARMA(1, 1) and GARCH(1, 1) are combined, both EWMA_m and EWMA_v are employed together (Reynolds and Stoumbos 2006). EWMA_m and EWMA_v are defined by equations (1) and (2), respectively:

$$\hat{\mu}_t = \lambda_m \hat{\mu}_{t-1} + (1 - \lambda_m) x_{t-1} \quad (1)$$

$$\hat{\sigma}_t = \sqrt{\lambda_v \hat{\sigma}_{t-1}^2 + (1 - \lambda_v) (x_{t-1} - \hat{\mu}_{t-1})^2} \quad (2)$$

where x_{t-1} is the realization of a time series at time $t - 1$ and $\lambda_m \in [0, 1]$ and $\lambda_v \in [0, 1]$ are the decay factors, representing the weight pertaining to past estimates for the processes of mean and volatility, respectively.

Let's assume for the moment that we know the relevant values for both λ parameters and hence can compute $\hat{\mu}_t$ and $\hat{\sigma}_t$ series⁸. We iteratively estimate these parameters on successive intervals $[1: t - 1]$ with $t = [t_{init}: T]$ ⁹. For each interval, we extrapolate the estimates of the central parameter, $\hat{\mu}_t$, and the volatility parameter, $\hat{\sigma}_t$, of the next period. The Tukey loss function, ψ_t , and the robust weight, w_t , are therefore computed as shown in equations (3) and (4). In order to introduce more flexibility in the filter by allowing for asymmetry, we consider two tuning constants $c_{w,U}$ and $c_{w,L}$ for the upper bound and the lower bound of the filter, respectively.

parsimonious and very simple to implement. In the context of our desmoothing model, in which we use relatively low frequency (quarterly) data, the simple specification of EWMA_v appears to be sufficient.

⁸ For example, Bollen (2015) suggests λ parameter values from around 0.70 to 0.90 for daily data.

⁹ For the initial periods of observations $t = [1:t_{init}]$ $\hat{\mu}_{t-1}$ and $\hat{\sigma}_{t-1}$ are estimated by the median and the median absolute deviation, respectively. We choose $t_{init} = 12$ quarters.

$$\psi_{U,t} = \begin{cases} \frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t} * \left(1 - \left(\frac{\frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t}}{c_{w,U}} \right)^2 \right)^2 & ; \text{if } 0 < x_t < \hat{\mu}_t + c_{w,U} * \hat{\sigma}_t \\ 0 & ; \text{if } x_t \geq \hat{\mu}_t + c_{w,U} * \hat{\sigma}_t \end{cases} \quad (3)$$

$$w_t = \frac{\psi_{U,t}}{\left(\frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t} \right)} = \left(1 - \left(\frac{\frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t}}{c_{w,U}} \right)^2 \right)^2$$

$$\psi_{L,t} = \begin{cases} \frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t} * \left(1 - \left(\frac{\frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t}}{c_{w,L}} \right)^2 \right)^2 & ; \text{if } 0 > x_t > \hat{\mu}_t + c_{w,L} * \hat{\sigma}_t \\ 0 & ; \text{if } x_t \leq \hat{\mu}_t + c_{w,L} * \hat{\sigma}_t \end{cases} \quad (4)$$

$$w_t = \frac{\psi_{L,t}}{\left(\frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t} \right)} = \left(1 - \left(\frac{\frac{x_t - \hat{\mu}_t}{\hat{\sigma}_t}}{c_{w,L}} \right)^2 \right)^2$$

Then, the original x_t value is replaced by its robust estimation computed as:

$$x_t^{rob} = w_t x_t + (1 - w_t) \hat{\mu}_t \quad (5)$$

We now discuss further the question of the λ_m and λ_v parameters, because instead of taking them as a priori values, we can estimate them. This is done by optimization of a forecast evaluation criterion. We consider criteria proposed by Andersen, Bollerslev and Lange (1999) and tested by Bollen (2015) for EWMA models. Since volatility clusters are expected, we rely on the Heteroscedasticity-adjusted Mean Absolute Error (HMAE) defined as:

$$HMAE = \frac{1}{T} \sum_{t=1}^T \left| 1 - \frac{\sigma_t^2}{\hat{\sigma}_t^2} \right| \quad (6)$$

where σ_t^2 is the realized volatility and $\hat{\sigma}_t^2$ is the estimated volatility. However, as we assume the existence of outliers, we propose a few adjustments to make this metric more robust. Instead of relying on the Mean Absolute Error (MAE) we prefer the MAD which is not sensitive to the presence of outliers in the data. Conserving the notion of adjustment for heteroscedasticity, the above criterion is then transformed as Heteroscedasticity-adjusted Median Absolute Deviation for volatility (HMADv) written as:

$$HMADv = Median \left(\left| 1 - \frac{\sigma_{t=1}^2}{\hat{\sigma}_{t=1}^2} \right|, \dots, \left| 1 - \frac{\sigma_{t=T}^2}{\hat{\sigma}_{t=T}^2} \right| \right) \quad (7)$$

and

$$\sigma_t^2 = (x_t - \hat{\mu}_t)^2 \quad (8)$$

HMADv is a function of λ_v because $\hat{\sigma}_t$ is expressed as in (2). Then the optimal λ_v is obtained by minimizing HMADv. Of course, an estimate for $\hat{\mu}_t$ is required. The estimation of the $\hat{\mu}_t$ series itself requires an estimate for λ_m which is obtained in the same manner as demonstrated above for λ_v . We need an equivalent criterion to HMADv called HMADm and defined as:

$$HMADm = Median \left(\left| 1 - \frac{x_{t=1}}{\hat{\mu}_{t=1}} \right|, \dots, \left| 1 - \frac{x_{t=T}}{\hat{\mu}_{t=T}} \right| \right) \quad (9)$$

Substituting $\hat{\mu}_t$ by (1) in (9), HMADm now only depends on λ_m and x_t , the observed series to be filtered, as the initial value, $\hat{\mu}_{t_{init}}$, is known. Then, HMADm is minimized with respect to λ_m . Once the optimal decay factor value, $\hat{\lambda}_m$, is obtained we use it in the EWMA_m formula (1) to get $\hat{\mu}_t$. The next step is to replace $\hat{\mu}_t$ in (2) and (8). Then, (2) in addition to (8) are substituted in (7) in order to obtain a HMADv criterion depending only on λ_v and the observed series x_t and the known initial value, $\hat{\sigma}_{t_{init}}$. Finally, HMADv is minimized with respect to λ_v . Following

this procedure, we get optimized λ parameters for EWMA_m and EWMA_v¹⁰. These series of time-varying parameters $\hat{\mu}_t$ and $\hat{\sigma}_t$ are used in the Tukey function as described above in equations (3) to (5).

The estimation of the tuning constants is detailed in Appendix B and indicates that the optimal values should be $c_{w,U} = 5$ and $c_{w,L} = -6.25$. The difference in absolute value between parameters is due to the slight negative skewness of real estate return distributions.

Note also that the robust filter is applied on the desmoothed series obtained with the AR, the TAR-TAR, and the GLM models. This is not the case for the series obtained with the MtV model, because as they result from a linear combination of factor loadings and observed factors series we believe that they do not suffer from artificially generated outliers¹¹. The unlevered REITs series are not filtered either as they are based on observed market returns.

Data

Regarding the choice of the transaction-based reference series, a number of conditions must be fulfilled. First, the availability period must be satisfactory with enough observations and has to correspond with the availability period of the appraisal-based index. Second, the method used to estimate the index must be reliable. Third, there must be adequacy in terms of inclusion criteria and property type mix between the appraisal-based index and the transaction-based index taken as benchmark. Hence, the NCREIF database appeared to be a natural candidate. Indeed, NCREIF transaction-based (NTBI) real estate series are available since 1984Q2 at the

¹⁰ Because EWMA processes, as all kinds of moving average processes, tend to lag the original series, we estimate the series of $\hat{\mu}_t$ and $\hat{\sigma}_t$ on the original return series and also on the reverse of the return series before taking the average value for each parameter at each point in time to obtain the $\hat{\mu}_t$ and $\hat{\sigma}_t$ series we actually use.

¹¹ In order to handle the effect of possible outliers affecting exogenous variables, factor loadings are estimated using robust regression. Comparison of results with those obtained when relying on classical regression estimation indicates that there is no significant change.

national level, while the appraisal-based (NPI) one is even older and both are produced at the quarterly frequency.

The method applied for building the NPI correspond to the standards of the industry. Regarding the estimation of the NTBI, as reviewed by Geltner (2011), it is currently done relying on a Sale Price Appraisal Ratio (SPAR) model, while until 2010Q4 the index was estimated with a hedonic model and produced by the MIT (MTBI hereafter). Geltner (2015) maintains that the hedonic approach does not suffer from the main biases affecting common alternative methods such as, for example, the repeat sales approach. Thus, we believe that a hedonic model would be the best method for building a relevant transaction-based index. However, some studies (Ling and Naranjo 2015) report that the current NTBI, restated for periods before 2011Q1, reproduces well the returns observed on the former MTBI. Comparing both the NTBI and the MTBI in Figure 2 we conclude to the similarity between these indices since 1997Q2. However, we recognize that prior to that period the NTBI seems to be more volatile, producing exaggerated returns compared with the MTBI, and hence is not reliable before 1997Q2. Further analyses reported in the Appendix A lead to the same conclusion.

As transactions belong to properties included in the NPI, the inclusion criteria are by construction the same. However, the NPI includes hotels, while the MTBI does not. Hence, the version of the NPI we use in this research does not consider hotels. In addition, the match of the property type mix is insured by building an appraisal-based series with the same proportions for each sector as the ones observed in the MTBI index. Moreover, as described by Fisher, Geltner and Pollakowski (2007), the hedonic model of the MTBI corrects for the possible selection bias of sold properties and filters the noise that may appear with quarterly frequency data¹².

¹² This is not the case with the SPAR NTBI and can explain the higher volatility of this series compared to that of the MTBI before the late 1990s.

According to the aforementioned discussion, we decide to focus our main analysis on the period 1984Q2-2018Q3, using for the transaction-based benchmark a series made of the hedonic MTBI until 2010Q4 spliced with the current SPAR NTBI. For robustness checks purposes, the analyses are also conducted on the MTBI only on the period 1984Q2-2010Q4.

[Figure 2 here]

As mentioned in Appendix B, the regimes of the TAR-TAR model are defined according to the U.S. MSCI total returns in excess of the 3-month T-Bill rate. For the multivariate desmoothing model we employ as exogenous factors the U.S. industrial production growth, core inflation, the term spread calculated as the 10-year Treasury rate minus the 3-month T-Bill rate, sourced from Thomson Reuters Datastream, as well as the credit spread computed as the AAA minus BAA corporate bond yields from Moody's, and the SMB, HML, WML and PS factors sourced from the WRDS database. Regarding the unlevered REIT series, we use total returns sourced from Thomson Reuters Datastream and debt balance sheet items required for the computation of unlevered return series are obtained from Compustat. Table 2 reports summary statistics for the aforementioned variables except direct real estate return series and unlevered REIT series as these are discussed in the section below. In particular, we observe high autocorrelation and low volatility for industrial production growth, core inflation, and the spreads, while this is the reverse for stock market returns and asset pricing factors.

[Table 2 here]

Results

For the period 1984Q2-2018Q3, the results pertaining to the AR model indicates for the smoothing equation an α_{AR} parameter¹³ of 0.85*** (0.02) and a sum of squared errors of 206.01, while the return generating process equation is:

$$R_t = \begin{matrix} 0.29 \\ (0.25) \end{matrix} - \begin{matrix} 0.18^{***} \\ (0.06) \end{matrix} R_{t-1} \quad (10)$$

Regarding the TAR-TAR model, for the smoothing equation, the alpha parameters are:

$$\begin{aligned} \alpha_L &= \begin{matrix} 0.45^{***} \\ (0.13) \end{matrix} ; \text{ when } Z_{s,t} < \begin{matrix} -13.72^{***} \\ (6.10) \end{matrix} \\ \alpha_H &= \begin{matrix} 0.87^{***} \\ (0.04) \end{matrix} ; \text{ when } Z_{s,t} \geq \begin{matrix} -13.72^{***} \\ (6.10) \end{matrix} \end{aligned} \quad (11)$$

The return generating process equation in the TAR-TAR model is:

$$R_t = \begin{cases} \begin{matrix} 2.41 \\ (2.54) \end{matrix} + \begin{matrix} 3.38^{***} \\ (1.78) \end{matrix} R_{t-1} ; \text{ when } Z_{r,t} < \begin{matrix} -10.44 \\ (8.27) \end{matrix} \\ \begin{matrix} 0.60 \\ (0.48) \end{matrix} - \begin{matrix} 0.23 \\ (0.30) \end{matrix} R_{t-1} ; \text{ when } Z_{r,t} \geq \begin{matrix} -10.44 \\ (8.27) \end{matrix} \end{cases} \quad (12)$$

The sum of squared errors of the TAR-TAR is 173.42, which is lower compared to the AR model suggesting that the regime-switching specification allows a better fitting of the data. The alpha parameter in the AR model is in-between the α_H and α_L parameters in the TAR-TAR model, which are 0.87 and 0.45, respectively. α_H and α_L values are significantly different from one another at the 1% confidence level.

