Filling the Gaps with MICE: Addressing Missing Data in Real Estate Price Indices

Miriam Steurer*[a] and Sabrina Spiegel[b]

[a]Department of Economics, University of Graz, Austria, miriam.steurer@uni-graz.at
[b]ZTdatenforum, Graz, Austria

March 30, 2026

Abstract

Missing data are a common feature of transaction-based real estate price indices, where key property characteristics required for hedonic quality adjustment often contain missing values. Standard approaches, such as complete-case analysis, can distort the measurement of price dynamics due to sample-selection and composition effects. This paper proposes multiple imputation as a way to handle missing characteristic values in index construction. The goal of this imputation process is not primarily to recover individual missing values, but to stabilize quality adjustments by integrating incomplete observations back into the estimation sample. We use multiple-imputation-by-chained-equations (MICE) as a flexible framework for this task. Because Rubin's rules are not consistent with the multiplicative structure of price indices, we develop an alternative aggregation procedure based on pooled growth rates. Empirically, we examine two applications. First, using a complete dataset of Vienna apartment transactions, we find that in a large and relatively homogeneous market, hedonic price indices are fairly robust to missing data under most missingness scenarios, so that differences between complete-case estimation and alternative imputation approaches remain limited, although MICE generally performs best. Second, for the much thinner and more heterogeneous market for Austrian office unit transactions, we find that imputation materially changes index dynamics, reflecting time-varying composition effects induced by extensive missingness in core descriptive variables. Overall, multiple imputation with MICE performs well when the observed data contain enough information to infer the missingness structure. However, it cannot eliminate bias when entire segments of the price-quality distribution are unobserved. Across MICE specifications, results are similar when flexible methods with rich predictor sets are used, whereas restricted linear specifications perform less well. More broadly, the paper argues that the treatment of missing data should be addressed more explicitly in the hedonic price-index literature.

Keywords: Real estate price indices; Missing data; Multiple imputation; Multiple Imputation by Chained Equations (MICE); Hedonic regression; Transaction-based indices; Commercial real estate; Official statistics
JEL codes: R31, E31, C81, R33, C55, E44

Acknowledgments: We thank *ZTdatenforum* for providing us with the data for this analysis.

*Corresponding author.

1 Introduction

Missing data are a common characteristic of micro-level transaction data used to construct property price indices. This situation arises because transaction data are mainly collected for legal and tax-related purposes, such as proof of ownership and taxation, rather than for statistical analysis. In practice, missingness typically arises in the characteristics that enter the right-hand side of hedonic models—such as size, location attributes, or building quality—rather than in transaction prices themselves. Since these characteristics are central to hedonic quality adjustment (Rosen, 1974; Diewert, 2002; Hill, 2013), missingness can directly affect the measurement of price dynamics.

From a price index perspective, the central issue is not the accurate prediction of missing characteristics at the micro level, but the stability of the quality adjustment underlying the index. Missing data matter primarily because they reduce the effective estimation sample and alter its composition. Methods that restore incomplete observations can therefore improve index stability even if individual imputations are imperfect. This shifts the focus from prediction accuracy at the transaction level to the effect of missing data on index construction. In practice, however, most official statistical work continues to rely on complete-case estimation, implicitly discarding observations with incomplete characteristics.

Guidance from the broader price index literature is of limited use for property price indices. In the context of consumer price indices (CPIs), methodological guidance primarily focuses on handling missing prices, where relatively simple methods such as imputation based on nearby observations or carry-forward rules are often adequate (Armknecht and Maitland-Smith, 1999). However, these approaches are not directly applicable to missing characteristics in hedonic models, where it is necessary to model the relationship between prices and characteristics. As a result, statistical agencies face a clear methodological gap. While missing characteristic values are common in property datasets, existing manuals (e.g., OECD et al. (2013)) offer little operational guidance on how to address them effectively.

Outside the price index literature, the treatment of missing data has been studied extensively, both in statistics and econometrics, starting with the seminal work of Rubin (1976, 1987). In fields such as epidemiology and medical research, multiple imputation (MI) is widely regarded as the gold standard for handling incomplete data on characteristics, particularly under the missing-at-random assumption. In particular, multiple imputation by chained equations (MICE) has emerged as a flexible and robust framework that performs well across a wide range of settings (see, e.g., van Buuren (2007, 2018); Erler et al. (2016)). MICE generates multiple completed datasets by iteratively sampling from conditional distributions, preserving relationships among variables while reflecting imputation uncertainty.

Despite its widespread adoption in other disciplines, the use of multiple imputation in the estimation of

hedonic property price indices remains limited.[1] Moreover, its direct application is not straightforward. Price indices are nonlinear objects derived from estimated models, and standard pooling procedures for multiple imputation, such as Rubin's rules (Rubin, 1987), do not carry over directly. Because price indices are constructed by chaining multiplicative growth rates, aggregation must instead be performed at the level of adjacent-period changes. This creates a methodological problem that is specific to price index construction and is not addressed by the standard missing-data literature. At the same time, the index-number literature emphasizes the multiplicative structure of price indices and the importance of consistent aggregation (Diewert, 2002; Hill, 2013). Against this background, three questions arise: (i) whether missing data materially affect hedonic price indices, (ii) whether multiple imputation can mitigate the effects of missing characteristic values on index trajectories, and (iii) how imputation-based estimates can be aggregated in a way that is consistent with the multiplicative structure of price indices.

This paper addresses these questions and makes a conceptual, a methodological, and an empirical contribution. Conceptually, we argue that the imputation of missing characteristics should be explicitly addressed in the hedonic price index literature. Methodologically, we show that standard pooling rules for multiple imputation results do not apply directly to chained price indices and propose an operational solution based on pooling adjacent-period growth rates. Empirically, we demonstrate how the consequences of missing data depend on market structure, missingness patterns, and the extent to which the missing region remains represented in the observed data. To make the framework operational for statistical practice, we also show how imputation uncertainty can be summarized using index dispersion bands.

To evaluate the empirical relevance of missing data and the performance of alternative imputation strategies, we consider two applications that cover different market conditions. The first application examines the Viennese apartment market, which features a high number of relatively similar transactions and is characterized by very moderate levels of missingness. Because the underlying dataset is essentially complete, we can introduce controlled missingness to evaluate the performance of alternative imputation approaches. The second application focuses on commercial office transactions, a substantially more heterogeneous setting with pronounced spatial and temporal variation and only around 1/20th as many transactions as the Viennese apartment market. This contrast allows us to assess how the impact of missingness and the benefits of imputation depend on the underlying market environment.

Our results yield three empirical insights. First, we show that in relatively homogeneous markets, imputation is of limited importance under most missingness scenarios. In these settings, missingness

[1] We alert the reader that the term "imputation" can be confusing in the hedonic price index literature because there already exists a specific hedonic regression technique called the "imputation method" (also known as the double imputation, double regression, or double prediction method) (OECD et al., 2013; Hill, 2013; Hill et al., 2018). This method is not the subject of this paper.

primarily increases dispersion, while price indices remain broadly robust. Consequently, differences across imputation methods are also small. Second, in heterogeneous markets with structured missingness, ignoring missing data can lead to substantial bias, and multiple imputation mitigates this bias by restoring the effective sample size. Third, while flexible MICE specifications (e.g., predictive mean matching and random forests) produce similar results, more restricted linear specifications perform less well, highlighting the importance of a flexible imputation algorithm and a sufficiently rich predictor set.

The remainder of the paper proceeds as follows. Section 2 outlines the methodological framework, including the multiple imputation procedure, the aggregation of price indices across imputations, and the construction of imputation bands. Section 3 presents the Vienna apartment application, in which controlled missingness enables direct evaluation against a benchmark index. Section 4 turns to the Austrian office market, where missingness is real, substantial, and structured. Section 5 discusses the implications for statistical practice and derives practical guidance. Section 6 concludes.

2 Methodological framework and practical guidance

This section introduces the methodological framework for handling missing characteristics in hedonic price index construction. We first explain why missing characteristic data matter for hedonic price indices, then introduce multiple imputation via MICE, and finally show how imputation can be integrated into price index construction in a way consistent with index-number theory.

2.1 Why multiple imputation is a natural solution

We are concerned about missing data because we aim to minimize the impact of missing characteristics on the final price index. Consequently, the key question is not how accurately we can recover missing values at a micro level, but rather which methods most effectively stabilize the quality adjustments that form the basis for hedonic index compilation. A priori, several approaches to handling missing data are possible. The most common approach in practice is to ignore missingness and perform the analysis on a complete-case basis. That is, only observations without missing values are retained for estimation. In most statistical software packages, complete-case analysis is the default behavior, and users are not always aware that observations are being dropped in this way. A second approach is single imputation, which replaces each missing value with a single estimate and treats the completed dataset as if fully observed. The methods used for single imputation range from very simple procedures (e.g., mean imputation) to more elaborate model-based approaches, but they all share the feature that exactly one value is imputed for each missing observation. While straightforward to implement, this approach treats imputed values as if they were actually observed and therefore ignores the uncertainty

introduced by missing data (Rubin, 1976, 1987; Little and Rubin, 2019; Schafer, 1999). As a result, it can lead to overconfident inference and potentially unstable estimates, particularly when the extent of missingness is substantial.

The approach we consider in this paper is multiple imputation. Multiple imputation replaces each missing value with several plausible draws from a predictive distribution, thereby generating multiple completed datasets that reflect different realizations of the missing information (Rubin, 1976, 1987). Each dataset is analyzed separately, and the resulting estimates are subsequently aggregated. Variation across the imputed datasets captures the uncertainty associated with the missing data. From a price index perspective, the appeal of multiple imputation is twofold. First, it restores incomplete observations to the estimation sample, thereby stabilizing the set of transactions used for quality adjustment. Second, it provides a natural way to quantify the sensitivity of the resulting index to missing data, for example, through dispersion bands across imputations.

In this paper, we implement multiple imputation using the Multiple Imputation by Chained Equations (MICE) framework (van Buuren, 2007, 2018). MICE is particularly well-suited to real estate applications, as it allows for variable-specific imputation models and accommodates different algorithms across variables. This flexibility is important in settings with heterogeneous property characteristics, where relationships may be nonlinear and involve complex dependence structures. Several imputation methods available within MICE can capture such features in a data-driven manner. In the empirical applications below, we focus on two widely used non-parametric approaches within the MICE framework: predictive mean matching (PMM) and random forests (RF).

The MICE framework is implemented via an iterative algorithm based on chained equations. Let $Y = (Y_1, \ldots, Y_p)$ denote the set of variables used in the imputation step, including transaction price, time indicators, and a set of property characteristics and auxiliary predictors. Partition the data into observed and missing components, $Y = (Y_{\text{obs}}, Y_{\text{mis}})$. For each incomplete variable Y_j (i.e., variable with missing data), MICE specifies a conditional model

$$P(Y_j \mid Y_{-j}, \theta_j)\,,$$

where Y_{-j} denotes the remaining variables used as predictors and θ_j are model parameters. Given these conditional models, the chained-equations algorithm proceeds as follows. The procedure begins with an initial set of imputations, typically obtained from simple random draws or preliminary estimates. It then cycles through the incomplete variables, updating them one at a time by drawing from their conditional predictive distributions. This iterative process is repeated until convergence or the maximum number of iterations is reached. The entire procedure is subsequently repeated m times, resulting in m completed datasets.

As in most imputation applications, MICE is motivated under a missing-at-random (MAR) assump-

tion: conditional on the observed variables included in the imputation model, missingness does not depend on the unobserved value itself (Rubin, 1976; Little and Rubin, 2019; Enders, 2022). The plausibility of this assumption increases when the imputation model includes both the variables used in the final analysis and additional auxiliary predictors that help explain missingness (Rubin, 1987; van Buuren, 2018).

2.2 Choice of imputation model within MICE

Within the MICE framework, the analyst must specify how each variable with missing values is imputed. The default implementation in the `mice` package (van Buuren and Groothuis-Oudshoorn, 2011) uses predictive mean matching (PMM) for continuous variables, logistic regression for binary variables, and polytomous logistic regression (polyreg) for unordered categorical variables with more than two categories. PMM, originally proposed by Little (1988b), is attractive for real estate applications because it imputes missing values using observed values from similar transactions. This approach avoids implausible extrapolation and ensures that imputed characteristics remain grounded in actual market observations. In relatively homogeneous markets, PMM tends to perform well and is also computationally efficient.

In addition to the default MICE specification, we also consider random-forest (RF) imputation. Random forests are flexible, nonparametric predictors that can capture nonlinear relationships and interactions without requiring explicit functional-form assumptions (Breiman, 2001). This flexibility is appealing in real estate data, where relationships among size, location, and quality characteristics are often complex. An additional advantage is that RF implicitly performs variable selection, which can be useful in high-dimensional settings. However, these gains come at the cost of a substantially higher computational burden.

Despite differences in the algorithms used within MICE, the empirical results presented in this paper indicate that the resulting price indices are similar as long as flexible specifications with sufficiently rich predictor sets are employed. In both empirical applications, MICE-default and MICE-RF yield closely aligned index trajectories. The differences become more pronounced when the imputation model is limited, either because of simplified specifications within MICE or because simpler imputation methods such as mean or linear regression imputation are used.

2.3 Hedonic price indices and aggregation across imputations

2.3.1 Hedonic price indices

Hedonic regression based on transaction data is a leading approach for constructing property price indices (Eurostat, 2013, 2017; Hill et al., 2018). We focus on the hedonic time-dummy (TD) method and its rolling-window variant, the rolling time-dummy (RTD) approach. For a given estimation sample, we estimate

$$\ln p_n^t = \beta_0 + \sum_{\tau=1}^{T} \delta^\tau D_n^\tau + \sum_{k=1}^{K} \beta_k z_{nk}^t + \epsilon_n^t, \tag{1}$$

where the time-dummy coefficients δ^τ capture quality-adjusted price changes. The corresponding index is given by $P_{0,t} = \exp(\hat{\delta}^t)$.

The rolling time-dummy (RTD) approach relaxes the assumption of constant coefficients over the full sample by estimating the hedonic regression over overlapping time windows. Specifically, for each period t, the model is estimated using transactions within a window of length W, covering periods $t - W + 1$ to t. The resulting time-dummy coefficients therefore reflect local comparisons within each window rather than a single global regression. As the window moves forward one period at a time, this procedure generates a sequence of overlapping local estimates of price changes, which are then chained to form a continuous price index. By allowing coefficients to vary gradually over time and by pooling observations within each window, the RTD approach can improve stability in thin or heterogeneous markets while retaining the interpretability of the time-dummy framework (Shimizu et al., 2010; Hill et al., 2022).

A key feature of this approach is that property characteristics enter the model as control variables, while the index itself is determined by the estimated time effects. Consequently, errors in imputed characteristics affect the index only indirectly through their impact on the quality adjustment. In particular, imputation influences the index primarily through its effect on sample composition rather than through the precise prediction of missing characteristics. This helps explain the empirical robustness of hedonic TD indices to different imputation methods documented below.

However, missing data can affect the index through a different mechanism: changes in the composition of the estimation sample. Complete-case estimation discards transactions with missing characteristics, potentially leading to a thinner and compositionally distorted sample. This reduction in observations increases estimation variance and may also cause the estimated time effects to reflect shifts in sample composition rather than true price dynamics. In relatively homogeneous settings, such as the Vienna apartment market considered below, the variance channel tends to dominate, whereas in more heterogeneous markets composition effects become more important.

2.3.2 Aggregation and Rubin's Rules

With multiple imputation, the hedonic model is estimated separately for each of the m completed datasets, producing m potential index paths. The key challenge is then how to aggregate these into a single price index. Standard pooling procedures such as Rubin's Rules (Rubin, 1987) are designed for aggregating regression estimates, not price indices. Price indices are nonlinear, multiplicative objects constructed by chaining adjacent-period growth rates. As a result, aggregating model parameters or averaging index paths generally fails to preserve the underlying index-number structure, particularly in the time-dummy setting considered here.

Aggregation must instead respect the multiplicative structure of index construction (Eurostat, 2013; Hill et al., 2018). We therefore pool price changes rather than levels. For each imputed dataset, we compute adjacent-period growth rates from the estimated time effects. These growth rates are then averaged across imputations and chained to obtain a single price index.

