UNIVERSITÄT
TRIER

Ludwig von Auer

The Gap Pattern Coefficient: Diagnosing Missing-Data Bias in Multilateral Price Level Measurement

Research Papers in Economics
No. 9/26

The Gap Pattern Coefficient: Diagnosing Missing-Data Bias in Multilateral Price Level Measurement

by Ludwig von Auer
Universität Trier

July 23, 2026

Abstract

Multilateral price indices are widely used to estimate and compare price levels across units such as time periods or regions using observed prices and quantities of individual items. Besides ensuring transitivity and reducing chain drift, these methods are often assumed to mitigate the effects of missing price observations by exploiting information from multiple comparison units. This paper argues that this advantage can only be realized if item-level missingness is independent of the sensitivity of item prices to overall unit price levels. Recent research has shown that this assumption is frequently violated in practice, resulting in biased price-level estimates.

To address this issue, the paper develops a simple diagnostic procedure. For each item, the procedure quantifies the sensitivity of its prices to unit price levels and measures missingness by the number of absent observations. The association between price-level sensitivity and missingness is then quantified by the gap pattern coefficient. A non-zero gap pattern coefficient indicates that standard multilateral index methods are likely to produce biased results. If desired, the analysis can be complemented by a permutation-based assessment of the statistical significance of the gap pattern coefficient. Monte Carlo simulations evaluate the performance of the proposed diagnostic procedure. The paper further presents simulation evidence showing that GEKS price-level estimates are more sensitive to missing observations than those obtained from CPD or GK.

JEL-Classification: C43, E31
Keywords: Measurement Bias, Monte Carlo Simulation, Price Index, Scanner Data, Test

1 Introduction

Multilateral price indices were originally developed to compare the price levels of different regions in a transitive manner. Over time, their scope of application was extended to intertemporal price level comparisons. For the purposes of the present analysis, the distinction between spatial and temporal price level comparisons is immaterial. Both the underlying problem and the proposed solution apply equally to comparisons across regions and across periods. Accordingly, the terms regions and periods are replaced by the more neutral concept of (comparison) units. Multilateral price indices thus compare the price levels of different units.

The most prominent multilateral price indices include the Gini (1931), Éltető-Köves (1964), and Szulc (1964) (GEKS) approach; the weighted country product dummy (CPD) regression (referred to as the TPD regression in an intertemporal context); and the Geary (1958) and Khamis (1970, 1972) (GK) index. For example, the International Comparison Program (2021, pp. 29-30) recommends aggregating prices within basic headings (the finest level of product classification for which expenditure weights are available) using the CPD regression and aggregating basic headings into overall price levels using the GEKS approach.

Advocates of multilateral indices argue that, besides ensuring transitivity and avoiding chain drift, these methods can also mitigate the information loss caused by missing data. In the present context, a missing observation is one for which a product was available and price and quantity information existed but did not enter the dataset. The missing information may consist of the price, the quantity, or both. Importantly, this definition also includes periods with zero sales. Although no transaction takes place in such periods, the prevailing price together with the zero quantity constitutes relevant information and should be transmitted to the dataset. If this information is omitted, the observation is regarded as missing. By contrast, products that were not yet available or were no longer available on the market do not constitute missing observations, because in these cases neither price nor quantity information exists. Missing observations may arise from several sources, including non-response, filtering, transmission errors, and the omission of intermittent zero-sales periods. Since multilateral indices draw on information from many units rather than only two, they are better equipped to cope with missing data.

This paper argues that while this claim is correct in principle, the GEKS approach, the CPD regression, and the GK index can only achieve this benefit if the number of missing observations of an item is uncorrelated with the sensitivity of the item's prices with respect to the price levels of the units.

This absence of correlation is implicitly assumed whenever these multilateral index methods are applied. However, Auer and Weinand (2025, 2026) demonstrate that this implicit assumption is frequently violated in practice and that such violations lead to biased CPD and GK price level estimates. They also explain why the GEKS approach suffers from the same problem.

It is therefore essential, when applying these methods, to examine whether item-level missingness is related to price-level sensitivity. To enable such an assessment, this study develops the *gap pattern coefficient*. It quantifies the association between price-level sensitivity and missingness. This diagnostic tool is the paper's primary methodological contribution.

The gap pattern coefficient is computed in two stages. First it computes for each item the extent to which its prices respond to variations in the overall price level of the unit, thereby yielding an item-specific measure of price-level sensitivity. Item-level missingness is then measured by the number of absent observations. Finally, the association between price-level sensitivity and missingness is computed.

If users additionally wish to assess whether the observed gap pattern could plausibly have arisen by chance, the analysis can be complemented by the *gap pattern test*, a permutation-based assessment of the statistical significance of the gap pattern coefficient.

Beyond developing the proposed diagnostic framework, the paper makes a second methodological contribution by investigating how incomplete data affect the robustness of widely used multilateral index methods. The analysis shows that GEKS price-level estimates are more sensitive to missing observations than those obtained from CPD or GK. This finding suggests that data completeness deserves particular attention when GEKS is the preferred index method. It also highlights the value of diagnostic tools for assessing missing-data patterns.

The remainder of the paper is organized as follows. Section 2 explains why an association between missingness and price-level sensitivity leads to biased price-level estimates. Section 3 introduces a pair-weighted extension of Kendall's rank concordance coefficient, τ_b, which is used in Section 4 to formalize the notion of price-level sensitivity and in Section 5 to define the gap pattern coefficient. Section 6 introduces the gap pattern test and discusses the applicability and reliability of the gap pattern coefficient and test. Section 7 presents a Monte Carlo simulation framework for evaluating the proposed diagnostic procedure. Section 8 investigates the robustness of widely used multilateral index methods to incomplete data. Section 9 concludes.

2 Bias Caused by Gap Patterns

If the number of missing observations differs across *units*, but within each unit observations are missing completely at random, this does not introduce bias in the measurement of price levels. Bias arises, however, when missing observations vary across *items* and the prices of these items differ in their sensitivity to price levels. This issue is illustrated below using a simple numerical example.

Table 1 reports the prices of the three items *Goods*, *Services*, and *Housing* for the four units, A, B, C, and D. For simplicity, all observations are assumed to be equally weighted. Four different scenarios are considered.

Mechanism:	Complete				MNAR neg.				MNAR pos.				MCAR			
Unit:	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Goods	5	5	6	5		5		5	5	5	6	5	5	5		5
Services	7	8	8	9	7	8	8	9	7	8	8	9	7	8	8	
Housing	3	4	6	8	3	4	6	8			6	8	3		6	8
$P^{AD}_{\text{GEKS-J.}}$	1.508				1.790				1.165				1.548			
P^{AD}_{CPD}	1.508				1.731				1.196				1.565			
P^{AD}_{GK}	1.467				1.626				1.225				1.549			

Table 1: A price scenario with four different gap patterns.

In the first scenario, price observations are complete. This scenario therefore serves as the benchmark against which the other three scenarios are evaluated. For the comparison between units A and D, the complete data yield the bilateral Jevons price index

$$P_J^{AD} = \left(\frac{5}{5} \cdot \frac{9}{7} \cdot \frac{8}{3}\right)^{1/3} = 1.508 \, .$$

Unit D therefore exhibits a price level that is 50.8 percent higher than that of unit A. This number, along with the other Jevons index numbers (P_J^{AB}, P_J^{AC}, P_J^{BC}, P_J^{BD}, and P_J^{CD}) is reported in Table 2.

