

WEATHER AND THE FASHION CALENDAR

Do temperature and rainfall anomalies shift the timing and intensity of garment search interest?

Evidence from search behaviour in forty-one markets, 2018–2026

PAU ALMAR

Barcelona · September 2026

DECLARATION

I declare that this thesis is my own work. The data were collected by pipelines I wrote, from two public sources, and both pipelines are reproducible without proprietary access or paid subscription. Every quantitative result reported here was produced by those pipelines and can be regenerated from them.

Where the work draws on my professional experience in fashion retail, that experience is identified as the source of a hypothesis and never as evidence for a conclusion. The distinction is maintained deliberately throughout: Chapter 1 explains why practitioner intuition motivates a test rather than substituting for one.

The approach was author-driven use of AI. A large language model performed the execution — it wrote the retrieval and estimation code, ran it, and drafted this document — while every criterion, every research decision, and all auditing and quality control were human. All hypotheses, all interpretation, all decisions on what to test and what to reject, and responsibility for every claim made here are mine. Chapter 12 reports, at length, four occasions on which that analytical process produced a wrong answer and how each was caught.

Pau Almar

Barcelona, 29 September 2026

ABSTRACT

Practitioners in fashion retail hold that weather moves the selling calendar: that a warm autumn postpones outerwear demand and a warm spring advances summer demand. The belief is widespread, it shapes commercial decisions worth a great deal of money, and it has almost never been tested at the resolution at which it is held. The stakes are rising rather than static. As climate change makes warm anomalies more frequent, shorter winters — largely taken for granted in Western markets — become a recurring exposure rather than an occasional one; and because winter is when spend per unit peaks, with outerwear priced well above the categories that compete with it for attention, a shift of interest away from outerwear carries a plausible consequence for average selling price, revenue growth and, in time, gross margin — one this thesis's search-interest results speak to without directly measuring. The empirical literature on weather and retail is substantial but concentrated on contemporaneous volume — how much is sold in a hot week — rather than on when a season begins.

This thesis addresses the timing question directly, using weekly Google Trends search interest for fifteen garment concepts as the object of study — queried as Knowledge Graph entity identifiers rather than text strings so that the series are disambiguated and comparable across languages — paired with detrended per-week-of-year weather anomalies derived from the ERA5 reanalysis. The panel covers forty-one fashion markets over 426 to 429 weeks each, from May 2018 to August 2026, across which 328 model specifications were estimated.

Eleven hypotheses were pre-specified and tested. Three principal findings emerge. First, weather reliably shifts the timing of seasonal search interest: a one-standard-deviation warmer autumn delays the onset of wintery garment interest by 4.18 days ($t = 2.95$, 224 market-years, correct sign in 20 of 28 temperate markets), and a warmer spring advances summery onset by 8.25 days ($t = -3.21$, 194 market-years, correct sign in 24 of 28). The wintery window contracts accordingly, by 10.67 days per standard deviation ($t = -2.50$): the season is set by the calendar, while interest in it is set by the climate. Both estimates survive deletion of any single market, and both market-level sign counts survive restriction to markets with a cold winter together with the study's own placebo rule. Second, and contrary to the study's own expectation, weather does not alter total garment interest: the aggregate series is null in 26 of 28 temperate markets, implying that weather redistributes interest between categories rather than changing its volume — a result that converts the commercial problem from forecasting into allocation. Third, the temperature response is threshold-shaped rather than linear, effectively confined below approximately 7.5 °C. The intensity response replicates with seasonal windows inverted in the two southern-hemisphere markets that pass their placebo controls, and is absent in most tropical markets, which have no thermal season — evidence that it is not an artefact of the northern calendar.

Four hypotheses were rejected, including two held confidently at the outset: that minimum temperature drives autumn outerwear search interest, and that milder climates respond more strongly than harsh ones. A fifth, an asymmetry hypothesis constructed from a single-market result, failed on the panel; the panel leans towards the simpler proposition that warmth is the more active channel, without establishing it.

The thesis reports its own methodological failures in a dedicated chapter rather than an appendix: an abandoned instrument, a statistically significant result of the wrong sign traced to a peak-sensitive onset metric, a placebo that proved not to be neutral, and an unresolved batching defect that

destroyed data in an identified set of markets and invalidates every null result in one of the three result families. The central limitation is that no transaction data was observed at any point: every result describes search interest, and the inferential step to purchase behaviour is untested here.

Keywords: weather and retail demand; seasonal timing; fashion retail; search data; Google Trends; ERA5 reanalysis; phenology; demand proxies; apparel supply chain.

ACKNOWLEDGEMENTS

The hypothesis in this thesis is not mine alone. It belongs to every store manager, country head and merchandiser who has stood in front of a sales report in October and said that the season had not started yet — and to the buying teams who had committed the units nine months earlier and could do nothing about it. The observation is theirs; the only thing added here is a test.

I am grateful to the institutions that make this kind of work possible without funding or affiliation. The European Centre for Medium-Range Weather Forecasts produces the ERA5 reanalysis and releases it openly; Open-Meteo maintains a free interface to it that is faster and more usable than most paid alternatives. Google publishes the Trends index without charge. That an independent study of forty-one international markets can be assembled at zero data cost is a fact worth stating.

The work also began as a personal question: whether sixteen years inside the fashion industry, a close reading of the research and case literature on Zara and the other large market-makers, and AI used as an analytical instrument could together produce an analysis that meets a doctoral standard. The reader is the judge of whether it does.

Finally, a note on what this document is. It uses the form of a doctoral thesis because that form imposes a standard — declare the hypothesis before the test, report the failures alongside the findings, and name the assumption that would undo everything. Commercial analysis rarely imposes that standard on itself. It should.

CONTENTS

The table below is a field. To populate its page numbers, select the whole document and press F9 in Microsoft Word, or use References → Update Table.

CHAPTER ONE	Introduction	2
1.1	Background.....	2
1.2	The practitioner’s observation	2
1.3	Why practitioner intuition is not evidence	3
1.4	Research questions	3
1.5	Aims and objectives	4
1.6	Scope and delimitations.....	4
1.7	Contributions.....	4
1.8	Structure of the thesis.....	5
CHAPTER TWO	Literature review	6
2.1	Introduction and scope of the review	6
2.2	Weather and consumer behaviour: mechanisms	6
2.2.1	The four channels.....	6
2.2.2	Mood and affect	6
2.2.3	Projection bias and salience.....	7
2.3	Weather and retail demand: empirical evidence	7
2.3.1	Aggregate and macroeconomic studies.....	7
2.3.2	Store-level and transaction studies.....	7
2.3.3	Apparel-specific evidence	8
2.3.4	The one direct antecedent.....	8
2.4	Fashion seasonality and the buying calendar.....	9
2.5	Search data as an economic proxy.....	9
2.5.1	The nowcasting programme.....	9
2.5.2	Instrument properties and known defects.....	10
2.5.3	Entities against search terms: a genuine gap.....	11
2.5.4	Search data in fashion.....	11
2.6	Phenology and the extraction of season onset.....	11
2.7	Synthesis and the gap	12
2.8	Conceptual framework.....	13
CHAPTER THREE	Hypothesis development	14
3.1	From observation to falsifiable proposition	14
3.2	Intensity: does weather move garment search at all?.....	14
3.3	Timing: the central propositions	14
3.4	The driving variable.....	15
3.5	Volume against mix	15
3.6	Functional form	16

3.7 Directional asymmetry	16
3.8 Moderators and boundary conditions	16
3.9 Summary of hypotheses.....	17
CHAPTER FOUR Research design	19
4.1 Philosophical position	19
4.2 Research strategy	19
4.3 Unit of analysis	19
4.4 Design of the panel	20
4.5 Ethics and data governance.....	20
4.6 Reproducibility	20
CHAPTER FIVE Data.....	22
5.1 Weather	22
5.1.1 Source and justification	22
5.1.2 Variables retrieved and derived	22
5.2 Search interest as the object of study	23
5.2.1 What is measured, and why	23
5.2.2 Entity identifiers rather than text	23
5.2.3 Curated against machine-generated entities	24
5.3 The entity set	24
5.3.1 Composition	24
5.3.2 An internal comparison of query types	24
5.4 The placebo controls.....	25
5.5 The market universe	25
5.6 Exclusions.....	26
5.7 The assembled panel.....	27
5.8 Relation to transaction data	27
CHAPTER SIX Analytical methods.....	28
6.1 Temporal alignment	28
6.2 Construction of weather anomalies	28
6.2.1 Why anomalies are necessary	28
6.2.2 Detrended per-week-of-year standardisation.....	28
6.2.3 Validation of the baseline definition	29
6.3 Construction of the search series.....	29
6.3.1 Window length.....	29
6.3.2 Chain-linking	30
6.3.3 Batch composition	30
6.3.4 Transformation and aggregation	30
6.3.5 Reference periods for the normal and the standard deviation	31
6.4 Model specifications	32
6.4.1 Intensity	32

6.4.2	Standard errors	32
6.4.3	Timing	32
6.4.4	Asymmetric specification	33
6.5	Validation	33
CHAPTER SEVEN	Intensity: does weather move garment search?	36
7.1	The enabling result	36
7.2	Magnitude and distribution.....	36
7.3	The precipitation channel.....	37
7.4	Which temperature variable governs the response	38
7.5	Summary of Chapter 7.....	39
CHAPTER EIGHT	Timing: does weather move the fashion calendar?.....	40
8.1	Warm autumn delays the wintery season.....	40
8.2	Warm spring advances the summery season.....	40
8.3	The two effects compared	41
8.4	Stability of the estimates.....	41
8.5	The shrinking estimate	42
8.6	Season width.....	42
8.7	Summary of Chapter 8.....	43
CHAPTER NINE	The structure of the response.....	44
9.1	Volume against mix: the principal finding	44
9.2	The shape of the temperature response.....	45
9.3	Level against rate of change.....	46
9.4	Directional asymmetry	46
9.5	Summary of Chapter 9.....	47
CHAPTER TEN	Robustness and boundary conditions	48
10.1	The southern panel	48
10.2	Does climate severity moderate the response?	49
10.3	Where the effect ceases.....	49
10.4	Placebo failures.....	50
10.5	The pandemic period	50
10.5.1	Extending the check to onset timing.....	51
10.6	A restriction to markets with a cold winter	52
10.7	Summary of Chapter 10	52
CHAPTER ELEVEN	Discussion	55
11.1	What has been established	55
11.2	Interpreting the timing effects.....	55
11.3	The reframing implied by the substitution result.....	56
11.4	Rainwear as the operational candidate.....	57
11.5	What should not yet be acted upon	57
11.6	Relation to the literature	57

