Online Appendix for:
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

April 13, 2026

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.

*Corresponding author.

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 (**?**).

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 B.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 B.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 B.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 B.2: Summary statistics: Apartment application

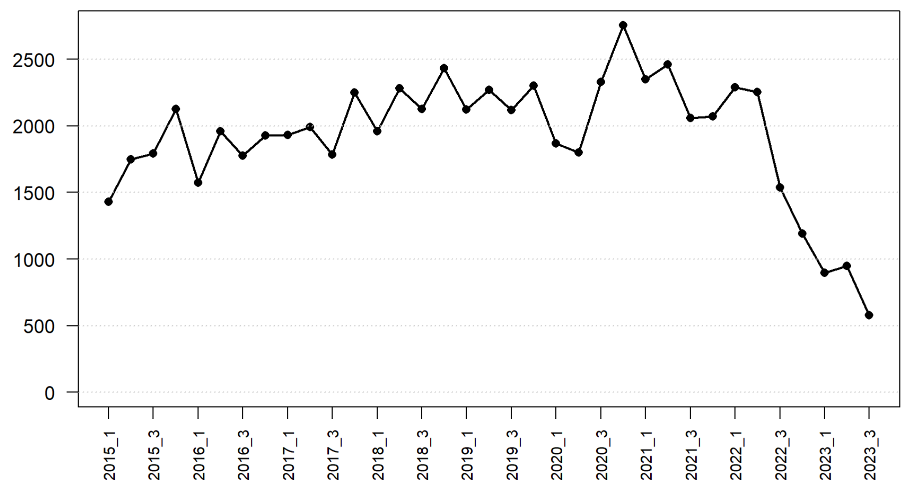
Variable	N / Share	Mean	SD	Median
Sample overview				
Number of transactions	67,292			
Number of quarters	35			
Categorical variables				
Builder (yes)	65.59%			
Builder (no)	34.41%			
Number of postcodes	23			
Continuous variables				
Price (€)	67,292	355,607	348,931	282,000
Price per sqm (€)	67,292	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 B.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.

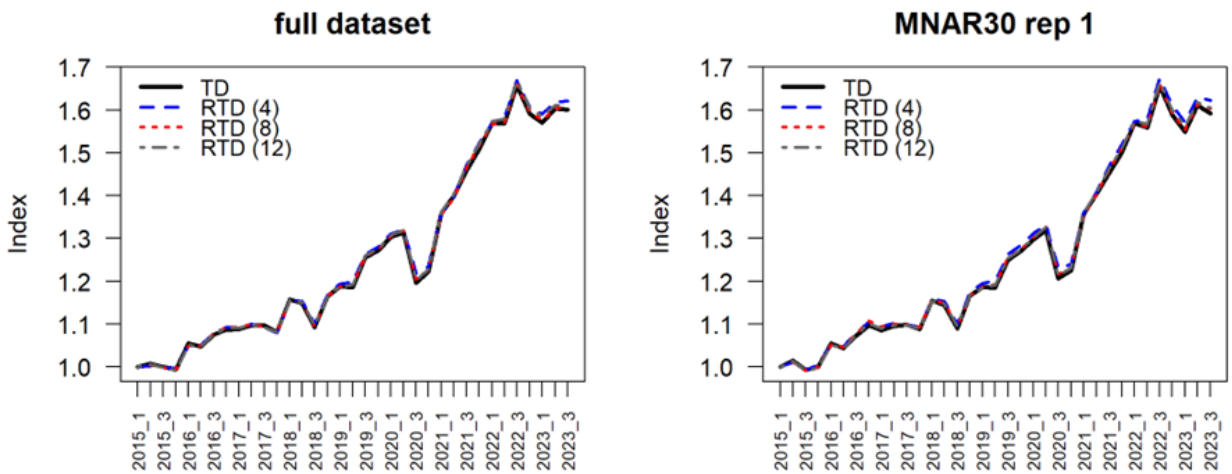
Figure B.1: Number of apartment transactions per quarter



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 B.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 B.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 (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.