We implement this aggregation via pooled growth rates as follows. Let $P_t^{\text{imp},i}$ denote the index level in period t obtained from imputed dataset i, $i = 1, \ldots, m$. Define the adjacent-period growth rate in each imputation as

$$1 + g_t^i = \frac{P_t^{\text{imp},i}}{P_{t-1}^{\text{imp},i}}, \qquad t = 1, \ldots, T. \tag{2}$$

We pool these growth rates across imputations using a geometric mean:

$$1 + g_t = \left(\prod_{i=1}^{m} (1 + g_t^i) \right)^{1/m}. \tag{3}$$

The pooled index is then obtained by chaining:

$$P_0^{\text{pool}} = 1, \qquad P_t^{\text{pool}} = \prod_{s=1}^{t} (1 + g_s), \qquad t = 1, \ldots, T. \tag{4}$$

Pooling at the level of adjacent-period growth rates ensures consistency with the multiplicative structure of price indices.

For comparison, we also compute a complete-case (CC) index by estimating the hedonic model using only observations with fully observed characteristics. Differences between the CC and pooled imputation indices therefore isolate the effect of missing-data treatment.

2.3.3 Imputation uncertainty and dispersion bands

Rather than producing a single completed dataset, the imputation procedure generates m plausible versions of the data, each yielding an index path specific to that version. The dispersion across these paths provides a natural measure of the index's sensitivity to missing data. To summarize this variation, we construct dispersion bands based on the cross-imputation distribution of index levels (e.g., 10–90 percentiles). These bands do not represent sampling uncertainty in the conventional sense. They are conditional on the observed dataset and capture only the additional variation arising from missing data and the imputation procedure. In particular, they do not reflect sampling variability that would arise from observing a different set of transactions.

From a price index perspective, the bands should therefore be interpreted as a diagnostic tool. They indicate how sensitive the estimated index is to alternative plausible completions of the missing data. Narrow bands suggest that missingness has little impact on the index, while wider bands indicate greater sensitivity of the inferred price dynamics to the imputation step.

2.4 Empirical Design: Two Contrasting Missingness Environments

In the next two sections, to assess how missing data affect hedonic price indices in practice, we consider two empirical settings that differ fundamentally in their data environment. The transaction data used for both settings are obtained from ZTdatenforum eGen (https://zt.co.at/), a commercial provider that collects and compiles official Austrian real estate transaction records from the registry offices. The first case study focuses on a large and relatively uniform dataset of apartment transactions in Vienna. In this dataset, missingness is minimal and can therefore be introduced in a controlled way. This creates a benchmark environment where the effects of missing data can be directly assessed. The second case study considers transaction data from the Austrian office market, which is substantially thinner and more heterogeneous, with high and systematically structured missingness in key characteristics. In this setting, no benchmark index is available, and the focus shifts to the stability and sensitivity to sample composition of the resulting indices.

Taken together, these two cases allow us to characterize how the impact of missingness depends on market structure, data availability, and the nature of the missingness process.

3 Application 1: An apartment price index for Vienna, Austria

Before turning to real-world missingness in Section 4, we first examine how missing data and alternative missing-data treatments affect hedonic price indices in a controlled setting. We begin with a clean, fully observed dataset of apartment transactions in Vienna and introduce synthetic missingness via various mechanisms and severity levels. Since the dataset is fully observed and relatively homogeneous, it serves as an effective benchmark environment where the effects of missing data can be clearly isolated. This enables us to differentiate between the variance and bias effects of missingness under controlled conditions.

The analysis proceeds in three steps. First, we introduce synthetic missingness under several mechanisms and missingness levels. Second, we construct hedonic price indices using complete-case estimation, single imputation, and multiple imputation. Third, we compare the resulting indices to the full-sample benchmark in order to assess their bias, stability, and overall accuracy.

3.1 Introducing synthetic missingness

We use a dataset of apartment transactions in Vienna covering the period from January 2015 to September 2023. The data were provided by *ZTdatenforum* and contain 67,292 transactions with complete information on all variables used in the analysis.[2] The data are further described in Appendix Section B.1.

The original dataset contains a large number of characteristics, but in this illustration we focus on a standard set of hedonic characteristics describing property price, location, and quality. These variables include transaction price (*price*), the time of sale (*Quarteryear*), postcode dummies (PLZ), internal size in square meters (*size*), an indicator for sales by a builder (*builder*), a variable describing if and how much was paid for a parking spot (*car_price*), and the distance to the nearest doctor's surgery (*doc_dist*) and the distance to the nearest pharmacy (*pharm_dist*).[3]

Starting from the complete baseline dataset, we impose synthetic missingness and study how the resulting index estimates change. Missing values are introduced only in the explanatory variables, while transaction price and the time variable remain fully observed throughout. We consider missingness levels ranging from 30% to 90%. For cell-wise designs, this refers to the share of missing entries across characteristics (excluding price and time), while for truncation designs it refers to the share of observations with missing size.

[2] We apply only minimal data cleaning, excluding transactions between related parties, transactions associated with bankruptcies, and apartments outside the size range of 20 to 400 square meters.

[3] Distance variables serve as proxies for urbanity and accessibility.

We implement three types of missingness mechanisms. First, under missing completely at random (MCAR), a fixed proportion of entries is deleted independently across observations and variables using the `mice::ampute` function with `mech = "MCAR"`.[4] By construction, missingness in one variable is independent of both observed and unobserved data.

Second, we generate non-random missingness (MNAR) using `mice::ampute` with `mech = "MNAR"` and `type = "LEFT"`. This specification makes missingness more likely for observations in the lower tail of each affected variable, thereby inducing systematic selection in the characteristics space.[5] Although the missingness rule is formally specified as MNAR, this does not imply that the resulting index problem is equally severe in every setting. In our Vienna application, the observed characteristics are strongly correlated with price and with one another. As a result, conditioning on the observed variables captures a substantial share of the variation associated with missingness. In that sense, the missing values remain partly recoverable from the observed data, even though the data-generating missingness rule itself depends on the unobserved value.[6] This differs fundamentally from the truncation designs considered below, where entire parts of the price–quality distribution are removed and therefore fall outside the support of the observed sample.

Third, we consider truncation-based missingness in the size variable. Transactions are ranked within each year by price, and size information is removed for observations in either the upper or lower part of the within-year price distribution (top or bottom truncation at 50%). This design mimics situations in which entire segments of the market are selectively unobserved. We focus on size because it is a key characteristic in hedonic models. Moreover, it mirrors the dominant missingness pattern in the Austria-wide office dataset considered in the next section, where size exhibits the highest rate of missingness. The truncation thresholds are calibrated so that the resulting share of missing size values approximately matches the target missingness level. In contrast to MCAR and MNAR designs, truncation directly removes parts of the price–quality distribution, thereby introducing a potential source of systematic bias in addition to increased sampling variability.

For MCAR and MNAR-induced missingness, we generate five independent incomplete datasets for each missingness level, which serve as Monte Carlo replicates of the missing-data process. These datasets differ only in the realization of missing entries. For truncated samples, we only have one incomplete dataset per missingness level. For each replicate, we construct a complete-case (CC) price index. We then apply multiple imputation within the MICE framework, generating $m = 5$ imputations for each replicate. This yields five completed datasets and corresponding index paths, which are aggregated

[4] As a diagnostic, we verify that Little's MCAR test does not reject the null of completely random missingness.

[5] Little's MCAR test strongly rejects MCAR under this design.

[6] This is consistent with the basic logic of hedonic modeling. Since Rosen (1974), the hedonic literature has treated transaction prices as systematically related to bundles of observed characteristics. If characteristics jointly explain a substantial share of price variation, it is not surprising that the same dependence structure can also provide predictive information for imputing missing characteristics. Of course, this does not eliminate the identification problem when missing observations fall outside the support of the observed sample.

into a single pooled index as described in Section 2.3.2.[7] Each replicate therefore produces both a complete-case index and a pooled imputation index. For comparison, we also construct indices based on single-imputation methods. This two-layer design separates two sources of variation: variation across realizations of the missing-data process (captured by the Monte Carlo replicates) and variation arising from the imputation procedure itself (captured by multiple imputations within each replicate). This distinction is central for interpreting differences in dispersion and bias across methods. In the baseline specification, imputation is implemented using the default MICE approach. Robustness to alternative imputation settings, in particular the use of random forests within MICE (MICE-RF), is examined separately. Appendix B.3 provides additional implementation details.

3.2 A benchmark apartment price index for Vienna

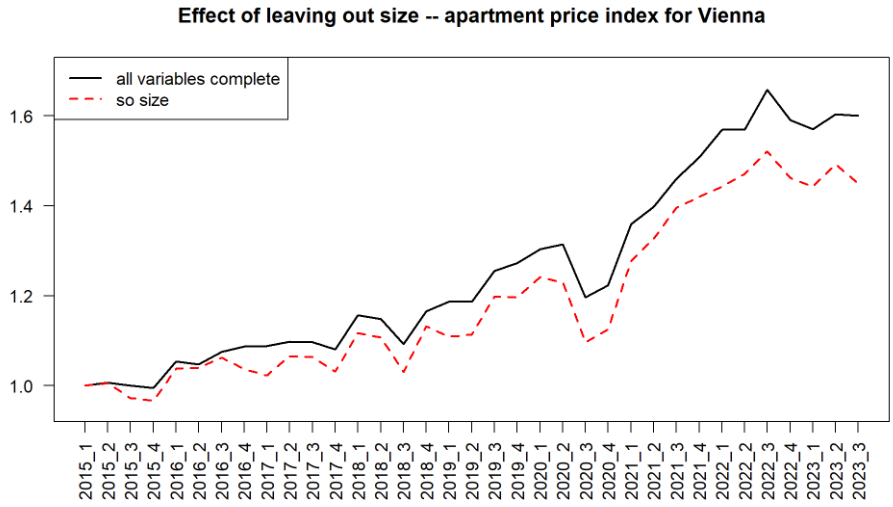
We construct a hedonic time-dummy price index according to equation (2), denoted by P_t^{bench}.[8] Constructed from the fully observed dataset, this benchmark index provides the reference path for all subsequent comparisons. This benchmark is constructed in a relatively homogeneous and transaction-rich market, providing a stable environment for evaluating the effects of missing data.

An important question is whether the benchmark index is sensitive to the choice of hedonic specification. Appendix Figure A.2 compares the time-dummy (TD) index to rolling time-dummy (RTD) indices constructed using different window lengths. The resulting trajectories are nearly identical, indicating that coefficient drift is limited in this dataset and that the benchmark is robust to alternative hedonic specifications when the full sample is used. The benchmark index is shown as the solid black line in Figure 1 and in all subsequent figures in this section. It exhibits sustained price growth over most of the sample period, with a temporary decline during 2020 associated with COVID-19 lockdown periods, followed by a period of stagnation from mid-2022 onward.

[7] The choice of $m = 5$ is at the lower end of typical practice but sufficient to illustrate qualitative differences across imputation methods. The primary objective of this section is to demonstrate the mechanics of MICE. In Section 4, we increase the number of imputations to $m = 100$.

[8] We intentionally do not apply influence-based data cleaning (e.g., via Cook's distance) in this section to avoid confounding the effects of imputation with post-estimation data treatment. However, we will evaluate the influence of removing influential observations in Section 4.

Figure 1: Why dropping a core descriptive variable is not a viable alternative



Note: This figure compares the full-sample hedonic index using the full set of core characteristics with an otherwise identical index that omits internal size (square meters). The divergence illustrates that dropping variables affected by missingness is not a credible strategy when key characteristics are involved. The full-sample index serves as the benchmark for all subsequent comparisons. **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations.

Figure 1 highlights an important practical issue. A potential response to missing data that we have not yet discussed is to exclude variables with incomplete information. However, when missingness affects a key characteristic such as size, this leads to substantial distortions in the resulting price index. This underscores the importance of principled missing-data treatments rather than ad hoc variable exclusion.

3.3 Simulation evidence

3.3.1 Simulation design and evaluation criteria

We evaluate the performance of price indices constructed under missing data relative to the full-sample benchmark. For each missingness mechanism and level, we generate multiple independent realizations of missing data. For each realization, we construct both a complete-case index and a pooled multiple-imputation index. The objective is not only to assess whether multiple imputation improves agreement with the benchmark on average, but also to understand how and why its effects differ from complete-case (CC) estimation.

We compare two estimators: a complete-case index and a multiple-imputation (MI) index based

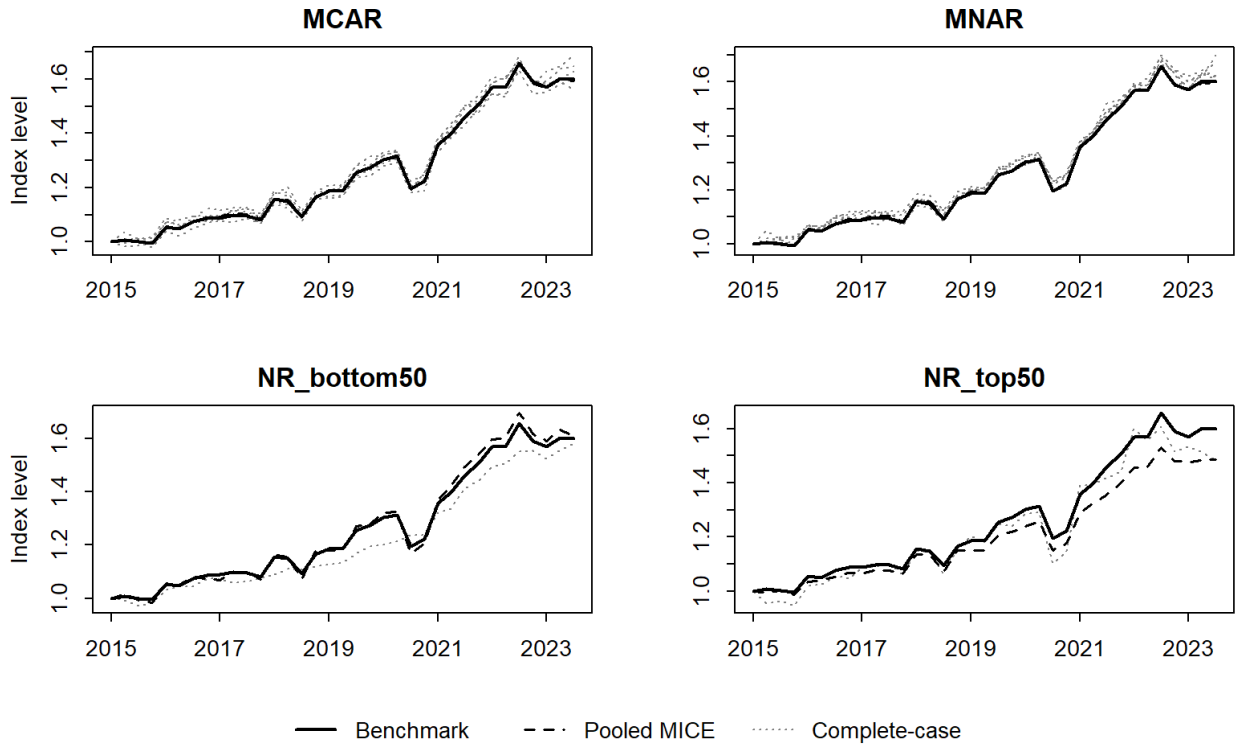
on datasets imputed using the MICE procedure, while holding the hedonic specification fixed. The simulation design incorporates two sources of variation: variation across realizations of the missing-data process and variation within each realization arising from the imputation procedure. Even under identical missingness mechanisms and rates, complete-case indices may differ across realizations because the usable subset of transactions changes.

Performance is evaluated using three complementary criteria: (i) visual inspection of index paths, (ii) signed quarterly growth differences, and (iii) root mean squared error (RMSE), which combines bias and dispersion across realizations. To ensure computational feasibility across a large number of simulation runs, the imputation model is based on a compact set of variables capturing size, location, and key quality characteristics of each property (see Appendix Table A.1). Although parsimonious, this set explains a substantial share of cross-sectional price variation in this relatively homogeneous market.

3.3.2 Visual evidence

Figure 2 illustrates how price indices constructed from incomplete data behave under different missingness mechanisms.

Figure 2: Finished hedonic price indices under selected missingness mechanisms at 50% missingness



Note: The solid black line denotes the full-sample benchmark index. Dashed lines show pooled MICE indices, while dotted lines represent complete-case indices across Monte Carlo realizations. All panels correspond to a missingness level of 50%. MCAR and MNAR denote random and non-random missingness mechanisms, respectively, while NR_bottom50 and NR_top50 refer to truncation of the lower and upper parts of the within-year price distribution. In the first two panels, the pooled MICE indices overlap completely with the full-sample benchmark. **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations.