Figure 3 depicts the appraisal-based NPI, the AR desmoothed, and the TAR-TAR desmoothed series, as well as the low regime phases of the desmoothing process for the TAR-TAR model. We observe that both desmoothed series treat smoothing which increases volatility and the TAR-TAR does not suffer the extreme drawdown reported by the AR in 2008Q4. This is due

¹³ The significance level of parameter estimates is indicated by the stars, with *, **, and *** corresponding to a p-value of 10%, 5% and 1%, respectively. Standard errors are indicated in brackets.

to the $\alpha_L = 0.45$ coefficient that is applied at that time being much smaller than $\alpha_H = 0.87$. The fact that α_H is larger than α_{AR} explains why the TAR-TAR is slightly more volatile during high regime phases. The TAR-TAR does not differ that much from the AR during the high volatility episode of the early 1990s. As can be seen in Figure 4 both desmoothed series overestimate the TBI volatility during these years.

[Figure 3 here]

[Figure 4 here]

In order to treat extreme values generated by the desmoothing process we now apply the robust filter on the AR and TAR-TAR series. For the TAR-TAR, the parameters of the EWMA_m and the EWMA_v processes are $\lambda_m = 0.85$ and $\lambda_v = 0.71$, respectively. For the AR, these parameters are $\lambda_m = 0.83$ and $\lambda_v = 0.67$. Figure 5 presents the desmoothed and the filtered series obtained from the TAR-TAR model, as well as key parameter series used by the robust filter. The grey curve is the EWMA_v series measuring the time-varying volatility of the TAR-TAR series. Volatility tends to decrease since the GFC. The grey bars are the robust weights applied to each observation. Weights are usually close to one and decrease when returns are relatively large in magnitude, especially during the early 1990s. The filtered series does not display the extreme returns observed for the unfiltered series, in particular during the early 1990s, when the TAR-TAR is the most erratic. The filter also mitigates the large rebound after the GFC.

[Figure 5 here]

In Figure 6, we compare the AR and TAR-TAR desmoothed series with their robust counterparts and the TBI. We observe that the robust filter treats the very large drawdown of 2008Q4 in the AR series. Actually, this strongly outlying observation receives a robust weight of about zero implying that for the robust AR series the corresponding value is near the time-varying average. Regarding the early 1990s higher volatility cluster, the largest changes are

shrunk toward TBI values. In general, it appears that the robust filter helps to correct both the AR and the TAR-TAR desmoothed series.

[Figure 6 here]

Regarding the GLM desmoothing model, the moving average order is two and the estimated parameters are $\theta_0 = 0.41^{***}$, $\theta_1 = 0.35^{***}$, and $\theta_2 = 0.25^{***}$. As for the AR and the TAR-TAR desmoothing model, it is possible that extreme values are generated through the desmoothing process. Thus, we also apply the robust filter¹⁴. The resulting series are reported in Figure 7 with the TBI. The GLM model produces series that are most of the time less volatile than those resulting from the AR and TAR-TAR models and that tend to underestimate transaction-based index volatility. Another observation is that the GLM does not produce striking outlying values. A consequence of the latter two observations is the only slight changes implied by the robust filter. Indeed, only three observations (1991Q4, 1992Q1, and 2008Q4) are substantially down-weighted.

[Figure 7 here]

Then, we turn to the MtV model whose estimated coefficients are reported in Table 3. Only the core inflation, the term spread, and the liquidity factor have significant effects, with negative coefficients for the former two and a positive coefficient for the latter one. The resulting series are reported in Figure 8 with the series of unlevered REITs and the TBI. The MtV series is different from the other two series during the early part of the period and has high autocorrelation. Unlevered REIT volatility appears usually comparable to that of the TBI.

[Table 3 here]

[Figure 8 here]

¹⁴ For the GLM, the robust filter parameters are $\lambda_m = 0.87$ and $\lambda_v = 0.88$

In order to emphasize differences of return distribution for each series with the transaction-based one, Figures 9 and 10 represent their kernel density estimations¹⁵. From Figure 9, we observe that the AR and the TAR-TAR distributions are in general close to that of the TBI and once the filter is applied the match is clearly better in the tails. In Figure 10, the deviation of the MtV distribution from the benchmark is marked. Regarding the GLM, it is not that close to the benchmark and its robust version produces a rather similar distribution, except maybe in the lower tail. The unlevered REIT distribution is not that different from the transaction-based one.

[Figure 9 here]

[Figure 10 here]

Summary statistics of quarterly total returns for the various methods are reported in Table 4. The differences between each series and the benchmark are more formally assessed through various statistical tests, whose results are reported in Table 5. The average and median returns are generally very close for all series (around two percent), with figures for REITs being slightly higher. Both the Welch T-test and the Kurskal-Wallis test conclude that in every case the hypothesis of equality in average return with the transaction-based series is not rejected¹⁶. This confirms, first, that the average returns measured on original appraisal-based series correspond to the ones measured on transaction-based series and, second, that the desmoothing process does not alter the average returns measured from the original series.

The standard deviations of desmoothed series are as expected larger than for the appraisal-based index and the autocorrelation is about zero, indicating that smoothing is actually treated. However, volatility figures for AR and TAR-TAR series over 8.0%, and of the MtV (5.44%), overstate the volatility of the transaction-based index (4.32%), while the GLM and robust GLM

¹⁵ This estimation is done using the Epanechnikov kernel.

¹⁶ We retain a critical confidence level of 95% and hence consider that the hypothesis made by each of the tests we apply is rejected if the p-value is below 5%.

series tend to understate volatility with values of 3.26% and 2.34%, respectively. Once the robust filter is applied to the AR and TAR-TAR series, the standard deviations are closer to that of the TBI, with values of 5.09% and 4.52%, respectively. The unlevered REIT series is also close to the TBI in terms of volatility. Actually, Bartlett and Fligner-Killeen tests both reject equality of variance with the reference series for the appraisal-based, the GLM, and the robust GLM series underestimating the variance, as well as the AR, the TAR-TAR, and the MtV series overestimating the variance. Conversely, the equality of variance is not rejected for the unlevered REITs series, as well as the robust AR and TAR-TAR series. For the latter two series, this improvement is achieved by correcting the extreme values with the robust filter.

Regarding the quantiles of the distributions, the AR and TAR-TAR match very well the benchmark once the robust filter is applied. This is especially true in the highest and the lowest quantiles; the part of the distribution that is precisely the segment where the robust filter is expected to enhance desmoothed series. Quantiles for unlevered REITs are also quite close to those of the benchmark in general. In contrast, the MtV, the GLM, and the robust GLM do not perform that well with respect to the quantiles matching. These aspects of the return distributions are tested by the Anderson-Darling test¹⁷, which rejects the equality with the transaction-based benchmark for the NPI, the GLM, the robust GLM, and the MtV. The unlevered REITs pass the test, like the robust AR and TAR-TAR series. For the latter two series, the p-values from the Anderson-Darling test are notably improved compared to the values returned for the non-filtered series¹⁸.

Skewness figures are close to the figure for the TBI, except for the AR and the GLM. In addition, all series but the MtV are leptokurtic, suggesting that dispersion of observations is larger than with a normal distribution. Before the robust filter has been applied, the AR, TAR-

¹⁷ The Anderson-Darling test works on cumulative distributions, like the Kolmogorov-Smirnov test, but it is more sensitive to differences in the tail quantiles.

¹⁸ The equality is rejected only at the 90% confidence level for the unfiltered TAR-TAR series.

TAR, and GLM series have kurtosis values overestimating the value displayed by the TBI. We would like to test differences in skewness and kurtosis in particular, but no satisfactory specific test exists. Hence, we propose simulating the confidence intervals of these statistics from the transaction-based benchmark and deriving the p-values for the figures measured on each series we investigate. Doing so, we conclude that the AR and the GLM series are different from the benchmark with respect to both skewness and kurtosis, as is the NPI. The TAR-TAR and the MtV differ only with respect to kurtosis. The unlevered REITs is not found to be different to the TBI. For the TAR-TAR, applying the robust filter adjusts the kurtosis, while it corrects both the skewness and kurtosis figures of the AR and GLM series.

[Table 4 here]

The autocorrelation is basically zero for most of the series except for the MtV, the robust GLM, the TAR-TAR, and obviously for the NPI. This high value for MtV can be explained by the fact that the series is mainly a linear combination of highly autocorrelated exogenous factor series. Furthermore, the correlation with the TBI is between 0.3 and 0.4 for the MtV, the AR, TAR-TAR, and GLM, as well as their robust versions, while it is only 0.19 for unlevered REITs and as much as 0.5 for the NPI.

We also report the maximum drawdown and the expected shortfall at 5% as alternative risk measures. The maximum drawdown appears particularly overestimated by the TAR-TAR, the AR, and the MtV series, with values between -46.79% and -78.64% compared to the -34.04% figure reported for the TBI. The TAR-TAR and the AR seem to overestimate the ES 5% too, with figures of -17.59% and -21.54%, respectively, while we measure -8.03% for the benchmark. The difference in ES 5% is also investigated with a simulated test. It returns that only the AR and the TAR-TAR series provide an ES 5% that is significantly different to that of the TBI. However, the robust filter corrects this discrepancy.

[Table 5 here]

As we are working with time series, investigating if they evolve closely together with the transaction-based series is important to understand whether series are able to reproduce accurately market dynamics and turning points. For this purpose, we measure the Mahalanobis distance between each series and the benchmark and test its significance¹⁹. We report a non-significant difference with the benchmark for the GLM and robust GLM series, as well as for the NPI. This indicates that these series actually capture well the evolution of the average, or short-term trend. In contrast, for the unlevered REIT, MtV, AR, and TAR-TAR series the difference with the benchmark is significant. Applying the robust filter leads to robust AR and TAR-TAR series having insignificant Mahalanobis distances. This result suggests that the robust filter is also able to improve the dynamics of the desmoothed series.

Then, we investigate how the different models are subject to generated outliers and how the robust filter helps dealing with them. The proportion of outliers with respect to the transaction-based benchmark is then computed²⁰. We observe that the AR, TAR-TAR, as well as to a lesser extent the MtV series, and the unlevered REITs, generate values that are considered as outliers when compared with the benchmark. The proportion of outliers is 7.97% for the TAR-TAR, 6.52% for the AR, and only 0.72% for the MtV and 1.45% for the unlevered REITs. The robust filter allows correcting all outliers, which highlights its ability in identifying and treating such deviating realizations.

Finally, we report the optimal allocation to real estate obtained while maximizing the Sharpe ratio of a portfolio containing stocks, bonds, and real estate for each desmoothed return series and the unlevered REIT return series. These figures can be compared with the value of 32.5%

¹⁹ The Mahalanobis distance is akin to the tracking error, but additionally takes into account the variances and the correlation structure between series. P-values for this test are obtained by bootstrapping.

²⁰ We define outliers as observations being among the 1% most extreme possible realizations for each point in time of an ARMA(1, 1)-GARCH(1, 1) model with skewed-Student distribution estimated from the transaction-based benchmark series. We also performed the same test with threshold values of 0.5%, 2.5%, and 5%. As expected, the lower the threshold value, the lower the number of identified outliers. However, the general conclusion concerning the usefulness of the robust filter does not change.

reported in Table 1 for the allocation to real estate obtained while relying on the TBI. Mainly because they overestimate the volatility, the allocation obtained when the TAR-TAR, the AR, and the MtV series are used is substantially lower (17.9%, 16.9% and 10.1%, respectively). However, once we apply the robust filter to the former two series the allocation (29.1% and 31.3%, respectively) is very close to the one reported with the TBI. The allocation reported for the GLM (46.5%) is clearly overstated, as several of this series' parameters do not match those of the TBI, in particular the volatility. Applying the filter on the GLM series amplifies the allocation (61.7%). Interestingly, the allocation reported with unlevered REIT returns is not that far from the one obtained with the TBI.