With complete data, the GEKS-Jevons approach yields exactly the same result as the Jevons index P_J^{AD}:

$$P^{AD}_{\text{GEKS-J.}} = \left[\left(P_J^{AA} P_J^{AD}\right) \cdot \left(P_J^{AB} P_J^{BD}\right) \cdot \left(P_J^{AC} P_J^{CD}\right) \cdot \left(P_J^{AD} P_J^{DD}\right)\right]^{(1/4)} = 1.508 \, .$$

The CPD regression also yields this value, whereas the GK index produces a value of 1.467 (see Table 1).

The scenario "MNAR neg." refers to "missing not at random" with a negative relationship between missingness and price-level sensitivity. In Table

Index	Complete	MNAR neg.	MNAR pos.	MCAR
P_J^{AB}	1.150	1.234	1.069	1.069
P_J^{AC}	1.340	1.512	1.171	1.512
P_J^{AD}	1.508	1.852	1.134	1.633
P_J^{BC}	1.216	1.225	1.095	1.000
P_J^{BD}	1.310	1.310	1.061	1.000
P_J^{CD}	1.077	1.225	1.077	1.333

Table 2: Jevons price index values for complete and incomplete data.

1, this scenario results in two missing observations for the item *Goods*, which has the lowest price-level sensitivity. Owing to these missing observations, all bilateral Jevons index values, except P_J^{BD}, are larger than in the case of complete data (see Table 2). As a result, the GEKS-Jevons approach produces a higher estimate than in the case of complete data. The findings stated in Table 1 show that the CPD regression and the GK index suffer from the same problem.

In the scenario "MNAR pos.", the sign of the relationship is reversed. Accordingly, the resulting index values are lower than in the case of complete data. In the MCAR (missing completely at random) scenario, deviations from the complete-data case may also occur. However, these deviations are not systematic.[1]

The numerical illustration demonstrates that multilateral price indices may be subject to bias when the number of gaps in the underlying dataset is systematically related to the items' price-level sensitivity. It is therefore important to assess whether such a relationship exists in practice. This requires a formal measure of price-level sensitivity and a framework for relating it to the extent of item-level missingness. Both objectives can be achieved by concepts derived from Kendall's (1945) rank concordance coefficient, τ_b.

3 Pair-Weighted Rank Concordance

Let $x = (x_g)_{g \in \mathcal{G}}$ and $y = (y_g)_{g \in \mathcal{G}}$ denote two variables observed on the common index set $\mathcal{G}$. For each $g \in \mathcal{G}$, the pair (x_g, y_g) represents one observation.

[1] In scanner datasets, periods with zero sales are typically not recorded, even though the corresponding price and weighting information for the item would be available. Consequently, these datasets contain missing observations. If zero-sales periods occur randomly, they do not introduce systematic bias. However, if they arise under an MNAR.pos (MNAR.neg) mechanism, the resulting multilateral price index is biased downward (upward) relative to the case with complete data.

For any two distinct observations indexed by $g, h \in \mathcal{G}$ with $g \neq h$, let $w_{gh} \geq 0$ denote the pair-specific weight that quantifies their joint importance.

This study introduces the coefficient $\tau_{b,w}(x,y)_{g\in\mathcal{G}}$, a pair-weighted extension of Kendall's (1945) $\tau_b(x,y)_{g\in\mathcal{G}}$, obtained by replacing the unweighted counts of concordant, discordant, and tied pairs with weighted pair totals. Their definition uses the sign functions

$$\mathrm{sgn}^{(x)}_{gh} = \mathrm{sgn}(x_g - x_h) \qquad \text{and} \qquad \mathrm{sgn}^{(y)}_{gh} = \mathrm{sgn}(y_g - y_h) \tag{1}$$

and the indicator function $1(\cdot)$, where $1(A)$ equals one if A is true and zero otherwise. The weighted concordant and discordant totals are

$$C = \sum_{g<h} w_{gh}\, 1\Big(\mathrm{sgn}^{(x)}_{gh} = \mathrm{sgn}^{(y)}_{gh} \neq 0\Big) \ , \tag{2}$$

$$D = \sum_{g<h} w_{gh}\, 1\Big(\mathrm{sgn}^{(x)}_{gh} = -\,\mathrm{sgn}^{(y)}_{gh}\Big) \ , \tag{3}$$

where $\sum_{g<h}$ denotes the sum over all unordered pairs of distinct elements g and h in $\mathcal{G}$ such that $g < h$. The weighted tie totals are

$$T_x = \sum_{g<h} w_{gh}\, 1\Big(\mathrm{sgn}^{(x)}_{gh} = 0,\ \mathrm{sgn}^{(y)}_{gh} \neq 0\Big) \ , \tag{4}$$

$$T_y = \sum_{g<h} w_{gh}\, 1\Big(\mathrm{sgn}^{(x)}_{gh} \neq 0,\ \mathrm{sgn}^{(y)}_{gh} = 0\Big) \ . \tag{5}$$

Definition 1 *The **pair-weighted rank concordance coefficient** is defined as*

$$\tau_{b,w}(x,y)_{g\in\mathcal{G}} = \frac{C - D}{\sqrt{(C + D + T_x)(C + D + T_y)}} \ , \tag{6}$$

where C, D, T_x, and T_y are defined by Eqs. (1) to (5).

When x and y exhibit a noisy monotonic relationship, randomly removing observations does not change the expected value of the coefficient $\tau_{b,w}(x,y)_{g\in\mathcal{G}}$, although it increases its sampling variability. By contrast, removing all observations above or below a threshold in either x or y reduces the range of variation and typically attenuates the absolute value of $\tau_{b,w}(x,y)_{g\in\mathcal{G}}$ toward zero.

Setting all pairwise weights w_{gh} $(g, h \in \mathcal{G})$ in Eqs. (2) to (5) equal to unity yields Kendall's ordinary, unweighted concordance coefficient $\tau_b(x,y)_{g\in\mathcal{G}}$. Whether the weighted or unweighted version is preferable depends on the context in which the coefficient is applied. In the analysis of a dataset's gap pattern, the coefficient is used at two distinct levels. At the lower level, where the items' price-level sensitivities are estimated, the unweighted version is generally preferable because it assigns equal importance to all units. At the upper level, where the relationship between price-level sensitivity and item-level missingness is assessed, the weighted version is often more attractive because it allows greater importance to be attached to economically more relevant items.

4 Price-Level Sensitivity of Prices

Let $r \in \mathcal{R} := \{1, \dots, R\}$ index units and let $i \in \mathcal{N} := \{1, \dots, N\}$ index items. The price level in unit r is denoted by P^r. The price of item i in unit r is denoted by $p_i^r > 0$, the quantity by $q_i^r > 0$, and the expenditure share within unit r by w_i^r. The set $\mathcal{R}_i$ contains all units in which item i is not missing.

Definition 2 *The estimated **price-level sensitivity** of item i is defined as*

$$S_i = \tau_b\big(\ln p_i^r, \ln P^r\big)_{r \in \mathcal{R}_i} \cdot \mathrm{sd}\big(\ln p_i^r\big)_{r \in \mathcal{R}_i} \,, \qquad i \in \mathcal{N} \,. \tag{7}$$

In the sensitivity measure S_i, Kendall's rank concordance coefficient $\tau_b(\cdot)$ and the standard deviation $\mathrm{sd}(\cdot)$ are computed across the set of units $r \in \mathcal{R}_i$, provided that the number of units in set $\mathcal{R}_i$ is at least as large as some minimum threshold. If $\mathcal{R}_i$ is smaller than this threshold, no S_i-value is computed. The only exception arises when at least two observations are available and all observed prices p_i^r are identical. In this case, the item's price-level sensitivity is defined as $S_i = 0$.