Under MCAR and MNAR mechanisms, complete-case indices exhibit noticeable dispersion across realizations. On the other hand, imputation-based indices, constructed using the growth-rate pooling approach described in Section 2, cluster tightly around a nearly identical path.[9] This suggests that in this relatively homogeneous context, the presence of missing data primarily increases the variability of estimated growth rates, rather than introducing systematic bias.

Under truncation designs, the results differ more significantly. When missing data affects the lower end of the market (bottom truncation), imputation effectively stabilizes the index and reduces deviations from the benchmark. This stabilizing effect can occur because the observed data retain some overlap with the lower end of the price–quality distribution, such as through relatively expensive transactions involving larger but lower-quality apartments. This overlap provides enough support for

[9] Although the design is formally categorized as MNAR, strong correlations among characteristics make the resulting patterns appear empirically closer to Missing at Random (MAR) in this dataset.

the imputation model to partially recover the missing segment.

Conversely, in the case of top truncation, the observed sample contains limited information about high-quality or luxury properties. Transactions at the lower end do not offer enough support to infer the relationship between premium characteristics and prices. Consequently, the imputation model must extrapolate beyond the observed data, making it impossible to reconstruct the missing upper tail of the market, which leads to a consistent downward bias in both complete-case and imputed indices.

This underscores a fundamental identification constraint. Even when missing values are partly predictable from observed characteristics, imputation cannot recover market segments that lie outside the support of the observed sample. The effectiveness of imputation depends therefore not only on the missingness mechanism, but also on whether the missing region remains represented in the observed price–quality distribution.

3.3.3 RMSE and interpretation

There are two main channels through which missing data affect price indices: (i) increased dispersion due to reduced sample size, and (ii) bias arising from systematic differences between observed and missing data. Both effects are captured jointly by the root mean squared error (RMSE). In our setting, RMSE is computed across Monte Carlo replicates of the missing-data process. Lower RMSE indicates that quarterly inflation estimates are both closer to the benchmark and more stable across samples.

To understand the differences in RMSE across various scenarios, we decompose RMSE into its bias and variance components (see Appendix Table A.3). For the Vienna apartment data, we find that under MCAR and the MNAR-left design, variance accounts for most of the increase in RMSE in the complete-case estimator, while bias remains small. Even with missingness rates of up to 90% in right-hand side variables, both complete-case and imputed indices remain approximately unbiased on average. However, the complete-case estimates become significantly more variable as the rate of missingness increases.

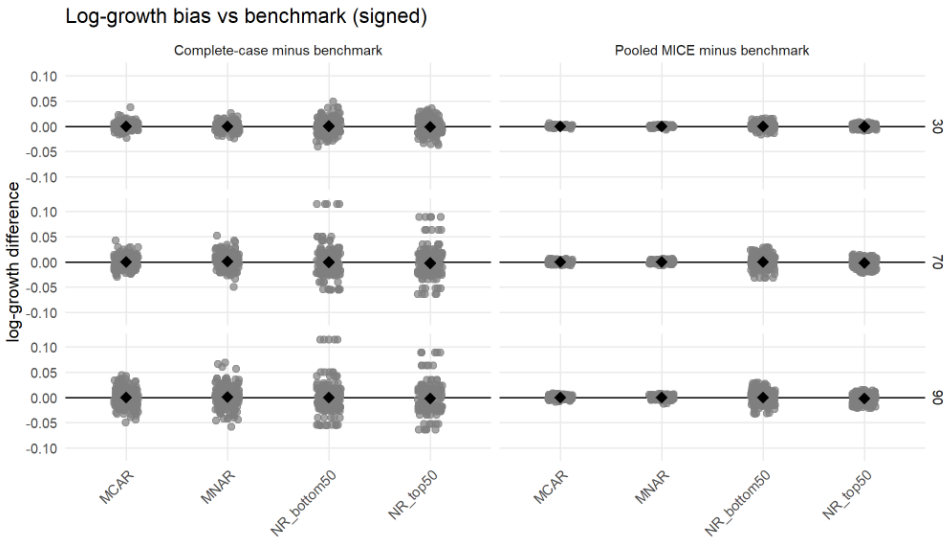
The situation differs under truncation-based missingness, where entire segments of the market remain unobserved. In these cases, bias becomes a more important source of error. The RMSE differences between complete-case and imputed indices reflect the extent to which the missing segment can be inferred from the remaining data. Consistent with the results above, imputation works well when the missing region lies within the support of the observed data, but cannot fully eliminate bias when entire market segments are unobserved.

The distribution of log-growth differences relative to the benchmark, shown in Figure 3, provides

a complementary perspective on the RMSE decomposition. Under MCAR and MNAR, both the complete-case and imputed indices are closely centered around zero. This indicates that, on average, the quarterly growth rates are approximately unbiased. However, the dispersion of the complete-case estimates is significantly larger, while the imputed indices show a tighter concentration. Under truncation-based missingness (NR_bottom50 and NR_top50), the distributions remain close to zero but exhibit greater dispersion and some asymmetry, indicating that selection primarily affects the stability of estimated growth rates rather than inducing large average biases at the quarterly frequency.

It is important to note that even small biases in period-to-period growth rates can accumulate over time. As a result, these biases become more apparent in cumulative inflation measures than in individual quarterly price changes.

Figure 3: Distribution of quarterly log-growth differences relative to the benchmark



The figure shows the distribution of signed quarterly log-growth differences relative to the full-sample benchmark across Monte Carlo realizations. The left panel reports complete-case estimates, while the right panel shows pooled MICE estimates. Rows correspond to different missingness levels (30%, 70%, and 90%), and columns distinguish missingness mechanisms (MCAR, MNAR, NR_bottom50, and NR_top50). Points represent individual realizations, and diamonds indicate average deviations. Under MCAR and MNAR, both estimators are centered close to zero, but complete-case estimates exhibit substantially greater dispersion. Under truncation-based missingness, dispersion increases and mild asymmetries emerge, while average quarterly growth differences remain small. **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations and simulations.

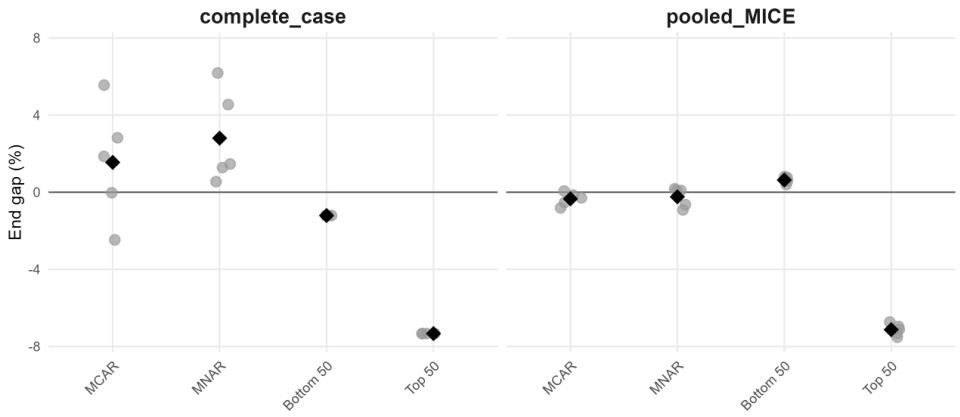
3.3.4 Cumulative inflation bias

One way to illustrate cumulative inflation differences is to examine end-of-sample cumulative inflation gaps relative to the benchmark. Figure 4 shows that complete-case indices exhibit greater dispersion

across realizations than indices based on MICE imputation. While different realizations of missingness lead to noticeable variation in complete-case indices, multiple imputation substantially reduces this dispersion, yielding much more stable end-of-sample outcomes across replicates. Appendix Figure A.3 shows that this pattern holds across different missingness levels.

However, under top truncation, both complete-case and imputed indices exhibit a persistent negative bias. As discussed above, this reflects the absence of support for high-end properties in the observed data.

Figure 4: End-of-sample index level gap relative to the benchmark (70% missingness)



Note: Each dot represents one simulation replicate; diamonds indicate the mean across replicates. The gap is defined as $100 \times (\text{Index}_{\text{end}}/\text{Benchmark}_{\text{end}} - 1)$, with negative values indicating underestimation. Complete-case indices exhibit substantial and mechanism-dependent biases, as well as noticeable dispersion across replicates due to differences in the underlying sample draws. In contrast, after MICE imputation, indices are tightly clustered around zero and show minimal cross-replicate variation. **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations and simulations.

3.3.5 Comparison with alternative missing-data treatments

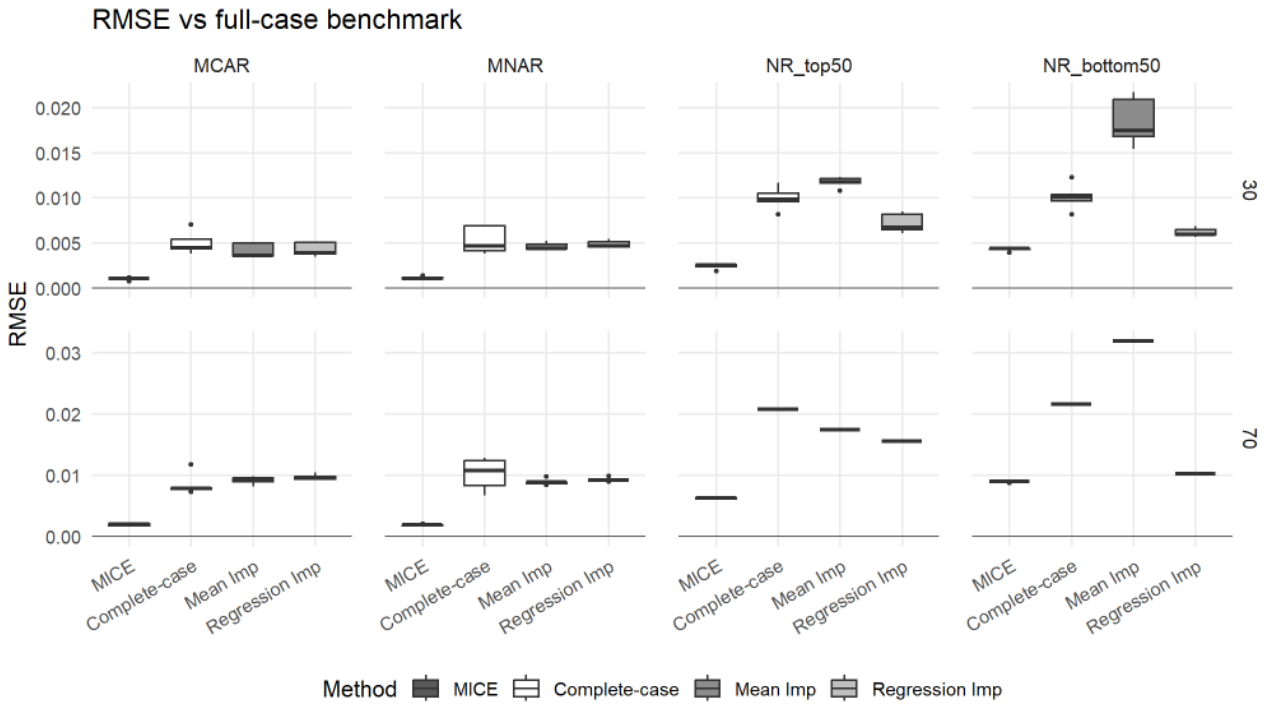
Next, we compare multiple imputation to alternative missing-data treatments. We consider two simpler approaches: mean imputation and single regression imputation. Mean imputation replaces missing values with the sample mean, retaining all observations but reducing cross-sectional variation and weakening identification of hedonic coefficients. Single regression imputation replaces missing values with predictions from auxiliary regressions, preserving some dependence structure but treating imputed values as known and thereby ignoring imputation uncertainty.

For each method, we aim to construct hedonic indices using the same specification as for the full-sample benchmark, where feasible. In practice, implementing regression imputation with the full set of predictors leads to numerical instability due to singularity issues (because of high correlation

between characteristics). We therefore use a reduced set of predictors for this method, further limiting its ability to capture the underlying dependence structure.

Figure 5 reports the RMSE of quarterly log price growth relative to the benchmark. Across all missingness mechanisms, multiple imputation using MICE consistently delivers lower RMSE and tighter dispersion across realizations, clearly outperforming both complete-case estimation and simpler imputation approaches. In contrast, mean and regression imputation reduce dispersion relative to complete-case estimation but exhibit similar bias. This reflects the fact that, in the Vienna apartment setting, bias in complete-case estimation is already limited. As shown in the next section, this conclusion does not hold in thinner, more heterogeneous markets, where simpler imputation methods can outperform complete-case estimation.

Figure 5: RMSE of quarterly log price growth relative to the full-case benchmark (all mechanisms)



The figure shows the root mean squared error (RMSE) of quarterly log price growth relative to the full-case benchmark index. Lower values indicate closer agreement with the benchmark. Across all missingness mechanisms, multiple imputation using the default MICE specification (or MICE-RF) yields the lowest RMSE and the most stable performance across replicates, clearly outperforming simpler imputation approaches and complete-case estimation. **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations and simulations.

3.3.6 Implications and transition to real data

The Vienna apartment market provided a controlled benchmark for evaluating missing-data treatments. Because the full dataset is observed prior to inducing missingness, index estimates can be directly compared to the benchmark. In practical applications, however, the true benchmark is not observed. Statistical agencies must construct price indices from datasets with genuinely missing characteristics. The relevant practical question is therefore whether multiple imputation yields more stable and more robust indices than complete-case estimation when the true benchmark is not available. Additionally, we must consider whether we can explain the observed differences in light of the empirically observed missingness structure.

The next section examines this question using transaction data from the Austrian office market. In contrast to the Vienna apartment market considered here, this setting is characterized by lower transaction volumes, greater heterogeneity, and weaker overlap in the support of observed characteristics. As a result, missingness is more likely to affect not only dispersion but also systematic bias, and the choice of imputation method becomes more consequential.

4 Application 2: A Commercial Office Price Index for Austria

We now turn to a setting with true missing data, where the underlying *true* price index is unobserved. We again use data from ztDatenforum[10], but now focus on office unit transactions. The data and underlying market environment differ markedly from the Vienna apartment case: the market is more heterogeneous, transaction volumes are lower, and there is substantial variation across properties and regions. A key feature of the data is that two important hedonic characteristics—size and age—are frequently missing. Unlike the simulation setting in Section 3, these missing values arise naturally in the data rather than being imposed experimentally.

In the absence of a benchmark index, the analysis focuses on how different methods for handling missing data affect the resulting price index, particularly its stability, sensitivity to sample composition, and consistency with the observed missingness structure.

The empirical analysis addresses three questions. First, is the observed missingness compatible with an MCAR structure? If so, complete-case estimation would primarily increase dispersion due to reduced sample size, while remaining approximately unbiased. Second, how does imputation with MICE affect the index trajectory? Third, does the imputed index yield a more credible representation of price dynamics, given the observed missingness patterns and the sensitivity of complete-case estimation to

[10] www.zt.co.at

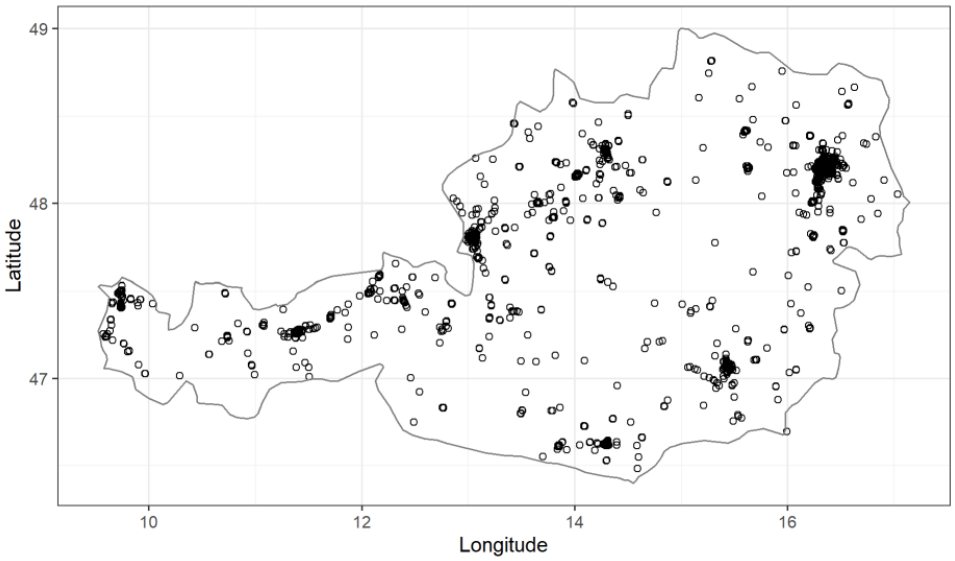
sample composition?