In sum, the analyses of characteristics of the various series, compared with those of the transaction-based series suggest the following. First, the original appraisal-based series deviates from the transaction-based benchmark mainly in terms of volatility, skewness, and kurtosis. It is however comparable to the benchmark with respect to mean returns as well as its dynamics and the measured expected shortfall at 5%. Second, the AR model tends to generate outliers, which leads to series overestimating the volatility, the negativity of the skewness, and the magnitude of the kurtosis, as well as of the expected shortfall. The dynamics of such series is also not in line with that of the benchmark. Series obtained with the TAR-TAR model suffer from similar drawbacks as the latter, except that the skewness figure is in line with the benchmark. Third, once we combine the robust filter with the latter two models, the outliers are identified and treated and the resulting series are fully in line with the transaction-based benchmark. Fourth, the GLM model generates series underestimating the volatility, while overestimating the kurtosis and the negativity of the skewness. Fifth, applying the robust filter to the latter series does not help with the volatility, but allows improving the skewness and kurtosis figures. Sixth, the MtV series is found to generate a few outliers and to overestimate the volatility and the kurtosis, and its dynamics deviate from that of the benchmark, while the unlevered REITs series does pretty well, except maybe with respect to the dynamics. Finally,

when applied on the TAR-TAR and the AR desmoothed series the robust filter enables reaching an allocation to real estate which is in line with the one obtained with the transaction-based series. The unlevered REITs series also provides good results in terms of portfolio allocation.

The analysis is repeated for robustness check purposes on the hedonic MTBI only from 1984Q2 until 2010Q4²¹. The general conclusions are the same as the ones reported for the main analysis²². In particular, the robust filter identifies and treats most of the outliers, which corrects the returns series and their distribution with respect to the characteristics we consider. The filter is especially helpful when combined with the AR and TAR-TAR desmoothing models. The robust filter also improves the distribution of returns obtained with the GLM model, despite the fact that it cannot correct the underestimated volatility.

Concluding Remarks

This research starts from the consideration that the classical desmoothing model presented by Geltner (1993b) and also its regime switching extension suggested by Lizieri, Satchell and Wongwachara (2012) are subject to generating unlikely extreme values when changes in the underlying original series are marked. These outlying values are problematic because they distort real estate return distribution features, which in turn leads to an erroneous assessment of real estate investment risk. This has important implications for portfolio management as the bias leads to overstated diversification benefits resulting from including real estate in mixed-asset portfolios.

We propose to enhance the quality of desmoothed series by applying robust statistical techniques. We design a robust time series filter that is able to identify and correct the outlying values so that their influence on the return distribution characteristics is suppressed. Results

²¹ Corresponding tables and figures can be provided on request.

²² Note that for the TAR-TAR model the main coefficients are $\alpha_H = 0.86$, $\alpha_L = 0.24$ and the threshold between regimes is -14.33% , while for the AR model $\alpha = 0.85$. Regarding the GLM model, coefficients are $\theta_0 = 0.39$, $\theta_1 = 0.36$ and $\theta_2 = 0.25$.

obtained with the classical desmoothing model, and its regime-switching version, are compared with those reached with the Getmansky, Lo and Makarov (2004) desmoothing model, a multivariate desmoothing model, as well as an alternative method relying on unlevered REIT returns.

Regarding the high autocorrelation found in appraisal-based series, all but one method (the multivariate model) are able to correct this issue and produce autocorrelation figures in line with the benchmark. About the dampened volatility figures displayed by appraisal-based series, every method we apply generates series with higher volatility. The unlevered REIT series volatility is generally in line with that of the benchmark, while the desmoothed series obtained with the AR and the TAR-TAR models, as well as the multivariate model tend to be too volatile. The GLM model underestimates the volatility. Overestimated volatility in the AR desmoothed series is due to the extreme values often generated during the desmoothing process. The outlier issue is to a lesser extent also encountered in series obtained with the regime-switching TAR-TAR model, although the greater flexibility allowed by its more sophisticated specification is helpful. We report that applying the robust filter proves useful in improving the general quality of desmoothed series by neutralizing the influence of most outlying values that inflate the volatility. Such filtered series notably display return distributions with volatility, skewness, and kurtosis figures that are more in line with the transaction-based ones and they evolve more closely to the benchmark over time. The filter also improves other risk measures such as the expected shortfall.

The unlevered REITs series, as well as the robust autoregressive desmoothing model and its robust regime-switching version in general produce the best results. The Getmansky, Lo and Makarov (2004) model underestimates volatility and despite the fact that the filter cannot improve this aspect of the distribution, it is able to enhance the skewness and kurtosis figures.

Series generated by the multivariate desmoothing model in general deviate from the benchmark whatever the criterion we consider.

The combination of the robust filter with the autoregressive model or the regime-switching model is appealing because it does not require relying on a lot of information. Indeed, the robust autoregressive model just requires the series to be desmoothed and the robust filter, while for the regime-switching model only an exogenous series for identification of the regimes is additionally necessary. In comparison, for the unlevered REIT index method that also performs well, information from the company balance sheets such as debt, interest expenses, dividends, and preferred shares are required and such information is not always readily available.

References

- Andersen, T.G., T. Bollerslev and S. Lange. 1999. Forecasting Financial Market Volatility: Sample Frequency Vis-à-vis Forecast Horizon. *Journal of Empirical Finance* 6(5): 457-477.
- Barkham, R. and D. Geltner. 1994. Unsmoothing British Valuation-Based Returns without Assuming an Efficient Market. *Journal of Property Research* 11(2): 81-95.
- Blundell, G.F. and C.W.R. Ward. 1987. Property Portfolio Allocation: A Multi-Factor Model. *Land Development Studies* 4(2): 145-156.
- Bollen, B. 2015. What Should the Value of Lambda Be in the Exponentially Weighted Moving Average Volatility Model? *Applied Economics* 47(8): 853-860.
- Bollerslev, T. 2010. Glossary to ARCH (GARCH*). In *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle*. T. Bollerslev, J. Russell and M. Watson, ed. U.K.: Oxford University Press. 137-163.
- Bond, S.A. and S. Hwang. 2007. Smoothing, Nonsynchronous Appraisal and Cross-Sectional Aggregation in Real Estate Price Indices. *Real Estate Economics* 35(3): 349-382.
- Bond, S.A., S. Hwang and G. Marcato. 2012. Commercial Real Estate Returns: An Anatomy of Smoothing in Asset and Index Returns. *Real Estate Economics* 40(4): 637-661.
- Carhart, M. 1997. On Persistence in Mutual Fund Performance. *Journal of Finance* 52(1): 57-82.
- Chaplin, R. 1997. Unsmoothing Valuation-Based Indices Using Multiple Regimes. *Journal of Property Research* 14(3): 189-210.
- Cho, Y., S. Hwang and Y.-K. Lee. 2014. The Dynamics of Appraisal Smoothing. *Real Estate Economics* 42(2): 497-529.
- Cho, H., Y. Kawaguchi and J.D. Shilling. 2003. Unsmoothing Commercial Property Returns: A Revision to Fisher-Geltner-Webb's Unsmoothing Methodology. *Journal of Real Estate Finance and Economics* 27(3): 393-405.
- Clayton, J., D. Geltner and S.W. Hamilton. 2001. Smoothing in Commercial Property Valuations: Evidence from Individual Appraisals. *Real Estate Economics* 29(3): 337-360.
- Cox, D.R. 1961. Prediction by Exponentially Weighted Moving Averages and Related Methods. *Journal of the Royal Statistical Society. Series B (Methodological)* 23(2): 414-422.
- Cumming, D., L.H. Hass and D. Schweizer. 2013. Private Equity Benchmarks and Portfolio Optimization. *Journal of Banking and Finance* 37(9): 3515-3528.
- Diaz, J. and M.L. Wolverton. 1998. A Longitudinal Examination of Appraisal Smoothing Hypothesis. *Real Estate Economics* 26(2): 349-358.

- Dimson, E. 1979. Risk Measurement when Shares are Subject to Infrequent Trading. *Journal of Financial Economics* 7(2): 197-226.
- Fama, E. and K.R. French. 1992. The Cross-Section of Expected Stock Returns. *Journal of Finance* 47(2): 427-465.
- Fama, E. and K.R. French. 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33(1): 3-56.
- Firstenberg, P.M., S.A. Ross and R.C. Zisler. 1988. Real Estate: The Whole Story. *Journal of Portfolio Management* 14(3): 22-34.
- Fisher, J. and D. Geltner. 2000. De-lagging the NCREIF Index: Transaction Prices and Reverse-engineering. *Real Estate Finance* 17(1): 7-22.
- Fisher, J., D. Geltner and H. Pollakowski. 2007. A Quarterly Transaction-based Index of Institutional Real Estate Investment Performance and Movements in Supply and Demand. *Journal of Real Estate Finance and Economics* 34(1): 5-33.
- Fisher, J., D. Geltner and R. Webb. 1994. Value Indices of Commercial Real Estate: A Comparison of Index Construction Methods. *Journal of Real Estate Finance and Economics* 9(2): 137-164.
- Fried, R. 2004. Robust Filtering of Time Series with Trends. *Journal of Nonparametric Statistics* 3(4): 313-328.
- Fried, R., J. Einbeck and U. Gather. 2007. Weighted Repeated Median Smoothing and Filtering. *Journal of the American Statistical Association* 102(480): 1300-1308.
- Fu, Y. 2003. Estimating the Lagging Error in Real Estate Price Indices. *Real Estate Economics* 31(1): 75-98.
- Geltner, D. 1989. Estimating Real Estate's Systematic Risk from Aggregate Level Appraisal-Based Returns. *Journal of the American Real Estate and Urban Economics Association* 17(4): 463-481.
- Geltner, D. 1991. Smoothing and Appraisal-Based Returns. *Journal of Real Estate Finance and Economics* 4(3): 327-345.
- Geltner, D. 1993a. Temporal Aggregation in Real Estate Return Indices. *Journal of the American Real Estate and Urban Economics Association* 21(2): 141-166.
- Geltner, D. 1993b. Estimating Market Values for Appraised Values without Assuming an Efficient Market. *Journal of Real Estate Research* 8(3): 325-346.
- Geltner, D. 2011. A Simplified Transaction-Based Index (TBI) for NCREIF Production, Cambridge, MA: MIT Center for Real Estate & Geltner Associates LLC.
- Geltner, D. 2015. Real Estate Price Indices and Price Dynamics: An Overview from an Investments Perspective, *Annual Review of Financial Economics* 7(1): 615-633.

- Geltner, D., B.D. MacGregor and G.M. Schwann. 2003. Appraisal Smoothing and Price Discovery in Real Estate Markets. *Urban Studies* 40(5-6): 1047-1064.
- Geltner, D., N.G. Miller, J. Clayton and P. Eichholtz. 2007. *Commercial Real Estate Analysis and Investments*. Mason, OH: Thomson South-Western.
- Getmansky, M., A.W. Lo and I. Makarov. 2004. An Econometric Model of Serial Correlation and Illiquidity in Hedge Funds Returns. *Journal of Financial Economics* 74(3): 529-609.
- Hoesli, M., L. Lekander and W. Witkiewicz. 2004. International Evidence on Real Estate as a Portfolio Diversifier. *Journal of Real Estate Research* 26(2): 161-206.
- Hoesli, M., E. Oikarinen and C. Serrano. 2015. Do Public Real Estate Returns Really Lead Private Returns? *Journal of Portfolio Management* 41(6): 105-117.
- Holt, C.C. 2004. Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. *International Journal of Forecasting* 20(1): 5-10.
- Huber, P.J. 1972. The 1972 Wald Lecture. Robust Statistics: A Review. *The Annals of Mathematical Statistics* 43(4): 1041-1067.
- Huber, P.J. 1981. *Robust Statistics*. New York, NY: Wiley.
- Ling, D.C. and A. Naranjo. 2015. Returns and Information Transmission Dynamics in Public and Private Real Estate Markets. *Real Estate Economics* 43(1): 163-208.
- Ling, D.C., A. Naranjo and M. Nimalendran. 2000. Estimating Returns on Commercial Real Estate: A New Methodology Using Latent-Variable Models. *Real Estate Economics* 28(2): 205-231.
- Lizieri, C., S. Satchell and W. Wongwachara. 2012. Unsmoothing Real Estate Returns: A Regime-Switching Approach. *Real Estate Economics* 40(4): 775-807.
- Lucas, J.M. and M.S. Saccucci. 1990. Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics* 32(1): 1-12.
- Marcato, G. and T. Key. 2007. Smoothing and Implication for Asset Allocation Choices. *Journal of Portfolio Management* 33(5): 85-99.
- Matysiak, G.A. and P. Wang. 1995. Commercial Property Prices and Valuation: Analysing the Correspondence. *Journal of Property Research* 12(1): 181-202.
- Pastor, L. and R.F. Stambaugh. 2003. Liquidity Risk and Expected Stock Returns. *Journal of Political Economy* 111(3): 642-685.
- Pedersen, N., S. Page and F. He. 2014. Asset Allocation: Risk Models for Alternative Investments. *Financial Analysts Journal* 70(3): 34-45.
- Quan, D.C. and J.M. Quigley. 1989. Inferring an Investment Return Series for Real Estate from Observations on Sales. *Real Estate Economics* 17(2): 218-230.