The measure S_i combines a directional component, captured by Kendall's rank concordance coefficient between log item prices and log price levels, with an amplification component, captured by the dispersion of log item prices. By construction, the measure assigns low sensitivity to items whose prices exhibit little systematic comovement with price levels, even if their idiosyncratic volatility is high, thereby avoiding spurious amplification driven by noise.

The use of logarithmic prices ensures invariance to changes in currency and measurement units, while the rank-based directional component provides robustness to outliers and nonlinear relationships. The resulting measure preserves an economically meaningful interpretation: positive (negative) values indicate systematic comovement (countermovement) with the price level, while the magnitude reflects the extent to which item-level price dispersion amplifies or dampens price-level variation. As a purely descriptive statistic,

S_i does not rely on parametric assumptions, functional-form restrictions, or distributional approximations.[2] It summarizes the association between an item's prices and the overall unit price level and should therefore not be interpreted as a causal effect or an elasticity.

The measure S_i treats all units equally and does not distinguish between units with larger and smaller importance. However, if different importance is to be attached to the units $r \in \mathcal{R}_i$, expression (7) can instead be based on the pair-weighted rank concordance coefficient (6). Then, the weighted standard deviation should be used, where the weights reflect the relative importance of the units for item i.

5 Gap Pattern Coefficient

The price-level sensitivity S_i can be embedded in a measure that formalizes whether item-level missingness, M_i, is systematically related to price-level sensitivity, S_i. Let

$$\widetilde{\mathcal{P}} := \{(p_i^r, q_i^r) : i \in \mathcal{N},\ r \in \mathcal{R}\}$$

denote the set of complete data, and let

$$\mathcal{P} \subseteq \widetilde{\mathcal{P}}$$

denote the subset of actually observed data. The mapping

$$\widetilde{\mathcal{P}} \mapsto \mathcal{P}$$

is assumed to be generated by a stochastic missingness mechanism that is unit-invariant but possibly item-specific. Let Z_i^r indicate whether (p_i^r, q_i^r) is observed in $\mathcal{P}$. Accordingly, the number of missing observations of item i,

$$M_i = \sum_{r \in \mathcal{R}} (1 - Z_i^r) \ , \tag{8}$$

may be associated with its price-level sensitivity S_i.

When multilateral indices such as GEKS, CPD, or GK are applied to incomplete data $\mathcal{P}$ while ignoring a possible relationship between item-level missingness M_i and price-level sensitivity S_i, the resulting price levels are

[2] An example of a parametric measure of price-level sensitivity is the slope coefficient in a linear regression of $\ln p_i^r$ on $\ln P^r$. Unlike S_i, such a measure captures amplification but does not separately account for the directional component through a rank-based concordance measure.

generally biased (as illustrated in Section 2). However, the price-levels' ordinal ranking is less affected. This robustness of the price-level ranking can be exploited to construct a measure for quantifying systematic deviations from missing completely at random (MCAR) mechanisms.

To this end, items are ranked according to their price-level sensitivity S_i and, separately, according to their number of missing observations M_i. The pair-weighted rank concordance coefficient, $\tau_{b,w}(S_i, M_i)_{i\in\mathcal{N}}$, is then used to quantify the dependence between S_i and M_i. The pairwise weights w_{ij} reflect the combined importance of items i and j. A natural definition of these weights is

$$w_{ij} = \sum_{r\in\mathcal{R}} \left(w_i^r + w_j^r\right) \ .$$

Definition 3 *The **gap pattern coefficient** is defined as*[3]

$$G = \tau_{b,w}(S_i, M_i)_{i\in\mathcal{N}} \ . \tag{9}$$

A value of $G = 0$ indicates the absence of a systematic relationship between price-level sensitivity and missingness in the observed dataset. In this case, the pattern of data gaps is consistent with random missingness and supports the reliability of price levels computed using standard multilateral index methods. A deviation from zero, by contrast, indicates that price-level sensitivity and missingness are associated in the observed dataset. Since such an association is precisely the mechanism through which data gaps can induce bias, standard multilateral index methods are likely to produce biased results and should therefore be interpreted with caution.

6 Statistical Significance

The gap pattern coefficient is primarily a descriptive diagnostic that characterizes the relationship between price-level sensitivity and missingness in the dataset at hand. Nevertheless, if desired, the analysis can be complemented by the *gap pattern test*. This test assesses how unusual the observed gap pattern coefficient, G, is under a null hypothesis of no systematic relationship between price-level sensitivity and missingness.

Let $G^{(a)}$ $(a = 1, \ldots, A)$ denote the A permutation replicates of the gap pattern coefficient. Each replicate is obtained by randomly permuting the observed missingness measures M_i across items while holding the estimated

[3] When all items have similar relevance, Kendall's ordinary, unweighted rank concordance coefficient appears preferable.

sensitivity measures S_i and the pairwise weights w_{ij} fixed. The permuted values are then used to compute a synthetic gap pattern coefficient $G^{(a)}$. The collection $G^{(1)}, \ldots, G^{(A)}$ provides an empirical reference distribution of the gap pattern coefficient under the null hypothesis of no systematic association between price-level sensitivity and missingness.

Thus, the default implementation of the test procedure conditions on the observed distributions of both the sensitivity measures S_i and the missingness measures M_i. The null hypothesis states that the observed pairing of the S_i- and M_i-values is random, implying that there is no systematic association between price-level sensitivity and missingness. Under this null hypothesis, any permutation of the observed M_i-values across items is equally likely.

As an alternative, the test can be implemented using an MCAR benchmark resampling procedure. This approach may be useful when the number of available items is small and the observed distribution of the missingness measures M_i is therefore itself strongly affected by sampling variability. For each replication, synthetic missingness counts $M_i^{(a)}$ are generated under an MCAR mechanism that preserves the observed overall missing rate by fixing the total number of missing item-unit observations at its observed value and randomly redistributing these missing observations across the complete item-unit matrix. The resulting $M_i^{(a)}$-values are then combined with the fixed S_i-values to obtain a synthetic gap pattern coefficient $G^{(a)}$.

Consequently, the MCAR benchmark version of the test conditions on the estimated sensitivity measures S_i, the pairwise weights w_{ij}, the numbers of items and units, and the observed overall missing rate λ. The null hypothesis states that missingness is generated by an MCAR mechanism with missing rate λ and is therefore independent of the item-specific price-level sensitivities. Under this null hypothesis, any observed association between the S_i- and M_i-values arises solely from random variation in the missingness process.

For both implementations of the gap pattern test, the p-value is given by

$$p = \frac{K+1}{A+1} \ ,$$

where

$$K = \sum_{a=1}^{A} 1\big(\left|G^{(a)}\right| \geq |G|\big) \ .$$

The null hypothesis is rejected whenever $p < \alpha$, where α denotes the chosen significance level.

The applicability and reliability of the gap pattern coefficient, G, and the accompanying gap pattern test depend on several factors. At the highest level, they depend on the number of items, N, as well as on the variation in the missingness measures, M_i, and the variation and reliability of the estimated sensitivity measures, S_i. The reliability of the S_i-values, in turn, depends on the number of units in the sets $\mathcal{R}_i$, the variation in the observed prices, p_i^r, and the variation and reliability of the estimated price levels, P^r. Finally, the reliability of the estimated price levels depends on the number of observed items in the respective unit, on the number of observed item-unit combinations, and on the reliability of the applied multilateral index method. Consequently, the reliability of the gap pattern coefficient is shaped by the entire structure of the underlying dataset and not merely by the number of observed item-unit combinations. The following section addresses these issues by introducing a Monte Carlo (MC) simulation framework.