4.1 Data

The dataset consists of transaction-level data for Austrian office unit sales between 2015 and 2024, obtained from *ZTdatenforum*. The raw data contain 3,372 transactions. After restricting the sample to office units (i.e., excluding entire office building sales and mixed-use properties), the estimation sample comprises 3,003 transactions. This is roughly one twentieth of the sample size in the Vienna apartment application (Section 3), highlighting the much thinner transaction environment in the office market. The market is also more heterogeneous than in the Vienna apartment case. Transaction prices and sizes vary widely, regional price differences between large cities and peripheral areas are substantial, and a small number of high-value transactions account for a disproportionate share of total transaction value.

Transactions are observed across all nine Austrian provinces, but are concentrated in and around Vienna and other regional capitals (Figure 6).

Figure 6: Geo-locations of Austrian office unit transactions (2015–2024)



Source: ZTdatenforum eGen (https://zt.co.at/)

In addition to price, transaction date, and location, the dataset contains several property characteristics. These include *size*, measuring interior floor area in square meters; *legal_age*, defined as the number of years since parification[11]; an indicator for whether the property was sold by a developer

[11] Parification refers to the legal subdivision of a building into separate ownership units. While it does not perfectly

or construction firm (*builder*); indicators for the inclusion of parking space (*hadPkwApPrice*) and inventory (*hadInventoryPrice*) in the transaction price; and distance-based measures of local amenities (e.g. distance to the city center, nearest doctor, or shops).

All variables are listed in Appendix Table A.10. We distinguish between core variables used in the hedonic index regression and additional auxiliary variables included in the imputation model. Using a richer set of predictors at the imputation stage is standard practice in the imputation literature, as it helps to make the missing-at-random (MAR) assumption more plausible and improves the quality of imputations (see, e.g., van Buuren, 2018; Little and Rubin, 2019). Appendices C.1–C.3 provide further details on the dataset and the missingness structure.

4.2 Missingness and why it matters for the office index

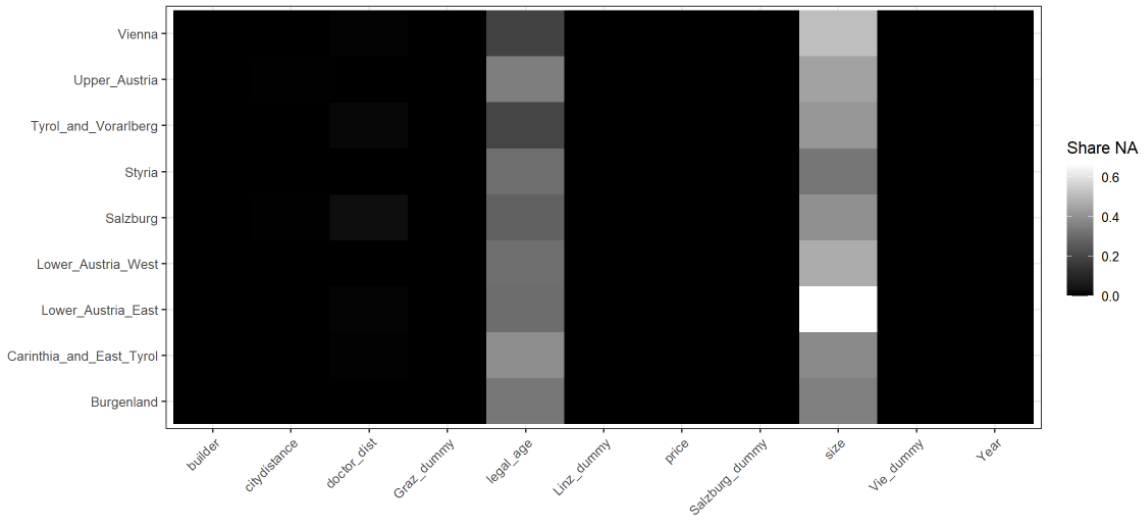
Most variables in the office dataset exhibit negligible missingness. However, two core hedonic controls are frequently missing: floor area (*size*) is missing in 45.2% of transactions and *legal_age* in 26.1%. Most other variables are almost fully observed. Because both size and age are central to hedonic quality adjustment, their absence substantially reduces the effective sample under complete-case estimation. Only 41.4% of transactions are complete in all baseline hedonic variables, implying that a complete-case index discards nearly 60% of the data. Given annual transaction counts of roughly 250–360, this substantially increases the price index's sensitivity to influential observations and thus its variability.

However, when missingness is not generated by an MCAR process, relying on a complete-case index is not only an issue of sample size but can also introduce bias. We find that missingness is clearly non-random in our dataset (Appendix C.4 reports Little's MCAR tests, which strongly reject MCAR). Furthermore, missingness varies across regions and over time. Specifically, Figure 7 shows that missingness in *size* and *legal_age* exhibits pronounced regional differences. High-price regions such as Vienna and its surrounding area (eastern Lower Austria) have higher rates of missing *size* than other regions. As a result, higher-priced market segments are disproportionately under-represented in the complete-case sample. Appendix Figure A.6 illustrates that there is also temporal variation in missingness in our sample. As a result, the composition of complete-case observations changes systematically over the sample period, which can distort measured price dynamics.

Taken together, missingness affects index construction through two channels. First, complete-case estimation reduces the effective sample size, increasing dispersion and sensitivity to influential observations. Second, and more importantly, structured missingness leads to selection bias. Multiple imputation directly addresses the first channel and mitigates the second under the MAR assumption.

correspond to construction year, it is strongly correlated with building age and renovation status.

Figure 7: Missingness heatmap by region for core hedonic variables



Note: The figure reports the share of missing observations by province and variable. Missingness is concentrated in *size* and *legal_age*, with clear regional variation. **Source:** ZTdatenforum eGen (https://zt.co.at/).

4.3 Empirical design

To assess how missing-data treatment affects index construction, we compare a set of estimators that differ only in how missing characteristics are handled, while keeping the hedonic specification fixed. In addition to the complete-case (CC) index, which discards all observations with missing characteristics, we consider four imputation approaches: (i) MICE with random forests (MICE-RF), our baseline; (ii) default MICE using predictive mean matching (PMM) with a rich set of predictors; (iii) regression-based multiple imputation with a reduced predictor set (MICE-linear-core); and (iv) mean imputation. These methods range from flexible multiple-imputation approaches using non-parametric algorithms to simple deterministic imputation rules.

We report both time-dummy (TD) and rolling time-dummy (RTD) indices as specified in Section 2.3. We focus on annual indices in the main text, as they most clearly reveal systematic level differences across missing-data treatments. A common hedonic specification is used throughout, so differences across estimators are driven entirely by how missing data are handled.

Because no benchmark index is available, we evaluate the estimators based on the stability of index paths, their sensitivity to alternative specifications, and their consistency with the observed missingness structure. The key question is whether different missing-data treatments lead to systematically different conclusions about price dynamics, and whether these differences can be linked to identifiable

selection patterns in the data.

4.4 Multiple imputation with MICE-RF

We implement multiple imputation using the MICE framework with a random forest (RF) algorithm, which flexibly captures nonlinearities and interactions, is robust to multicollinearity, and provides implicit variable selection. The imputation model includes all variables used in the hedonic regression as well as additional auxiliary predictors shown in Table A.10. This richer specification helps to make the missing-at-random (MAR) assumption more plausible and allows the imputation model to exploit a broader information set than the index regression. Appendix Figure A.8 illustrates that the RF imputations are driven by economically meaningful predictors, including price, location, and accessibility measures.

Our baseline specification uses $m = 100$ imputations and $maxit = 5$ iterations with `meth="rf"` (via the `ranger` backend), and we set $ntree$=20 per forest. Convergence diagnostics are reported in Appendix C.5. After imputation, we apply influence diagnostics at the estimation stage. Observations are retained during imputation to preserve the joint distribution of characteristics, and Cook's distance (Cook, 1977) is used to identify influential observations using the standard $4/n$ threshold. Influence measures are computed within each imputed dataset and averaged to determine a common exclusion set.

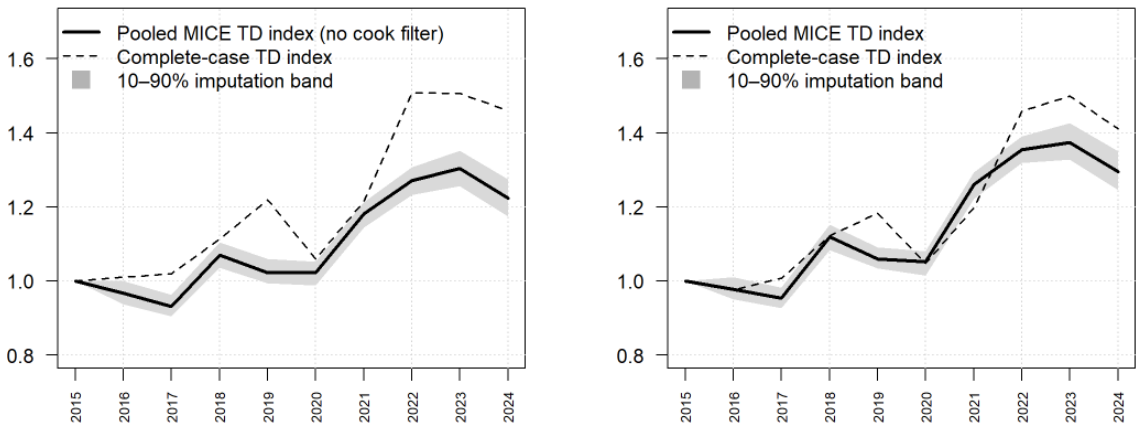
We estimate hedonic TD and RTD models (as specified in Section 2) separately in each completed dataset using the core variables. Pooled coefficient estimates (reported in Appendix C.6) exhibit economically plausible magnitudes and remain stable across imputations. As in the Vienna apartment application, we pool adjacent-period growth rates rather than index levels when combining results across imputations. Let $g_{t,i}$ denote the growth rate between periods $t-1$ and t in imputation i. The pooled index is constructed as

$$P_0^{\text{pool}} = 1, \qquad P_t^{\text{pool}} = \prod_{s=1}^{t} (1 + \bar{g}_s)\,, \tag{5}$$

where $\bar{g}_s$ is the geometric mean of price relatives across imputations.

Figure 8 compares two annual TD office price indices for Austria over 2015–2024: (i) a complete-case (CC) index estimated on transactions with fully observed hedonic characteristics; and (ii) a pooled imputation index constructed from $m = 100$ multiply imputed datasets using the growth-rate pooling approach described in Section 2. Both indices are estimated after Cook's-distance influence filtering, ensuring that differences are not driven by a small number of high-leverage observations.

Figure 8: Complete-case versus pooled multiple-imputation TD index (annual, with Cook filtering)



Note: Solid line: pooled TD index constructed from $m = 100$ multiply imputed datasets (baseline: MICE-RF). Dashed line: complete-case TD index. Shaded areas represent the 10–90% cross-imputation dispersion of chained index levels. **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations and simulations.

The indices exhibit a persistent level difference throughout most of the sample, with the complete-case index generally lying above the pooled imputation index, particularly in the post-2020 period. The complete-case index continues to increase strongly through 2022, whereas the pooled imputation index flattens earlier and partially reverses toward the end of the sample. This pattern is consistent with the structured missingness documented in Section 4.2. Because high-value transactions are disproportionately excluded under complete-case estimation—particularly at the start of the sample—the composition of the CC sample shifts over time toward higher-value segments, mechanically amplifying measured price growth. By contrast, multiple imputation restores the effective estimation sample and attenuates this composition effect.

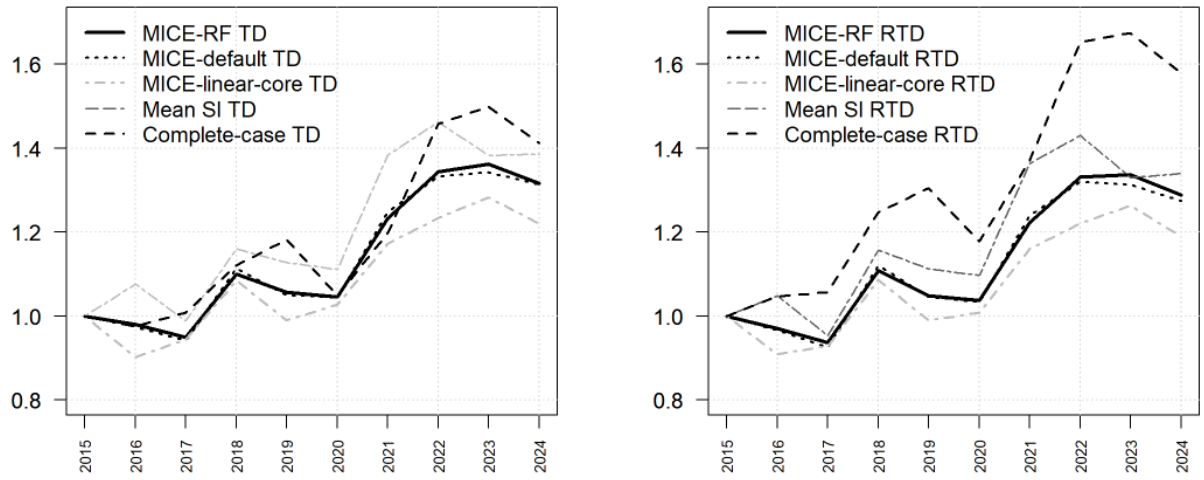
Dispersion across individual imputations is somewhat larger toward the end of the sample, reflecting lower transaction volumes and more pronounced selection effects, which reduce effective support for the imputation model. Appendix C.8 reports the corresponding dispersion of year-to-year index growth rates. These measures capture the uncertainty arising from imputation variability across completed datasets.

Overall, multiple imputation produces lower and more stable estimates of price dynamics than complete-case estimation. This pattern is consistent with the view that restoring incomplete observations reduces sample-composition distortions induced by structured missingness.

4.5 How Sensitive Are Price Indices to the Imputation Method?

We next assess the sensitivity of the results to the choice of imputation method. We compare four approaches: (i) MICE with random forests (MICE-RF), our baseline; (ii) default MICE using predictive mean matching for numeric and polytomous regression for categorical variables (MICE-default); (iii) regression-based multiple imputation with a reduced predictor set (MICE-linear-core); and (iv) mean imputation. All methods use a common set of variables where feasible, although the linear specification requires a reduced predictor set due to singularity issues arising from (near-)perfect multicollinearity (see Appendix C.7). Figure 9 compares the resulting indices after Cook's-distance filtering. The left panel shows time-dummy (TD) indices, and the right panel shows rolling time-dummy (RTD) indices.

Figure 9: Comparison of price indices under alternative imputation methods



Note: Impact of different imputation methods. The left panel shows time-dummy (TD) indices and the right panel rolling time-dummy (RTD) indices. The solid line represents the pooled MICE-RF index, while alternative line types correspond to default MICE, MICE-linear-core, mean imputation, and the complete-case benchmark. All indices are estimated after outlier removal using Cook's distance.

Across both specifications, the two rich MICE implementations (MICE-RF and default MICE) produce very similar index trajectories.[12] By contrast, simpler or more restricted approaches show greater differences. Mean imputation yields index paths that deviate more strongly, while the MICE-linear-core specification also differs noticeably from the richer MICE implementations. In the latter case, this reflects the significantly reduced predictor set required for numerical stability, which limits the imputation model's ability to recover the data's joint distribution.

[12]The maximum deviation between the two series is below 3 percent across all specifications and falls to below 2 percent after Cook-distance filtering.

Differences across methods are more pronounced under the rolling time-dummy (RTD) specification. Especially the complete-case index diverges strongly from the other indices (and also from its TD version). This reflects the smaller effective sample sizes in RTD estimation, where each regression is based on a shorter time window and is therefore more sensitive to missing data. Differences in indices due to imputation methods are even more pronounced if we do not filter for influential observations (see Appendix Figure A.9). In particular, the complete-case index exhibits greater volatility and stronger upward movements, especially under RTD. Thus, applying Cook's distance filtering reduces dispersion across series but does not fundamentally alter the ranking of the methods. In particular, the close alignment between MICE-RF and default MICE, and the weaker performance of mean imputation and the linear-core specification, remain unchanged.