CHAPTER TWELVE Methodological failures	59
12.1 An abandoned instrument.....	59
12.2 A significant result of the wrong sign	59
12.3 A placebo that was not neutral	60
12.4 Calibration on a single market.....	60
12.4.1 Distinguishing destroyed data from genuine sparsity	61
12.4.2 Which conclusions this contaminates.....	62
12.4.3 The remedy, and why it was not applied.....	62
12.5 An asymmetry that did not generalise.....	62
12.6 What the corrections cost, and what they bought	63
CHAPTER THIRTEEN Limitations.....	64
13.1 The governing limitation	64
13.2 Geographic representation	64
13.3 Panel length	64
13.4 Composite construction and completeness	65
13.5 Properties of the instrument.....	65
13.6 What would overturn these findings	66
CHAPTER FOURTEEN Further work.....	67
14.1 Per-market batch calibration and re-estimation	67
14.2 Validation against sell-through	67
14.3 Purchase-proximity elasticity	67
14.4 Extension of the panel to 2004.....	67
14.5 Full-price and markdown decomposition	67
14.6 ENSO regimes.....	68
14.7 Onset-metric sensitivity analysis.....	68
CHAPTER FIFTEEN Conclusion	69
15.1 Answering the research questions	69
15.2 Summary of hypothesis outcomes.....	69
15.3 Contribution restated	70
15.4 A closing statement about what this is	70
References	71
Appendices	75
A Entity identifiers.....	75
B Markets.....	75
C Excluded markets.....	77
D Model output.....	77
E Placebo audit	77
F Reproduction and reporting standard.....	77
G Glossary.....	77

LIST OF TABLES

Table 1	Hypotheses, their sources, and where each is tested	Ch. 3
Table 2	Composition of the entity set.....	Ch. 5
Table 3	The market universe by seasonal amplitude.....	Ch. 5
Table 4	Markets excluded, with the measurement that excluded them	Ch. 5
Table 5	Descriptive statistics of the assembled panel.....	Ch. 5
Table 6	Comparison of three climatology definitions	Ch. 6
Table 6a	Reference periods for the normal and the standard deviation	Ch. 6
Table 7	Validation gates and what each is designed to catch	Ch. 6
Table 8	Intensity of the wintery response, strongest markets	Ch. 7
Table 9	Intensity of the rainwear response, strongest markets	Ch. 7
Table 10	Competitive test of thermal drivers.....	Ch. 7
Table 11	The two timing effects compared	Ch. 8
Table 12	Jackknife stability of the pooled timing estimates.....	Ch. 8
Table 13	Wintery composite by temperature band.....	Ch. 9
Table 14	Directional decomposition of the seasonal response.....	Ch. 9
Table 15	The southern panel, seasonal windows inverted	Ch. 10
Table 16	Sensitivity to the pandemic period	Ch. 10
Table 16b	Market-level results under a cold-winter restriction	Ch. 10
Table 17	Markets failing the placebo gate	Ch. 10
Table 18	Raincoat series, queried alone against queried in batch	Ch. 12
Table 19	Entity identifiers and their resolutions	App. A
Table 20	Summary of hypothesis outcomes	Ch. 15
Table 21	The market roster: all fifty cities by group	App. B

LIST OF FIGURES

Figure 1 Comparison of three climatology baseline definitions	Ch. 6
Figure 2 Competitive test of minimum, mean and maximum temperature	Ch. 7
Figure 3 Wintery intensity across the strongest markets.....	Ch. 7
Figure 4 The two timing effects with 95 per cent intervals	Ch. 8
Figure 5 History of the autumn-delay estimate across three datasets.....	Ch. 8
Figure 6 The threshold profile of the wintery response	Ch. 9
Figure 7 Directional decomposition of the seasonal response	Ch. 9
Figure 8 The southern panel.....	Ch. 10
Figure 9 The batching defect: destroyed data against genuine sparsity	Ch. 12

ABBREVIATIONS AND NOTATION

SYMBOL	MEANING
ENSO	El Niño–Southern Oscillation
ERA5	Fifth-generation ECMWF atmospheric reanalysis
HAC	Heteroskedasticity- and autocorrelation-consistent (standard errors)
OOS	Out-of-sample
SD	Standard deviation, computed within week-of-year across years
β	Regression coefficient, summed across distributed lags 0–3
$\varepsilon(w, y)$	Residual for week w of year y after removal of level and trend
z	A standardised, deseasonalised search series
/m/xxxx	A curated Google Knowledge Graph entity identifier
t	A HAC t statistic unless otherwise stated
q	A p value adjusted for multiplicity by the Benjamini–Hochberg procedure

A note on two words. Throughout this thesis, garment categories are called *wintery* and *summery* rather than warm and cold. The reason is that a coat is a warm garment worn in cold weather, so calling either category "warm" is ambiguous in exactly the place where precision matters most. *Wintery* and *summery* name the season the garment belongs to, which is unambiguous.

PART I

Foundations

CHAPTER ONE

Introduction

1.1 Background

Fashion retail commits its inventory long before it knows its weather. A garment sold in a European store in October was specified in January, sampled in March, ordered in April and manufactured over the summer. By the time the first cold week arrives, the decision that matters — how many coats exist, in what sizes, in which countries — has been fixed for months and cannot be revised. The industry's response to this structural problem has been to shorten the commitment horizon, and a substantial operations literature documents the gains available from doing so (Fisher and Raman, 1996; Cachon and Swinney, 2011). Downstream of that commitment, a separate literature addresses how to allocate what has already been produced across a store network as the season unfolds (Caro and Gallien, 2010). But shortening the horizon does not eliminate the problem; it relocates it. Whatever remains committed still has to meet a season whose start date nobody controls.

This creates an asymmetry that anyone who has worked a selling season recognises. The plan is calendar-driven — week 40 is when outerwear goes to the floor, because week 40 is when outerwear went to the floor last year. The demand is not calendar-driven. It is weather-driven, and the weather moves. In a warm October the coats sit; in a cold September they were not there yet. Both outcomes look like a forecasting error and neither is: the volume forecast may have been perfectly correct while the timing assumption underneath it was wrong.

The length of the season sets the cost of such an error. A fashion season lives for roughly six months before the next one replaces it, so a start delayed by a week does not move the season intact: it removes a material share of the full-price selling window, and if the markdown calendar holds, it can compromise the season as a whole.

The question this thesis asks is whether that intuition is true, and if so by how much. It is a narrow question deliberately. Not whether weather affects retail sales, which is established beyond reasonable dispute, but whether weather moves the date on which a season starts, which is a different claim requiring different evidence.

1.2 The practitioner's observation

The hypothesis examined here did not originate in a dataset. It originated on a shop floor, in a commercial role at Inditex, where weekly sell-through across a large store estate was a routine object of attention over many years. The observation was that garment categories did not turn on the calendar. They turned on the weather, and they turned earlier or later than the buying plan assumed, in ways that recurred often enough to look like a pattern rather than noise.

Three propositions followed from that experience, and they are the propositions this study was built to test. First, that an autumn running warmer than normal postpones the point at which customers begin buying outerwear. Second, that a spring running warmer than normal brings forward the point at which they begin buying summer garments. Third, and following from the

first two, that the wintery selling window is not fixed but compresses in warm years — arriving late and departing early.

A fourth proposition was more specific and, as Chapter 7 reports, wrong. The practitioner view held that *minimum* temperatures and rainfall drove the autumn turn — that cold nights and wet weeks were what pushed customers into coats — while *maximum* temperatures governed the summer categories. The second half of this is supported. The first half is contradicted decisively.

A fifth observation concerned Brazil, where the effect appeared more pronounced in El Niño and La Niña years. This study does not test that claim, and Section 14.6 sets out what testing it would require. It is worth flagging early that the one published study closest to this design does find exactly such a regime dependence (Oh, Jo and Ha, 2021), which raises the value of the untested claim considerably.

1.3 Why practitioner intuition is not evidence

Intuition of this kind has three well-known failure modes, and each is a reason to test rather than assert.

1. **Selective memory.** Warm autumns in which coats sold late are more memorable than warm autumns in which they did not, because the former generated a problem that required explanation and the latter generated nothing. The recalled sample is conditioned on the outcome.
2. **Unreliable attribution.** A late season has many candidate causes — the buying calendar, the markdown calendar, a competitor’s timing, the position of a public holiday, the weather — and a person standing inside the business observes all of them simultaneously and none of them cleanly. The literature on weather and retail is itself partly a literature about disentangling these (Roth Tran, 2023; Badorf and Hoberg, 2020).
3. **Unobservable counterfactual.** Nobody sees what the season would have done under different weather. The practitioner compares what happened to what was planned, which is a comparison against a forecast rather than against an alternative world.

The value of a formal test is not that it is cleverer than the practitioner. Chapter 2 will show that the practitioner is, on the central point, ahead of the published literature. The value is that a formal test can be wrong in public. This study was designed so that each proposition could fail, and Chapters 7 to 10 report four that did.

1.4 Research questions

The study addresses four questions, in descending order of importance.

- **RQ1.** Do deviations of temperature and rainfall from local seasonal norms shift the *timing* of seasonal garment search interest — the date on which a season begins, ends, and its width?
- **RQ2.** Do those deviations change the *total quantity* of garment search interest, or do they redistribute a fixed quantity between categories?
- **RQ3.** What is the *functional form* of the response — linear in degrees, threshold-shaped, or asymmetric between warming and cooling — and which temperature measure drives it?

- **RQ4.** Under what *boundary conditions* does the effect hold, and where does it cease?

1.5 Aims and objectives

The aim is to establish, at weekly resolution and across an international panel, whether and by how much weather anomalies shift the fashion calendar. Six objectives follow.

