

Checking Whether a Recorded Value Exists in the Original Record

— Between a population matching at **55--65%** and one at **1--8%**, the median difference is **0.000** against **18.05**, more than **70** times apart — scanning across shifted times, one shows a spike and the other is flat everywhere —

[Paper 428]

Yuuki YAMAGISHI (山岸 勇輝)

Yamagishi Complexity Torus Lab, Japan

September 2026

— *From thinking to syncing.* —

Abstract

A secondary record can be checked against the original. This paper shows that such a check separates “the clock is off” from “the value is not there”—and that what separates them is the shape of a scan across shifted times. No new theorem or law is claimed. First, the match rate divides into **55--65%** and **1--8%**, more than **6.9** times apart (Section 2). Second, the median difference divides into **0.000--0.05** and **3.5--18.05**, more than **70** times apart (Section 2). Third, this is the core of the paper. Scanning across shifted times, one shows a spike at zero standing **3.73** times its neighbours, and the other is flat wherever it is shifted (Section 3). Fourth, the flat one’s maximum of **0.09** does not reach even the neighbours of the spike at **0.15** (Section 3). Fifth, the discriminant is height, not ratio: by ratio the flat one comes out larger (Section 3). Sixth, the separator is whether the time differs or the value is absent (Section 4).

Scope of this paper (scope note): No new theorem or law is claimed—record linkage, verification of clock synchronisation, and delay estimation by cross-correlation are all standard. Nobody and nothing is named—no person, organisation, product or case is the subject, and none appears. Only the procedure of checking a secondary record against an original is treated. No wrongdoing is discussed—what causes a mismatch cannot be settled by this procedure. Rounding, a difference of units, a value from another feed, a transcription error—many explanations are possible. Only whether the values match is measured. The original is not warranted correct—gaps or errors in the original lower the match rate too. This paper has no ground for treating either side as true. No threshold is set—what match rate counts as passing is decided by the use. No particular field is discussed—the figures come from checking a time series, but the subject is not. Relation to earlier papers: Paper 426 showed that look-ahead survives a split sample and shows only under a shift in time—the same operation, shifting the clock, separates something else here. Paper 420 showed that a gap creating no bias still moves the sign—there absence from the record mattered; here disagreement within it is measured. Paper 300 showed that common root and distinct root can be decided—“the clock is off” and “the value is nowhere” are distinct roots by that criterion. Paper 421 showed that what looks like leading is not leading—there too the shape of a lag scan gave the answer. What is added is setting the match rates and differences of two populations side by side, separating spike from flat by a shift scan, showing that the discriminant is height rather than ratio, putting a genuine and a fabricated record through the same procedure synthetically, and placing the separator at “time or absence”.

1. Notation and the Question

Setting: The original record is a sequence of values by time, $\{x_t\}$. The secondary record is a set of pairs (τ_i, p_i) . For a tolerance ε and a shift s , define the match rate

$$H(s) = \frac{1}{n} \# \{i : |x_{\tau_i+s} - p_i| \leq \varepsilon\}$$

Observation 1.1

Secondary records are usually not checked against originals. Consistency is assumed—because checking needs the original, and one often does not have it.

Observation 1.2

With the original in hand, the check is one line. Look up the time and compare the value—no convention need be assumed.

Question: When they do not match, is the clock off or is the value not there.

2. Both the Match Rate and the Difference Divide by Orders

Computation 2.1: Check two populations at the same tolerance.

Population	Count	Match rate	Median difference
First	2087	0.55--0.65	0.000--0.05
Second	3403	0.01--0.08	3.5--18.05

Observation 2.2

The match rates are more than **6.9** times apart.0.55 against 0.08—and that is the comparison least favourable to the first.

Observation 2.3

The differences are more than **70** times apart.3.5 against 0.05—not a question of whether they match but of orders of magnitude.

Observation 2.4

In the first population the lower end of the difference is exactly **0.000**.Exactly matching pairs exist—direct evidence that the checking machinery works.

Observation 2.5

And its match rate is not 100% either.55--65 per cent—the original has gaps too, so unmatched pairs are expected. A low match rate alone therefore says nothing.

3. Under a Shift Scan, the Shapes Divide

Computation 3.1: Measure the match rate while shifting the time backwards and forwards.

Population	Unshifted	Neighbouring shifts
First	0.56--0.63	0.05--0.15
Second	0.01--0.09 (everywhere)	the same

The first shows a spike standing **3.73** times its neighbours. The second is flat wherever it is shifted.

Observation 3.2

A spike proves the clock is right. Shift it and the rate falls—had the clock been off, the spike would stand elsewhere.

Observation 3.3

Flatness means the clock cannot explain it. No shift raises the rate—which closes the explanation “fix the clock and they will match”.

Observation 3.4

The discriminant is height, not ratio. Measuring the flat one’s variation as max over min gives 9.00, larger than the spike’s 3.73—ratios of small numbers are unstable; what should be compared is the spike’s height 0.56 against the flat one’s maximum 0.09.

Observation 3.5

And that 0.09 does not reach even 0.15, the spike’s neighbours. The best of one is below the worst of the other—they do not overlap, so they separate.

Computation 3.6: Put two records through the same procedure synthetically.

Shift	Copied from the original	Values not in the original
−40	0.018	0.020
−10	0.067	0.074
0	0.957	0.001
+10	0.058	0.062
+40	0.016	0.022

Observation 3.7

The same shape appears synthetically. One spikes to 0.957 at zero; the other never exceeds 0.074—confirming that the procedure separates them.