Taken together, the evidence suggests three main conclusions. First, addressing missing data through imputation is essential, as complete-case indices are substantially more volatile and can diverge markedly from imputation-based estimates. Second, multiple imputation using a rich set of predictors—implemented via MICE-RF or default MICE—produces stable, closely aligned index trajectories. Third, simpler or more constrained approaches, such as mean imputation or regression-based imputation with a reduced predictor set, lead to less satisfactory index behavior.

4.6 Interpretation and plausibility

The results above raise the question of how to interpret the differences between complete-case and imputation-based indices in this setting.

Overall, the pooled MICE-RF and MICE-default indices provide a more stable representation of estimated office price dynamics than the complete-case alternative. Given the structured missingness documented in Section 4.2, this pattern is consistent with the view that imputation helps to reduce sample-composition distortions. In the absence of a benchmark index that reveals the true price trajectory, we cannot formally verify which index is closer to the truth. However, several pieces of evidence suggest that the divergence between complete-case and imputed indices is not arbitrary, but closely related to the observed missingness structure.

First, in our case, missingness in *size* and *legal_age* varies systematically across regions and over time (Section 4.2). Under such conditions, complete-case estimation can induce both variance inflation (through sample thinning) and bias (through segment-specific over-representation and time-varying selection). The divergence observed between MICE-imputed and complete-case indices at the beginning of the sample and after 2021 is consistent with the time-varying sample selection rather than purely stronger underlying price growth. This selection mechanism directly links the observed divergence in index trajectories to missing-data patterns.

Second, beyond selection, missing data amplify the influence of high-leverage observations in this thin-market setting. When the effective sample is reduced, individual transactions—particularly high-value or atypical ones—can exert a disproportionate impact on estimated price dynamics. Multiple imputation mitigates this effect both by restoring sample size and by stabilizing the estimation base, which reduces sensitivity to such observations. Consequently, we found that MICE-imputed indices were significantly less sensitive both to the removal of influential observations and to whether indices were compiled using the TD or RTD method.

Third, in our sample, missingness was particularly pronounced in the early part of the sample and for transactions in Vienna and Lower Austria East, which are among the highest-priced regions. The under-representation of the higher-end market in the initial periods can mechanically lower the starting level of the complete-case index and thereby generate spurious subsequent price growth. This mechanism can explain the steeper price trajectory of the CC index, particularly in the RTD method that uses shorter data windows.

Fourth, the extremely strong price growth in the CC office price index between 2020 and 2022 appears difficult to reconcile with broader structural shifts in office markets following the COVID-19 pandemic. Although this comparison cannot validate the imputed series, it does cast doubt on interpreting the CC trajectory at face value. Internationally, the post-2021 period was marked by the spread of remote work and a reassessment of office demand, which reduced demand for office properties in many markets (Gupta et al., 2022; Deghi et al., 2022; Duca and Ling, 2024).

Finally, the performance of imputation methods depends on the richness of the information set used in the imputation step. MICE specifications based on a rich set of predictors and flexible algorithms produce nearly identical and comparatively stable index trajectories. By contrast, simpler approaches such as mean imputation or multiple imputation with a restricted predictor set yield less stable and more variable results. This reinforces the interpretation above: the gains from multiple imputation arise not only from increasing sample size, but also from more accurately reconstructing the joint distribution of observed characteristics.

Taken together, the evidence provides fairly clear answers to the three questions posed at the beginning of this section. First, missingness in our data is not consistent with MCAR and induces systematic selection. Second, imputation materially alters the index trajectory by restoring incomplete observations and thereby reducing sample-composition distortions. Third, the indices compiled after multiple imputation with MICE appear more credible, as they align more closely with known market developments and are less sensitive to the selection effects induced by complete-case estimation. Importantly, this interpretation does not rely on a single imputation algorithm: the two rich MICE specifications (MICE-RF and MICE-default) produce closely aligned series, whereas the complete-case index diverges most strongly precisely in settings such as RTD estimation, where sample thinning should matter most.

5 Discussion of Results

5.1 Comparison across missingness environments

The two empirical settings considered in this paper demonstrate that the effect of missing data depends on market structure and the nature of the missingness process. In the Vienna apartment market, transaction volumes are large and properties are relatively homogeneous. Provided that missingness does not arise from severe truncation, complete-case indices remain fairly stable even at high levels of missing data. Differences across imputation methods are correspondingly small. While both MICE-default and MICE-RF perform best overall, the differences between imputation methods—and between imputation and complete-case estimation—are minor. In this setting, the primary benefit of imputation is therefore a reduction in volatility through restoration of the effective estimation sample. In contrast, the Austrian office market is more heterogeneous, thinner, and characterized by strongly structured missingness. Imputation can therefore play a more central role by restoring underrepresented market segments. Multiple imputation with rich and flexible models (MICE-default and MICE-RF) are very similar to each other, even though they use quite different internal algorithms. These indices also remain stable across hedonic models. Complete-case estimation differs markedly from these MICE-imputed indices, while simpler imputation approaches yield results that generally lie between the complete-case and MICE-based estimates.

Overall, the comparison of the two applications highlights three mechanisms through which missing data affect hedonic price indices. First, discarding incomplete observations can alter the composition of the estimation sample, so that measured price dynamics partly reflect changing sample composition rather than underlying market movements. Second, missing data amplify the influence of individual transactions when the effective sample becomes small, particularly in thin markets and under RTD estimation. Third, the effectiveness of imputation depends on the richness of the information set used to approximate the joint distribution of the data. These mechanisms are relatively weak in the homogeneous and transaction-rich Vienna apartment market, but become economically important in the thinner and more heterogeneous Austrian office market.

We next translate these findings into practical guidance for statistical agencies compiling transaction-based price indices.

5.2 Practical guidance for handling missing data

1. Avoid relying mechanically on complete-case estimation when missingness is substantial or systematically structured. In large and relatively homogeneous markets, complete-case indices may remain reasonably stable. However, in thinner or more heterogeneous settings, they can

become highly sensitive to sample-composition effects and influential observations.

2. The primary benefit of imputation is sample restoration. The main gain from imputation is not the precise recovery of missing values, but the ability to retain incomplete observations in the estimation sample. This stabilizes the quality adjustment and reduces excess variability in index movements.

3. Use multiple imputation to quantify uncertainty. Multiple imputation provides a principled way to account for uncertainty associated with missing data. Dispersion across imputed index paths can be used to construct uncertainty bands, which serve as a diagnostic for how sensitive the index is to the imputation step.

4. Use imputation models that exploit a rich set of predictors. In relatively homogeneous settings, different flexible imputation approaches—such as predictive mean matching or random forests within MICE—often yield very similar index trajectories. In contrast, in thinner and more heterogeneous markets, methods that rely on restricted predictor sets or simple rules (such as mean imputation) can produce more volatile and less consistent indices. This reflects the importance of adequately capturing the joint dependence structure of the data when imputing missing values.

Where feasible, the imputation model should also include transaction price as a predictor. Since price is strongly related to key characteristics such as size, location, and building attributes, including it can substantially improve the quality of imputations. In practice, however, this may be constrained by numerical instability or multicollinearity in parametric specifications, which is one reason why more flexible non-parametric approaches such as random forests can be advantageous.

5. Imputation cannot recover missing market segments. When entire segments of the market are systematically absent from the data, no imputation method can reconstruct them reliably from observed data. In such cases, bias may persist even after imputation.

6. Aggregate imputation results in a way consistent with index-number theory. Because price indices are multiplicative objects constructed from chained growth rates, standard pooling rules such as Rubin's rules do not apply. Imputation results should instead be combined at the level of adjacent-period price relatives before chaining.

6 Conclusion

Missing property characteristics are common in real estate transaction datasets and complicate the construction of hedonic price indices. This paper examined how missing data affect transaction-based

property price indices and whether imputing missing characteristics using multiple imputation, implemented via MICE, improves index reliability. The focus is not on imputing values for their own sake, but on stabilizing the quality adjustment in hedonic regressions when key characteristics are missing. Across both empirical settings, a consistent result emerged: discarding incomplete observations through complete-case estimation reduces the effective estimation sample and can lead to unstable and potentially misleading price indices. This is particularly the case in thin markets, when missingness affects entire market segments, or when missingness patterns vary over time. Restoring incomplete observations through multiple imputation stabilizes the estimation base and can yield more robust index trajectories, provided that the missing region remains sufficiently represented in the observed data.

More generally, the impact of missing data depends on the data environment. In relatively homogeneous and transaction-rich markets, missingness primarily increases dispersion, and multiple imputation mainly improves stability by restoring the effective estimation sample. In thinner and more heterogeneous markets, where missingness is structured, imputation becomes much more important, as it mitigates composition effects and helps reduce distortions induced by selective sample loss. At the same time, the effectiveness of multiple imputation depends on the information available in the data. Flexible imputation approaches that exploit a rich set of predictors tend to produce stable and closely aligned results, whereas simpler methods based on restricted information can yield more volatile and less consistent indices. However, when entire segments of the price–quality distribution are unobserved, imputation cannot fully eliminate bias, as the missing information cannot be recovered from the observed data.

The paper makes four main contributions. First, it provides systematic evidence on how different missingness mechanisms affect hedonic real estate price indices in a setting where a benchmark index is observed. Second, it clarifies the bias–variance trade-off associated with discarding incomplete observations. Third, it proposes an index-consistent method for integrating multiple imputation by pooling adjacent-period price relatives. Fourth, it shows that hedonic indices are relatively robust to moderate imputation errors, but that the choice of imputation model becomes more important in thin and heterogeneous markets.

These findings have direct implications for statistical practice. Relying on complete-case estimation can lead to unstable and potentially biased indices when missingness is systematic. Multiple imputation provides a practical and implementable alternative that can be integrated into existing workflows and supplemented with transparent diagnostics.

Overall, missing data should be treated as an integral component of index construction rather than as an overlooked issue. Explicitly incorporating multiple imputation into the index compilation process can materially improve the stability and interpretability of transaction-based real estate price indices.

References

Armknecht, P. A. and Maitland-Smith, F. (1999). Price imputation and other techniques for dealing with missing observations, seasonality and quality change in price indices. IMF Working Papers, 1999(078):A001.

Breiman, L. (2001). Random forests. Machine learning, 45(1):5–32.

Cook, R. D. (1977). Detection of influential observation in linear regression. Technometrics, 19(1):15–18.

Deghi, A., Natalucci, F., and Qureshi, M. S. (2022). Commercial real estate prices during covid-19: What is driving the divergence? Global Financial Stability Notes 2022/02, International Monetary Fund.

Diewert, W. E. (2002). Similarity and Dissimilarity Indexes: An Axiomatic Approach. Discussion Paper 02-10, Department of Economics, University of British Columbia, Vancouver, Canada.

Duca, J. V. and Ling, D. C. (2024). Work-from-home, covid-19, and online retail effects on commercial real estate prices. Working Paper.

Enders, C. (2022). Applied Missing Data Analysis. Methodology in the Social Sciences Series. Guilford Publications.

Erler, N. S., Rizopoulos, D., Rosmalen, J. v., Jaddoe, V. W. V., Franco, O. H., and Lesaffre, E. M. E. H. (2016). Dealing with missing covariates in epidemiologic studies: a comparison between multiple imputation and a full bayesian approach. Statistics in Medicine, 35(17):2955–2974.

Eurostat (2013). Handbook on residential property price indices. https://ec.europa.eu/eurostat/documents/3859598/5925925/KS-RA-12-022-EN.PDF.

Eurostat (2017). Eurostat, commercial property price indicators - sources, methods and issues.

Gupta, A., Mittal, V., and Van Nieuwerburgh, S. (2022). Work from home and the office real estate apocalypse. Working Paper 30526, National Bureau of Economic Research.

Hill, R. J. (2013). Hedonic price indexes for residential housing: A survey, evaluation and taxonomy. Journal of Economic Surveys, 27(5):879–914.

Hill, R. J., Scholz, M., Shimizu, C., and Steurer, M. (2018). An evaluation of the methods used by european countries to compute their official house price indexes. Economie et Statistique, 500-502:221–238.

Hill, R. J., Scholz, M., Shimizu, C., and Steurer, M. (2022). Rolling-time-dummy house price indexes: Window length, linking and options for dealing with low transaction volume. Journal of Official Statistics, forthcoming.

Little, R. J. (1988a). A test of missing completely at random for multivariate data with missing values. Journal of the American statistical Association, 83(404):1198–1202.

Little, R. J. and Rubin, D. B. (2019). Statistical analysis with missing data, volume 793. John Wiley & Sons.

Little, R. J. A. (1988b). Missing-data adjustments in large surveys. Journal of Business Economic Statistics, 6(3):287–296.

OECD, Eurostat, Organization, I. L., Fund, I. M., Bank, T. W., and for Europe, U. N. E. C. (2013). Handbook on Residential Property Price Indices.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. Journal of Political Economy, 82(1):34–55.

Rubin, D. B. (1976). Inference and missing data. Biometrika, 63(3):581–592.

Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. Wiley.

Schafer, J. L. (1999). Multiple imputation: a primer. Statistical Methods in Medical Research, 8(1):3–15. PMID: 10347857.

Shimizu, C., Takatsuji, H., Ono, H., and Nishimura, K. G. (2010). Structural and temporal changes in the housing market and hedonic housing price indices: a case of the previously owned condominium market in the tokyo metropolitan area. International Journal of Housing Markets and Analysis.

van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. Statistical Methods in Medical Research, 16(3):219–242. PMID: 17621469.

van Buuren, S. (2018). Flexible Imputation of Missing Data. Chapman and Hall/CRC, Boca Raton, FL, 2nd edition.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. Journal of Statistical Software, 45(3):1–67.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. Journal of statistical software, 45:1–67.

Appendix

A Appendix related to Section 2

A.1 Implementation details for MICE

Implementing MICE requires selecting several tuning parameters that affect both computational cost and the properties of the imputed datasets. We briefly discuss the most relevant choices.

Number of iterations (`maxit`). The parameter `maxit` determines how often the chained equations are iterated when generating each completed dataset. Setting `maxit` > 1 allows the imputation system to update conditional models repeatedly, so that imputations for different variables become mutually consistent. This is particularly important when multiple variables exhibit missingness and are jointly related through the underlying data-generating process. In contrast to joint-model multiple imputation approaches, where imputations are drawn directly from a specified multivariate distribution, MICE relies on iterated updates of conditional models. Multiple iterations are therefore an integral component of the MICE procedure, as they allow the conditional models to become mutually consistent across variables. In our applications, key characteristics such as size and building age are missing for a substantial share of observations. As they are correlated with other variables, using multiple iterations allows the imputation procedure to better approximate the joint distribution of the data. We therefore set `maxit` $= 5$ as a practical compromise between convergence of the chained equations and computational cost.

Number of imputations (m). The parameter m determines how many completed datasets are generated. A larger number of imputations reduces Monte Carlo variation arising from the imputation procedure and leads to more stable pooled estimates. This is particularly relevant in our setting, where index construction involves pooling adjacent-period price relatives across imputations and chaining them over time. In addition, we use the dispersion across imputed index paths to construct uncertainty bands. Lower values of m typically have only a modest effect on the mean index path, but can lead to noisier estimates of imputation uncertainty and less stable index trajectories in later periods, where small differences in growth rates accumulate through chaining. We therefore use $m = 100$ imputations to ensure stable estimation of both the index and the associated uncertainty.

Number of trees (`ntree`). The parameter `ntree` is relevant only for random forest–based imputation. It controls the number of trees used to estimate each conditional model. A larger number of trees generally improves the stability of the imputation model at the cost of increased computation time. In our implementation, we set `ntree` $= 20$, which provides a balance between computational feasibility and sufficient stability of the imputed values in this high-dimensional and relatively sparse

setting.

A.2 Random forest conditional models within MICE

Random forests are used inside MICE to predict missing values conditional on the observed variables. They are well suited to real estate data because they automatically capture nonlinearities, interactions, and mixed data types without requiring manual model specification (Breiman, 2001).

For completeness, we summarize the random forest algorithm used for conditional prediction. This algorithm is applied separately for each variable with missing values.