1. To construct a search-interest series at weekly resolution that is disambiguated, comparable across languages and markets, and reproducible from public sources.
2. To construct weather anomalies that isolate deviation from local seasonal norm, free of both the annual cycle and the secular warming trend.
3. To specify eleven falsifiable hypotheses in advance of estimation, derived from practitioner observation and from the literature.
4. To estimate each hypothesis across the panel with inference robust to serial correlation and corrected for multiplicity.
5. To subject every result to validation gates — placebo, out-of-sample, jackknife — specified before results were seen.
6. To report the study's own failures with the same prominence as its findings.

1.6 Scope and delimitations

Four boundaries are drawn explicitly, and each excludes something a reader might reasonably expect.

- **Search interest, not transactions.** No transaction data was used. Search interest is treated as the object of study in its own right — an aggregate signal spanning every retailer and channel in a market, rather than one firm's own pricing and assortment decisions — not as a stand-in for a sales figure this study lacks.
- **Timing and mix, not volume.** The study does not attempt to forecast sales volume, and the finding at Chapter 9 suggests volume is the wrong target in any case.
- **Temperate markets carry the argument.** Thirteen of the forty-one markets have insufficient seasonal amplitude for a season onset to be defined (tropical or close to tropical cities). They are retained as a boundary condition rather than dropped, but they contribute nothing to the central estimates.
- **Eight seasons.** The panel spans 2018 to 2026. This is short for a phenological analysis, and Section 13.3 argues that extending it is the single most cost-effective improvement available.

1.7 Contributions

The thesis makes four contributions, of which the third is arguably the most useful.

1. **An empirical estimate of the timing effect.** The literature on weather and retail has concentrated on contemporaneous volume. This study estimates the shift in season-onset date directly, across forty-one markets, and reports it with confidence intervals and market-level sign counts.

2. **A substitution result.** The finding that weather moves search interest between categories without changing its total reframes the commercial problem from forecasting to allocation. Its zero-net-volume component corroborates, on independent data and at a much larger scale, a null effect on whole-season sales volume that Bahng and Kincade (2012) found at a single retailer.
3. **A reflexive account of method failure.** Chapter 12 documents occasions on which the analysis produced a wrong answer, including one statistically significant result of the wrong sign, and shows in each case which validation gate caught it. Studies that report only what survived give the reader no basis for judging how carefully the analysis was checked.
4. **A methodological argument for entity-based querying.** The study queries Google Trends by Knowledge Graph entity identifier rather than by text string, and Section 2.5.4 shows that the literature contains a normative preference for this with no empirical evaluation behind it.

1.8 Structure of the thesis

The thesis is in four parts. Part I establishes foundations: this introduction, a review of the four literatures the study sits between (Chapter 2), and the derivation of eleven hypotheses (Chapter 3). Part II sets out method: research design and philosophical position (Chapter 4), the data and its construction (Chapter 5), and the analytical procedures (Chapter 6). Part III reports results in four chapters, organised by the kind of claim rather than by hypothesis number: intensity (Chapter 7), timing (Chapter 8), the structure of the response (Chapter 9), and robustness and boundary conditions (Chapter 10). Part IV synthesises: discussion and commercial interpretation (Chapter 11), methodological failures (Chapter 12), limitations (Chapter 13), further work (Chapter 14) and conclusion (Chapter 15).

Chapter 12 is placed in the main body rather than in an appendix. This is a deliberate structural choice and the reasoning is given at the head of that chapter.

CHAPTER TWO**Literature review****2.1 Introduction and scope of the review**

This study sits at the intersection of four literatures that rarely cite one another. The first is the economics and marketing literature on weather and consumer demand, which is seventy years old and largely concerned with volume. The second is the operations literature on fashion supply chains and seasonality, which is concerned with lead times and markdowns and treats weather as noise. The third is the literature on internet search data as an economic proxy, which is concerned with nowcasting macroeconomic aggregates and has been through a well-documented crisis of confidence. The fourth is land-surface phenology, which is not a social science at all but has spent thirty years solving the precise technical problem this thesis faces: how to define the start of a season from a noisy annual curve.

The review proceeds through each in turn, then identifies the gap at Section 2.7 and sets out the conceptual framework at Section 2.8.

2.2 Weather and consumer behaviour: mechanisms**2.2.1 *The four channels***

The foundational study is Steele (1951), which analysed daily sales at a Des Moines department store and related them to weather conditions. More durable than any single estimate from it is Steele's taxonomy of four transmission channels: physical comfort, physical prevention of travel, psychological effect on shopping inclination, and the desirability of particular products under particular weather. Seventy-five years later this remains the organising frame, and the literature since has largely consisted of separating the channels and establishing which dominates in which setting.

The distinction matters here because only the fourth channel — product desirability — implies a timing effect of the kind this thesis tests. The first three imply that weather changes whether and how much people shop. The fourth implies that weather changes what they shop for, which is a claim about mix and calendar rather than about volume.

2.2.2 *Mood and affect*

A substantial strand attributes weather effects to mood rather than to physical need. Parsons (2001), studying shopper counts at a New Zealand mall, discusses mood alongside the pure physical-deterrent explanation of rainfall effects without fully adjudicating between them. Murray, Di Muro, Finn and Popkowski Leszczyc (2010) established the mechanism experimentally across three laboratory and field studies: sunlight raises consumer spending, and the effect is mediated by reduced negative affect. This is a real and well-identified phenomenon, but it is a volume mechanism, and it predicts nothing about which category a consumer buys.

2.2.3 *Projection bias and salience*

The most theoretically interesting mechanism for present purposes is projection bias — the tendency to over-weight current state when predicting future preferences. Conlin, O'Donoghue and Vogelsang (2007) tested this on catalogue orders of cold-weather items and found that a 30 °F decline in temperature on the order date raised the probability of the item being returned by 3.95 per cent. Consumers, in other words, bought winter clothing because it was cold today, not because they had correctly forecast their need for it.

Busse, Pope, Pope and Silva-Risso (2015) confirmed the pattern at scale on more than forty million vehicle transactions: a 20 °F temperature rise raises convertible sales by 8.5 per cent, and four-wheel-drive share rises 6 per cent for two to three weeks after a heavy snowstorm. The persistence of the snowstorm effect beyond the snow itself is the signature of salience rather than need.

Projection bias is the mechanism that makes this thesis's hypothesis plausible at the level of the individual consumer. If purchases of seasonal clothing are triggered by current weather rather than by anticipated need, then the date of the first sufficiently cold week should determine the date the season starts — which is precisely the proposition under test.

2.3 Weather and retail demand: empirical evidence

2.3.1 *Aggregate and macroeconomic studies*

Starr-McCluer (2000), writing for the Federal Reserve, found that unusual weather has a modest but statistically significant role in monthly retail sales, and made an observation central to this thesis: lagged effects offset contemporaneous ones, so the influence largely washes out at quarterly frequency. This is intertemporal substitution — demand is moved in time rather than created or destroyed — and it is the macroeconomic ancestor of the substitution result reported at Chapter 9.

Roth Tran (2023) provides the most careful modern treatment, using daily store-level sales to separate short-run weather shocks from long-run climate adaptation. The distinction matters for the present study because it identifies exactly the confound that a per-market analysis must handle: a market's response to a cold week depends on what that market's weather is normally like, which is the reason this thesis standardises anomalies within market and within week-of-year.

2.3.2 *Store-level and transaction studies*

The recent literature is rich and consistent in finding effects, across settings that share little beyond the presence of a weather signal: daily weather effects across 673 brick-and-mortar stores (Badorf and Hoberg, 2020), extreme-temperature effects on Seoul credit-card transactions (Yoo, Eom and Zhou, 2024), and rain-driven shifts in basket composition toward fewer, higher-priced items across six million transactions at 146 Chinese convenience stores (Tian, Cao and Song, 2021). Rose and Dolega (2022) add a distinction the others do not draw: high-street sales respond to weather more strongly than out-of-town locations, and wind, not temperature, is the most influential variable.

This literature is overwhelmingly concerned with daily or weekly volume, and overwhelmingly cross-sectional in weather — wet days against dry days, rather than warm seasons against cold ones. Neither framing addresses the question of when a season begins.

2.3.3 *Apparel-specific evidence*

A smaller literature addresses clothing directly, and it is closer to the present question.

Travers and Payne (1998) is probably the earliest explicit statement of the timing hypothesis in the fashion literature. Examining high-street womenswear in England, they argued that hot dry summers and warm wet winters had blurred the distinction between seasons, requiring transseasonal fabrics and threatening the four-season buying calendar. The paper is qualitative and now nearly thirty years old, but it names the phenomenon precisely.

Bahng and Kincade (2012) is the closest published antecedent for the substitution result's volume component. Analysing sales data from a branded womenswear retailer, they found that day-to-day and week-to-week temperature changes did not significantly alter whole-season sales volume — a null result that anticipates, on a single retailer's data, the panel-wide finding at Chapter 9 that weather redistributes garment search interest rather than expanding or contracting it.

Martínez-de-Albéniz and Belkaid (2021) provide the cleanest separation of the channels. At a large fashion apparel retailer they show that rain drives footfall — up in malls, down on streets — while temperature drives conversion, raising sales of weather-appropriate categories: positive temperature shocks sell summer items more and winter items less. This is Steele's fourth channel, isolated and measured.

On the operations side, Bertrand, Brusset and Fortin (2015) estimate apparel sales impacts from unexpected deviations of monthly temperature from seasonal norms and design weather derivatives to hedge them. Steinker, Hoberg and Thonemann (2017) show at a large European online fashion retailer that sunshine, temperature and rain significantly affect daily sales and that week-ahead weather-informed forecasting yields measurable workforce savings. Babongo and colleagues (2018) report forecast-error reductions of up to 45 per cent for seasonal sports equipment using meteorological predictors. Štulec, Petljak and Naletina (2019), studying Croatian convenience-store beverage sales, propose weather derivatives as a hedge against adverse weather deviations — a further instance of the operations literature's preference for treating weather as a source of variance to be priced rather than a signal to be acted on, the distinction Section 11.3 returns to.

2.3.4 *The one direct antecedent*

Oh, Jo and Ha (2021) is the study closest to the present design, and it deserves separate treatment. Working on New York between 2014 and 2019, they combine Google Trends search interest for winter jackets with cumulative-sum statistics on temperature to locate the timing of seasonal clothing demand. They find that search activity begins after approximately six consecutive days of rapid cooling — a threshold rather than a linear response — and, remarkably, that the CUSUM threshold shifts sharply between ENSO regimes, taking values of 217 to 248 in El Niño years, 281 in normal years, and 118 to 128 in La Niña years.