4. The Separator

The separator: does the time differ, or is the value absent.

Observation 4.1

A difference of time is repaired by shifting. The spike stands somewhere else—and once corrected, a high match rate appears.

Observation 4.2

Absence is not repaired by shifting. No position produces a spike—that value is at no time at all.

Observation 4.3

These are distinct roots by the criterion of Paper 300. Both are called “it does not match”—one a problem of correspondence, one a problem of existence.

Observation 4.4

Why it does not match is not settled by this procedure. Rounding, a difference of units, a value from another feed, a transcription error—many explanations are possible. What is measured is only whether they match.

Observation 4.5

Nor is there ground here for treating the original as true. Gaps or errors on that side lower the rate too—which is part of why one population does not reach 100%.

Observation 4.6

To be honest—a low match rate alone says nothing. Gaps in the original lower it as well—what does the work is the shape of the shift scan: spike or flat.

Observation 4.7

One last thing—records that can be checked are worth checking. It assumes no convention—and for the part that matches, everything computed downstream can be trusted.

5. Roll Call and Closing

1. Match rates more than 6.9 times apart (Observation 2.2): **0.55 against 0.08.**
2. Differences more than 70 times apart (Observation 2.3): **3.5 against 0.05.**
3. The lower end is exactly 0.000 (Observation 2.4): **exact matches exist.**
4. Not 100% either (Observation 2.5): **the original has gaps.**
5. The spike stands 3.73 times its neighbours (Computation 3.1): **the clock is right.**
6. The other is flat everywhere (Computation 3.1): **the clock cannot explain it.**
7. By ratio the flat one gives 9.00 (Observation 3.4): **the discriminant is height.**
8. 0.09 is below 0.15 (Observation 3.5): **they do not overlap.**
9. Synthetically 0.957 against 0.074 (Computation 3.6): **the procedure separates.**

10. The separator is correspondence or existence (Observation 4.3): distinct roots by Paper 300.

Closing: Secondary records are usually not checked against originals—consistency is assumed, because checking needs the original and one often does not have it. But with the original in hand the check is one line: look up the time and compare the value, assuming no convention at all. Checked, they divide by orders of magnitude—the match rates are 0.55 against 0.08, more than **6.9** times apart, and the median differences are 3.5 against 0.05, more than **70**. In one population the lower end of the difference is exactly **0.000**: exactly matching pairs exist, which is direct evidence that the machinery works. Its match rate is nevertheless 55–65 per cent rather than 100, because the original has gaps too—so a low match rate alone says nothing. What separates them is the shape of a scan across shifted times—one shows a spike where it is unshifted, standing **3.73** times its neighbours; since shifting lowers it, the clock is right. The other stays at 0.01–0.09 wherever it is shifted; since no position raises it, the explanation “fix the clock and they will match” is closed. The discriminant is height rather than ratio: measuring the flat one’s variation as max over min gives 9.00, larger than the spike’s 3.73, because ratios of small numbers are unstable. What should be compared is the spike’s height **0.56** against the flat one’s maximum **0.09**. And that 0.09 does not reach even 0.15, the spike’s neighbours—the best of one lies below the worst of the other, so they do not overlap. Synthetically the same shape appears: a record copied from the original spikes to **0.957** at zero, while one built from values not in the original never exceeds **0.074**. One thing separates them—does the time differ, or is the value absent. A difference of time is repaired by shifting, and the spike stands somewhere else; absence is not repaired by shifting, and no position produces a spike, because that value is at no time at all. They are distinct roots by the criterion of Paper 300, one a problem of correspondence and one of existence. To be honest—why it does not match is not settled by this procedure. Rounding, a difference of units, a value from another feed, a transcription error—many explanations are possible, and only whether they match is measured. Nor is there ground here for treating the original as true; gaps or errors on that side lower the rate as well, which is part of why one population does not reach 100%. One last thing—records that can be checked are worth checking. It assumes no convention, and for the part that matches, everything computed downstream can be trusted.

References

1. Fellegi, I. P., & Sunter, A. B. (1969). A Theory for Record Linkage. *Journal of the American Statistical Association*.
2. Christen, P. (2012). *Data Matching*. Berlin: Springer.
3. Redman, T. C. (1998). The Impact of Poor Data Quality on the Typical Enterprise. *Communications of the ACM*.
4. Knuth, D. E. (1998). *The Art of Computer Programming: Sorting and Searching* (2nd ed.). Reading: Addison-Wesley.
5. Yamagishi, Y. (2026). *The Three Hundredth Census: Whether Two Things Share a Root Is Decidable* [Paper 300]. Zenodo. 10.5281/zenodo.22171914
6. Yamagishi, Y. (2026). *A Gap That Creates No Bias Still Moves the Sign* [Paper 420]. Zenodo. 10.5281/zenodo.22699994
7. Yamagishi, Y. (2026). *What Looks Like Leading Is Not Leading* [Paper 421]. Zenodo. 10.5281/zenodo.22700022

8. Yamagishi, Y. (2026). *Look-Ahead Survives Both a Split Sample and a High t* [Paper 426]. Zenodo. 10.5281/zenodo.22700184

— *From thinking to syncing.* —

Yuuki YAMAGISHI (山岸 勇輝)

© 2026 Yamagishi Complexity Torus Lab, Japan · Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0)