7 Monte Carlo Simulation

The simulation serves two purposes. First, it evaluates the ability of the gap pattern coefficient to detect and characterize systematic relationships between price-level sensitivity and missingness. Second, it examines the finite-sample size and power properties of the optional gap pattern test that is based on this coefficient.

Simulation Design

In real-world settings, the relationship between prices p_i^r and quantities q_i^r depends on both the context of the comparison (intertemporal or spatial) and the nature of the dataset. For example, survey data are likely to exhibit different price-quantity relationships than data collected through in-store visits or scanner records, while web-scraped data typically contain no quantity information at all. In view of this heterogeneity, the controlled data-generating process (DGP) of the MC simulation deliberately avoids imposing a specific relationship between prices and quantities. Instead, all pairwise weights w_{ij} ($i, j \in \mathcal{N}$) are assumed to be equal to unity and Kendall's unweighted rank concordance coefficient is applied.

The simulation assesses the distribution of the coefficients G under alternative N-R combinations, missingness mechanisms (MCAR, MNAR.pos, MNAR.neg), missing rates, and constructions of the reference price levels (GEKS, CPD, GK, and a benchmark case). The numbers of items and units each take the values

$$N, R \in \{10, 20, 60, 180\} .$$

while the overall missing rate varies from 5 percent to 85 percent in steps of 10 percentage points,

$$\lambda \in \{0.05, 0.15, \ldots, 0.85\} \,.$$

For every combination of these design parameters, $L = 2,000$ independent MC replications are performed.

Each replication proceeds in three stages: (1) a complete synthetic price tableau is generated from unit-specific synthetic price levels, (2) missing observations are introduced, and (3) the gap pattern coefficient and the corresponding permutation-test p-value are computed. If the incomplete tableau becomes too sparse, the replication yields no result and is treated as missing for that parameter configuration. The collection of valid p-values across replications forms the basis for the empirical assessment of size and power reported in the next section. The Appendix describes the three stages of the simulation in detail.

For the gap pattern coefficient, G, the key question is whether its value reflects the underlying missingness mechanism. Under MCAR, the coefficient should be close to zero because price-level sensitivity and missingness are independent by construction. Under MNAR.pos and MNAR.neg, the coefficient should move away from zero and correctly reflect the direction of the underlying relationship. The gap pattern test builds on this coefficient and therefore inherits its strengths and limitations.

The gap pattern test (permutation approach) examines the null hypothesis that an item's price-level sensitivity S_i is not related to the number of missing observations M_i. Inference is based on a permutation distribution of G, yielding one p-value. This value is compared with the chosen significance level, which is set to $\alpha = 5\%$. When $p < \alpha$, the null hypothesis is rejected. Each application of the gap pattern test yields one permutation-based p-value. The share of replications leading to a rejection is the rejection rate.

MCAR Results

The MC simulation results relating to the MCAR missingness mechanism provide evidence on the size properties of the test. Rejections of the null hypothesis correspond to type I errors. Since the p-values should be approximately uniformly distributed on $[0, 1]$, the rejection rate should be close to the chosen significance level a.

Monte Carlo results presented in Figure 1 show that rejection rates are generally close to the nominal significance level of 5% across a wide range of parameter configurations, including different numbers of items, reference units, and overall missing rates. However, there are two peculiarities that deserve further discussion.

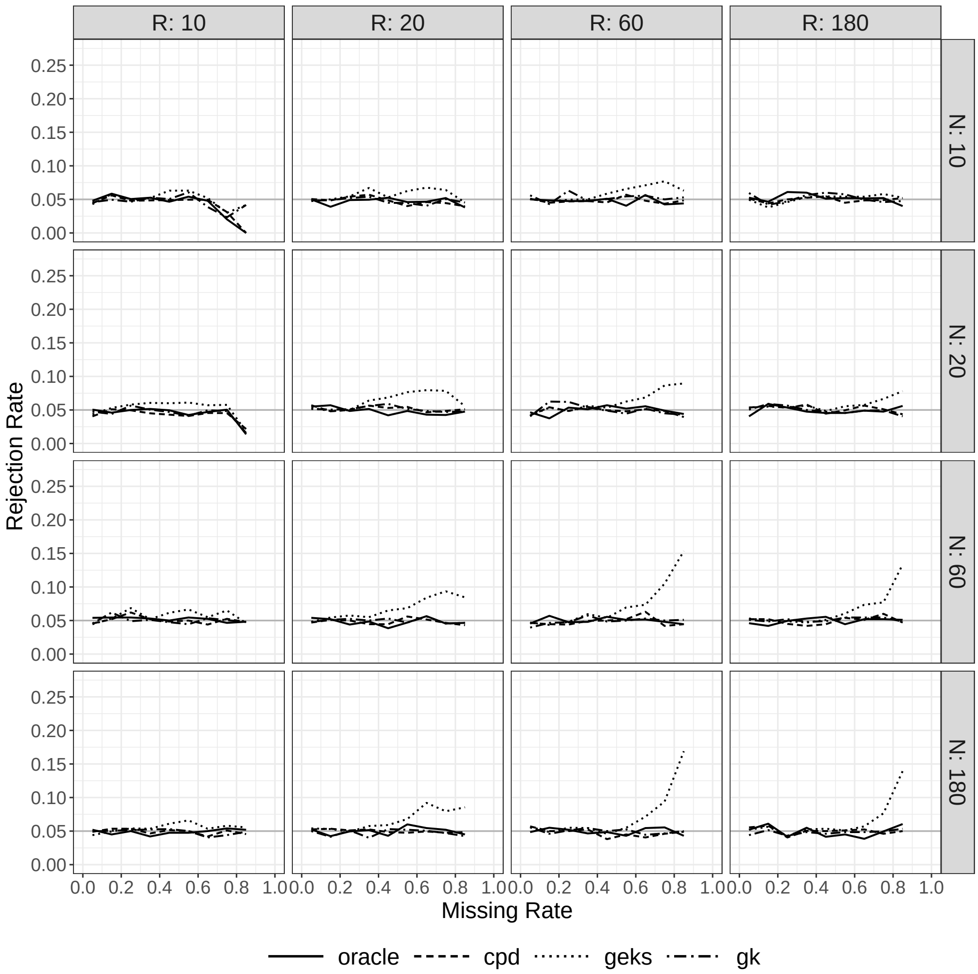

Figure 1: Monte Carlo rejection rates under MCAR.

The two upper left panels of Figure 1 refer to cases with only $R = 10$ units and $N \leq 20$ items. In these cases, high missing rates lead to undersized test results. This observation is likely related to the small number of units in $\mathcal{R}_i$. If this number is smaller than the minimum threshold, Kendall's rank concordance coefficient, $\tau_b \left(\ln p_i^r, \ln P^r\right)_{r\in\mathcal{R}_i}$, and therefore the price-level sensitivity S_i, cannot be computed. As a result, fewer items can be used for the computation of G, reducing its reliability. In addition, the probability that an item's S_i-value is not available increases with the item's missingness measure M_i. This reduces the variation in the M_i-values and attenuates the absolute value of G toward zero, reducing the rejection rates below the significance level.

Furthermore, Figure 1 indicates that sensitivities S_i derived from GEKS-based reference price levels tend to render the test oversized, whereas sensitivities based on CPD- or GK-based reference price levels maintain approximately correct test size even for missing rates as high as 85%. Section 8 investigates this phenomenon in greater detail. It shows that the estimated sensitivities S_i of items with low missingness M_i tend to be inflated relative to those of items with high missingness and that this effect is stronger for GEKS than for CPD and GK. This mechanism gives rise to a spurious negative relationship between estimated price-level sensitivity and missingness even when no such relationship exists in the underlying data-generating process. This problem becomes more pronounced as missing rates increase.