Three things follow. First, the core design of this thesis — search proxy against temperature anomaly, targeting timing — has published precedent and is not novel in kind. Second, the threshold finding at Chapter 9 of this thesis corroborates theirs on a much larger panel. Third, and most striking, the ENSO regime dependence that Oh and colleagues report is the published counterpart of the Brazil observation recorded at Section 1.2, which this thesis does not test. That an independent study found regime dependence in New York substantially raises the prior on the untested Brazilian claim, and Section 14.6 treats it accordingly.

The follow-up papers (Oh, Ha and Jo, 2022a; 2022b) extend the predictor set, finding wind chill and month of year to be the significant predictors, and review the Korean clothing-and-weather literature since 2000.

2.4 Fashion seasonality and the buying calendar

The operations literature explains why timing has commercial consequences. Şen (2008) reviews the US apparel supply chain and documents the structural fact underlying this thesis: the buying calendar is fixed months ahead of the selling season, so the assortment that meets the weather was determined before the weather was knowable. Fisher and Raman (1996) formalised the response — accurate response, in which lead times are shortened so a larger share of production is committed after early demand signals arrive — and demonstrated at a skiwear firm that it raised profits by roughly 60 per cent. Cachon and Swinney (2011) show that quick response and enhanced design are complements and that their joint value rises sharply when consumers behave strategically about waiting for markdowns. Caro and Gallien (2010) model the in-season shipment allocation problem at Zara, which is precisely the decision layer an intra-season weather signal would feed.

What this literature does not do is treat weather as a signal. It treats early sales as the signal and weather as unmodelled variance. If the timing effect this thesis tests is real, then weather is an additional and earlier signal than early sales, available before the season starts rather than after — which is the argument developed at Section 11.3.

2.5 Search data as an economic proxy

2.5.1 *The nowcasting programme*

The use of search data to measure contemporaneous economic activity begins with Ginsberg and colleagues (2009), who estimated regional influenza rates from 45 search queries roughly a week ahead of official surveillance. The following decade generalised the result well beyond epidemiology: adding Trends variables to simple autoregressive models improved near-term estimates of automobile sales, unemployment claims, travel planning and consumer confidence in the United States (Choi and Varian, 2012), of unemployment in Germany (Askatas and Zimmermann, 2009), and of private consumption, where a Trends-based indicator outperformed both the Michigan and Conference Board sentiment indices in nearly all forecasting experiments (Vosen and Schmidt, 2011). Woloszko (2020) extended the approach to a weekly GDP tracker for 46 OECD and G20 countries that performed well through the COVID-19 collapse, and Varian (2023) revisited the programme a decade on.

Validation specifically against retail sales exists. Lind and Ramesh (2025) build a search-intensity index for a panel of publicly traded US retailers and show it nowcasts firm sales and predicts analyst forecast errors after controlling for past sales and fixed effects. Golovanova and Zubarev (2021) show Trends categories improve out-of-sample nowcasts of aggregate retail sales and of food and non-food subcategories.

2.5.2 *Instrument properties and known defects*

The programme's clearest cautionary case is Google Flu Trends, which overestimated influenza-like illness by more than a factor of two; Lazer, Kennedy, King and Vespignani (2014) trace the failure to two general causes, big-data hubris and undocumented algorithm changes in Google's own product. The diagnosis motivates the validation apparatus used throughout this thesis — placebo controls, out-of-sample gates and jackknife tests at every specification — and the related point that an instrument maintained by a commercial party for its own purposes is not guaranteed stationary over an eight-year observation window (Section 13.5 returns to this as a limitation).

A methodological literature has grown up around the instrument's quirks, and it is unusually candid. Eichenauer, Indergand, Martínez and Sax (2022) show that raw Trends data are simultaneously frequency-inconsistent — daily pulls miss the long-run trend — and sampling-noisy, and give a repeated-sampling and chain-linking procedure that addresses both. This is the direct methodological warrant for the chain-linking described at Section 6.4.

Cebrián and Domenech (2024) model the sampling process and derive how many independent samples a stable series requires; their earlier paper (2023) assesses Trends against standard data-quality dimensions and documents measurement problems that propagate downstream. Gummer and Oehrlein (2025) establish a reliability-frequency continuum: reliability degrades systematically as search frequency falls, which is the empirical justification for imposing a minimum-volume screen on low-traffic garment entities.

The debate has a sharp edge. Franzén (2023) reports an unplanned natural experiment in which identical queries returned non-replicable series and argues that scientists should refrain from using Trends altogether. Raubenheimer (2024) replies that the blanket rejection overreaches and that the reliability problems are manageable with correct sampling practice. This thesis sides with Raubenheimer, and Section 6.4 states what "correct practice" was taken to mean here.

Two normalisation results bear directly on a defect this study did not resolve. Springer, Strzelecki and Zieger (2023) show that because the 0–100 index is normalised within each query, cross-query comparison is invalid without an external anchor. Springer and colleagues (2025) treat the relative-search-volume normalisation and its limits systematically. Chapter 12 reports a batching defect that is, precisely, a failure to respect this constraint: entities were grouped into shared-scale queries using magnitudes measured in one market, and the normalisation crushed low-volume companions elsewhere.

Finally, Hölzl, Keusch and Sajons (2025) review 360 studies using Trends and find that the large majority test neither internal validity nor reliability across samples nor generalisability. Mavragani and Ochoa (2019) propose a reporting framework requiring keyword, region, period

and category to be documented for every Trends series. This thesis attempts to meet that standard, and Appendix F states where it falls short.

2.5.3 *Entities against search terms: a genuine gap*

This study queries Trends by Knowledge Graph entity identifier — the curated /m/ topics — rather than by text string. The justification is disambiguation and cross-language comparability, and it has a normative precedent. Woloszko (2020) states the argument directly:

Queries based on keywords are language-specific and subject to ambiguity. Google Trends series for the keyword 'apple' mixes up searches for the fruit and the company. Both categories and topics are harmonised across languages, and queries based on categories or topics yield series comparable across countries.

— Woloszko (2020), OECD Working Paper No. 1634

The same paper supplies the caveat that must be cited alongside it: the content of topics is opaque, and the algorithm allocating keywords to them can make arbitrary choices — Woloszko's example is a French-dominated "unemployment benefits" topic in Canada that is informative about Québec rather than the country. Narita and Yin (2018) provide an early topic-based application.

What does not exist is any empirical evaluation. There is no published study comparing whether entity-identifier queries and text-term queries yield different measured demand signals, and none on the relative sampling stability of the two query types. The field holds a normative preference with no evidence underneath it, and Hölzl and colleagues (2025) suggest this is characteristic of query-construction choices generally. Section 5.3 of this thesis reports a small internal comparison but does not close the gap.

2.5.4 *Search data in fashion*

Applications to apparel are sparse. Silva, Hassani, Madsen and Gee (2019) forecast fashion consumer behaviour from Trends using Burberry as a case, finding no single univariate model dominates across horizons and documenting a strong, persistent annual seasonality in apparel search — evidence that the underlying phenomenon this thesis measures is stable enough to model. Skenderi, Joppi, Denitto and Cristani (2024) combine Trends series with product images to forecast sales of new fashion items with no sales history, validated against real retailer sales. No peer-reviewed paper was located whose primary subject is search data and seasonal goods as a general class.

2.6 *Phenology and the extraction of season onset*

The technical problem this thesis faces — given a noisy annual curve, determine the date the season starts — was solved thirty years ago in a discipline that has nothing to do with retail. Land-surface phenology extracts green-up dates from satellite vegetation indices, and its methods transfer directly.

Reed and colleagues (1994) introduced early quantitative methods for extracting onset and offset dates from smoothed vegetation time series. Jönsson and Eklundh (2004), in the TIMESAT program, fit smooth functions to noisy seasonal series and extract season start and

end as user-specified fractions of the seasonal amplitude — the half-amplitude rule this thesis adopts, and the closest formal analogue to the procedure at Section 6.7. Zhang and colleagues (2003) offer the main alternative: piecewise logistic fits with transition dates at curvature extrema — a peak- and curvature-sensitive class of metric.

That distinction is not academic. White and colleagues (2009) applied ten different start-of-season extraction methods to one identical dataset and found they produced materially different onset dates and, critically, different trend signs. This is the single most important methodological citation in this thesis, because Chapter 12 reports exactly that failure occurring: a peak-sensitive onset metric produced a large, significant, wrongly-signed result that a threshold metric did not.

The substantive phenological findings also frame the expected magnitude. Menzel and Fabian (1999) found spring events in Europe advanced six days and autumn events delayed 4.8 days since the early 1960s, lengthening the growing season by 10.8 days. Menzel and colleagues (2006), across 125,000 series in 21 countries, found spring and summer events advancing 2.5 days per decade with the spatial pattern of the shift matching the spatial pattern of warming. Linderholm (2006) reviews season-length change and attends to whether extension comes from earlier onset or later senescence — the same decomposition this thesis applies to the wintery window.

Plants, then, shift their seasons by days per decade in response to warming. The question here is whether consumers shift theirs by days per standard deviation in response to weather.

2.7 Synthesis and the gap

Four observations follow from the review.

1. The weather–retail literature is large, consistent and almost entirely about volume. Where timing appears it is incidental (Starr-McCluer, 2000) or qualitative (Travers and Payne, 1998) or confined to a single retailer (Bahng and Kincade, 2012).
2. The one study that targets timing with a search proxy (Oh, Jo and Ha, 2021) covers one city, one garment concept and five years. Its findings — a threshold response and ENSO regime dependence — are important and untested at scale.
3. The search-proxy literature has developed sophisticated diagnostics for its own instrument, but the fashion applications do not use them, and the entity-versus-term question is settled by assertion rather than evidence.
4. Phenology has solved the onset-extraction problem and demonstrated that the choice of extraction method can reverse a result, but no retail study appears to have imported either the methods or the warning.

The gap this thesis addresses is therefore specific: no study estimates the weather-driven shift in seasonal garment search-interest timing across an international panel, at weekly resolution, with the instrument diagnostics that the search-data literature demands and the onset-extraction discipline that the phenology literature demands. That is what follows.