- Quan, D.C. and J.M. Quigley. 1991. Price Formation and the Appraisal Function in Real Estate Markets. *Journal of Real Estate Finance and Economics* 4(2): 127-146.
- Reynolds, M.R. Jr. and Z.G. Stoumbos. 2006. Comparisons of Some Exponentially Weighted Moving Average Control Charts for Monitoring the Process Mean and Variance. *Technometrics* 48(4): 550-567.
- Riddiough, T.J., M. Moriarty and P.J. Yeatman. 2005. Privately versus Publicly Held Asset Investment Performance. *Real Estate Economics* 33(1): 121-146.
- Rousseeuw, P.J. 1984. Least Median of Squares Regression. *Journal of the American Statistical Association* 79(388): 871-880.
- Rousseeuw, P.J. and C. Croux. 1993. Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association* 88(424): 1273-1283.
- Rousseeuw, P.J. and M. Hubert. 1999. Regression Depth. *Journal of the American Statistical Association* 94(446): 388-402.
- Scholes, M. and J. Williams. 1977. Estimating Betas from Nonsynchronous Data. *Journal of Financial Economics* 5(3): 309-327.
- Shepard, P. and L. Liu. 2014. The Barra Private Equity Model (PEQ2). MSCI Research Paper, MSCI.
- Siegel, A.F. 1982. Robust Regression Using Repeated Medians. *Biometrika* 68(1): 242-244.
- Wang, P. 2006. Errors in Variables, Links between Variables and Recovery of Volatility Information in Appraisal-Based Real Estate Return Indexes. *Real Estate Economics* 34(4): 497-518.

Appendix A: Comparison between the MTBI and NTBI indices

Appendix A compares the hedonic MTBI with the SPAR NTBI on various periods, in order to assess how comparable they are and particularly if the NTBI is similar enough to the MTBI for using it as a relevant benchmark after 2010Q4, the last date of availability of the MTBI. Details of these analyses are reported in Table A1, where we produce statistics for total return distributions of these series, and in Table A2, where we display p-values of several tests aiming to compare the moments and other key characteristics of both series distributions. Further details on the tests we employ are provided in the results section of this article.

From these analyses, we report that for the 1984Q2-2010Q4 period the MTBI and the NTBI differ. Indeed, according to the Bartlett and Fligner-Killeen tests, equality of variance is rejected at a 95% confidence level basis. Hence, the volatility of the NTBI of 6.38% is significantly larger than the 4.5% reported for the MTBI. In addition, tests indicates that both series display significantly different 5% expected shortfall and that 6.54% of NTBI returns could be considered as outliers with respect to the MTBI distribution. Actually, differences between both series are the strongest for the 1984Q2-1997Q1 period, where the MTBI and NTBI series are significantly different from almost all points of view, except regarding the average total returns and to a lesser extent the skewness. The difference in terms of volatility is particularly pronounced, with 3.49% for the MTBI and a more than twice larger volatility for the NTBI (7.21%). Conversely, for the 1997Q2-2010Q4 period, no test rejects the equality between the two series. More interestingly, comparing the NTBI from 1997Q2 until 2018Q3 with the MTBI on a period of same length from 1989Q3 to 2010Q4, we conclude with respect to all tests that both series are not statistically different. The most striking feature is the same volatility figure of 4.8% reported for both series.

From these analyses, we conclude that for the early period 1984Q2-1997Q1, the MTBI and the NTBI cannot be considered as comparable, while since 1997Q2 they can be considered as

comparable, because statistical tests do not reject equality between both series from the perspectives of the different main moments and characteristics of these return distributions. Thus, we are confident in using in our main analysis a spliced series made of the returns of the MTBI from 1984Q2 until 2010Q4 and of those of the NTBI from 2011Q1 to 2018Q3.

Table A1 Summary Statistics of Total Return Series for MTBI and NTBI by period

| | MTBI 1984Q2- 2010Q4 | NTBI 1984Q2- 2010Q4 | MTBI 1984Q2- 1997Q1 | NTBI 1984Q2- 1997Q1 | MTBI 1997Q2- 2010Q4 | NTBI 1997Q2- 2010Q4 | MTBI 1989Q3- 2010Q4 | NTBI 1997Q2- 2018Q3 |
|----------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Max | 19.05 | 18.14 | 8.02 | 15.24 | 19.05 | 18.14 | 19.05 | 18.14 |
| q99 | 12.74 | 15.21 | 7.54 | 14.96 | 15.76 | 15.51 | 13.88 | 14.00 |
| q95 | 7.84 | 11.68 | 5.89 | 12.35 | 8.52 | 10.36 | 7.99 | 9.76 |
| q90 | 6.94 | 9.48 | 4.83 | 9.54 | 7.68 | 9.17 | 7.43 | 7.52 |
| q75 | 4.55 | 5.87 | 3.75 | 6.39 | 5.51 | 5.24 | 4.59 | 5.19 |
| Median | 2.26 | 2.29 | 1.93 | 2.22 | 2.73 | 2.47 | 2.12 | 2.56 |
| Average | 1.97 | 2.10 | 1.42 | 1.59 | 2.48 | 2.58 | 2.08 | 2.65 |
| q25 | -0.29 | -0.24 | -0.05 | -2.57 | -0.31 | 0.14 | -0.41 | 0.11 |
| q10 | -3.71 | -5.20 | -4.26 | -9.29 | -3.24 | -2.44 | -3.71 | -1.58 |
| q05 | -5.04 | -11.55 | -5.02 | -12.36 | -4.78 | -5.03 | -5.04 | -3.24 |
| q01 | -9.65 | -15.95 | -7.30 | -15.28 | -12.94 | -13.94 | -10.79 | -12.69 |
| Min | -16.69 | -16.12 | -8.06 | -16.03 | -16.69 | -16.12 | -16.69 | -16.12 |
| St.dev. | 4.57 | 6.38 | 3.49 | 7.21 | 5.38 | 5.50 | 4.80 | 4.80 |
| Skewness | -0.37 | -0.57 | -0.78 | -0.54 | -0.43 | -0.44 | -0.32 | -0.38 |
| Kurtosis | 3.49 | 1.09 | 0.39 | 0.13 | 3.34 | 2.86 | 3.45 | 3.55 |
| Autocor. | 0.04 | -0.14 | -0.12 | -0.19 | 0.08 | -0.06 | 0.08 | -0.13 |
| MaxDD | -34.04 | -27.00 | -9.90 | -16.03 | -34.04 | -27.00 | -34.04 | -27.00 |
| ES 5% | -8.54 | -13.92 | -6.57 | -14.35 | -10.51 | -11.14 | -8.63 | -8.35 |

Note: Total returns are expressed in percent. The reported kurtosis is actually the excess kurtosis.

Table A2 Tests of Similarity Between MTBI and NTBI by period

| | MTBI vs. NTBI 1984Q2-2010Q4 | MTBI vs. NTBI 1984Q2-1997Q1 | MTBI vs. NTBI 1997Q2-2010Q4 | MTBI 1989Q3-2010Q4 vs. NTBI 1997Q2-2018Q3 |
|-------------------------|--------------------------------|--------------------------------|--------------------------------|--|
| T-test | 0.74 | 0.82 | 0.77 | 0.43 |
| Kruskal-Wallis | 0.63 | 0.50 | 0.96 | 0.54 |
| Bartlett | 0.00 | 0.00 | 0.87 | 1.00 |
| Fligner-Killeen | 0.01 | 0.00 | 0.96 | 0.90 |
| Anderson-Darling | 0.17 | 0.04 | 0.98 | 0.73 |
| Simulated Skewness | 0.46 | 0.13 | 0.92 | 0.90 |
| Simulated Kurtosis | 0.51 | 0.01 | 0.98 | 0.58 |
| Simulated ES 5% | 0.01 | 0.00 | 0.58 | 0.96 |
| Mahalanobis Distance | 0.75 | 0.00 | 0.96 | - |
| Outliers Proportion (%) | 6.54 | 13.46 | 0.00 | - |

Note: Table A2 reports the p-values of various tests used to compare the hedonic MTBI with the SPAR NTBI. It also reports the proportion of outliers found in the NTBI compared to what would be expected with the MTBI. The Welch T-test and Kruskal-Wallis test investigate if realizations of compared groups could be considered as coming from the same distribution on average. The Bartlett and the Fligner-Killeen tests assess if both groups have the same variance. The Anderson-Darling tests check if the cumulative distribution of compared groups is the same. The latter test is similar to the Kolmogorov-Smirnov but it is more sensitive to differences in tail quantiles. No satisfactory statistical test exists for comparing the skewness and kurtosis of two distributions. Hence, we propose estimating confidence intervals through a simulation process and derive the corresponding p-values. We apply the jackknife method for resampling the benchmark series, with resampled series length being 90% of the benchmark series length and the number of samples being 10,000. The moments of interest are then computed on each resampled series and the 10,000 measures allow defining the confidence intervals. In addition to skewness and kurtosis, we also simulated the confidence intervals for the expected shortfall at 5%. As we use time series we also perform one test taking into account the temporal dimension of the data, by measuring the Mahalanobis distance between each series and the benchmark. The latter is similar to the well-known tracking error, but it additionally takes into account the variances and correlation structure of the compared series. The significance is estimated relying on bootstrapped critical values. Finally, the outlier proportion is the number of observations relative to the series length being over given upper and lower thresholds defined as a specific quantile of the time-varying MTBI series distribution. Time-varying parameters of the latter distribution are obtained by fitting an ARMA(1, 1)-GARCH(1, 1) model with skewed-Student distribution on the transaction-based series. The threshold quantiles are defined such that observations are taken as outliers if they belong to the 1% most extreme realizations of the estimated model at a given point in time. Note also that, as the periods differ in the last comparison case, it is not relevant to run the Mahalanobis distance test, as well as the count of outliers with respect to the MTBI.

Appendix B: Details on Desmoothing Models

Appendix B presents the computational details of the various desmoothing models we apply and of the unlevered REITs index, as well as computational details of the estimation of tuning constants of the Tukey function used in the robust filter.

The AR Desmoothing Model

The Geltner (1993b) desmoothing model considers that the current observed appraised value depends partly on the past appraised values and on the current public information available. This feature comes from the fact that appraisers tend to rely largely on the values they produced during the previous periods, adjusting them with respect to the current market environment. This leads to the appraisal rule or transfer function we mentioned above that defines the current appraised value as:

$$V_t^a = \alpha V_{t-1}^a + (1 - \alpha)P_t \quad (\text{B1})$$

where V_t^a is the current appraised value, V_{t-1}^a the previous appraised value, P_t the unobserved current market equilibrium price, and α the proportion of the current appraised value that relies on the previous appraised value (the so-called (de)smoothing parameter). One obtains the smoothing equation by rewriting equation (B1) in terms of returns as:

$$r_t^a = \alpha r_{t-1}^a + (1 - \alpha)r_t \quad (\text{B2})$$

with r_t^a the appraised return observed at time t , r_{t-1}^a the appraised return for the previous period and, r_t , the true return that should prevail with respect to the variation of the unobserved market price. This equation is rearranged in order to express the true return:

$$r_t = \frac{(r_t^a - \alpha r_{t-1}^a)}{(1 - \alpha)} \quad (\text{B3})$$

According to this equation, the only unknown that must be estimated is α . Estimating α requires making some assumptions that particularly pertain to the way the actual unobserved

market return r_t is generated. As mentioned, it can be assumed to follow an AR(1) process²³, such as:

$$r_t = \gamma + \varphi r_{t-1} + \varepsilon_t \quad (\text{B4})$$

with φ the autoregressive coefficient of the model, γ the constant term, and ε_t the *iid* error term at period t . By substituting equation (B3) in equation (B4) one obtains an AR(2) representation of the appraised return process:

$$r_t^a = \gamma(1 - \alpha) + (\alpha + \varphi)r_{t-1}^a + \alpha\varphi r_{t-2}^a + \omega_t \quad (\text{B5})$$

with $\omega_t = (1 - \alpha)\varepsilon_t$. Given any values of γ and φ one can estimate $\hat{\alpha}$ as:

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \sum_{t=1}^T \omega_t^2(\alpha; \gamma, \varphi) \quad (\text{B6})$$

Once the estimate is obtained, one computes the r_t series using $\hat{\alpha}$ in equation (B3). Then, with this r_t series and relying on equation (B4) we estimate new values for $\hat{\gamma}$ and $\hat{\varphi}$ as:

$$(\hat{\gamma}, \hat{\varphi}) = \underset{\gamma, \varphi}{\operatorname{argmin}} \sum_{t=1}^T \varepsilon_t^2(\gamma, \varphi; \hat{\alpha}) \quad (\text{B7})$$

These last two estimates feed into equation (B5) and the procedure is repeated until convergence. Note that for the AR models, as well as for the TAR-TAR and the Getmansky, Lo and Makarov (2004) models presented below, we transform the price return series and not directly the total return series as we assume smoothing impacts the price series only, not the income series. Once the desmoothed price return series is obtained, we add the income returns to get the desmoothed total return series. Decomposing returns in that way is not necessary when considering the multivariate model which deals with total returns directly.