MNAR.pos Results

Under MNAR.pos, price-level sensitivity, S_i, and missingness, M_i, are positively related. Thus, the gap pattern coefficient is expected to be positive. The simulation therefore examines whether the gap pattern coefficient correctly identifies this relationship and whether the corresponding test is able to reject the null hypothesis. Rejections reflect the power of the gap pattern test, while non-rejections correspond to type II errors.

Figure 2 reports the rejection rates. They reveal three clear patterns that warrant further discussion.

First, the test becomes increasingly effective at detecting systematic relationships between price-level sensitivity and missingness as the amount of information increases. The test relies on estimating product-specific price-level sensitivities. When only a small number of items or units is available, these sensitivities are estimated imprecisely and only few informative comparisons contribute to the gap pattern coefficient. Consequently, the power of the test is limited for designs with few items and units, even though the empirical size remains close to the nominal level under MCAR. With sufficiently large numbers of items and units and moderate missing rates, rejection rates increase and approach unity.

Second, rejection rates are generally highest at moderate levels of missingness. When the missing rate is too low, the relationship between missingness and sensitivity is relatively weak and therefore difficult to detect. As the missing rate increases, the signal becomes stronger and rejection rates rise. Beyond a certain point, further increases in missingness reduce the number of observed price comparisons available for estimating the product-specific sensitivities S_i. Consequently, the gap pattern coefficient becomes noisier

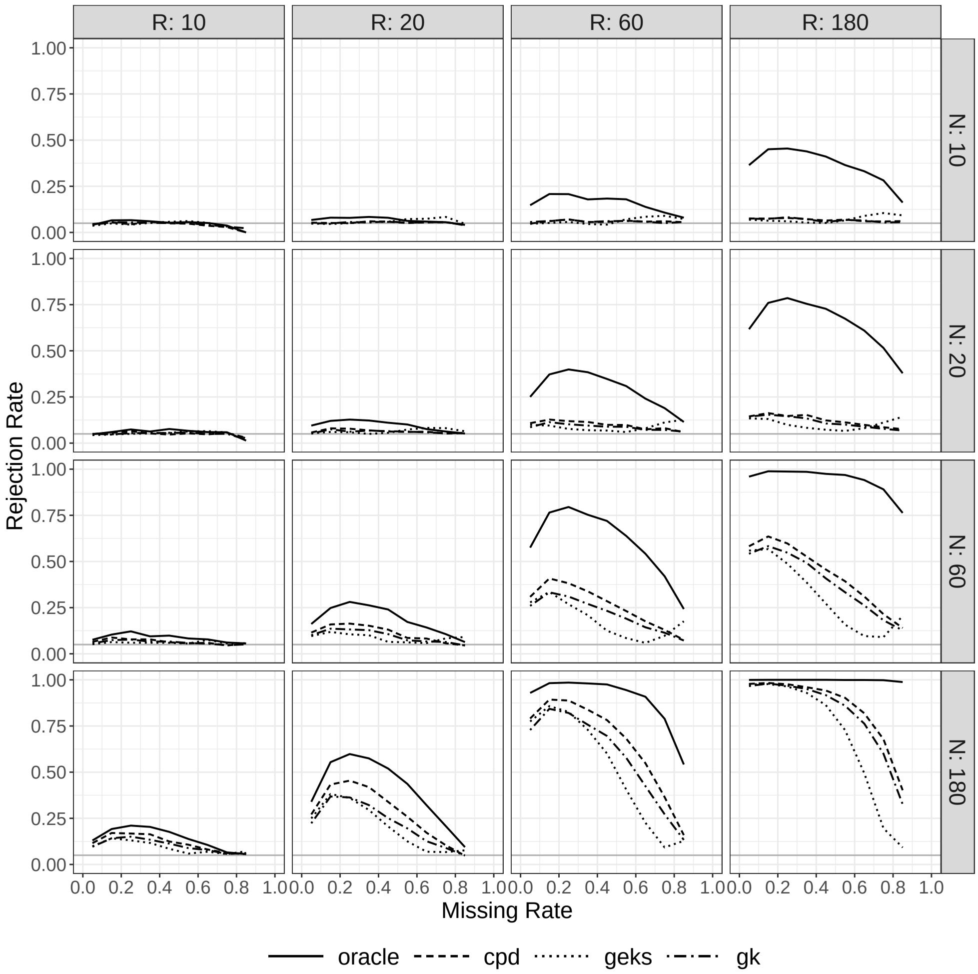

Figure 2: Monte Carlo rejection rates under MNAR.pos.

and rejection rates decline.

Third, the oracle-based version of the test consistently achieves the highest rejection rates. Since the oracle price levels are observed without estimation error, the corresponding sensitivity measures are estimated most accurately. However, the oracle-based procedure serves solely as a benchmark because the true price levels are unavailable in empirical applications. Among the feasible approaches, the CPD-based version generally performs best, followed by the GK-based version, while the GEKS-based version tends to exhibit the lowest rejection rates. Part of this performance gap reflects the spurious negative relationship between the estimated sensitivity measures S_i and the missingness measures M_i that became visible under the MCAR mechanism. Under the present MNAR.pos mechanism, this artificial negative component counteracts the true positive relationship between S_i and M_i and thereby reduces the power of the GEKS-based test.

MNAR.neg Results

Under MNAR.neg, the coefficient is expected to be negative because items with higher price-level sensitivity tend to exhibit fewer missing observations. The simulation therefore evaluates whether the coefficient correctly captures this inverse relationship and whether the corresponding test possesses sufficient power.

Figure 3 reports the results. The general patterns are similar to those observed under the positive MNAR mechanism. Rejection rates increase substantially with both the number of items, N, and the number of units, R, indicating that the gap pattern test is able to detect systematic negative relationships between sensitivity and missingness with increasing reliability as the amount of available information grows. As before, rejection rates tend to be highest at intermediate missing rates.

The oracle-based version of the test again provides the natural benchmark. Among the feasible approaches, the GEKS-based version yields higher rejection rates than the CPD- and GK-based versions. This apparent advantage of GEKS should be interpreted with caution. Because the oracle price levels are free of estimation error, the resulting sensitivity measures are not affected by distortions arising from the estimation of reference price levels. In contrast, GEKS-based price levels tend to induce a spurious negative relationship between the estimated sensitivity measures S_i and the missingness measures M_i (see Section 8). Under the present MNAR.neg mechanism, where the true relationship is also negative, this artificial component reinforces the genuine signal and thereby increases rejection rates. Consequently, the superior performance of the GEKS-based test in this setting does not necessarily reflect a more accurate recovery of the underlying gap pattern.