2.8 Conceptual framework

The framework the study tests has four elements, and each maps to a literature above.

- **A trigger.** Weather that deviates from the local seasonal norm alters the perceived appropriateness of a garment category. This is Steele’s fourth channel, isolated by Martínez-de-Albéniz and Belkaid (2021), and it operates through projection bias at the individual level (Conlin et al., 2007; Busse et al., 2015).
- **A threshold.** The trigger is not linear in degrees. Below some temperature the response is strong and above it absent, consistent with the consecutive-cooling-days threshold found by Oh, Jo and Ha (2021).
- **A redistribution, not a creation.** Because the underlying need for clothing is not weather-dependent, the trigger moves search interest between categories and across time rather than changing its total — the intertemporal substitution of Starr-McCluer (2000) and the whole-season invariance of Bahng and Kincade (2012).
- **A calendar consequence.** Because the assortment was committed months earlier (Şen, 2008), a shift in the trigger date produces a mismatch that cannot be corrected within the season except through the allocation and markdown levers (Caro and Gallien, 2010; Cachon and Swinney, 2011).

Chapter 3 converts this framework into eleven testable propositions.

CHAPTER THREE

Hypothesis development

3.1 From observation to falsifiable proposition

Eleven hypotheses were specified before estimation. Six derive from the practitioner observations at Section 1.2, three from the literature reviewed in Chapter 2, and two were formulated during the study in response to intermediate results — a provenance that is recorded honestly in Table 1 because a hypothesis generated after seeing data carries less evidential weight than one specified before.

Each is stated so that it can fail. Where a hypothesis is directional, the direction is given; where it concerns functional form, the alternative is named. Chapters 7 to 10 report the outcomes and Chapter 15 tabulates them.

3.2 Intensity: does weather move garment search at all?

H1. *Weekly deviations of temperature and precipitation from local seasonal norms are associated with contemporaneous deviations in garment search interest, negatively for wintery categories against temperature and positively for rainwear against precipitation.*

This is the enabling hypothesis rather than an interesting one. If it fails, nothing else in the study can be estimated, because there is no signal in the proxy to work with. Its function is to establish that the instrument detects the phenomenon at all, and to locate the markets where it does so most cleanly. The literature makes failure unlikely — Martínez-de-Albéniz and Belkaid (2021) and Steinker and colleagues (2017) both find the effect in apparel sales — but the demonstration is required because those studies measured sales and this one measures search.

3.3 Timing: the central propositions

These two hypotheses carry the thesis. Both derive directly from Section 1.2, and both concern the date of a transition rather than the level of a series.

H2. *The onset of wintery garment search interest occurs later in years when the preceding autumn is warmer than normal.*

H3. *The onset of summery garment search interest occurs earlier in years when the preceding spring is warmer than normal.*

The mechanism is projection bias. If consumers buy seasonal clothing in response to weather they are currently experiencing rather than weather they anticipate (Conlin et al., 2007), then the season begins when the weather crosses whatever threshold triggers the purchase, and a warm shoulder season postpones that crossing. The phenological literature provides the expected order of magnitude: plants shift by days per decade under sustained warming (Menzel et al., 2006), and consumers responding to a single season's anomaly should plausibly shift by days per standard deviation.

Both hypotheses are one-sided in prediction but were estimated two-sided, and the per-market sign counts are reported alongside the pooled estimate precisely so that a pooled result driven by a few markets can be distinguished from a general one.

H2a. *Following from H2 and H3 jointly, the wintery selling window compresses in warm years: it begins later and ends earlier, so its width contracts.*

This is the compression proposition of Section 1.2 and the retail counterpart of the growing-season length literature (Menzel and Fabian, 1999; Linderholm, 2006). It is treated as a corollary rather than an independent hypothesis because its two components are H2 and the mirror of H3.

3.4 The driving variable

H4. *Minimum temperature is the principal driver of wintery garment search interest, while maximum temperature governs summery search interest.*

This is the fourth practitioner proposition, and it is the one held most confidently going into the analysis. The reasoning is intuitive: it is the cold night that makes a person conclude they need a coat, and rainfall reinforces it. The hypothesis is tested competitively — minimum, mean and maximum temperature entered against the same outcome — rather than by testing minimum alone, because the three measures are highly collinear and a significant coefficient on any one of them would prove nothing about the others.

The literature offers no clear prior. Rose and Dolega (2022) find wind the most influential weather variable in UK retail sales; Oh, Ha and Jo (2022a) find wind chill dominant, which is a function of temperature and wind rather than of the minimum. Neither adjudicates the minimum-versus-maximum question directly.

3.5 Volume against mix

H5. *Weather anomalies change the total quantity of garment interest, not merely its distribution across categories.*

This hypothesis is stated in the affirmative and expected to be confirmed, which makes its rejection the most consequential result in the thesis. Two literatures predict rejection. Starr-McCluer (2000) found weather effects on retail sales offsetting across lags at quarterly frequency; Bahng and Kincade (2012) found week-to-week temperature changes leaving whole-season volume unaltered. Against them stands the entire mood-and-affect strand (Murray et al., 2010), which predicts a genuine volume effect through spending propensity.

The test is direct: an aggregate composite of all garment entities, excluding placebos, regressed on the temperature anomaly. If the category-level responses are large and the aggregate is null, search interest is being redistributed rather than created.

An instrument caveat stated in advance. Google Trends reports a normalised share of total search rather than an absolute volume. A null result on an aggregate composite is therefore consistent both with genuine substitution and with the normalisation mechanically bounding

the aggregate. This cannot be resolved from inside the instrument, and Section 13.6 treats it as one of the conditions that would overturn the finding. The result is reported with that qualification attached rather than without it.

3.6 Functional form

H6. *The response of wintery search interest to temperature is nonlinear, with a threshold below which sensitivity is concentrated and above which it is absent.*

The practitioner formulation of this is that the season turns on a switch rather than a slope: there is a week in which it becomes cold enough, and before that week the temperature does not much matter. Oh, Jo and Ha (2021) find a threshold of approximately six consecutive days of rapid cooling in New York, which is a threshold in duration rather than level, but establishes that the response is not linear.

H10. *The rate of temperature change matters more than its level: a sharp drop triggers search interest more than sustained cold does.*

H10 is the natural competitor to H6 and is tested by entering level and first difference jointly. Its plausibility rests on salience: Busse and colleagues (2015) found four-wheel-drive purchases elevated for weeks after a snowstorm, which is a response to an event rather than to a state.

3.7 Directional asymmetry

H7. *Season entry is driven by the arriving weather and season exit by the departing weather, such that autumn cold pulls the wintery season in while spring warmth pushes it out.*

This hypothesis has an unusual provenance and it is recorded accurately. It was not specified at the outset. It was constructed after a single-market analysis of Paris appeared to support it strongly, and it was then tested on the remaining markets. Its status is therefore exploratory, and Chapter 9 reports that it did not survive.

The refined proposition it was built to capture is that a warmer January means fewer people looking for coats, and a warmer March means more people looking for shorts — an asymmetry in which warmth is the active agent and cold does comparatively little. That narrower claim is separable from H7 and is tested alongside it.

3.8 Moderators and boundary conditions

H8. *Markets with milder winters respond more strongly than markets with harsh ones, because a larger share of consumers lack adequate outerwear and must acquire it.*

This is a saturation hypothesis: in Munich everyone already owns a coat, so a cold week generates no purchase, whereas in Lisbon it does. It is tested by regressing each market's estimated coefficient on that market's seasonal temperature amplitude. Roth Tran (2023) provides the theoretical frame in distinguishing weather shocks from climate adaptation.

H9. *Cold and wet conditions compound, producing a response greater than either alone.*

The practitioner intuition is that a cold dry week is tolerable and a cold wet week is what sends people to buy a coat. It is tested with an interaction term between temperature and precipitation anomalies.

H11. *Consumers habituate within a season, so an identical anomaly produces a smaller response later in the season than earlier.*

If the trigger is a threshold crossing, the first crossing should matter more than the tenth. This is tested by interacting the anomaly with elapsed season progression.

3.9 Summary of hypotheses

Table 1 records each hypothesis, its provenance, the chapter reporting its outcome, and — stated here rather than after the fact — what evidence would count as failure.

Table 1 — Hypotheses, their sources, and where each is tested

	PROPOSITION, IN BRIEF	SOURCE	TESTED IN
H1	Weather moves garment search	Practice + literature	Ch. 7
H2	Warm autumn delays wintery onset	Practice	Ch. 8
H2a	Warm years compress the wintery window	Practice	Ch. 8
H3	Warm spring advances summery onset	Practice	Ch. 8
H4	Minimum temperature drives wintery interest	Practice	Ch. 7
H5	Weather changes total garment interest	Literature (contested)	Ch. 9
H6	The response is threshold-shaped	Practice + Oh et al. (2021)	Ch. 9
H7	Entry driven by cold, exit by warmth	Formed mid-study	Ch. 9
H8	Milder climates respond more strongly	Practice	Ch. 10
H9	Cold and wet compound	Practice	Ch. 10
H10	Rate of change beats level	Literature (salience)	Ch. 9
H11	Consumers habituate within season	Formed mid-study	Ch. 10

Provenance is recorded because it bears on evidential weight. H7 and H11 were formulated after the study began, in response to intermediate results, and are exploratory. The remaining ten were specified before estimation.

PART II

Method

CHAPTER FOUR

Research design

4.1 Philosophical position

The study is positivist in the ordinary working sense: it treats the relationship between weather and garment search interest as a regularity existing independently of the observer, approachable through measurement, and knowable only provisionally. It is not naïvely so. The central quantity is not observed directly but inferred from a proxy, and the proxy is produced by a commercial instrument whose internal workings are undisclosed and whose behaviour has changed before (Lazer et al., 2014). A position that treated the measurements as transparent windows onto reality would be untenable here.

The working stance is therefore closer to what might be called disciplined fallibilism: every quantity is a construct with known limitations, those limitations are enumerated, and the design is arranged so that the most likely failure modes would announce themselves. Placebo controls, out-of-sample validation and jackknife resampling are carried through every specification for this reason, and Chapter 12 sets out, in detail, the occasions on which the analytical process nonetheless produced a wrong answer.