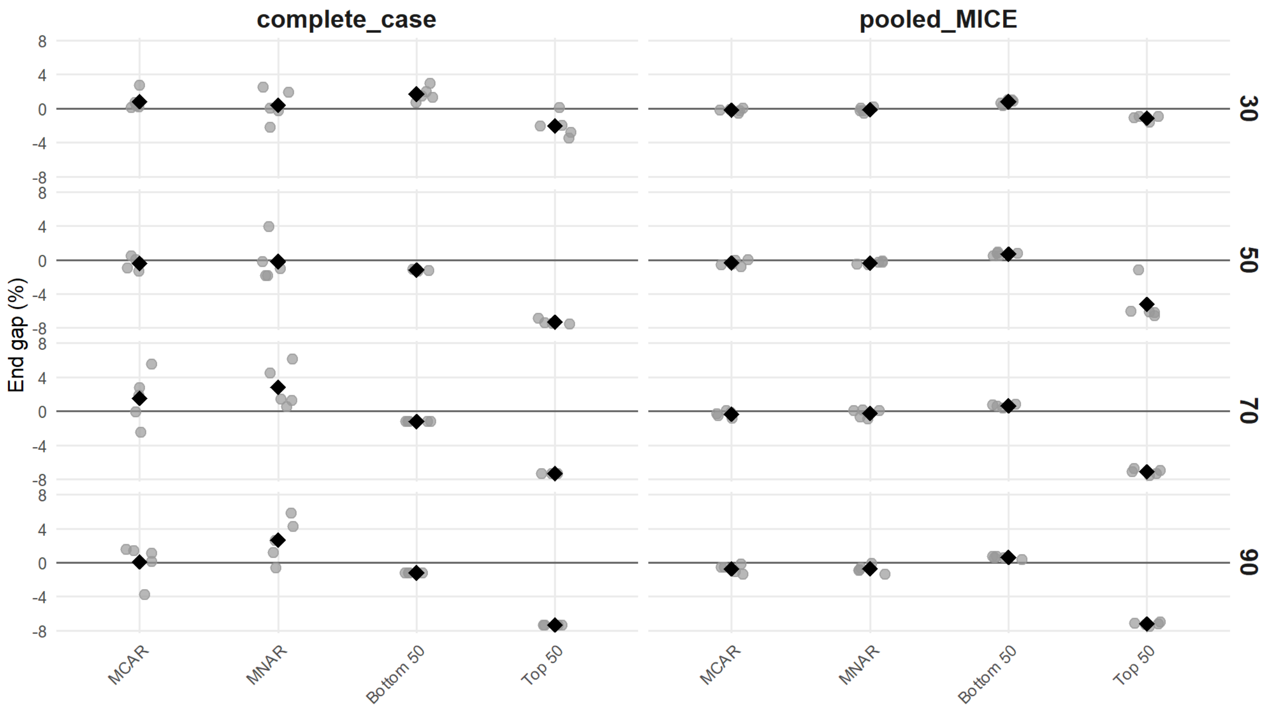
For each mechanism and each missingness level, we generate five independent realizations of the incomplete dataset. Each realization corresponds to a different random draw under the same design. This allows us to assess the stability of index estimates across samples.

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 **??** 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 B.3: End-of-sample index level gap relative to the benchmark across missingness level



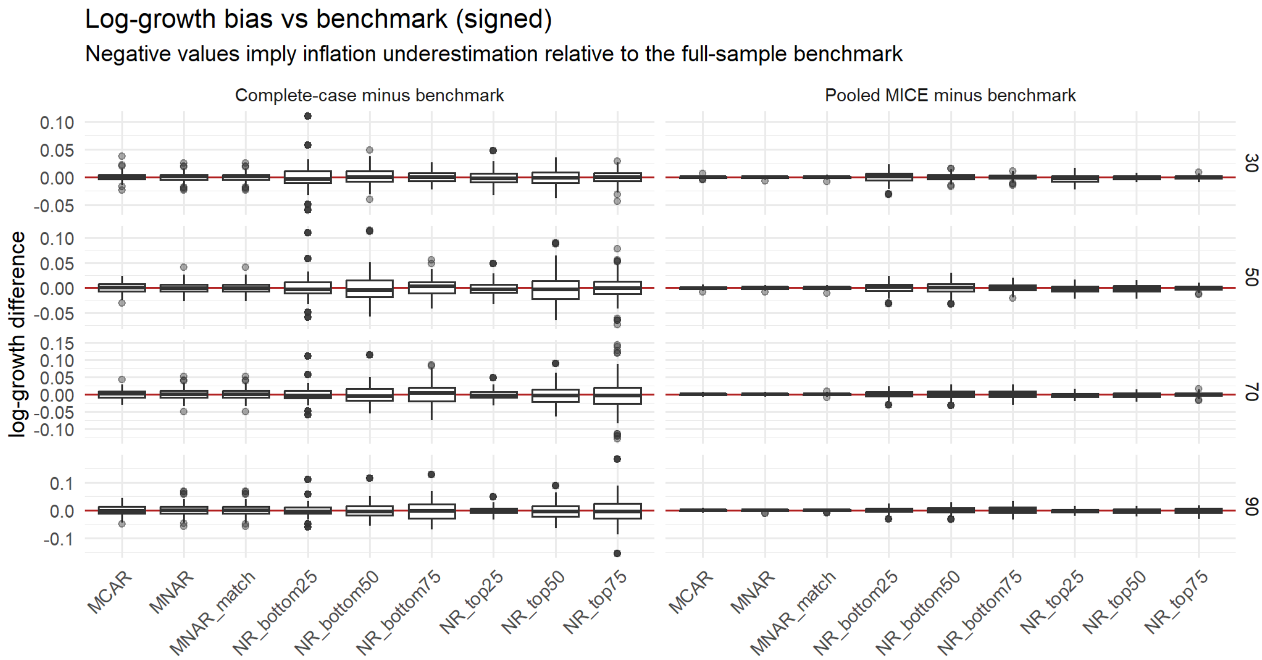
Note: Each dot represents one simulation replicate; diamonds indicate the mean across replicates. The gap is defined as $100 \times (\text{Index}_{\text{end}}/\text{Benchmark}_{\text{end}} - 1)$. Complete-case estimation leads to increasingly large and systematic deviations from the benchmark, along 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 B.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 B.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 (https://zt.co.at/), authors' calculations and simulations.