Further Results

A comparison of Figures 2 and 3 reveals a striking asymmetry. The oracle-based curves under MNAR.pos and MNAR.neg are very similar, whereas for each of the feasible price-level methods (GEKS, CPD, and GK) the rejection-rate curve under MNAR.neg consistently lies above the corresponding curve under MNAR.pos. This difference is consistent with the spurious negative relationship between estimated sensitivity and missingness that arises when estimated rather than oracle price levels are used. The effect is particularly pronounced for GEKS but is also visible, albeit to a lesser extent, for CPD

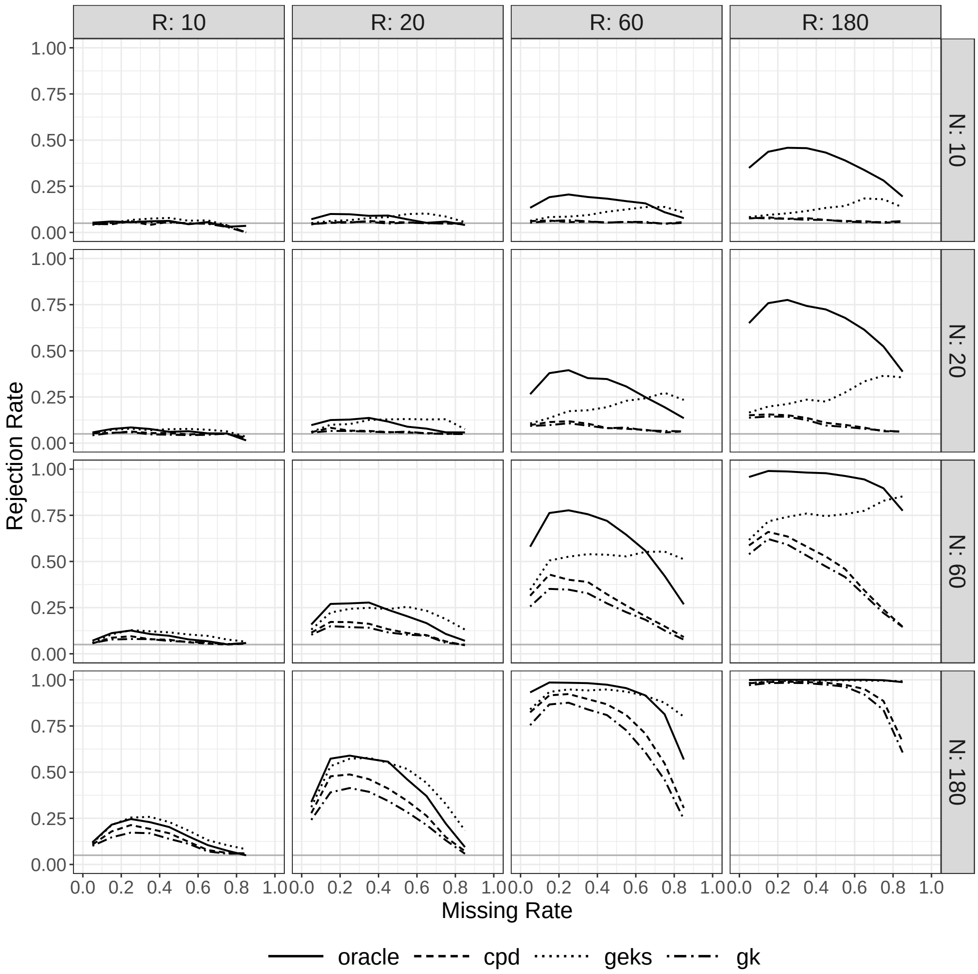

Figure 3: Monte Carlo rejection rates under MNAR.neg.

and GK (see Section 8).

Overall, the simulation study demonstrates that the proposed gap pattern test has well-controlled size under MCAR and substantial power under both MNAR mechanisms, provided that the dataset contains sufficiently many items and reference units. Among the feasible approaches, CPD- and GK-based reference price levels generally appear preferable to GEKS-based reference price levels, as they are less prone to induce a spurious negative association between estimated price-level sensitivity and missingness and therefore provide more balanced size and power properties.

Beyond evaluating the proposed gap pattern methodology, the MC simulations also reveal systematic differences between the multilateral estimation methods. These results indicate that GEKS, CPD, and GK differ in their sensitivity to incomplete data. The following section investigates this issue by comparing the influence of individual items on the estimated price levels obtained from the three methods.

8 Robustness of Multilateral Price-Level Estimators under Incomplete Data

Items with relatively few missing observations exert a stronger influence on the estimated price-level vector than items with many missing observations. In the following simulation experiment, the systematic difference in influence between well-observed items and poorly observed items (hereafter referred to as *influence asymmetry*) turns out to be more pronounced for GEKS(-Jevons) than for CPD and GK. This finding implies that GEKS price-level estimates are more sensitive to the loss of item information than those obtained from CPD or GK.

The simulation experiment starts from a complete price tableau generated according to the linear data-generating process

$$\ln p_i^r = \alpha + \beta_i \ln P^r + \varepsilon_i^r \ ,$$

where P^r denotes the true unit price level, β_i determines the price-level sensitivity of item i, and ε_i^r is an idiosyncratic disturbance term. The parameters were set to $N = 60$ items, $R = 60$ units, $\alpha = 0.5$, $\beta_i \sim N(1, 0.5^2)$, and $\varepsilon_i^r \sim N(0, 0.3^2)$. Missing observations were then introduced by an MCAR mechanism with missing rates ranging from 0.15 to 0.85 in increments of 0.10.

For each missing-rate scenario and for each index method (GEKS, CPD, and GK), price levels were first estimated using the incomplete dataset. Subsequently, a leave-one-item influence analysis was carried out. Specifically, for each item i, the corresponding observations were removed completely from the dataset, the price levels were re-estimated, and the resulting change in the price-level vector was recorded.

Influence can be assessed from different perspectives. The first measure quantifies the effect of removing an item on the numerical values of the estimated price levels. The second measure evaluates how strongly the removal of an item alters the ranking of the estimated price levels.

First Measure: For each item i, the influence measure

$$\mathrm{sd}_i^m = \mathrm{sd}\left(\ln P^{r,m} - \ln P_{-i}^{r,m}\right)_{r\in\mathcal{R}}$$

was computed, where $\ln P^{r,m}$ denotes the estimated log price level of unit r obtained with method $m \in$ (GEKS, CPD, GK) and $\ln P_{-i}^{r,m}$ denotes the corresponding estimate obtained after removing item i. Larger values of sd_i^m indicate a stronger influence of item i on the estimated price levels.

For each missing-rate scenario, Kendall's rank concordance coefficient between the influence measure sd_i^m and the item-specific missingness level M_i was computed: $\tau_b\left(\mathrm{sd}_i^m, M_i\right)_{i\in\mathcal{N}}$. These coefficients were averaged across Monte Carlo replications.

Figure 4 reports the results. The coefficient $\tau_b\left(\mathrm{sd}_i^m, M_i\right)$ is negative for all three methods, indicating that items with fewer missing observations generally exert a larger influence on the estimated price levels. However, the relationship is substantially stronger for GEKS than for CPD and GK. Moreover, the difference becomes increasingly pronounced as the overall missing rate rises. At a missing rate of 0.85, the GEKS coefficient exceeds -0.5, whereas the corresponding CPD and GK coefficients remain just below -0.3. This suggests that the influence of well-observed items on the estimated price levels is larger under GEKS than under CPD or GK.

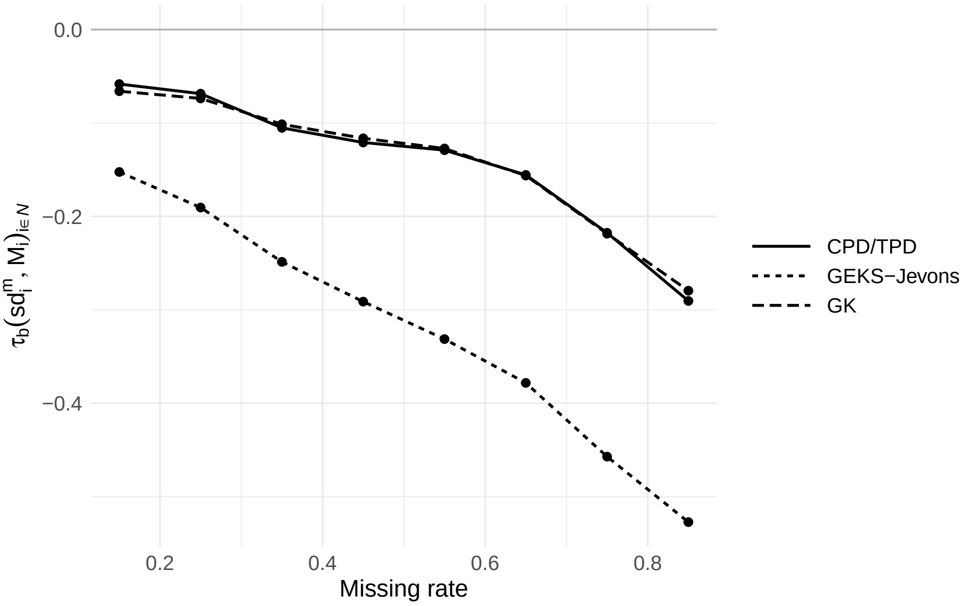

Figure 4: Comparison of the relationship between item missingness and item influence on the estimated price levels P^r across alternative index methods, measured by Kendall's rank concordance coefficient $\tau_b\left(\mathrm{sd}_i^m, M_i\right)_{i\in\mathcal{N}}$.