One consequence deserves stating. A study in this position cannot claim to have established a causal mechanism. It can claim to have established a robust conditional association, to have excluded several classes of artefact, and to have failed to falsify a directional prediction across an international panel. That is a weaker claim than causation and a stronger one than correlation, and it is the claim made throughout.

4.2 Research strategy

The strategy is quantitative, observational and panel-based. No experiment is possible: weather cannot be assigned, and the counterfactual season cannot be run. What compensates is the panel structure. Forty-one markets experience partially independent weather while sharing an instrument, a method and a definition of the outcome. A result that appears in one market is a finding about that market and quite possibly about its data. A result that appears with the same sign in twenty-four markets out of twenty-eight is harder to attribute to artefact, because the artefact would have to operate identically across languages, hemispheres, climates and Trends geographies.

This is why per-market sign counts are reported alongside every pooled estimate in Part III, and why the binomial probability of the observed sign pattern is given. A pooled coefficient can be produced by a handful of influential markets; a sign count cannot.

4.3 Unit of analysis

Two units are used, and conflating them would be an error.

- **The market-week** is the unit for intensity models (H1, H4, H5, H6, H9, H10, H11). There are 426 to 429 weeks per market, the three-week difference reflecting the lag structure of

the asymmetric specifications. The question at this unit is whether an unusual week produces unusual search.

- **The market-year** is the unit for timing models (H2, H2a, H3, H7). There are at most eight per market for the wintery season and seven for the summery. The question at this unit is whether an unusual season produces a shifted onset date.

The disparity in sample size between the two is the reason the timing results are the more fragile of the two families, and the reason Section 13.3 identifies panel extension as the highest-value improvement. The intensity models have tens of thousands of observations; the timing models have 224 and 194 respectively.

4.4 Design of the panel

The panel is unbalanced by construction. Markets differ in the number of entities that survived volume screening, in the length of usable Trends history, and in whether a season onset can be defined at all. Rather than trimming to a balanced panel — which would have meant discarding the strongest markets to match the weakest — the analysis reports coverage per market and treats completeness as a covariate to be examined. Section 13.4 reports that composite strength does vary with the number of surviving entities, which is a limitation this design choice makes visible rather than hides.

Market fixed effects are used in all pooled specifications. This absorbs everything time-invariant about a market — its climate, its language, its search penetration, its retail structure — and identifies the effect from within-market variation across years. It is the appropriate estimator here because the cross-sectional differences between Munich and Jakarta are not the object of interest and would dominate any pooled estimate that did not remove them.

4.5 Ethics and data governance

The study uses no personal data. Google Trends returns an aggregated, normalised index computed over an undisclosed sample of queries; no individual query, user or session is observable, and the instrument is designed so that it cannot be. ERA5 is a physical reanalysis product containing no human subject data. No consent process was required and none was sought.

Two governance points nonetheless apply. First, the Trends interface used is undocumented — the endpoints backing the public web application rather than a published API. Retrieval was rate-limited and cached to avoid imposing load, and the study makes no claim that this access route is guaranteed to persist. Second, the study is unfunded and unaffiliated, and has no commercial relationship with any party named in it, including the former employer whose operations motivated the hypothesis.

4.6 Reproducibility

Both data pipelines are documented and reproducible without proprietary access, paid subscription or credentialed API.

The reporting standard adopted is that of Mavragani and Ochoa (2019), which requires keyword, region, period and category to be documented for every Trends series; this thesis additionally logs retrieval date for every series, a discipline the sampling literature makes necessary even though it is not part of Mavragani and Ochoa's own framework. Appendix F records where this study meets these standards and where it does not. The principal shortfall is that retrieval dates vary across markets by up to several weeks, because the panel was assembled incrementally, and the sampling literature establishes that the same query retrieved on different dates can return different series (Franzén, 2023; Cebrián and Domenech, 2024).

CHAPTER FIVE

Data

5.1 Weather

5.1.1 *Source and justification*

Weather data are drawn from the ERA5 reanalysis (Hersbach et al., 2020), accessed through the Open-Meteo archive interface (Zippenfenig, 2023). ERA5 is a physically consistent reconstruction of the atmospheric state produced by the European Centre for Medium-Range Weather Forecasts, available globally from 1940 at hourly resolution on an approximately 31 km grid, with a latency of about five days.

Reanalysis was preferred to station observation for three reasons. It is spatially continuous, so every market can be sampled at a comparable location rather than at whatever station happens to exist. It is free of the artefacts that afflict long station records — relocation, instrument change, urbanisation of the surroundings — which would otherwise contaminate the secular trend that this study explicitly removes. And it is internationally consistent, which matters for a panel spanning four continents and several national meteorological services with different practices.

The cost is that reanalysis is a model output constrained by observations rather than an observation. Its fidelity for near-surface temperature has been evaluated: Liu and colleagues (2024) benchmark ERA5 globally against station references over roughly seventy years and find good overall agreement, with accuracy weakest in regions of sparse weather-station coverage and temperature consistently outperforming precipitation. Tarek, Brissette and Arsenault (2020) show ERA5 performs comparably to station observations as hydrological model forcing across most of 3,138 North American catchments, with station-driven modelling retaining a clear edge only in the eastern United States. Since every market in this panel is a low-elevation urban centre, and since the analysis uses standardised anomalies rather than absolute levels, the residual error is small relative to the quantity of interest.

5.1.2 *Variables retrieved and derived*

For each market, daily series were retrieved from 2006 to 2026 for maximum, minimum and mean temperature, and precipitation total. From these, weekly aggregates were constructed together with nine derived quantities: heating degree-days on a 15.5 °C base, cooling degree-days on an 18 °C base, rain-day counts at a 1 mm threshold, coldest night, hottest day, apparent temperature, precipitation hours, sunshine hours and maximum wind speed.

The derived variables exist because the human response to temperature is not linear in degrees. A week averaging 8 °C with three nights below freezing is a different experience from a week averaging 8 °C with none, and a count of days beyond a threshold captures that where a weekly mean does not. This is also the form in which the closest antecedent found its effect: Oh, Jo and Ha (2021) identify a threshold in consecutive cooling days rather than in temperature level.

5.2 Search interest as the object of study

5.2.1 *What is measured, and why*

This thesis measures weekly Google Trends search interest for a defined set of garment concepts, and treats it as the object of study in its own right rather than as an approximation of a sales figure. Two properties recommend it beyond availability. It is behavioural rather than declared — it records what a market was actually searching for, week by week, rather than a stated intention. And it aggregates across every retailer, channel and price point operating in a market simultaneously, which a single firm's sales record cannot do: a retailer's own sales reflect that firm's pricing, assortment and promotional calendar as much as they reflect underlying seasonal interest, whereas an aggregate search series is largely free of any one firm's decisions.

Prior literature has separately validated search interest as tracking retail sales, which supports its use here without this thesis needing to establish it again. Lind and Ramesh (2025) show a search-intensity index nowcasts firm-level retail sales for a panel of US retailers; Golovanova and Zubarev (2021) show Trends categories improve out-of-sample nowcasts of aggregate retail sales. Within fashion specifically, Skenderi and colleagues (2024) forecast sales of new items from Trends validated against real retailer data, and Oh, Jo and Ha (2021) use winter-jacket Trends for exactly the timing question addressed here.

The limitation this framing does not remove is timing: search precedes any resulting purchase by an unknown and probably variable interval, and includes engagement that never converts at all. Section 13.1 treats this as the central limitation of the thesis, and Section 14.3 proposes a partial test that does not require proprietary data.

5.2.2 *Entity identifiers rather than text*

A methodological choice with substantial consequences was made at the outset: queries were submitted as Knowledge Graph entity identifiers rather than as text strings.

The disambiguation case is easily demonstrated. The entity `/m/ohnnb` denotes the umbrella as an object; a text query for "umbrella" in an English-language market returns a series contaminated by a television series, a corporation in a video-game franchise and a popular song. The entity `/m/o1bfm9` denotes shorts as a garment, distinct from an actor of that surname and from a category of confectionery. Woloszko (2020) makes the identical argument with the fruit-versus-company ambiguity of "apple".

The cross-language case is stronger still, and it is what makes an international panel feasible at all. The same identifier resolves to the appropriate lexical item in every language: `/m/ohnnb` returns results labelled *Umbrella* under an English interface and *Parapluie* under a French one, and the underlying series is identical. This removes translation from the pipeline entirely, and with it a failure mode in which a poorly chosen translation is indistinguishable from genuinely absent search interest — a failure that would be invisible in the output and fatal to a cross-market comparison.

The caveat, which Woloszko states and which this thesis adopts, is that the content of an entity is opaque. Google's clustering of queries into topics is algorithmic and undisclosed, it can make

arbitrary choices, and it may change. The study cannot verify what is inside /m/01xygc beyond observing that the series behaves as a coat series should.

An untested preference. As Section 2.5.4 established, no published study compares entity-identifier and text-term queries empirically. The preference adopted here rests on a normative argument (Woloszko, 2020) and on the internal comparison at Section 5.3.2, not on evidence from the literature. A reader is entitled to regard this as an assumption rather than a settled methodological point.

5.2.3 *Curated against machine-generated entities*

Two families of identifier exist. Curated entities carry /m/ prefixes; machine-generated ones carry /g/. The study uses /m/ exclusively, and a direct comparison motivated the rule. The machine-generated entity /g/11h79tz_t9, labelled "Winter Coat", returned zero in 147 of 209 Paris weeks. The curated /m/01xygc, labelled "Coat", returned no empty weeks over the same period. Machine-generated entities are sparse enough to be unusable at weekly resolution in all but the largest markets.

5.3 The entity set

5.3.1 *Composition*

Fifteen entities were selected across four categories: six wintery, five summery, two rainwear and two placebo. The full identifiers and their resolutions are given in Appendix A; Table 2 records the design logic.

Table 2 — Composition of the entity set

CATEGORY	N	MEMBERS	ROLE IN THE DESIGN
Wintery	6	coat, sweater, boots, gloves, scarf, beanie	Primary outcome for H2, H4, H6
Summery	5	shorts, sunglasses, swimwear, sandals, flip-flops	Primary outcome for H3
Rainwear	2	umbrella, raincoat	Precipitation channel; independent confirmation
Placebo	2	shoes, belt	Control that must return null

Categories are named wintery and summery rather than cold and warm because a coat is a warm garment worn in cold weather, and the ambiguity would fall exactly where precision is needed.