B.6 RMSE decomposition

Table B.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 $\text{RMSE}^2 = \text{bias}^2 + \text{variance}$. Bias is reported in quarterly basis points (qbps); bias^2, variance, and RMSE are in log-growth units. **Source:** ZTdatenforum (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 B.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 B.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 B.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 (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 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 C.1: 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

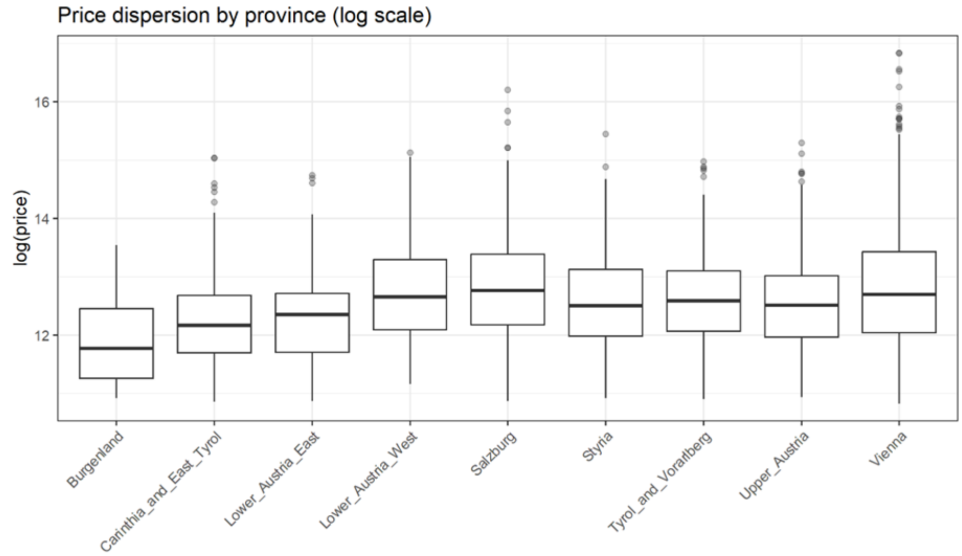
Notes: The table reports the number and percentage of missing observations by variable. Missingness is concentrated in the core quality characteristics, particularly unit size and legal age. All remaining variables used in the hedonic specification are fully observed. **Source:** ZTdatenforum (https://zt.co.at/).

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 C.1: Price dispersion by province (log scale)



Source: ZTdatenforum (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 C.2 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 C.2: 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 (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 C.3: 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 (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.