Second Measure: The influence of removing item i was computed by Kendall's rank concordance coefficient between the original and leave-one-item-out price-level vectors:

$$\tau_{b,i}^m = \tau_b\left(\ln P^{r,m}, \ln P_{-i}^{r,m}\right)_{r\in\mathcal{R}}.$$

Since these coefficients are expected to be close to 1, the ranking influence of item i was then measured by $(1 - \tau^m_{b,i})$. Again, larger values indicate a stronger influence of item i on the estimated price-level ranking.

For each missing-rate scenario, Kendall's rank concordance coefficient $\tau_b\left((1 - \tau^m_{b,i}), M_i\right)_{i \in \mathcal{N}}$ was computed and averaged across Monte Carlo replications. Figure **??** shows the results. Again, the relationship is markedly more negative for GEKS than for CPD and GK confirming the conjecture that GEKS assigns disproportionately greater influence to items with low missingness.

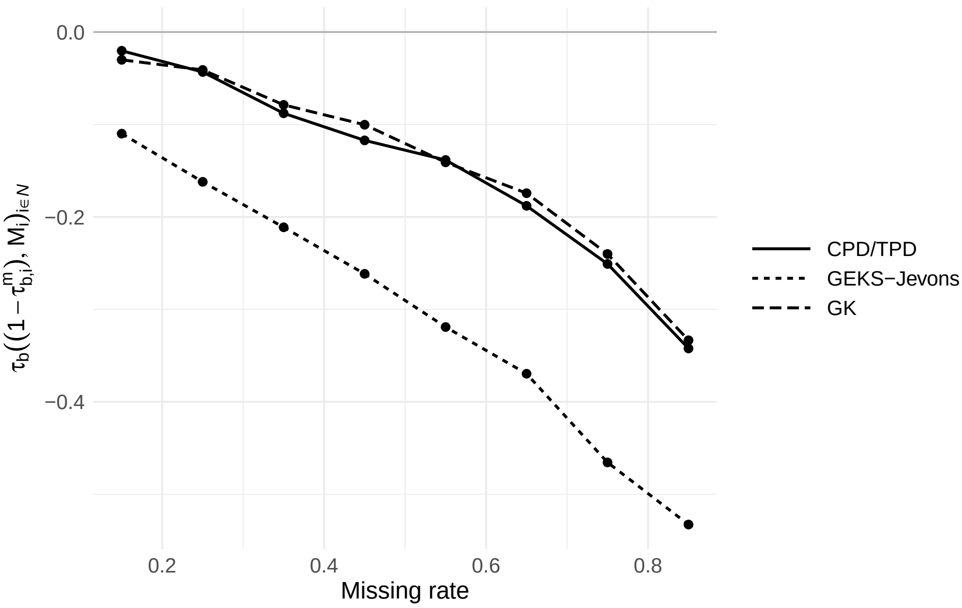

Figure 5: Comparison of the relationship between item missingness and item influence on the estimated price levels P^r across alternative index methods, measured by Kendall's rank concordance coefficient $\tau_b\left(\left(1 - \tau^m_{b,i}\right), M_i\right)_{i \in \mathcal{N}}$.

The two influence measures quantify different aspects of the effect of item removal—changes in the numerical values and changes in the ranking of the estimated price levels. Despite their conceptual differences, both measures lead to the same conclusion: GEKS exhibits substantially stronger influence asymmetry than CPD and GK, implying that the estimated GEKS price levels depend more heavily on well-observed items and are more sensitive to the loss of information from individual items. From a practical perspective, this suggests that data completeness deserves particular attention when GEKS is the preferred index method and that diagnostic tools for assessing missing-data patterns are especially valuable in conjunction with GEKS-based analyses.

The degree of influence asymmetry between GEKS on the one hand and CPD and GK on the other hand is also relevant for the gap pattern coefficient because the prices of influential items tend to be more closely aligned with the estimated price levels than the prices of less influential items. This closer alignment may result in a larger sensitivity measure, S_i, thereby potentially inducing a spurious negative relationship between item-level missingness, M_i, and estimated price-level sensitivity, S_i, even under an MCAR mechanism. Accordingly, under an MCAR mechanism, a stronger influence asymmetry should induce a more pronounced negative bias in the gap pattern coefficient and, consequently, a higher rejection rate of the null hypothesis. Under MNAR.neg, the same mechanism should reinforce the true negative relationship between missingness and price-level sensitivity, thereby increasing the power of the gap pattern test. Conversely, under MNAR.pos, the spurious negative relationship should partially offset the true positive relationship, resulting in a lower rejection rate. The Monte Carlo simulations reported in Figures 1 to 3 of Section 7 are consistent with this interpretation.

9 Concluding Remarks

This paper has argued that the ability of multilateral price index methods to mitigate the effects of missing price observations depends critically on an assumption that is usually left implicit: item-level missingness must be independent of the sensitivity of item prices with respect to overall unit price levels. When this condition is violated, systematic bias may arise in the estimated price levels obtained from methods such as GEKS, CPD, and GK.

To diagnose such situations, the paper introduced the gap pattern coefficient, a descriptive measure that quantifies the association between item-level missingness and price-level sensitivity. A value close to zero indicates that missingness and sensitivity are largely unrelated, whereas larger positive or negative values indicate increasingly systematic gap patterns and therefore a greater potential for bias in multilateral price-level estimates. The sign of the coefficient additionally reveals whether missingness is concentrated among items with relatively high or relatively low price-level sensitivity.

The paper also proposed the gap pattern test, a permutation-based procedure for assessing the statistical significance of the gap pattern coefficient. For practical price measurement, however, statistical significance is often of secondary importance. The primary question is whether the observed dataset exhibits a systematic association between price-level sensitivity and missingness and therefore contains a gap pattern that may distort the estimated price levels. This information is provided directly by the gap pattern coefficient.

The Monte Carlo simulations provide evidence on both the diagnostic performance of the gap pattern coefficient and the finite-sample properties of the accompanying gap pattern test. The results show that the coefficient is capable of detecting both positive and negative relationships between price-level sensitivity and missingness and that its reliability improves substantially as the numbers of items and units increase. The simulations further demonstrate that GEKS price-level estimates are intrinsically more sensitive to missing observations than those obtained from CPD and GK because they assign relatively greater influence to well-observed items. As a consequence, GEKS-based reference price levels are more prone to inducing a spurious negative association between estimated price-level sensitivity and missingness, particularly at higher missing rates.

More generally, the gap pattern coefficient provides a practical diagnostic tool for evaluating whether the missing-data structure of a dataset is likely to compromise the reliability of multilateral price comparisons. By making the relationship between missingness and price-level sensitivity explicit, the coefficient offers a simple way to assess an important source of bias that is otherwise difficult to detect.