The composite is formed as the unweighted mean of member series after standardisation. Unweighted averaging is a deliberate choice over principal components or volume weighting: it keeps the composite interpretable, prevents a single high-volume entity such as boots from dominating, and avoids fitting the weighting to the outcome, which would compromise the out-of-sample gate.

5.3.2 *An internal comparison of query types*

Since the literature offers no empirical evaluation of entity against text queries, a small internal comparison was run on Paris. Text queries for the French lexical equivalents were retrieved alongside the entity series for the same period. Two differences emerged: the text series carried visible contamination in three of the fifteen slots, and the text series required a translation

decision per market that the entity series did not. This is a demonstration rather than an evaluation — it establishes that the choice matters, not that entities are superior on any general criterion — and it is reported as such.

5.4 The placebo controls

Two entities, shoes and belts, were designated in advance as weather-neutral. They are worn year-round and there is no plausible mechanism by which temperature or rainfall should move interest in them. They are carried through every specification as a control that must return a null result. Where it does not, the specification is treated as contaminated by residual seasonality and its substantive result is set aside — a rule applied in Section 10.4, under which eleven market-driver combinations fail the placebo criterion and are excluded from interpretation, including the specification with the second-largest coefficient in the panel.

This is not a formality, and the reason is instructive. The initial placebo set included socks. Socks proved genuinely cold-sensitive in the French market at $t = -3.84$, which caused the composite control to appear to fail and cast doubt on an otherwise sound specification. The episode is reported in full at Section 12.3. Its lesson is that a placebo must be verified empirically rather than nominated on grounds of plausibility, since plausibility is precisely the judgement that failed.

The design follows the logic Lazer and colleagues (2014) recommend after Google Flu: a search-based instrument requires continuous validation against something whose answer is known in advance. Where no ground truth exists — and here none does — a control that must return zero is the closest available substitute.

5.5 The market universe

The study began with fifty cities selected for their significance to the global fashion industry across design, retail, manufacturing and consumption. Nine were removed for cause, leaving forty-one. Table 3 gives the resulting distribution by seasonal amplitude, which is the variable that determines whether the central hypotheses are testable in a market at all; Table 21, in Appendix B, lists all fifty cities by name with the group to which each is assigned.

Table 3 — The market universe by seasonal temperature amplitude

BAND	N	CHARACTER	ROLE IN THE ANALYSIS
$\geq 20\text{ }^{\circ}\text{C}$	21	Continental, and three hot-desert markets	Northern panel
$12 - 20\text{ }^{\circ}\text{C}$	7	Maritime, Mediterranean and coastal	Northern panel
$6 - 12\text{ }^{\circ}\text{C}$	7	Southern extratropical, highland, subtropical	Southern panel; boundary
$< 6\text{ }^{\circ}\text{C}$	6	Equatorial	No definable season; boundary
Total	41		

The $12\text{ }^{\circ}\text{C}$ amplitude threshold divides the panel into 28 markets with a defined season and 13 without one.

Amplitude alone conflates two different reasons a market falls below the threshold, and the panel is reported throughout this thesis as four groups rather than two. The 28 markets above

the threshold are all in the northern hemisphere and are referred to as the northern panel; they carry every pooled estimate in Chapters 7 to 9. Three markets below the threshold — São Paulo, Johannesburg and Sydney, at 6.5, 9.9 and 11.1 °C respectively — sit at extratropical latitudes (23.5° S or further from the equator) despite their gentler amplitude, most plausibly because of oceanic moderation rather than the absence of a season; they are reported separately as the southern panel and are examined in Section 10.1. The remaining ten markets, including Jakarta at 3.9 °C and 6.2° from the equator, are genuinely tropical: too close to the equator for a thermal season to exist at all, and treated in Section 10.3 as the boundary condition beyond which the effect is expected to disappear.

5.6 Exclusions

Nine markets were removed, and the measurements that removed them are given in Table 4 rather than asserted.

Table 4 — Markets excluded, with the measurement that excluded them

MARKET	GROUND	MEASUREMENT
Shanghai	No usable data	mean index 0.5; 206–208 of 209 weeks exactly zero
Beijing	No usable data	as above
Guangzhou	No usable data	as above
Shenzhen	No usable data	as above
Chengdu	No usable data	as above
Hangzhou	No usable data	as above
Porto	Insufficient volume	below screening threshold at regional and national scope
Hong Kong	No definable winter	wintery composite $t = -0.95$; onset wrongly signed
Taipei	No definable winter	onset wrongly signed

The six mainland Chinese markets are inaccessible because Google services are unavailable there; the measured series confirms this rather than assuming it. Reaching these markets would require Baidu Index data, which is not freely available.

The exclusion of Hong Kong and Taipei is substantive rather than technical and deserves a defence. Both are subtropical and possess no winter of sufficient amplitude for a seasonal onset to be defined. They measure the absence of the phenomenon rather than the phenomenon, and their inclusion would depress the panel estimate without informing it. This reasoning could be abused — excluding markets that fail to confirm a hypothesis is the classic route to a spurious result — so the discipline applied was that exclusion required a criterion independent of the outcome. Seasonal amplitude is such a criterion; it is a property of the climate, fixed before any model was estimated. Thirteen further markets with low amplitude were retained precisely because they were not excluded on this basis, and they are reported at Section 10.3 as a boundary condition.

5.7 The assembled panel

Table 5 — Descriptive statistics of the assembled panel

DIMENSION	VALUE
Markets	41
Temperate markets (amplitude ≥ 12 °C)	28
Weeks per market, maximum	429
Period covered	27 May 2018 – 9 August 2026
Entities per market, maximum	15
Markets retaining all 15 entities	13
Model specifications estimated	328
Market-years, wintery onset	224
Market-years, summery onset	194
Weather series, first year retrieved	2006
Weather variables per market	13 (4 retrieved, 9 derived)

The gap between 41 markets and 13 markets retaining all entities is the batching defect reported at Section 12.4.

5.8 Relation to transaction data

No transaction data of any kind — sell-through, units, revenue, footfall or basket — was used in this study; every result in Part III describes search interest. The inferential step from search interest to purchase behaviour is untested here and is treated formally as the central limitation of the thesis at Section 13.1, which also identifies what would be required to test it.

CHAPTER SIX

Analytical methods

6.1 Temporal alignment

Google Trends reports weekly data on a Sunday-to-Saturday convention. Weather data were therefore aggregated to the identical convention rather than to ISO weeks, which begin on Monday.

This is not cosmetic. Section 6.3 reports a direct test in which a climatology built on calendar dates — one that ignores the weekly structure entirely — loses between ten and fifteen per cent of explanatory power relative to the aligned specification. A one-day misalignment between two weekly series introduces an averaging error in every observation, and at a resolution where the effect of interest is measured in days, that error is not negligible.

6.2 Construction of weather anomalies

6.2.1 *Why anomalies are necessary*

Garment search interest and temperature both follow a strong annual cycle. Correlating them in levels returns a coefficient near -0.9 which reflects nothing but the shared seasonality of the two series, and would be obtained equally by correlating either against a sine wave. The quantity of interest is not whether coats are searched for in winter — they are — but whether an unusually cold week produces unusually high coat interest relative to what that week of the year normally delivers.

6.2.2 *Detrended per-week-of-year standardisation*

For each of fifty-two week-of-year buckets, and separately for each variable and each market, the expected value was estimated by ordinary least squares on the calendar year:

$$\text{value}(w, y) = \alpha(w) + \beta(w) \cdot y + \varepsilon(w, y) \quad (6.1)$$

and the anomaly defined as the standardised residual within that bucket:

$$\text{anomaly}(w, y) = \varepsilon(w, y) / \text{SD}[\varepsilon(w)] \quad (6.2)$$

Two features of this construction do necessary work.

The linear term in year removes the secular warming trend. Without it, a flat multi-year mean would be systematically too cold for recent years and would manufacture an apparent warm anomaly in every modern observation — a bias that would grow monotonically across the panel and would be indistinguishable from a genuine effect if the outcome also trended. Given that the observation window is eight years within a warming climate, this is not a hypothetical concern.

The per-bucket standard deviation addresses heteroskedasticity across the calendar. In Paris, late-January temperature varies 1.9 times more between years than mid-August temperature: residual standard deviations of 3.48°C and 1.85°C respectively. A single pooled standard deviation across all weeks would be 6.28°C — a figure dominated by the seasonal cycle rather

than by year-to-year variation — and would render almost every genuine anomaly invisible by dividing it by a quantity an order of magnitude too large.

6.2.3 Validation of the baseline definition

Three candidate climatologies were constructed and compared directly against the same search data: a calendar-date baseline grouping observations by month and day; a day-position baseline grouping by day-of-week within week-of-year; and the weekly-aggregate baseline described above.

Table 6 — Comparison of three climatology definitions, Paris, 555 weeks

SPECIFICATION	A: CALENDAR DATE	B: DAY-IN-WEEK	C: WEEKLY MEAN
Wintery ~ maximum temperature	0.326	0.357	0.354
Wintery ~ mean temperature	0.282	0.311	0.311
Rainwear ~ precipitation	0.497	0.519	0.514
Summery ~ maximum temperature	0.210	0.216	0.222

Values are R^2 . Definitions B and C are statistically indistinguishable and correlate above 0.99; A is consistently inferior across all four outcomes. Definition C was retained on grounds of parsimony, requiring no daily-resolution intermediate step.