²³ Of course, a more sophisticated return generating process could be chosen among the ARIMA(p, d, q) family.

The TAR-TAR Desmoothing Model

As proposed by Lizieri, Satchell and Wongwachara (2012) more sophistication can be brought into the AR model by allowing for regime switching features in the smoothing equation (TAR-AR) or in the return process (AR-TAR) or in both equations (TAR-TAR)²⁴. Regime switching in the smoothing equation reflects the fact that appraisers work differently according to the economic context, while the regime switching in the return process reflects how it is altered across market situations.

In the case of the regime switching smoothing equation, (B2) becomes:

$$r_t^a = (\alpha_H 1_{z_s > c_s} + \alpha_L 1_{z_s \leq c_s}) r_{t-1}^a + (1 - (\alpha_H 1_{z_s > c_s} + \alpha_L 1_{z_s \leq c_s})) r_t \quad (\text{B8})$$

where α_H and α_L are the desmoothing parameters prevailing in high and low regimes, respectively, and $1_{z_s > c_s}$ and $1_{z_s \leq c_s}$, indicate if the process is in the high or the low regime according to z_s , the exogenous variable for determining smoothing regimes, and c_s , the threshold separating smoothing regimes.

In the same way, the return generating process (B4) becomes:

$$r_t = (\gamma_H 1_{z_r > c_r} + \gamma_L 1_{z_r \leq c_r}) + (\varphi_H 1_{z_r > c_r} + \varphi_L 1_{z_r \leq c_r}) r_{t-1} + \varepsilon_t \quad (\text{B9})$$

where (γ_H, φ_H) and (γ_L, φ_L) are the return equation parameters prevailing in high and low regimes, respectively, and $1_{z_r > c_r}$ and $1_{z_r \leq c_r}$, indicate if the process is in its high or low regime according to z_r , the exogenous variable used to determine return process regimes, and, c_r , the threshold separating return process regimes. The exogenous variable used for determining regime thresholds can be the same for the smoothing equation and the return generating process.

We choose the excess stock market returns as the exogenous variable for both processes²⁵.

²⁴ As the classical desmoothing model relies on AR processes in both the smoothing equation (B2) and return process (B4), we could refer to it as the AR-AR model.

²⁵ We also tested with economic activity and interest rates as exogenous variables and results indicated that stock market returns were the best choice, according to the optimization outcomes and the proximity reached with the

The Getmansky, Lo and Makarov (2004) Model

The GLM model is actually close to the classic AR model. However, instead of working on an autoregressive representation of the smoothing process it relies on a moving average representation. In order to estimate parameters and practically compute the unsmoothed values a moving average process of order k (MA(k))²⁶ is fitted on the observed return series under the constraints $\theta_j \in [0, 1] \forall j = 0:k$ and $\sum \theta_j = 1$. Then, the unobserved market returns are estimated through the following formula:

$$r_t = \frac{r_t^\alpha - \sum_{j=1}^k \theta_j r_{t-j}}{\theta_0} \quad (\text{B10})$$

Note also that from the θ_j parameters one can compute the so-called smoothing index that indicates the intensity of smoothing the series suffers. It is defined as:

$$\xi = \sum_{j=0}^k \theta_{adj..j}^2 \quad (\text{B11})$$

The closer ξ to 0 the more smoothed the series and conversely.

The Multivariate Desmoothing Model

This model, based on multivariate regressions, is particularly inspired by the one presented by Pedersen, Page and He (2014) and the time invariant one of Cho, Hwang and Lee (2014). As in the classical desmoothing model, the multivariate model (MtV hereafter) recognizes the fact that the current appraisal-based return depends partly on the past appraised return and partly on the current unobserved true return. The true returns are also assumed to depend on exogenous

corresponding transaction-based series. Note also that unlike Lizieri, Satchell and Wongwachara (2012), using British data, we do not lag by one period the exogenous variable but we take its contemporaneous value as our tests indicate better results this way. Another difference is that we prefer to take the excess stock market return with respect to the three-month interest rate, taken as the risk-free rate which has changed significantly during the relatively long time period that we cover, to insure that threshold values are consistent over the whole period, while Lizieri and his colleagues use the raw stock market returns.

²⁶ Regarding the optimal number of lags, k , Getmansky, Lo and Makarov (2004) suggest $k = 2$. However, our results indicate that such desmoothed series clearly underestimate the volatility with respect to the benchmark transaction-based index. Hence, we also apply the GLM model with k selected such s all non-zero θ_j parameters are present in the model. This implies $k = 8$.

market and macroeconomic factors. This is represented by equation (B12) below which is an adaptation of equation (B2):

$$r_t^a = (1 - \omega)r_{t-1}^a + \omega r_t + \varepsilon_t = (1 - \omega)r_{t-1}^a + \omega \sum_{j=1}^J \beta_j f_{j;t} + \varepsilon_t \quad (\text{B12})$$

where ω is the weight of current unobserved true returns in the current appraisal-based return, $f_{j;t}$ is the j^{th} factor entering the equation explaining the current true return and β_j is the corresponding factor loading²⁷. Actually, most of the factors proposed by Cho, Hwang and Lee (2014) and Pedersen, Page and He (2014) are themselves highly autocorrelated. Hence, having together in the equation the realizations of the set of factors f at time t and r_{t-1}^a , which are by construction assumed to depend on all past realizations of these factors, may lead to severe multicollinearity and estimation issues. Thus, we propose to focus on the second part of equation (B12), which is also more parsimonious. Equation (B12) is simply rewritten as:

$$r_t^a = \omega \sum_{j=1}^J \beta_j f_{j;t} + \varepsilon_t \quad (\text{B13})$$

Following Pedersen, Page and He (2014), we use the autoregressive coefficient of order one, φ , of the appraisal-based return series process in order to estimate ω as $1 - \varphi$. We now slightly modify equation (B13) as we transform the set of factors f as $g = \omega f$:

$$r_t^a = \sum_{j=1}^J \delta_j g_{j;t} + \varepsilon_t \quad (\text{B14})$$

with δ_j the response to the j^{th} transformed factor. Finally, the estimated coefficients δ_j are applied to the original set of factors f in order to get an estimate of the true unobserved returns. Combining some results of Cho, Hwang and Lee (2014) and Pedersen, Page and He (2014) we select as exogenous factors industrial production, the core inflation, the stock market returns, the term spread, the credit spread, the Fama and French (1992, 1993) size factor (SMB), the

²⁷ Unlike Cho, Hwang and Lee (2014) our goal is not to estimate equation (B12) in order to define the desmoothing parameter of a classical desmoothing model as $(1 - \omega)$, which is simply the autoregressive coefficient of the appraisal-based series filtered by a given set of exogenous factors.

value/growth factor (HML), the Carhart (1997) momentum factor (WML), and the Pastor and Stambaugh (2003) liquidity factor (PS)²⁸.

REIT-based Indices

For each direct real estate index that we investigate, we build a corresponding index of REIT returns. As is the case in Fisher, Geltner and Webb (1994), Ling and Naranjo (2015) as well as in Hoesli, Oikarinen, and Serrano (2015), the effect of leverage is removed from REIT returns because direct investment performance is measured without debt. For computing the unlevered REIT returns we need to retrieve the returns generated by total assets. To do so, we apply the simple weighted average cost of capital (WACC) model as reported in Ling and Naranjo (2015):

$$r_{i,t}^{TA} = (r_{i,t}^E \vartheta_{i,t}^E) + (r_{i,t}^D \vartheta_{i,t}^D) + (r_{i,t}^P \vartheta_{i,t}^P) \quad (\text{B15})$$

where $r_{i,t}^{TA}$ is the unlevered return on total asset, $r_{i,t}^E$, the levered return on equity observed on the stock market, $r_{i,t}^D$, the interest rate paid to the firm's debt lenders computed as the interest expense over the book value of debt at the previous period ($i_{i,t}^D/bval_{i,t-1}^D$), and $r_{i,t}^P$, the dividend rate paid to preferred equity holders computed as the total preferred dividends over the estimated liquidation value of outstanding preferred shares at the previous period ($pdiv_{i,t-1}^P/lval_{i,t-1}^P$). Defining the market capitalization of a firm during the previous period as $mcap_{i,t-1}^E$, the total asset value at that time is $TA_{i,t-1} = mcap_{i,t-1}^E + bval_{i,t-1}^D + lval_{i,t-1}^P$ and the weights are $\vartheta_{i,t}^E = mcap_{i,t-1}^E/TA_{i,t-1}$, $\vartheta_{i,t}^D = bval_{i,t-1}^D/TA_{i,t-1}$, and $\vartheta_{i,t}^P = pdiv_{i,t-1}^P/TA_{i,t-1}$. Once returns have been unlevered at the REIT level, indices are built in order to match with the corresponding direct index in terms of property type mix. This insures that the proportion of each sector (apartment, industrial, office, and retail) is the same in both indices. Another adjustment is related to the fact that direct real estate performance is reported

²⁸ The stock market returns are the U.S. MSCI total returns, the term spread is the 10-year Treasury rate minus the 3-month T-Bill rate, the credit spread is the AAA – BAA corporate bond yields from Moody's and the SMB, HML, WML, and PS factors are sourced from the WRDS database.

without accounting for management fees while they are accounted for in reported REIT returns. We hence add management fees to REIT returns to make them more comparable with direct market returns. As Ling and Naranjo (2015) and Hoesli, Oikarinen and Serrano (2015), we follow Riddiough, Moriarty and Yeatman (2005) and apply 20 basis points management fees to quarterly returns. Finally, as several studies (Ling and Naranjo 2015, Hoesli, Oikarinen and Serrano 2015) suggest that the REIT market commonly leads the direct market, we decide to lag our REIT indices by one quarter²⁹.

Robust Filter Optimal Tuning Constants Estimation

We estimate the optimal tuning constants $c_{w,U}$ and $c_{w,L}$ entering the robust filter that are relevant in our time series context. We estimate them by simulation which consists first in fitting the MTBI index with an ARMA(1, 1)-GARCH(1, 1) process having a skewed-Student residual distribution in order to allow for time-varying parameters of the mean and standard deviation as well as an asymmetric and leptokurtic distribution. Then, series of 10,000 observations are generated using the parameters estimated during the previous step. Because these series are assumed to be outlier free, we add a given proportion of outliers by replacing randomly chosen observations in the simulated series by unlikely values. Outliers are defined as being equal to the 1st or the 99th percentile of the distribution at the point in the series where it replaces the previously simulated value. For each generated series, we apply the robust filter and a given observation is recognized as an outlier if its robust weight is less than 0.5. Ideally, the filter should be able to detect all imposed outliers and should not indicate an originally simulated observation to be an outlier. In order to evaluate the ability of the filter to detect the true outliers and not wrongly indicate a correct simulated value, with respect to the $c_{w,U}$ and $c_{w,L}$ parameters, we consider minimizing the Anderson-Darling test statistic. This test, similarly to the Kolmogorov-Smirnov test, aims at measuring the difference between two distributions in terms

²⁹ We also perform the analysis without lagging unlevered REIT series by one quarter and with lagging them by two quarters, respectively, and find that the best match is obtained with a one-quarter lag.

of quantiles repartition. However, the former test is more sensitive to differences in distribution tails than the latter one, what is precisely the part we want to focus on. The simulation is repeated with proportions of outliers being 0, 1, 2.5, 5, and 10 percent, respectively. The simulation results indicate that the optimal values should be $c_{w,U} = 5$ and $c_{w,L} = -6.25$. The absolute value of these parameters is larger than the 4.685 value usually retained in the literature. This can be explained by the fact that the distributions we consider are leptokurtic while the usual value pertains to the normal distribution case. The difference in absolute value between both parameters we estimate is due to the slight negative skewness of real estate return distributions.