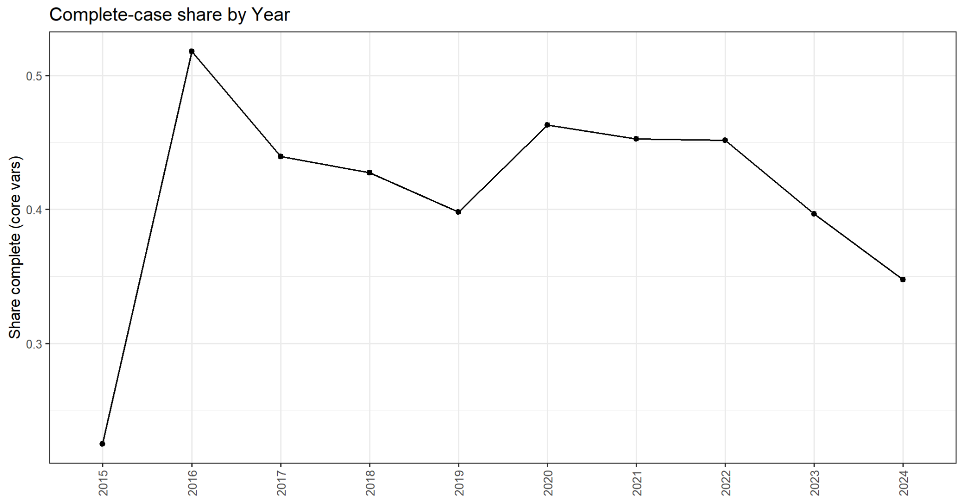
Table C.4: Summary statistics by region and size availability

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 (https://zt.co.at/).

Temporal variation in missingness are shown in Table C.2.

Figure C.2: 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 (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 C.5: 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 (?) is applied to nested blocks of covariates to assess whether missingness can be considered missing completely at random (MCAR). Table C.6 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 C.6: 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

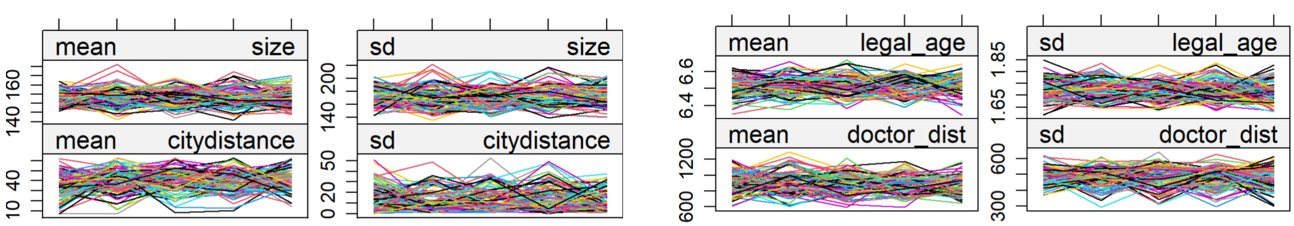
Notes: Each row reports Little's MCAR test for an increasingly rich set of covariates. Rejection of MCAR indicates that missingness is systematically related to observed characteristics. **Source:** ZTdatenforum (https://zt.co.at/), authors' calculations.

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 C.3: Imputation convergence diagnostics (trace plots)



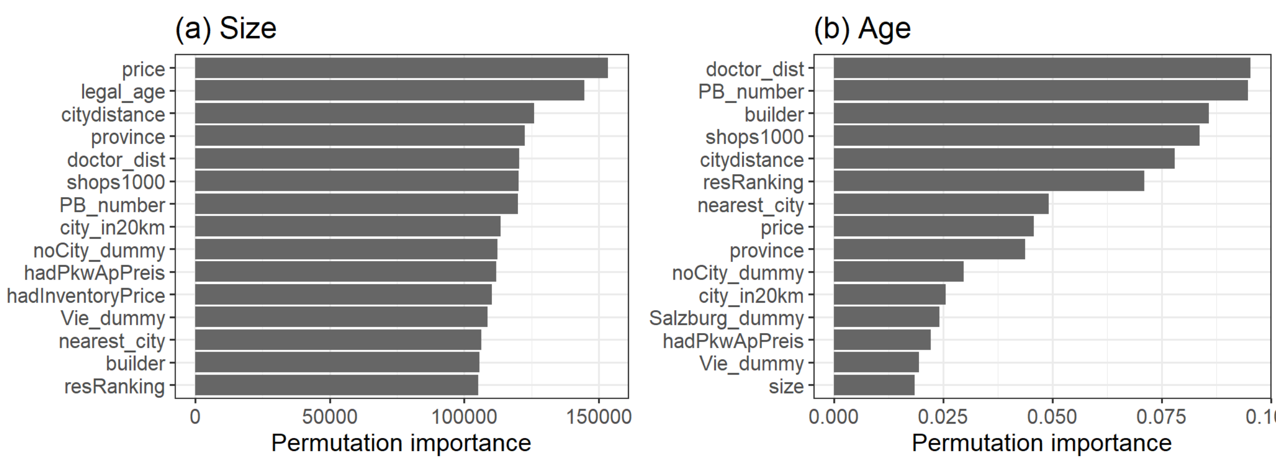
Note: Convergence diagnostics indicate stable imputation streams after five iterations. Between-stream variance does not exceed within-stream variance (?). **Source:** ZTdatenforum (https://zt.co.at/), authors' calculations.