Input. A training dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, where $\mathbf{x}_i \in \mathbb{R}^p$ denotes the predictor variables and y_i the target variable. The number of trees is denoted by T, and the number of candidate predictors at each split by m.

Algorithm.

1. For $t = 1, \ldots, T$:
   (a) Draw a bootstrap sample $\mathcal{D}_t$ of size n from $\mathcal{D}$.
   (b) Grow a decision tree $\mathcal{T}_t$ on $\mathcal{D}_t$:
      - At each node, randomly select m predictors from the full set of p predictors.
      - Choose the split that minimizes squared error for continuous variables or classification error for categorical variables.
      - Continue splitting until a stopping rule is met (e.g. minimum node size).
   (c) Store the fitted tree $\mathcal{T}_t$.
2. Aggregate predictions across trees:
   - For continuous variables, predictions are averaged across trees.
   - For categorical variables, predictions are determined by majority vote.

Randomness and multiple imputation. Although a single random forest model is deterministic conditional on a given bootstrap sample and feature selection sequence, stochasticity arises from three sources. First, each tree is grown on a bootstrap resample of the data. Second, a random subset of predictors is considered at each split. Third, the MICE algorithm iterates through variables and cycles repeatedly. Together, these features generate multiple plausible imputations across iterations and across imputed datasets.

Role of random forests within MICE. Within MICE, random forests are used to approximate the conditional predictive distributions of variables with missing values. They do not rely on parametric assumptions and can adapt flexibly to complex dependence structures. This makes Random Forest–based MICE particularly attractive for real estate transaction data, where characteristics are highly correlated and relationships with price are nonlinear.

B Appendix related to Section 3 (Vienna apartment market)

This appendix reports additional results for Application 1, covering the full set of missingness mechanisms and severity levels. Unless otherwise noted, all comparisons use the full-sample hedonic index as the benchmark.

B.1 Summary statistics of Vienna apartment market data

This subsection describes the transaction-level dataset used in Application 1. The data comprise apartment transactions in Vienna over the period 2015–2023 and serve as the benchmark setting for the missing-data experiments in Section 3.

Table A.1 lists the variables used in the imputation and hedonic regression processes. The benchmark hedonic specification includes core structural characteristics, location controls, and local amenity measures, together with time fixed effects at the quarterly level. In addition to these variables, the distance to the nearest pharmacy (*pharm_dist*) is included as an auxiliary predictor in the imputation models.

Table A.1: Variables used in the Vienna apartment application

Variable	Description
Core property characteristics	
size	Apartment size (floor area, in square metres)
builder	Indicator for developer / builder transaction
Location controls	
postcode	Postal-code fixed effect capturing 23 intra-city locations
Local amenities	
doc_dist	Distance to nearest doctor in meters
pharm_dist	Distance to nearest pharmacy in meters
Price and transaction variables	
price	Transaction price in euros
car_price	Parking price variable / parking-related transaction component
Time controls	
Quarteryear	Quarter-year of transaction

Notes: Variables shown in bold are included in the hedonic regressions for the Vienna apartment market. All listed variables are used as predictors in the missing-data analysis.

Table A.2 reports summary statistics for the complete-case benchmark sample. The dataset contains 67,292 transactions over 35 quarters, corresponding to an average of roughly 1,900 transactions per quarter. Prices exhibit some dispersion, reflecting variation in property characteristics and location within the city. The median transaction price (€282,000) lies well below the mean (€355,607), indicating a right-skewed distribution typical for housing markets.

Table A.2: Summary statistics: Apartment application

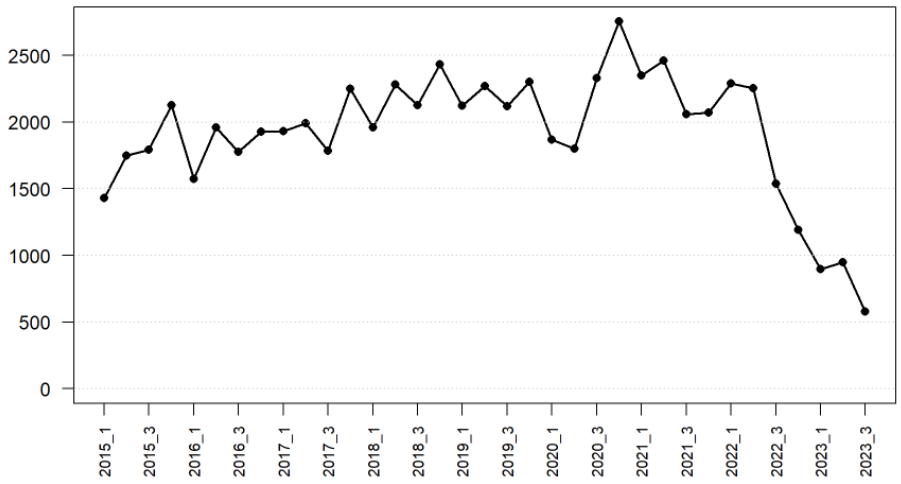
Variable	N / Share	Mean	SD	Median
Sample overview				
Number of transactions	67,292			
Number of quarters	35			
Categorical variables				
Builder (yes)	65.59%			
Builder (no)	34.41%			
Number of postcodes	23			
Continuous variables				
Price (€)	67,292	355,607	348,931	282,000
Price per sqm (€)	67,292	4,742	2,200	4,426
Size (sqm)	67,292	72.2	33.6	65.0
Doctor distance (m)	67,292	303.9	384.5	183.7
Pharmacy distance (m)	67,292	312.8	251.9	245.4
Car price (€)	67,292	6,253	13,832	0

Notes: For categorical variables, shares refer to the percentage of total transactions.

The Vienna apartment market represents a relatively homogeneous setting with standardized property types and a dense urban structure.

Figure A.1 shows the number of transactions per quarter. Transaction volumes are relatively stable over most of the sample, but decline markedly in later periods, particularly from 2022 onwards.

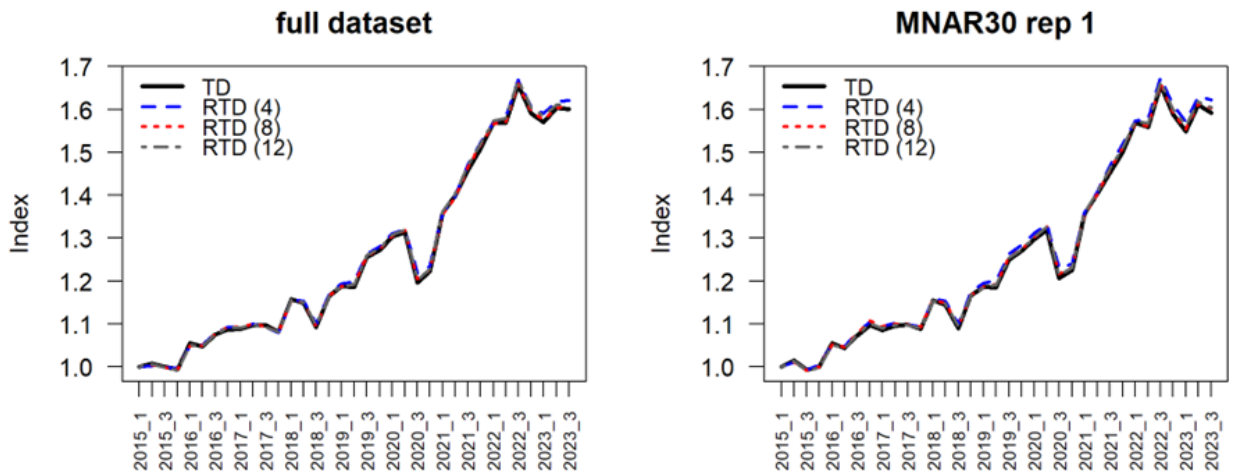
Figure A.1: Number of apartment transactions per quarter



B.2 Robustness of hedonic index specification in the Vienna apartment market

An important question is whether the benchmark index is sensitive to the choice of hedonic specification, in particular to potential time variation in hedonic coefficients. To assess this, Figure A.2 compares the time-dummy (TD) index to rolling time-dummy (RTD) indices constructed using different window lengths. The resulting index paths are nearly identical across window settings. This indicates that coefficient drift is limited in this dataset and does not materially affect the estimated price trajectory, supporting the use of the TD specification as a reliable benchmark. We therefore focus on the TD hedonic index throughout the Vienna apartment application.

Figure A.2: Robustness of hedonic index specifications



Note: The figure compares the benchmark time-dummy (TD) hedonic index with rolling time-dummy (RTD) indices using different window lengths. The left panel is based on the full dataset, while the right panel illustrates a sample with 70% missingness generated under an MNAR mechanism, using the same hedonic specification. The trajectories of indices compiled with different window lengths are nearly identical, indicating that coefficient drift has only a minor effect on index construction in this dataset. **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations and simulations.

B.3 Missingness designs

B.3.1 Random missingness (MCAR)

To generate missing completely at random (MCAR) patterns, we randomly delete a fixed share of entries in the explanatory variables. Transaction price and transaction date are never deleted. In practice, these variables are almost always recorded in transaction price data because they are required for tax and registration purposes.

Missingness rates range from 10% to 90% of cells in the explanatory variables. Deletions are independent across variables and observations. Missingness in one variable therefore provides no information about missingness in another. Little's MCAR tests do not reject MCAR in these samples, which is consistent with the design.

B.3.2 Correlated non-random missingness (MNAR-left)

To generate non-random missingness across characteristics, we use `mice::ampute` and apply an MNAR-left mechanism. Observations with lower values of an affected variable are more likely to be missing.

Missingness probabilities differ across variables and are correlated across characteristics. Transaction price and transaction date are again excluded from deletion. Missingness rates range from 10% to 90%.

The algorithm is designed to induce MNAR at the variable level. In practice, however, the resulting missingness behaves closer to MAR once we condition on observed characteristics. Real estate attributes are strongly correlated, and observed characteristics proxy for latent quality. Conditional on the variables used in the hedonic model, selection on unobservables is therefore reduced.

B.3.3 Truncation-based missingness in size

In the third design, missingness is concentrated in property size, while all other variables remain observed. We focus on size because this pattern is common in commercial transaction data and closely resembles the Austrian office application in Section 4.

Transactions are ranked by purchase price within each year. Size is then removed systematically from either the upper or lower part of the price distribution. This mechanism represents a severe form of non-random missingness. It removes information from one side of the price–quality distribution and is therefore challenging for both complete-case estimation and imputation.

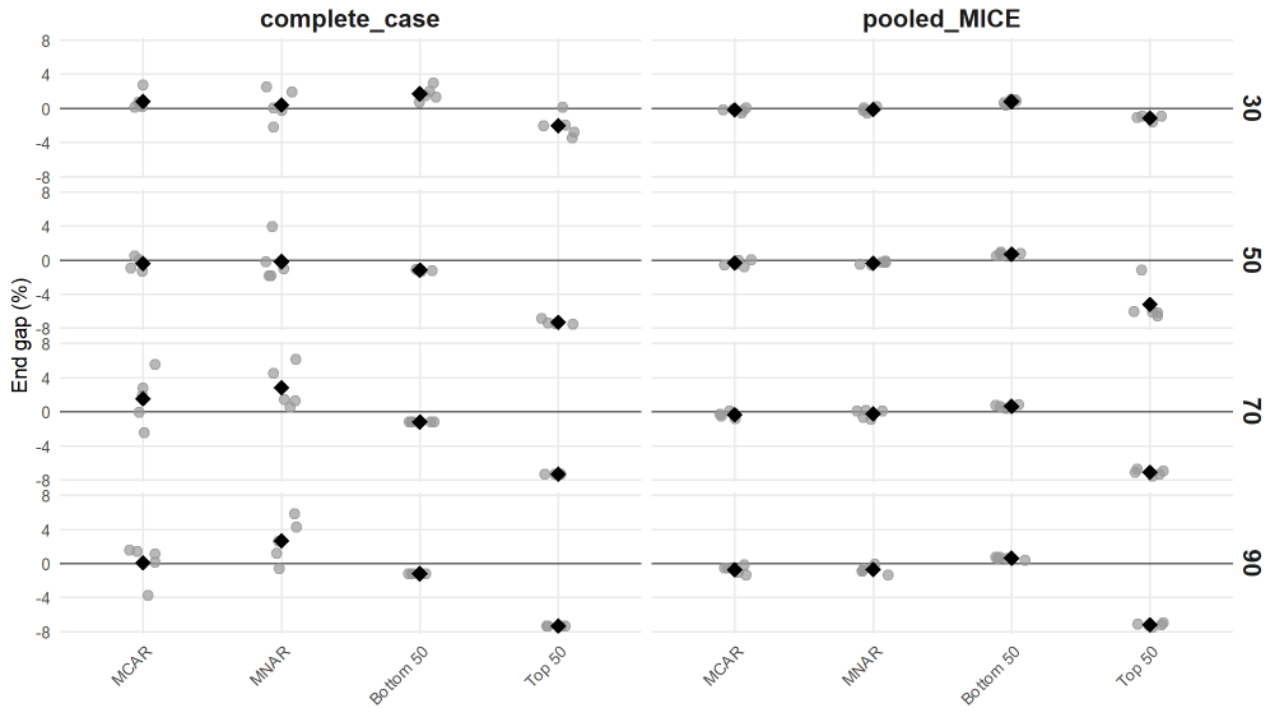
For each mechanism and each missingness level, we generate five independent realizations of the incomplete dataset. Each realization corresponds to a different random draw under the same design. This allows us to assess the stability of index estimates across samples.

At high missingness levels, especially under truncation designs, the pool to draw observations from can become very small. Complete-case samples can then become nearly identical across realizations. This mechanically reduces across-draw variability. We take this feature into account when interpreting variance and RMSE results.

B.4 End-of-sample inflation gaps under all mechanisms

This section extends Figure 4 in the main text. It reports end-of-sample cumulative inflation gaps for all missingness mechanisms and severity levels considered in the simulation design.

Figure A.3: End-of-sample index level gap relative to the benchmark across missingness level



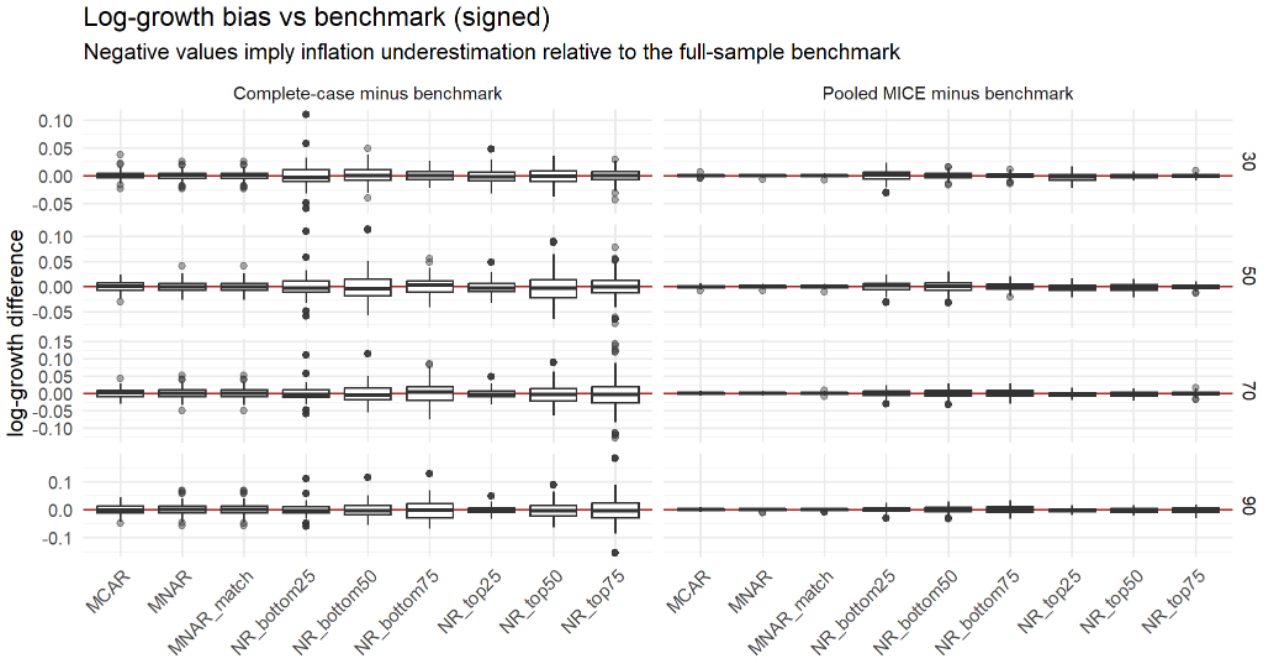
Note: Each dot represents one simulation replicate; diamonds indicate the mean across replicates. The gap is defined as $100 \times (\text{Index}_{\text{end}}/\text{Benchmark}_{\text{end}} - 1)$. Complete-case estimation leads to increasingly large and systematic deviations from the benchmark, along with substantial variation across replicates reflecting sensitivity to the realized sample. In contrast, multiple imputation produces index estimates that are both unbiased and tightly clustered across replicates, even at high levels of missingness.