Tables and Figures

Table 1 Appraisal- and Transaction-Based Indices Comparison – 1984Q2-2018Q3

| | NPI | TBI | Desmoothed NPI alpha = 0.8 |
|------------------------|--------|--------|----------------------------------|
| Maximum | 5.42 | 19.05 | 20.80 |
| Median | 2.36 | 2.32 | 2.17 |
| Average | 1.99 | 2.15 | 1.96 |
| Minimum | -8.24 | -16.69 | -40.79 |
| Std Deviation | 2.07 | 4.32 | 6.22 |
| Skewness | -2.36 | -0.35 | -2.78 |
| Kurtosis | 8.07 | 3.61 | 18.41 |
| ES 5% | -4.91 | -8.03 | -15.35 |
| MaxDD | -23.79 | -34.04 | -49.22 |
| Sharpe Ratio | 1.12 | 0.61 | 0.36 |
| Real estate allocation | 68.0 | 32.5 | 22.6 |

Note: Table 1 reports summary statistics of total returns, as well as the optimal allocation to real estate achieved by maximizing the Sharpe ratio of a portfolio made of real estate, stocks, and bonds. These figures are in percent, except for skewness, kurtosis, and the Sharpe ratio. Note also that the kurtosis is actually the excess kurtosis.

Table 2 Summary Statistics of Macroeconomic and Financial Series, 1984Q2-2018Q3

| | Industrial Production growth | Core Inflation | MSCI Total Returns | MSCI excess Total Returns | Term Spread | Credit Spread | SMB | HML | WML | PS |
|----------|------------------------------------|-------------------|--------------------------|------------------------------------|----------------|------------------|--------|-------|--------|--------|
| Max | 2.51 | 1.52 | 21.90 | 19.71 | 3.92 | 3.00 | 13.86 | 7.37 | 16.60 | 9.36 |
| q75 | 1.13 | 0.84 | 7.65 | 5.43 | 2.80 | 1.17 | 2.14 | 1.40 | 3.45 | 2.28 |
| Median | 0.63 | 0.59 | 3.53 | 1.64 | 1.94 | 0.91 | 0.51 | 0.25 | 1.47 | 0.21 |
| Average | 0.51 | 0.67 | 2.93 | 0.66 | 1.89 | 1.00 | 0.54 | 0.23 | 1.70 | 0.03 |
| q25 | 0.21 | 0.49 | -1.17 | -3.45 | 1.01 | 0.75 | -0.90 | -1.24 | -0.22 | -1.86 |
| Min | -11.24 | 0.07 | -28.80 | -30.03 | -0.37 | 0.57 | -16.88 | -9.86 | -11.54 | -10.78 |
| St.Dev. | 1.27 | 0.33 | 8.43 | 8.41 | 1.19 | 0.40 | 2.94 | 2.62 | 3.82 | 3.54 |
| Autocor. | 0.72 | 0.65 | 0.07 | 0.07 | 0.88 | 0.89 | -0.34 | -0.09 | -0.00 | 0.19 |

Table 3 MtV Model Estimated Parameters, 1984Q2-2018Q3

| | Coefficients | T-value |
|-----------------------|--------------|---------|
| Constant | 4.83*** | 8.92 |
| Industrial Production | 0.01 | 1.40 |
| CPI core | -0.12*** | -7.75 |
| MSCI | 0.00 | 0.38 |
| Term Spread | -0.03*** | -6.79 |
| Credit Spread | -0.02 | 0.08 |
| SMB | 0.00 | 0.21 |
| HML | -0.00 | 0.72 |
| WML | -0.00 | 0.29 |
| PS | 0.01*** | 2.03 |

| | |
|----------------|------|
| Adj. R squared | 0.41 |
|----------------|------|

Note: *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Table 4 Summary Statistics of Total Return Series, 1984Q2-2018Q3

| | NPI | TAR-TAR | TAR-TAR robust | AR | AR robust | GLM | GLM robust | REITs unlevered | MtV | TBI |
|------------------|--------|---------|-------------------|--------|-----------|--------|---------------|--------------------|--------|--------|
| Max | 5.42 | 35.85 | 13.70 | 31.20 | 11.48 | 10.60 | 6.76 | 18.16 | 12.66 | 19.05 |
| q99 | 5.09 | 22.94 | 12.41 | 20.18 | 11.02 | 8.59 | 5.89 | 13.90 | 11.61 | 12.46 |
| q95 | 4.48 | 13.21 | 10.39 | 12.06 | 9.38 | 5.68 | 5.40 | 9.52 | 9.89 | 7.75 |
| q90 | 3.79 | 10.90 | 8.10 | 9.72 | 7.45 | 4.66 | 4.39 | 8.28 | 8.75 | 6.91 |
| q75 | 3.09 | 5.80 | 5.28 | 5.44 | 4.42 | 3.54 | 3.32 | 5.38 | 5.92 | 4.56 |
| Median | 2.36 | 2.24 | 2.14 | 2.27 | 2.22 | 2.39 | 2.40 | 2.61 | 2.44 | 2.32 |
| Average | 1.99 | 2.34 | 1.98 | 1.95 | 1.89 | 1.98 | 1.98 | 2.68 | 1.92 | 2.15 |
| q25 | 1.64 | -0.62 | -0.48 | -0.36 | -0.22 | 1.21 | 1.20 | 0.27 | -1.91 | 0.03 |
| q10 | -0.05 | -5.61 | -4.26 | -4.78 | -3.39 | -0.68 | -1.20 | -2.15 | -5.83 | -2.98 |
| q05 | -1.58 | -8.94 | -7.34 | -7.85 | -6.01 | -2.75 | -2.77 | -5.26 | -7.57 | -4.76 |
| q01 | -6.61 | -18.97 | -12.42 | -28.30 | -11.80 | -9.29 | -5.59 | -12.97 | -11.00 | -9.13 |
| Min | -8.24 | -39.07 | -13.84 | -55.76 | -12.46 | -19.88 | -6.23 | -17.48 | -11.73 | -16.69 |
| St.dev. | 2.07 | 8.08 | 5.09 | 8.61 | 4.52 | 3.26 | 2.34 | 4.91 | 5.44 | 4.32 |
| Skewness | -2.36 | -0.47 | -0.44 | -2.43 | -0.53 | -2.73 | -1.15 | -0.69 | -0.38 | -0.35 |
| Kurtosis | 8.07 | 6.76 | 0.71 | 16.88 | 0.85 | 15.80 | 2.03 | 3.01 | -0.45 | 3.61 |
| Autocor. | 0.80 | -0.33 | -0.09 | -0.17 | -0.12 | 0.03 | 0.40 | 0.06 | 0.72 | -0.02 |
| Correl. with TBI | 0.50 | 0.33 | 0.35 | 0.37 | 0.34 | 0.38 | 0.40 | 0.19 | 0.30 | 1.00 |
| MaxDD | -23.79 | -46.79 | -35.38 | -64.38 | -36.06 | -27.49 | -19.84 | -30.97 | -78.64 | -34.04 |
| ES 5% | -4.91 | -17.59 | -10.18 | -21.54 | -9.03 | -7.79 | -4.56 | -10.37 | -9.59 | -8.03 |

Note: Total returns are in percent. The reported kurtosis is actually the excess kurtosis.

Table 5 Tests of Similarity Between All Series and National Transaction-Based Series, 1984Q2-2018Q3

| | NPI | TAR-TAR | TAR-TAR robust | AR | AR robust | GLM | GLM robust | REITs unlevered | MtV |
|----------------------------|------|---------|-------------------|------|-----------|------|---------------|--------------------|------|
| T-test | 0.62 | 0.78 | 0.71 | 0.77 | 0.55 | 0.64 | 0.62 | 0.29 | 0.65 |
| Kruskal-Wallis | 0.64 | 0.85 | 0.84 | 0.91 | 0.73 | 0.83 | 0.63 | 0.20 | 0.89 |
| Bartlett | 0.00 | 0.00 | 0.06 | 0.00 | 0.59 | 0.00 | 0.00 | 0.13 | 0.01 |
| Fligner-Killeen | 0.00 | 0.00 | 0.09 | 0.01 | 0.43 | 0.00 | 0.00 | 0.35 | 0.00 |
| Anderson-Darling | 0.00 | 0.06 | 0.43 | 0.17 | 0.82 | 0.01 | 0.00 | 0.26 | 0.04 |
| Skewness simu. | 0.00 | 0.79 | 0.83 | 0.00 | 0.72 | 0.00 | 0.21 | 0.57 | 0.91 |
| Kurtosis simu. | 0.01 | 0.02 | 0.24 | 0.00 | 0.28 | 0.00 | 0.51 | 0.93 | 0.01 |
| ES 5% simu. | 0.08 | 0.00 | 0.20 | 0.00 | 0.43 | 0.80 | 0.07 | 0.18 | 0.31 |
| Mahalanobis Distance | 0.99 | 0.00 | 0.07 | 0.00 | 0.15 | 0.85 | 0.96 | 0.02 | 0.02 |
| Outliers Proportion (%) | 0.00 | 7.97 | 0.00 | 6.52 | 0.00 | 0.00 | 0.00 | 1.45 | 0.72 |
| Real estate allocation (%) | 68.0 | 17.9 | 29.1 | 16.9 | 31.3 | 46.5 | 61.7 | 36.8 | 10.1 |

Note: Table 5 reports the p-values of various tests used to compare each desmoothed series and unlevered REIT series with the reference transaction-based series. It also reports the proportion of outliers found in the series. The Welch T-test and Kruskal-Wallis test investigate if realizations of compared groups could be considered as coming from the same distribution on average. The Bartlett and the Fligner-Killeen tests assess if both groups have the same variance. The Anderson-Darling tests check if the cumulative distribution of compared groups is the same. The latter test is similar to the Kolmogorov-Smirnov but it is more sensitive to differences in tail quantiles. No satisfactory statistical test exists for comparing skewness and kurtosis of two distributions. Hence, we propose estimating confidence intervals through a simulation process and derive the corresponding p-values. We apply the jackknife method for resampling the benchmark series, with resampled series length being 90% on the benchmark series length and the number of sample being 10,000. The moments of interest are then computed on each resampled series and the 10,000 measures allow defining the confidence intervals. In addition to skewness and kurtosis, we also simulated the confidence intervals for the expected shortfall at 5%. As we use time series, we also perform one test taking into account the temporal dimension of the data, by measuring the Mahalanobis distance between each series and the benchmark. The latter is similar to the well-known tracking error, but it additionally takes into account the variances and correlation structure of the compared series. The significance is estimated relying on bootstrapped critical values. Finally, the outlier proportion is the number of observations relative to the series length being over given upper and lower thresholds defined as a specific quantile of the time-varying transaction-based series distribution. Time-varying parameters of the latter distribution are obtained by fitting an ARMA(1, 1)-GARCH(1, 1) model with skewed-Student distribution on the transaction-based series. The threshold quantiles are defined such that observations are taken as outliers if they belong to the 1% most extreme realizations of the estimated model at a given point in time. Finally, the optimal allocation to real estate achieved by maximizing the Sharpe ratio of a portfolio made of real estate, stocks and bonds is reported for each investigated series in order to compare with the 32.5% allocation achieved with the TBI as reported in Table 1.