If an association between price-level sensitivity and missingness is detected, the resulting price-level estimates should be interpreted with caution. Alternative multilateral index methods that explicitly account for such dependence structures may then be preferable. For example, Auer and Weinand (2025, 2026) propose a nonlinear extension of the CPD regression that accommodates systematic relationships between missingness and price-level sensitivity. A simpler alternative is to use the gap pattern coefficient to partition items into groups with similar price-level sensitivities and then apply the chosen multilateral index method separately within each group. Provided that the resulting group-level price estimates contain few or no missing observations, or that the sensitivity of the group-level price estimates with respect to the overall price levels is uncorrelated with the number of missing observations within the groups, the same multilateral index method can subsequently be used to aggregate the group-level results into an overall estimate. In this way, the gap pattern coefficient can serve not only as a diagnostic tool but also as a guide for constructing more robust multilateral price-level estimates in the presence of non-random missingness.

Appendix: Three Stages of the MC Simulation

This appendix spells out the three stages of the MC simulation.

Stage 1: DGP

The starting point of the DGP are synthetic unit-specific price levels $\bar{P}^r$ for $r \in \mathcal{R}$. They serve as latent oracle price levels and are drawn independently from the following log-normal distribution:

$$\ln \bar{P}^r \sim N\left(-\frac{\sigma_P^2}{2}, \sigma_P^2\right) , \qquad r \in \mathcal{R} .$$

To ensure that $E(\bar{P}^r) = 1$, the normal distribution's variance is set to $\sigma_P^2 = 0.2$.

For generating item-level prices, a linear DGP is used where the log price levels $\ln \bar{P}^r$ enter directly into the generation of item-level prices. The log price of item i in unit r is generated according to

$$\ln p_i^r = \alpha_i + \beta_i \ln \bar{P}^r + \varepsilon_i^r , \tag{10}$$

with the item-specific intercepts, α_i, and price-level elasticities, β_i, stochastically generated by

$$\alpha_i \sim N(\alpha_{\text{mean}}, \alpha_{\text{sd}}^2) \qquad \text{and} \qquad \beta_i \sim N(\beta_{\text{mean}}, \beta_{\text{sd}}^2) . \tag{11}$$

The item-specific idiosyncratic error term in Eq. (10), ε_i^r, is determined by

$$\varepsilon_i^r \sim N(0, \sigma_{\varepsilon,i}^2), \qquad r = 1, \ldots, R , \tag{12}$$

where

$$\sigma_{\varepsilon,i} = \sigma_\varepsilon \, u_i \tag{13}$$

and

$$u_i \sim U(m_{\text{min}}, m_{\text{max}}) . \tag{14}$$

The noise factor σ_ε in Eq. (13) is constant across units and items, while the parameters m_{min} and m_{max} denote the lower and upper bounds of the item-specific multiplicative noise factor u_i. Thus, the variance $\sigma_{\varepsilon,i}^2$ of the error term, ε_i^r, may differ across items, but is constant across units for a given item.

The observed price is then obtained as

$$p_i^r = \exp(\ln p_i^r) ,$$

with a negligible lower bound imposed in the code to avoid numerical problems. This specification implies that items differ in the strength with which their prices respond to the latent unit-specific price level, $\bar{P}^r$.

The following parameter values are used in the linear DGP: $\alpha_{\text{mean}} = 0$, $\alpha_{\text{sd}} = 0.5$, $\beta_{\text{mean}} = 1$, $\beta_{\text{sd}} = 0.5$, $\sigma_\varepsilon = 0.6$, $m_{\min} = 0.3$, and $m_{\max} = 2.3$.

Stage 2: Missingness Mechanisms

Starting from the complete synthetic tableau of $N \cdot R$ prices, missing observations are introduced in a second stage. Three missingness mechanisms are considered: missing completely at random (MCAR), missing not at random with positive dependence on price-level sensitivity (MNAR.pos), and missing not at random with negative dependence on price-level sensitivity (MNAR.neg).

Let Z_i^r indicate whether the price of item i in unit r is observed. Under MCAR, all observations are retained independently with common probability $(1 - \lambda)$, where $\lambda \in (0, 1)$ denotes the target missing rate.

Under the two MNAR mechanisms, the probability that an observation is missing depends on the item-specific sensitivity parameter β_i. More precisely, the missing probability is generated according to

$$\pi(\beta_i) = \frac{1}{1 + \exp[-(a + b\beta_i)]} ,$$

where $b > 0$ under MNAR.pos and $b < 0$ under MNAR.neg. This specification is item-specific but unit-invariant. Conditional on β_i, the same observation probability applies to item i in every unit. The intercept a is chosen such that the resulting average missing rate equals the target value λ. Choosing the intercept a in this way ensures that all three mechanisms generate the same average missing rate, thereby facilitating a direct comparison of their effects.

Observation indicators are then generated according to

$$\Pr(Z_i^r = 1) = 1 - \pi(\beta_i).$$

Under MNAR.pos, items with higher price-level sensitivity are more likely to be missing, whereas under MNAR.neg, items with lower price-level sensitivity are more likely to be missing.

Stage 3: Gap Pattern Coefficient and Test

The gap pattern coefficient requires, in addition to the incomplete price tableau, a reference vector of price levels, P^r, against which item-specific sensitivities of the prices p_i^r are measured. For each of the three incomplete tableaus (MCAR, MNAR.pos, and MNAR.neg), four alternative reference vectors are considered: (1) the oracle price level vector given by the latent synthetic price levels ($P^r = \bar{P}^r$); (2) the GEKS-Jevons-based price levels; (3) the CPD-based price levels; and (4) the GK-based price levels. This setup yields twelve cases. For each case, the gap pattern coefficient is computed and the associated test is conducted. This approach provides insight into whether the performance of the procedure depends on whether the reference vector of price levels is observed without error (oracle) or is instead estimated from incomplete data.

References

Auer, L. von, S. Weinand (2025). "The Country-Product-Dummy Method With Product-Specific Spatial Price Variation". *Review of Income and Wealth*, 71 (1). `https://doi.org/10.1111/roiw.70005`

Auer, L. von, S. Weinand (2026), *Multilateral Index Approaches In the Presence of Product-Specific Price Trends and Data Gaps*, Research Papers in Economics No. 08/26, Universität Trier. `https://www.uni-trier.de/index.php?id=51049`

Éltető Ö., P. Köves (1964) "On an index computation problem in international comparisons (in Hungarian)," *Statiztikai Szemle*, 507-518.

Geary, R. G. (1958) "A Note on Comparisons of Exchange Rates and Purchasing Power Between Countries," *Journal of the Royal Statistical Society, Series A*, 121, 97–99.

Gini, C. (1931) "On the Circular Test of Index Numbers," Metron, 9, 3-24.

International Comparison Program (2021), *A Guide to the Compilation of Subnational Purchasing Power Parities (PPPs).* `https://thedocs.worldbank.org/en/doc/5064f2288436664bc8f9811c8a5b8c55-0050022021/original/Guide-Subnational-PPPs.pdf`

Kendall, M. G. (1945), "The Treatment of Ties in Ranking Problems," *Biometrika*, 33(3), 239-251.

Khamis, S. H. (1970) "Properties and conditions for the existence of a new type of index number," *Sankhya B*, 32, 81-98.

—— (1972) "A new system of index numbers for national and international purposes," *Journal of the Royal Statistical Society, Series A*, 135, 96-121.

Szulc, B. (1964) "Index numbers of multilateral regional comparisons (in Polish)," *Przeglad Statystyczny*, 3, 239-254.