The extended results confirm the patterns discussed in the main text. Deviations from the benchmark remain small under random or non-random missingness as long as the information content of the missing data can be inferred from the existing sample. Strong asymmetric missingness induces larger deviations. This is particularly visible in extreme truncation designs, where information about one segment of the price–quality distribution is largely absent. These cases illustrate the limits of any estimation approach when the missing segment cannot be reliably inferred from the observed data.

B.5 Distribution of quarterly growth bias across mechanisms

Figure A.4 plots signed quarterly log-growth differences relative to the benchmark. Negative values indicate inflation underestimation. The left panel shows complete-case estimates. The right panel shows pooled multiple-imputation estimates. The figure is a distributional complement to the end-of-sample gaps. Small quarterly biases can accumulate into larger level gaps over time.

Figure A.4: Signed quarterly log-growth bias relative to the full-case benchmark



Note: The figure plots signed quarterly log-growth differences relative to the full-case benchmark. Negative values indicate inflation underestimation. The left panel shows complete-case estimates; the right panel shows pooled multiple-imputation estimates. **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations and simulations.

B.6 RMSE decomposition

Table A.3: RMSE decomposition for quarterly log-growth errors relative to the benchmark (selected mechanisms)

		Complete-case (CC)				Pooled MI			
Mechanism	Level	Bias (qbps)	Bias^2	Var	RMSE	Bias (qbps)	Bias^2	Var	RMSE
MCAR	30	2.45	0.00000006	0.0000545	0.00739	-0.54	0.00000000	0.0000023	0.00150
	50	-1.21	0.00000001	0.0001110	0.01055	-1.09	0.00000001	0.0000053	0.00231
	70	4.41	0.00000020	0.0001540	0.01243	-1.01	0.00000001	0.0000080	0.00283
	90	0.22	0.00000000	0.0003250	0.01802	-2.15	0.00000005	0.0000113	0.00337
MNAR	30	1.21	0.00000001	0.0000614	0.00784	-0.45	0.00000000	0.0000027	0.00164
	50	-0.60	0.00000000	0.0001130	0.01064	-1.03	0.00000001	0.0000053	0.00231
	70	8.06	0.00000065	0.0002240	0.01499	-0.71	0.00000001	0.0000070	0.00265
	90	7.60	0.00000058	0.0004520	0.02128	-2.19	0.00000005	0.0000110	0.00332
NR_bottom50	30	4.99	0.00000025	0.0002120	0.01457	2.39	0.00000006	0.0000378	0.00615
	50	-3.51	0.00000012	0.0009290	0.03048	2.06	0.00000004	0.0001650	0.01287
	70	-3.58	0.00000013	0.0009490	0.03080	1.86	0.00000003	0.0001640	0.01282
	90	-3.58	0.00000013	0.0009490	0.03080	1.81	0.00000003	0.0001650	0.01286
NR_top50	30	-6.06	0.00000037	0.0002050	0.01432	-3.30	0.00000011	0.0000124	0.00353
	50	-22.33	0.00000499	0.0008740	0.02965	-15.83	0.00000251	0.0000604	0.00793
	70	-22.39	0.00000501	0.0008770	0.02970	-21.75	0.00000473	0.0000760	0.00899
	90	-22.39	0.00000501	0.0008770	0.02970	-22.01	0.00000485	0.0000772	0.00906

Note: The table reports the mean signed bias, its squared component (bias^2), the variance component, and RMSE, where RMSE^2 = bias^2 + variance. Bias is reported in quarterly basis points (qbps); bias^2, variance, and RMSE are in log-growth units. **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations and simulations.

B.7 RMSE comparison across imputation engines

Here we compare the performance of the default MICE imputation procedure with random forest–based MICE (MICE-RF) for the Vienna apartment simulation at 70% missingness.

Table A.4: Differences between default MICE and random forest MICE

Mechanism	Typical difference (%)	Upper bound (%)
MCAR	< 0.2	< 0.3
MNAR	< 0.2	< 0.3
NR_bottom50	0.2–0.4	< 0.5
NR_top50	0.3–0.5	< 0.5

Note: Differences are expressed in percentage deviations of the resulting price indices. Values are inferred from RMSE comparisons of log price levels at 70% missingness.

Even under high missingness (70%), differences between default MICE and random forest MICE remain small across all considered mechanisms. Based on RMSE comparisons of log price levels, implied differences in index levels are generally below 0.5%, and typically closer to 0.2–0.3%. No clear

dominance between imputation algorithms is visible and these differences are economically negligible. Overall, the results indicate that default MICE and MICE-RF perform almost equally well in this setting. However, regression-based imputations within the MICE framework perform substantially worse in this application. In the presence of strong multicollinearity and sparse factor levels, linear models require a reduced predictor set to avoid near-singular estimation problems, which leads to less stable and less accurate index estimates. We discuss this issue further in Appendix C.7.

B.8 Full RMSE decomposition for all mechanisms

Table A.5 decomposes RMSE into squared bias and variance components for all mechanisms. The MICE results correspond to the default MICE implementation used in the Vienna simulation. All components are computed relative to the full-case benchmark.

Table A.5: RMSE decomposition for quarterly log-growth errors relative to the benchmark (all mechanisms, default MICE)

		Bias (qbps)		Bias^2		Variance		RMSE	
Mechanism	Level	CC	MICE	CC	MICE	CC	MICE	CC	MICE
MCAR	30	2.45	-0.54	0.0000000602	0.0000000030	0.0000557466	0.0000023070	0.00729173	0.00150635
	50	-1.21	-1.09	0.0000000145	0.0000000118	0.0001140018	0.0000054259	0.01063123	0.00231661
	70	4.41	-1.01	0.0000001946	0.0000000103	0.0001574571	0.0000082023	0.01235120	0.00285924
	90	0.22	-2.15	0.0000000005	0.0000000464	0.0003322337	0.0000115472	0.01809936	0.00337609
MNAR	30	1.21	-0.45	0.0000000146	0.0000000020	0.0000626775	0.0000027598	0.00768869	0.00164412
	50	-0.60	-1.03	0.0000000036	0.0000000106	0.0001155727	0.0000054388	0.01070495	0.00232086
	70	8.06	-0.71	0.0000006500	0.0000000050	0.0002291634	0.0000071796	0.01478509	0.00267195
	90	7.60	-2.19	0.0000005782	0.0000000478	0.0004628363	0.0000112571	0.02119211	0.00333113
NR_bottom50	30	4.99	2.39	0.0000002494	0.0000000571	0.0002171630	0.0000386883	0.01462367	0.00621896
	50	-3.51	2.06	0.0000001232	0.0000000425	0.0009516973	0.0001695013	0.03085127	0.01302051
	70	-3.58	1.86	0.0000001280	0.0000000345	0.0009715900	0.0001683391	0.03117239	0.01297532
	90	-3.58	1.81	0.0000001280	0.0000000327	0.0009715900	0.0001692500	0.03117239	0.01301016
NR_top50	30	-6.06	-3.30	0.0000003670	0.0000001088	0.0002096043	0.0000126463	0.01439980	0.00354601
	50	-22.33	-15.83	0.0000049867	0.0000025074	0.0008952238	0.0000614382	0.03000325	0.00800161
	70	-22.39	-21.75	0.0000050115	0.0000047291	0.0008986017	0.0000778817	0.03006016	0.00908628
	90	-22.39	-22.01	0.0000050115	0.0000048452	0.0008986017	0.0000791015	0.03006016	0.00915735

Note: Bias is reported in quarterly basis points (qbps). Bias^2, variance, and RMSE are in log-growth units. The identity $\text{RMSE}^2 = \text{bias}^2 + \text{variance}$ holds by construction. **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations and simulations.

Interpretation. Across most mechanisms, RMSE differences are driven mainly by the variance component rather than by systematic bias. Complete-case variance rises rapidly with missingness because the usable sample varies substantially across realizations. Multiple imputation substantially reduces this dispersion by restoring the effective sample size. Under truncation designs, bias becomes more important because one side of the price–quality distribution is systematically missing. In these cases, imputation reduces volatility but cannot fully recover the missing market segment.

C Appendix C. Additional material for the office application

This appendix provides additional evidence for the Austrian office-market application. It documents (i) sample characteristics and missingness patterns, (ii) composition effects under complete-case estimation, (iii) variable definitions and imputation inputs, (iv) diagnostics for the imputation procedure, (v) pooled regression summaries, and (vi) robustness checks for imputation methods and temporal aggregation.

C.1 Sample definition and missingness summary

The raw dataset contains 3,003 office unit transactions between 2015 and 2024. We restrict attention to office units within multi-unit buildings and exclude complete building sales. Only 1,244 observations (41.4%) are complete for the full hedonic specification. When restricting attention to observations with non-missing price and size, 1,646 observations (54.8%) remain.

Missingness is concentrated in two core characteristics: *size* (45.2%) and *legal_age* (26.1%). All remaining variables exhibit negligible missingness.

Table A.6: Missingness in Austrian office unit transaction data

Variable	Missing observations	Missing share (%)
size	1,357	45.2
legal_age	783	26.1
doctor_dist	24	0.8
citydistance	2	0.1
city_in20km	2	0.1
nearest_city	2	0.1
All remaining variables	0	0.0

Notes: The table reports the number and percentage of missing observations by variable. Missingness is concentrated in the core quality characteristics, particularly unit size and legal age. All remaining variables used in the hedonic specification are fully observed. **Source:** ZTdatenforum eGen (https://zt.co.at/).

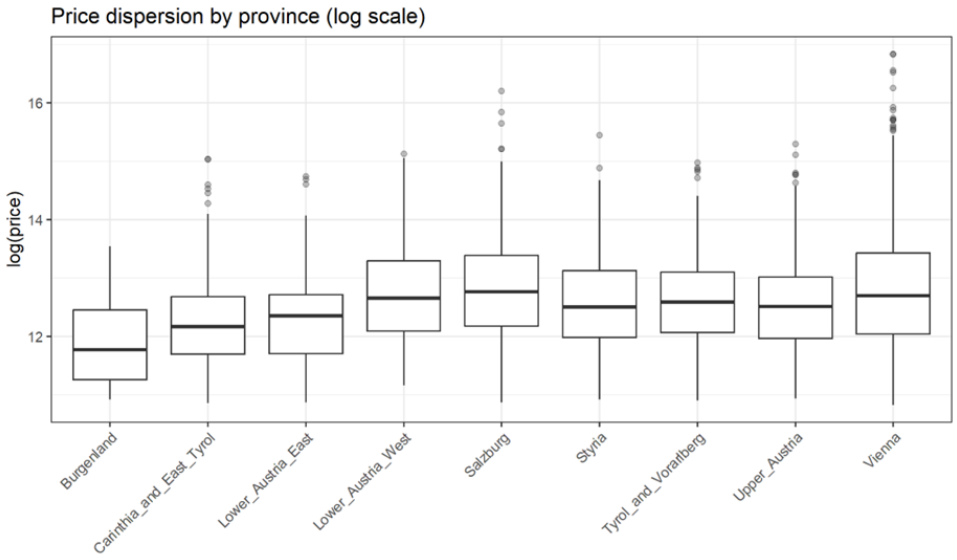
C.2 Composition effects under complete-case estimation

Complete-case estimation changes the effective composition of the estimation sample in economically meaningful ways. This matters because office prices vary strongly across regions and transaction types.

Regional composition. The complete-case sample slightly under-represents Vienna transactions

(29.7%) relative to incomplete observations (33.5%). Given that Vienna prices are substantially higher than elsewhere, this compositional shift is economically relevant.

Figure A.5: Price dispersion by province (log scale)



Source: ZTdatenforum eGen (https://zt.co.at/).

The pronounced cross-provincial dispersion underscores the importance of controlling for regional composition in hedonic estimation.

Complete versus incomplete transactions. Table A.7 compares summary statistics across complete and incomplete observations. While the complete-case sample exhibits a slightly lower mean price, it has a higher median and higher mean log price, indicating non-linear selection effects.

Table A.7: Complete versus incomplete transactions

Sample	N	Vienna share	Mean price (€)	Mean log price
Incomplete (non-CC)	1,759	0.335	530,634	12.6
Complete (CC)	1,244	0.297	507,087	12.7

Source: ZTdatenforum eGen (https://zt.co.at/).

Builder composition. The share of properties sold by builders varies substantially over the sample period. Builder-sold properties exhibit substantially lower missingness in *size* and *legal_age*. Time

variation in builder shares therefore mechanically affects the fraction of transactions retained under complete-case estimation.

Table A.8: Number and share of builder versus non-builder properties by year

Category	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Builder	79	61	140	105	83	113	129	157	137	77
Nonbuilder	223	189	198	185	206	195	224	202	152	133
Total	302	250	338	290	289	308	353	359	289	210
Share_Builder	0.26	0.24	0.41	0.36	0.29	0.37	0.36	0.44	0.47	0.37

Source: ZTdatenforum eGen (https://zt.co.at/).

Price differences conditional on size availability. Transactions with missing *size* differ systematically from those with size observed, and these differences vary by region. In Vienna, transactions with missing size are on average more expensive than those with size recorded. Outside Vienna, missing size is associated with lower prices and greater distance from major urban centres.

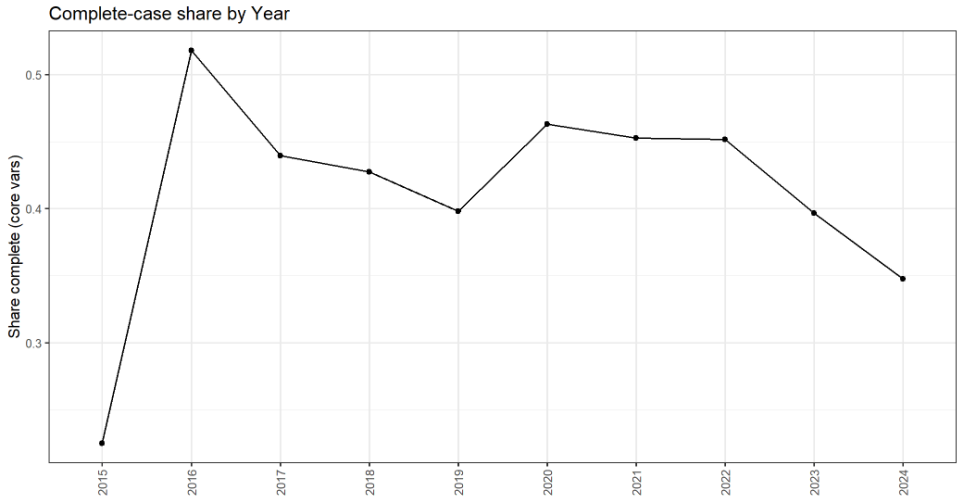
Table A.9: Summary statistics by region and size availability

Type	Variable	Rest_Size_no	Rest_Size_yes	Vienna_Size_no	Vienna_Size_yes
Mean	price	357994	489294	804191	612216
Mean	constructionSize	1063	1265	1196	980
Mean	plotSize	3123	3132	1989	1570
Mean	citydistance	24.94	23.12	3.93	4.15
Mean	doctor_dist	474.87	306.9	150.7	160.8
Mode	builder	0	0	0	1
Mode	legal_age	2000s	2015–2020	2000s	2015–2020
Mode	province	Salzburg	Salzburg	Vienna	Vienna

Source: ZTdatenforum eGen (https://zt.co.at/).

Temporal variation in missingness are shown in Table A.6.

Figure A.6: Temporal evolution of missingness in core variables



Note: The figure shows the share of missing observations over time for key hedonic variables. Missingness declines over the sample period, particularly for *size*. **Source:** ZTdatenforum eGen (https://zt.co.at/).

Taken together, these diagnostics indicate that complete-case estimation relies on a thinned and compositionally altered subsample.