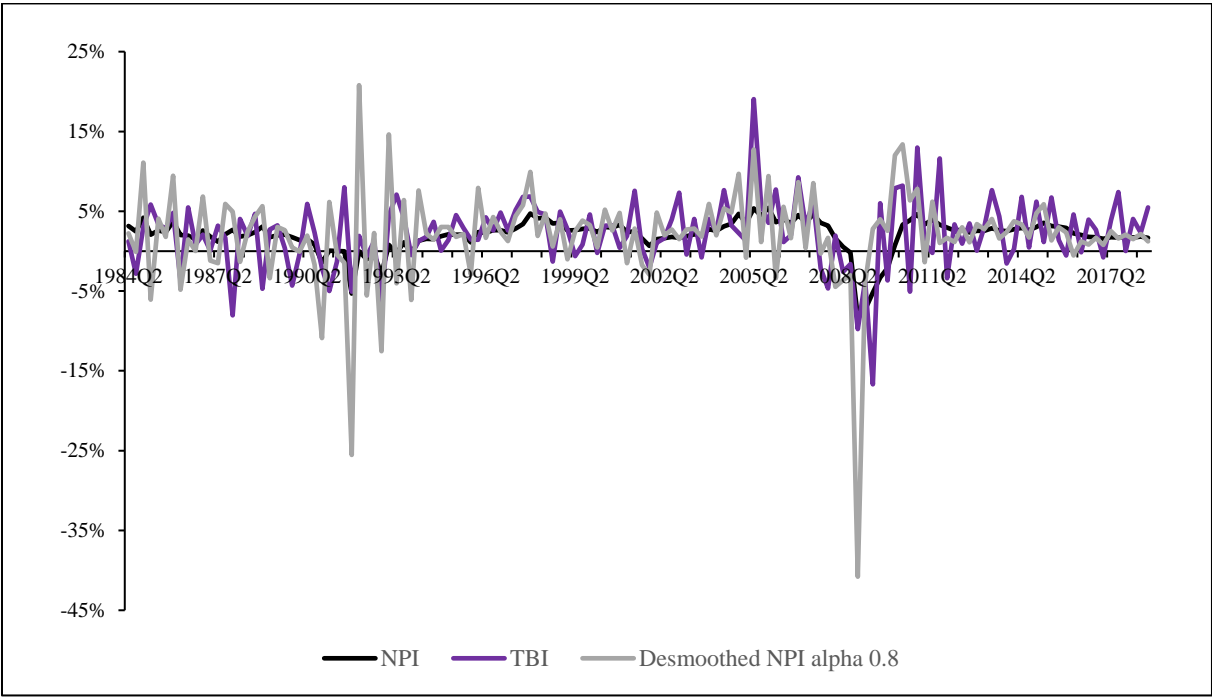


Figure 1 NPI, TBI, and desmoothed NPI Total Returns, 1984Q2 – 2018Q3

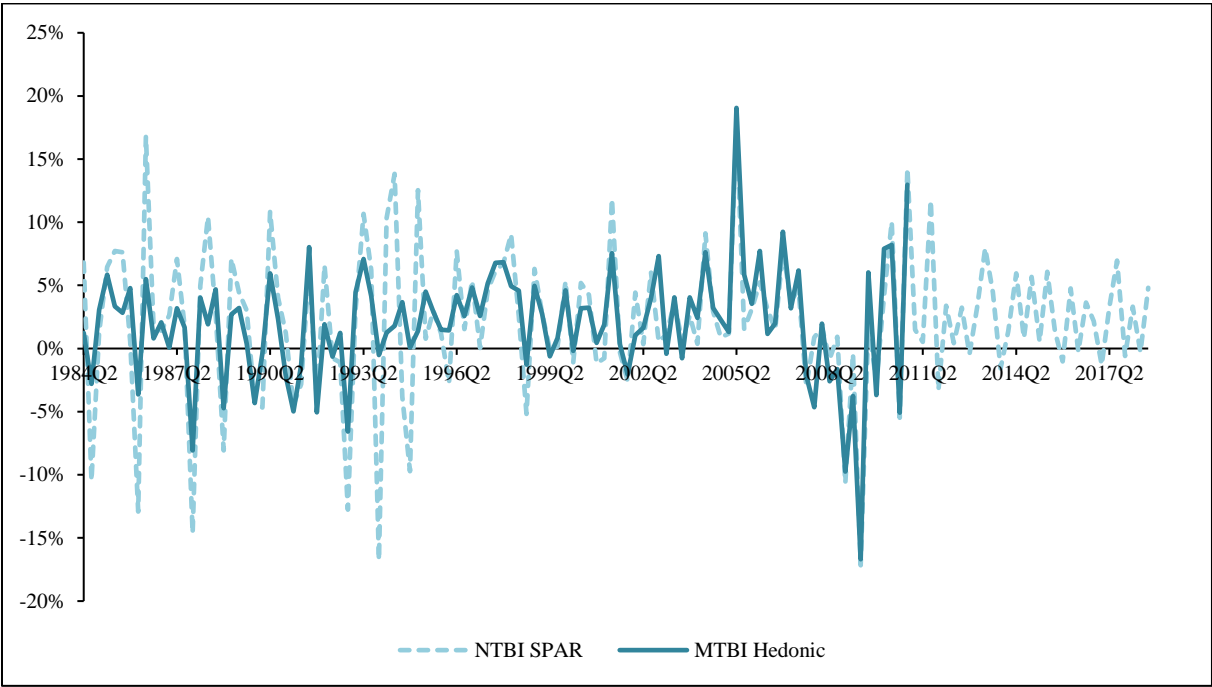


Figure 2 Hedonic MTBI and SPAR NTBI, 1984Q2 – 2018Q3

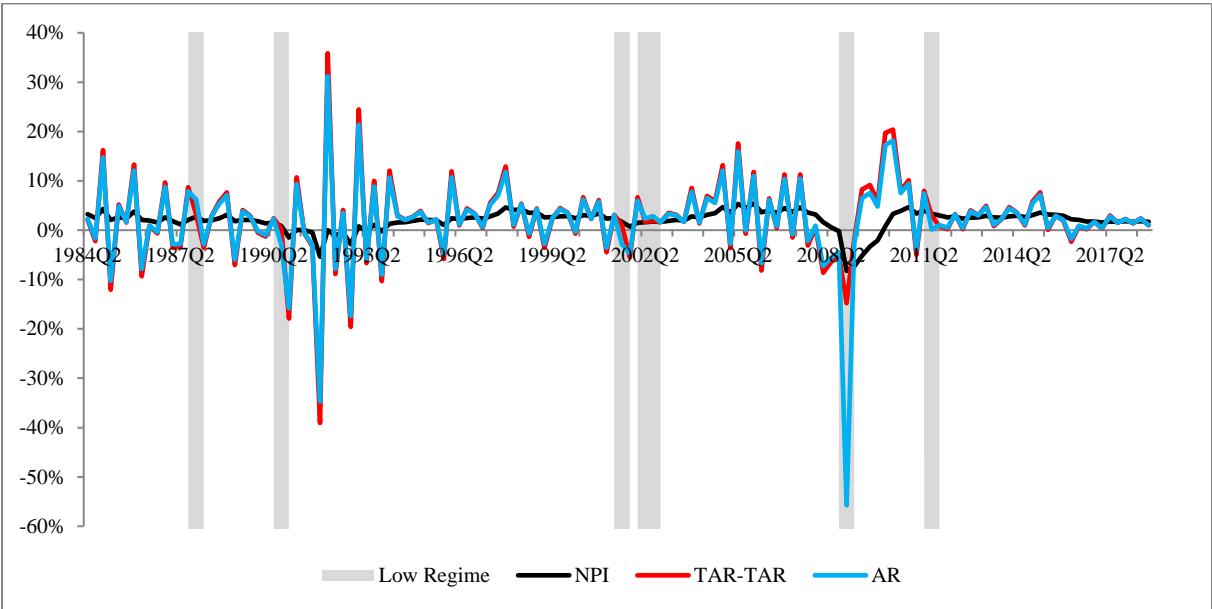


Figure 3 NPI, TAR-TAR, and AR Returns with Regimes, 1984Q2 – 2018Q3

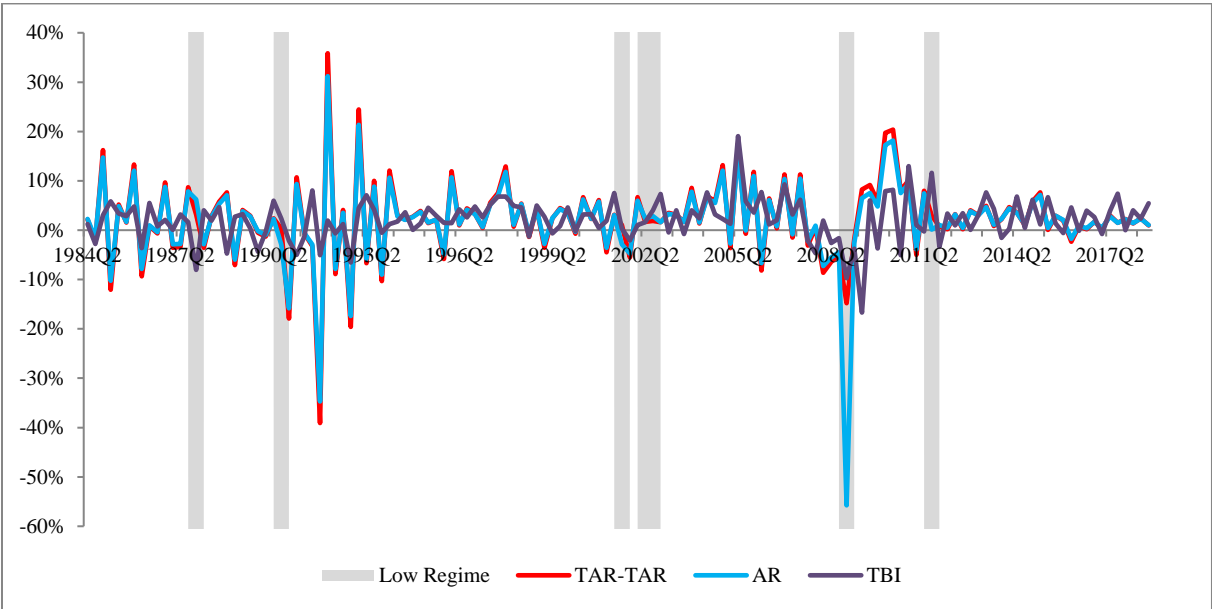


Figure 4 TBI, TAR-TAR, and AR Returns with Regimes, 1984Q2 – 2018Q3

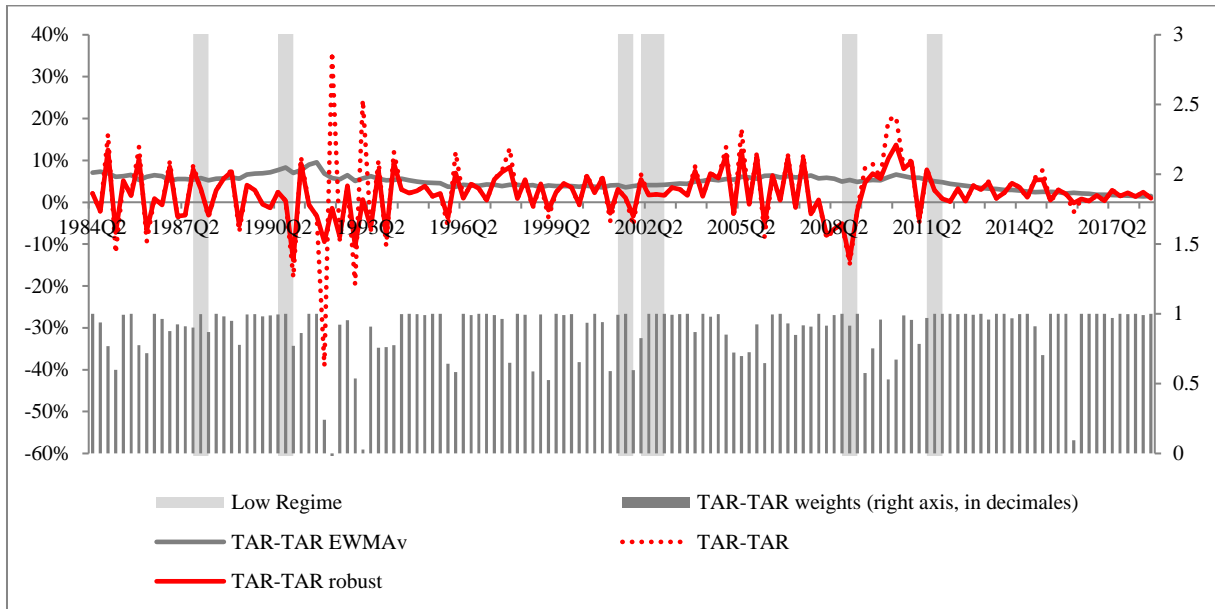


Figure 5 TAR-TAR, and Robust TAR-TAR Returns with Regimes and Robust Weights

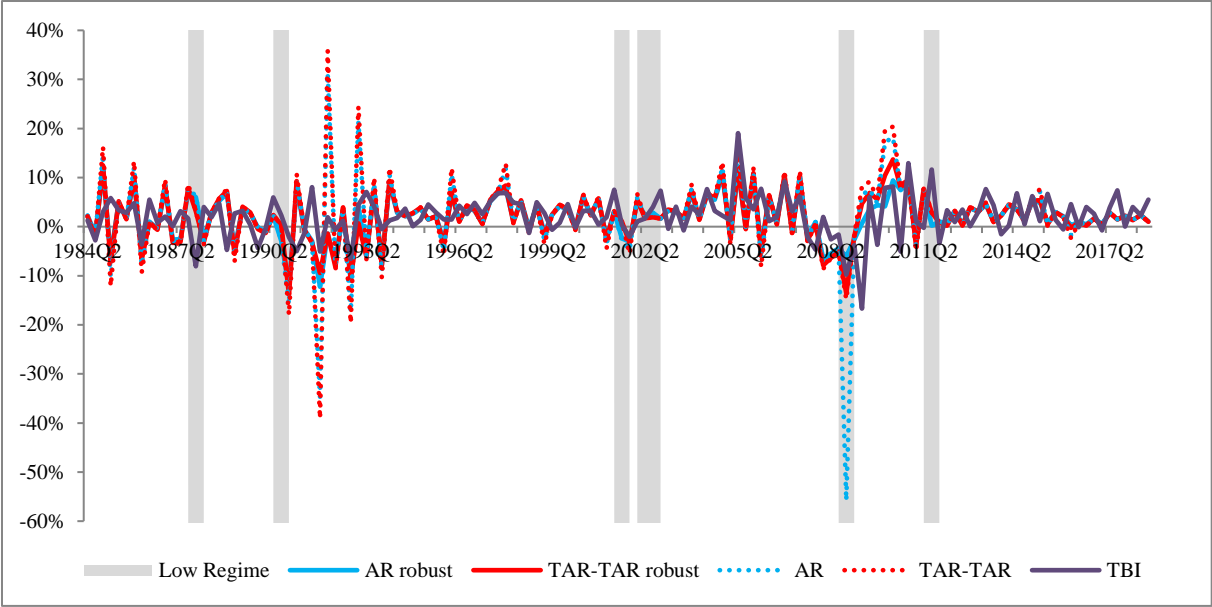


Figure 6 TBI, TAR-TAR, and AR Total Returns with Robust Series and Regimes, 1984Q2 – 2018Q3