Variable importance in the imputation model

Figure C.4 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 C.4: 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 (https://zt.co.at/), authors' calculations.

C.6 Pooled coefficient summary

Table C.7 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 C.7: 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 (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 **??**.

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.

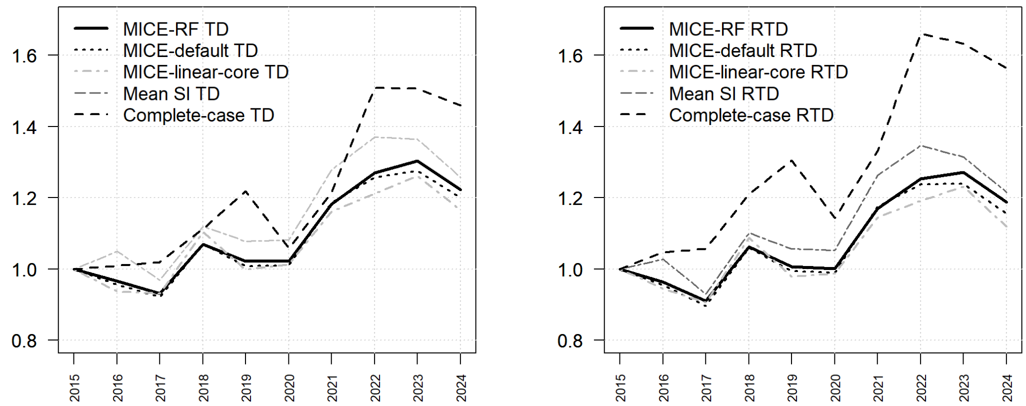
Linear MICE with reduced predictor set (MICE-linear-core). As an additional benchmark, we implement linear imputation within the MICE framework using standard regression models for conditional prediction. In principle, we aimed to use the same rich predictor set as in the baseline MICE specifications. 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 C.6 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 C.5: Comparison of price indices across imputation methods (no influence filtering)

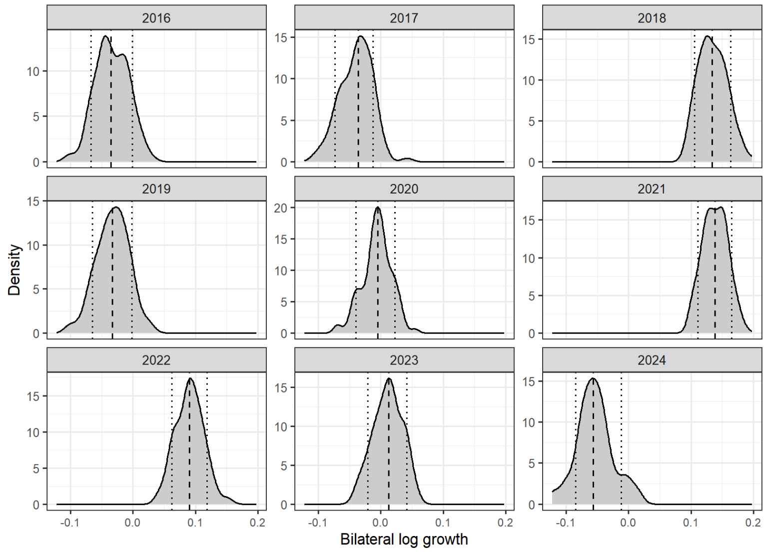


Note: Annual time-dummy (TD, left panel) and rolling time-dummy (RTD, right panel) indices constructed using alternative missing-data treatments prior to Cook's-distance filtering. The solid line denotes MICE-RF, the dotted line default MICE with PMM, the light dashed line MICE with a reduced linear 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 (https://zt.co.at/), authors' calculations.

C.8 Imputation uncertainty and temporal aggregation

Figure C.6 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 C.6: 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 (https://zt.co.at/), authors' calculations.