C.3 Variable definitions and construction

The imputation model includes all variables used in the hedonic regression as well as additional auxiliary predictors, ensuring that the MAR assumption is as plausible as possible.

Table A.10: Variables used in the imputation model

Category	Variable	Description
Core property characteristics	**size**	Property size (e.g. floor area)
	legal_age	Building age category
	builder	Indicator for developer/builder transaction
Location and accessibility	**citydistance**	Distance to regional or capital city center
	city_in20km	Indicator: within 20km of regional or capital city center (Y/N)
	nearest_city	Nearest major city
	province	Federal state location
	Vie_dummy	Vienna indicator
	Graz_dummy	Graz indicator
	Salzburg_dummy	Salzburg indicator
	Linz_dummy	Linz indicator
	noCity_dummy	Outside major cities indicator
Local amenities	**doctor_dist**	Distance to nearest doctor
	shops1000	Number of shops within 1000m
Price and transaction variables	price	Transaction price
	hadPkwApPreis	Parking price indicator (Y/N)
	hadInventoryPrice	Inventory price indicator (Y/N)
Time controls	**Year**	Transaction year
	Halfyear	Half-year period
	Quarteryear	Quarter-year period
Spatial / regional identifiers	**PB_number**	District-level identifier
	PB_number_sel	Selected regional grouping
Additional controls	resRanking	Ordinal ranking indicating relative residential price level at district (PB) level

Note: Variables shown in bold are included on the right-hand side in the hedonic price regression. All listed variables are used as auxiliary predictors in the imputation model, including price where feasible. In practice, this is feasible for flexible specifications such as random forests, while parametric models may face numerical stability constraints when incorporating a similarly rich predictor set.

Legal age. The variable *legal_age* refers to the number of years since a property unit was officially parified (Parifizierung). Because major renovations or structural reconfigurations that alter ownership shares trigger a new parification, *legal_age* serves as a proxy for effective building age and renovation status.

Geolocation and accessibility measures. Property geolocation is determined using recorded transaction addresses. For a small subset of transactions where geocoding failed due to ambiguity, centroid coordinates of the corresponding postcode area are substituted. Distance-based accessibility measures are computed using these coordinates. Derived locational measures include *citydistance*, *doctor_dist*, and *shops1000*. Missingness in locational variables is negligible.

C.4 Little's MCAR test

Little's MCAR test (Little, 1988a) is applied to nested blocks of covariates to assess whether missingness can be considered missing completely at random (MCAR). Table A.11 reports the test statistics for increasingly rich specifications.

When only price, time, and size are included, the null hypothesis of MCAR cannot be rejected. However, once core quality characteristics and locational controls are added, MCAR is decisively rejected. Missingness is therefore systematically related to observed characteristics.

Table A.11: Little's MCAR test for nested covariate blocks (Austrian office transactions)

Covariate block	χ^2	df	p-value	Patterns
Price, time, size	0.76	2	0.684	2
+ Core quality characteristics	255.0	11	< 0.001	4
+ Parking indicator	273.4	14	< 0.001	4
+ Accessibility variables	394.9	63	< 0.001	10
+ Regional and city indicators	530.3	108	< 0.001	10
+ Full operational covariate set	559.8	115	< 0.001	10

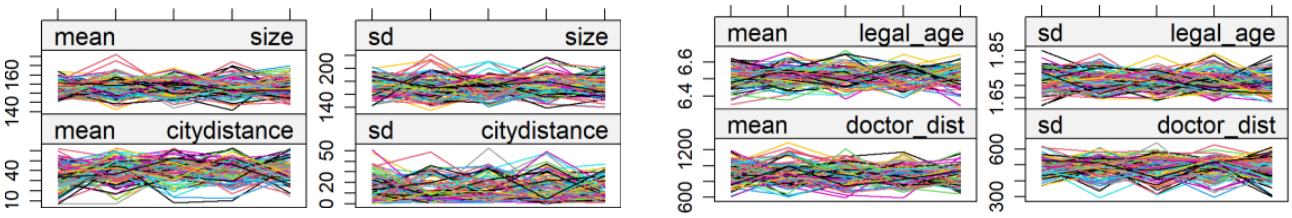
Notes: Each row reports Little's MCAR test for an increasingly rich set of covariates. Rejection of MCAR indicates that missingness is systematically related to observed characteristics. **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations.

C.5 Imputation diagnostics

Convergence diagnostics

The trace plots show the evolution of imputed values across iterations and chains for the main variables with missingness. The trajectories stabilise rapidly and exhibit no systematic trends, while variation across chains remains comparable to within-chain variation. This pattern indicates satisfactory convergence of the chained-equations algorithm and suggests that the imputation draws are sampled from a stable distribution.

Figure A.7: Imputation convergence diagnostics (trace plots)



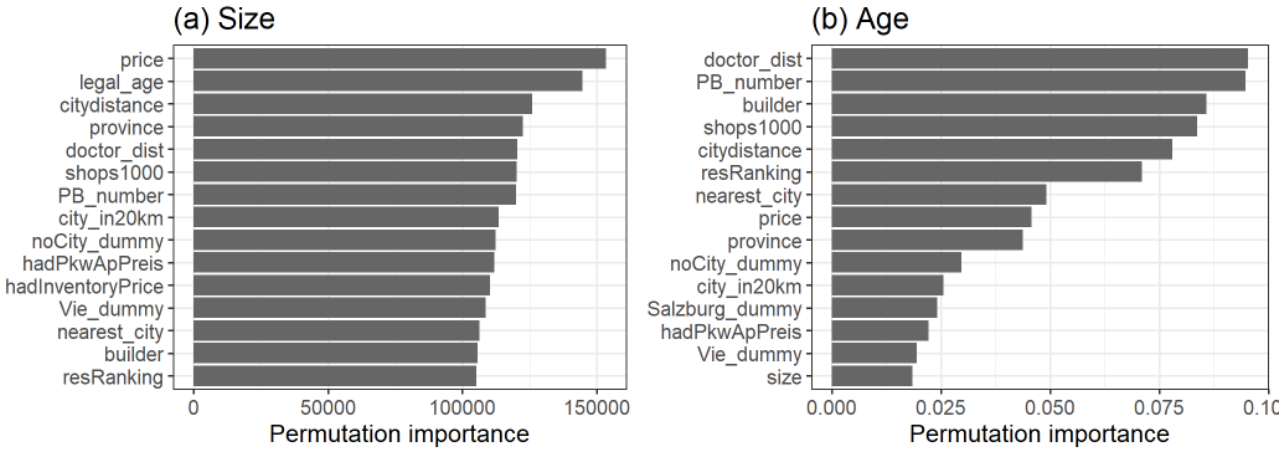
Note: Convergence diagnostics indicate stable imputation streams after five iterations. Between-stream variance does not exceed within-stream variance (van Buuren and Groothuis-Oudshoorn, 2011). **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations.

Variable importance in the imputation model

Figure A.8 reports permutation-based variable importance from the random-forest imputation models for the two main variables with missingness (*size* and *legal_age*). The results indicate that imputations are driven by economically meaningful predictors, including price, location, and accessibility measures. This suggests that the imputation procedure exploits systematic relationships in the data rather than relying primarily on mechanical or noise-driven predictions.

These patterns differ across target variables. For *size*, the most important predictors include transaction price and core structural characteristics, reflecting that imputation is anchored in overall property scale and quality. By contrast, the imputation of *legal_age* is driven more strongly by locational and accessibility variables, implying that building age is inferred primarily from spatial and neighbourhood characteristics rather than from price alone. This contrast highlights that the imputation model adapts to the economic nature of the missing variable.

Figure A.8: Random-forest variable importance for imputed variables



The figure reports permutation-based variable importance for the imputation of *size* (left panel) and *legal_age* (right panel). Higher values indicate greater predictive contribution in the random-forest imputation models. **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations.

C.6 Pooled coefficient summary

Table A.12 reports pooled coefficient estimates from the hedonic time-dummy regression corresponding to the baseline specification (MICE-RF imputation with Cook's-distance filtering). Coefficients are combined across $m = 100$ imputations using Rubin's rules.

The table serves two purposes. First, it provides a benchmark summary of the underlying hedonic model used to construct the price index. The estimated coefficients exhibit economically plausible signs and magnitudes, with strong and precisely estimated effects for key characteristics such as size, location, and building attributes. Second, the results confirm that the hedonic structure is stable across imputations. Despite substantial missingness in core variables, the pooled estimates are well-behaved and statistically precise for the main covariates. This suggests that multiple imputation restores the effective estimation sample without generating visible instability in the underlying regression model.

Importantly, Rubin's rules are used here only for coefficient inference. The construction of the price index itself does not rely on Rubin's pooling, but instead follows the growth-based aggregation of adjacent-period price relatives described in Sections 2 and 4.

Table A.12: Pooled hedonic time-dummy regression results (MICE-RF, Cook-filtered)

Variable	Estimate	SE	Lower CI	Upper CI
(Intercept)	8.525***	0.319	7.899	9.151
log(size)	0.553***	0.027	0.499	0.607
log(citydistance)	-0.082**	0.037	-0.154	-0.009
builder (yes)	0.210***	0.032	0.148	0.272
legal_age1960s	0.305	0.240	-0.165	0.776
legal_age1970s	0.262	0.232	-0.192	0.717
legal_age1980s	0.469*	0.233	0.011	0.926
legal_age1990s	0.324	0.225	-0.118	0.765
legal_age2000s	0.402	0.226	-0.042	0.845
legal_age2010–15	0.517**	0.228	0.071	0.963
legal_age2015–2020	0.454**	0.225	0.012	0.896
legal_age2020+	0.505**	0.226	0.063	0.948
doctor_dist	-0.00006**	0.00002	-0.00011	-0.00001
PB_number109	0.777**	0.268	0.251	1.302
PB_number208	0.986***	0.299	0.400	1.571
PB_number324	1.185***	0.239	0.716	1.654
PB_number704	1.487***	0.244	1.008	1.966
PB_number9001	2.336***	0.234	1.878	2.795
PB_numberOther	0.647***	0.219	0.218	1.076
		⋮		
hadPkwApPreis	0.288***	0.037	0.216	0.360
Year2016	-0.023	0.060	-0.141	0.094
Year2017	-0.047	0.056	-0.157	0.062
Year2018	0.112*	0.058	-0.001	0.226
Year2019	0.058	0.057	-0.054	0.171
Year2020	0.050	0.057	-0.062	0.163
Year2021	0.231***	0.056	0.122	0.340
Year2022	0.304***	0.055	0.195	0.413
Year2023	0.317***	0.061	0.199	0.436
Year2024	0.259***	0.067	0.127	0.390

Notes: Coefficients are pooled across $m = 100$ multiply imputed datasets using Rubin's rules. Estimation is based on the Cook-filtered baseline specification. For brevity, the table reports core hedonic variables, selected location effects, and time dummies; remaining district and city indicators are included but not shown. **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

C.7 Imputation method comparison: RF, linear, and mean imputation

This section documents the alternative imputation approaches used in Section 4.5.

MICE-default. As a further benchmark, we implement the default MICE specification, which uses predictive mean matching (PMM) for numerical variables and logistic or multinomial regression for

categorical variables. PMM estimates predicted values from a regression model, finds observed cases with similar predicted means, and then imputes missing values by drawing an observed value from that donor set.

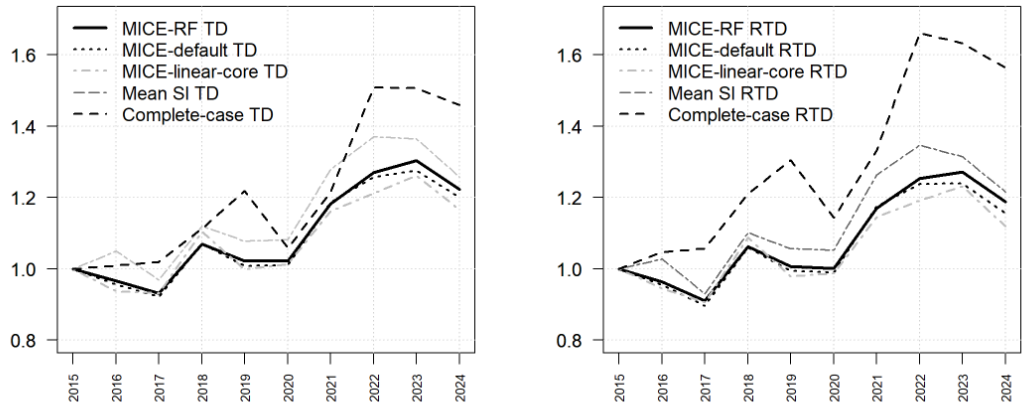
Linear MICE with reduced predictor set (MICE-linear-core). As an additional benchmark, we implement linear imputation within the MICE framework using standard regression models for conditional prediction. In principle, we aimed to use the same rich predictor set as in the baseline MICE specifications. However, fully specified linear chained-equations models proved numerically unstable due to strong multicollinearity and sparse factor levels, leading to near-singular design matrices in several conditional models. We therefore restrict the predictor set for the linear specification to a core subset of variables, excluding high-dimensional location controls and other auxiliary predictors. Differences relative to the baseline thus reflect both the linear functional form and the reduced predictor set.

Mean imputation. As a simple reference, missing values in *size* and *legal_age* are replaced by their sample means. This approach ignores both uncertainty and conditional relationships and is included solely as a mechanical benchmark.

Across all specifications, imputation-based indices are constructed by separately estimating the hedonic model for each completed dataset and pooling adjacent-period growth rates, as described in Section 2.

Figure A.10 compares index trajectories across imputation methods prior to influence filtering. Approaches based on restricted predictor sets or simple imputation rules tend to track the complete-case index more closely, while richer MICE specifications yield systematically lower and smoother index paths.

Figure A.9: Comparison of price indices across imputation methods (no influence filtering)

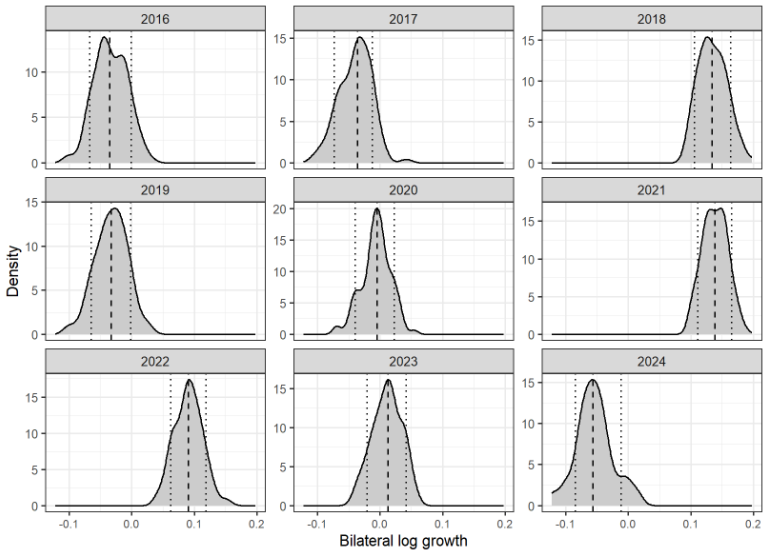


Note: Annual time-dummy (TD, left panel) and rolling time-dummy (RTD, right panel) indices constructed using alternative missing-data treatments prior to Cook's-distance filtering. The solid line denotes MICE-RF, the dotted line default MICE with PMM, the light dashed line MICE with a reduced linear predictor set (MICE-linear-core), the dash-dotted line mean imputation, and the long-dashed line the complete-case index. MICE-RF and default MICE produce very similar trajectories, whereas MICE-linear-core and mean imputation lie closer to the complete-case index. **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations.

C.8 Imputation uncertainty and temporal aggregation

Figure A.10 shows the distribution of annual bilateral log growth rates across the $m = 100$ imputed datasets prior to influence cleaning. Dispersion is relatively high in the earliest years of the sample, narrows in the middle period, and widens again toward the end of the sample. This pattern is consistent with more severe missingness in the beginning of the sample, and with lower transaction counts and greater market heterogeneity towards the end.

Figure A.10: Distribution of annual bilateral log growth rates across imputations



Note: Kernel densities summarize annual bilateral log growth rates across the $m = 100$ imputed datasets. Dispersion reflects imputation-to-imputation variability prior to influence cleaning. **Source:** ZTdatenforum eGen (https://zt.co.at/), authors' calculations.