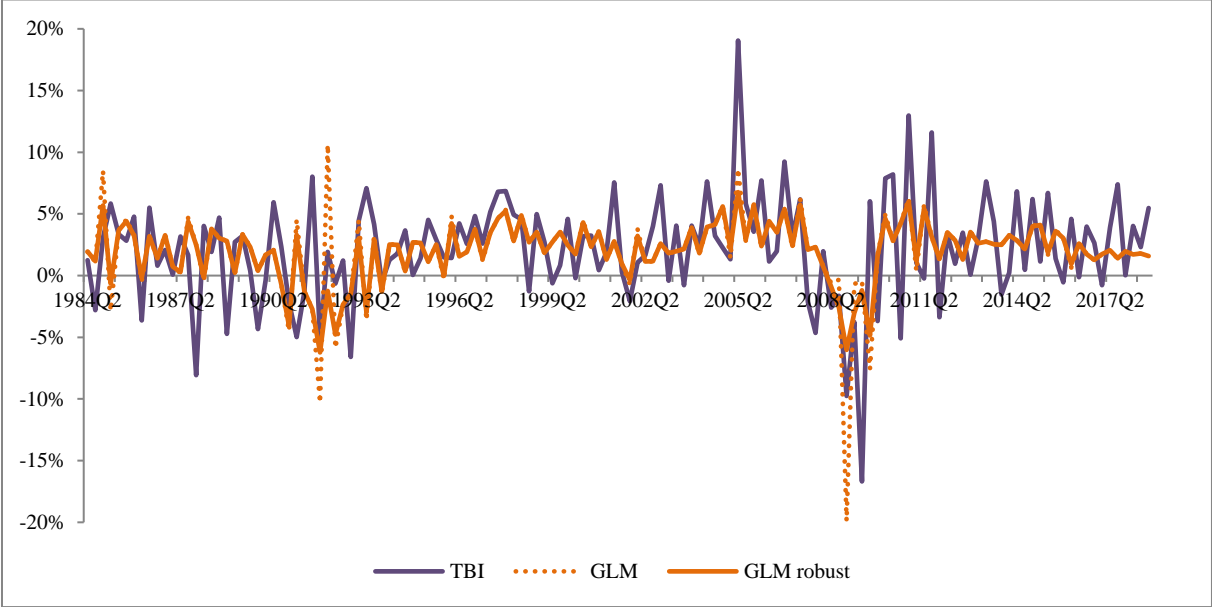


Figure 7 TBI, GLM, and Robust GLM Returns, 1984Q2 – 2018Q3

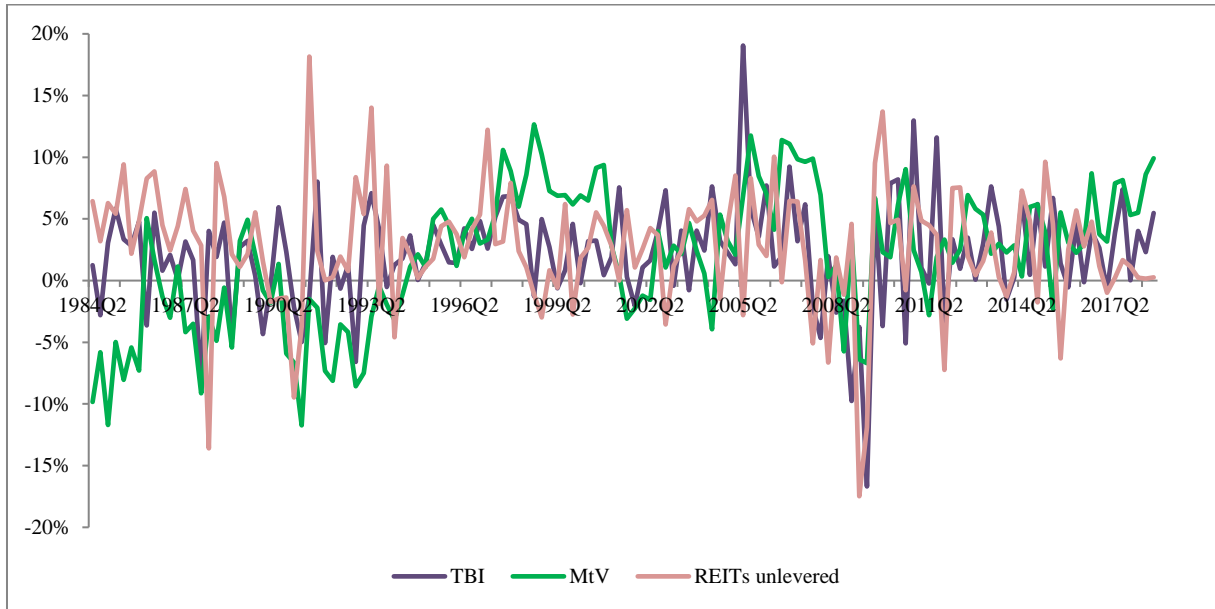


Figure 8 TBI, MtV, and Unlevered REIT Returns, 1984Q2 – 2018Q3

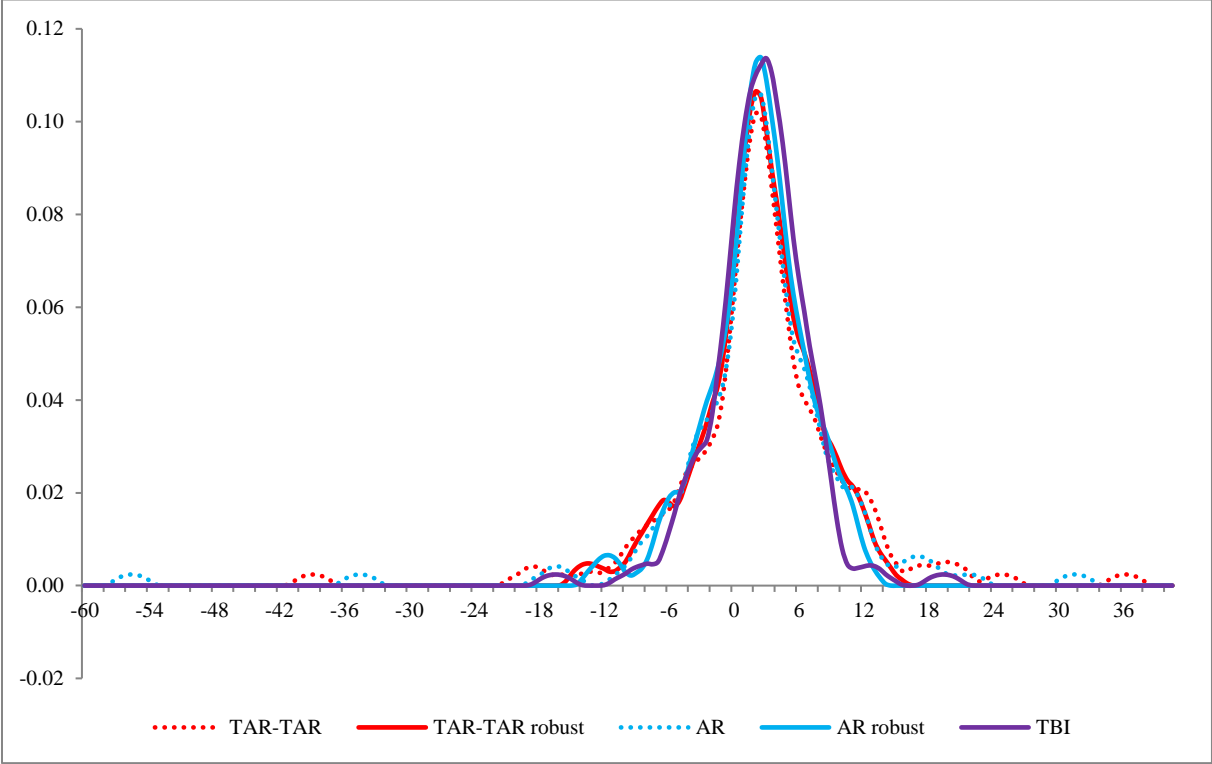


Figure 9 Kernel Density Estimation of Total Return Distributions, 1984Q2 – 2018Q3

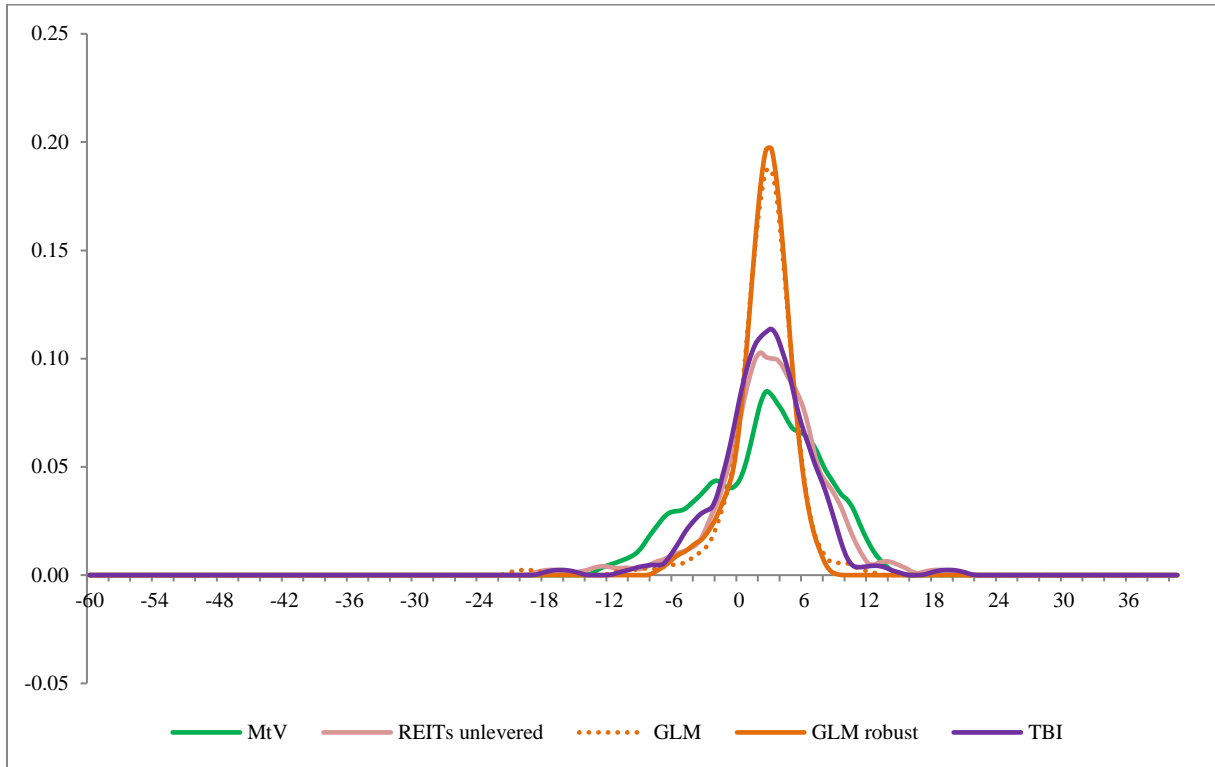


Figure 10 Kernel Density Estimation of Total Return Distributions cont'd, 1984Q2 – 2018Q3