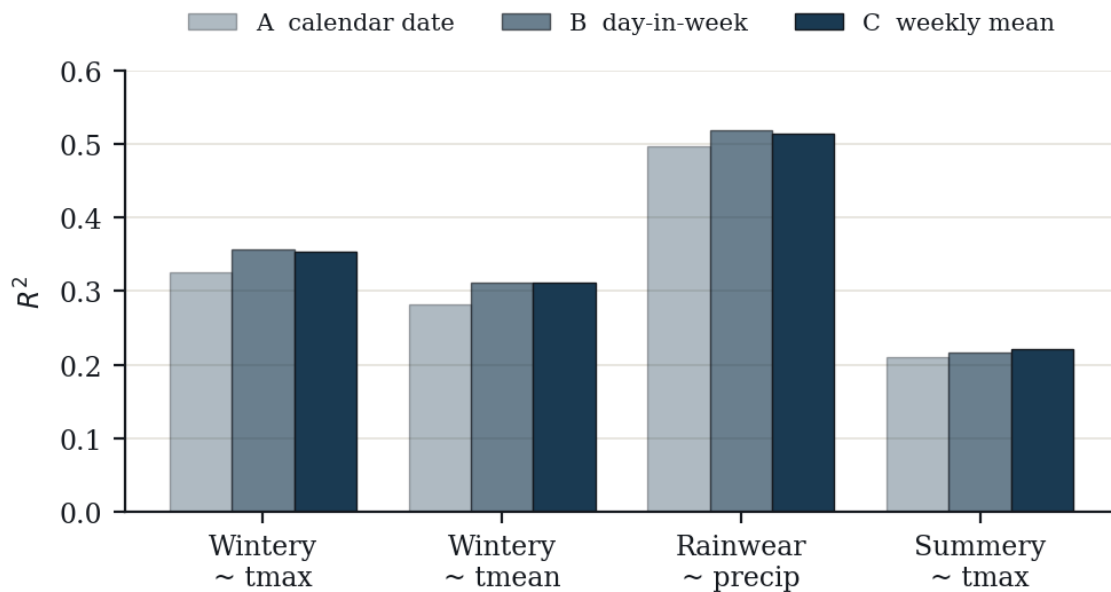


Figure 1 — Comparison of three climatology baseline definitions, Paris. The weekly-aligned definitions (B and C) dominate the calendar-date definition (A) on every outcome tested. The margin is between ten and fifteen per cent of explanatory power.

6.3 Construction of the search series

6.3.1 Window length

Google Trends returns weekly resolution only for query windows shorter than approximately five years; longer requests are silently downgraded to monthly. The word silently matters: no error is raised and no flag is set, so a pipeline that requests a decade receives monthly data that looks like a successful retrieval.

Monthly resolution cannot detect a season-onset shift of one to two weeks, which is the order of magnitude the phenological literature and the practitioner's own observation would lead one to expect (Section 3.3). An early version of this study was compromised by exactly this failure. History was therefore assembled from overlapping windows of under four years.

6.3.2 Chain-linking

Each Trends query is normalised to its own maximum, so adjacent windows arrive on incompatible scales and cannot be concatenated directly. Windows were spliced by the ratio of sums across their shared weeks, requiring a minimum overlap of ten observations.

$$\begin{aligned}
 \text{scale} &= \sum \text{value_A}(t) / \sum \text{value_B}(t) && \text{over shared weeks } t \\
 \text{linked}(t) &= \text{value_A}(t) && \text{for } t \text{ in window A} \\
 &= \text{value_B}(t) \cdot \text{scale} && \text{for } t \text{ in window B}
 \end{aligned} \tag{6.3}$$

The procedure follows the approach of Eichenauer, Indergand, Martínez and Sax (2022), who show that raw Trends data are simultaneously frequency-inconsistent and sampling-noisy and give a repeated-sampling and chain-linking method addressing both.

It was validated against synthetic series constructed with known scale factors of 0.6, 1.0 and 1.7. Chain-linking recovered a single consistent scale with a residual standard deviation of 0.0014 and produced no discontinuity at either seam. This is a validation of the arithmetic rather than of the instrument: it establishes that the splice does not introduce error, not that the underlying samples were drawn identically.

6.3.3 Batch composition

Up to five entities may be submitted in a single query and thereby share a common scale, which is desirable because it makes their levels comparable. However, Trends normalises the batch to its own maximum, so a high-volume entity compresses its low-volume companions toward zero. Springer, Strzelecki and Zieger (2023) formalise this: because the index is normalised within each query, cross-query comparison is invalid without an external anchor.

Entities were therefore grouped by expected magnitude rather than by category, using volumes measured in Paris:

batch 1	shoes · raincoat · belt · flipflops	1855 → 982	
batch 2	sandals · gloves · boots · scarf	761 → 477	
batch 3	coat · sweater · swim · shorts	366 → 192	
batch 4	sunglasses · umbrella · beanie	135 → 69	(6.4)

Section 12.4 reports that this calibration was performed on one market and did not transfer to the other forty. It is the most serious unresolved defect in the study, and it is a direct instance of the constraint Springer and colleagues describe.

6.3.4 Transformation and aggregation

Each series was log-transformed with a continuity correction, then deseasonalised by regression on week-of-year indicator variables and a quadratic time trend, then standardised:

$$\begin{aligned}
x(t) &= \log(\text{value}(t) + 0.5) \\
x(t) &= \gamma_0 + \sum \gamma_w \cdot \text{woy}_w(t) + \delta_1 \cdot t + \delta_2 \cdot t^2 + u(t) \\
z(t) &= u(t) / \text{SD}[u]
\end{aligned}
\tag{6.5}$$

The additive constant of 0.5 rather than 1.0 preserves weeks recorded as zero rather than dropping them. In an early version, Paris lost approximately half its observations to zero-value drops before this was corrected — an error that would have biased the sample toward high-volume weeks and therefore toward the seasonal peak, which is the worst possible selection for a study about season boundaries.

The quadratic trend term removes drift in the underlying popularity of a search concept over the eight-year window, which is distinct from the seasonal cycle and would otherwise contaminate the anomaly.

6.3.5 Reference periods for the normal and the standard deviation

Every anomaly in this thesis is a deviation from a normal, divided by a standard deviation, and the period over which each is estimated determines what an anomaly means. The weather and search series use different reference periods, for a practical reason: weather reanalysis is available for decades, while weekly search data is available only in windows of under five years and was assembled for the panel period alone. Table 6a sets out the three reference periods.

Table 6a — Reference periods used to define the normal and the standard deviation

SERIES	REFERENCE PERIOD	THE NORMAL	THE STANDARD DEVIATION
Weather (weekly)	2006–2026, about 20 years, per market	Trend line on calendar year, fitted separately for each week of the year (equation 6.1)	Residual spread around that trend, separately for each week of the year; about 20 observations each (equation 6.2)
Search (weekly)	27 May 2018 – 9 August 2026, 426–429 weeks, per market and entity	Week-of-year pattern plus a quadratic trend over the whole period (equation 6.5)	One value over the whole period, not per week of the year (equation 6.5)
Shoulder season (timing models)	The autumn or spring preceding each season, within 2018–2026	Mean of the weekly weather anomalies in that season	Not documented: see below

Southern-hemisphere markets have their seasonal windows displaced by six months.

Three consequences follow. First, a weather anomaly in the panel years is measured against a twenty-year trend, so a week is warm relative to what that week of the year has recently been, not relative to a pre-warming climate. Second, the search normal is estimated on the same eight years in which the effects are measured, so a search anomaly is relative to the market's own recent history. Third, the two standard deviations are built differently — per week of the year for weather, one value for search — so a one-standard-deviation weather anomaly is larger in winter weeks, when temperature varies more between years, than in summer weeks.

The timing models of Chapter 8 use a third quantity, and its definition was not preserved. The predictor is the mean of the weekly weather anomalies across the shoulder season preceding the season being timed, but the retained records do not state which weeks define each shoulder season in the 41-market run, nor whether the seasonal mean was re-standardised before

estimation. The Paris pilot defined autumn as September to October. The distinction matters for the unit of the headline results. If the seasonal mean was re-standardised, “one standard deviation warmer autumn” means an autumn unusually warm by the standard of autumns. If it was not, averaging several weekly anomalies produces a narrower spread, one unit corresponds to a rarer and warmer autumn than that phrase suggests, and the effect per ordinary seasonal standard deviation is smaller than the reported 4.18 and 8.25 days. The sign of the timing results does not depend on this choice, and a common rescaling of the predictor would leave their t statistics unchanged; their magnitude does depend on it. Section 13.3 records it as a limitation, and none of the timing estimates is expressed here in degrees Celsius for the same reason.

6.4 Model specifications

6.4.1 Intensity

The contemporaneous response was estimated as a distributed lag over four weeks:

$$\text{composite}(t) = \alpha + \sum \beta(k) \cdot \text{anomaly}(t - k) + e(t), \quad k = 0 \dots 3 \quad (6.6)$$

The four-week window is a compromise. A purely contemporaneous specification would miss a response that takes a fortnight to reach a search query; a longer window would consume degrees of freedom and begin to overlap the seasonal structure the deseasonalisation was meant to remove. The reported coefficient is the sum across lags, with its variance computed from the full submatrix of the covariance estimator rather than by summing individual variances, which would ignore the covariances between lags and understate the standard error.

6.4.2 Standard errors

Standard errors are Newey–West heteroskedasticity- and autocorrelation-consistent at six lags (Newey and West, 1987). Both corrections are necessary. Weekly search residuals are serially correlated because interest in a garment persists across weeks; and their variance is not constant, being visibly larger in the seasonal shoulder than at the peak or trough. An ordinary least squares standard error would be too small on both counts, and the effect would be to manufacture significance.

The choice of six lags is conventional rather than derived. It exceeds the four-week lag structure of the model, which is the minimum defensible requirement.

6.4.3 Timing

Season boundaries were located on the undeseasonalised composite, since here the seasonal shape is the object of interest rather than a nuisance to be removed. The composite was smoothed, then evaluated over a full annual cycle from July to June, and onset and offset defined as the upward and downward crossings of the half-amplitude level:

$$\begin{aligned} \text{half}(y) &= \min(S_y) + 0.5 \cdot (\max(S_y) - \min(S_y)) \\ \text{onset}(y) &= \text{first } t \text{ where } S_y(t) \text{ crosses half}(y) \text{ upward} \\ \text{offset}(y) &= \text{first } t \text{ after the peak where } S_y(t) \text{ crosses half}(y) \text{ downward} \\ \text{width}(y) &= \text{offset}(y) - \text{onset}(y) \end{aligned} \quad (6.7)$$

This is the half-amplitude rule described in Section 2.6, from the amplitude-fraction family formalised by Jönsson and Eklundh (2004) in TIMESAT. Southern-hemisphere markets have their annual window displaced by six months so that the wintery season is not split across the boundary.

Onset, offset and width were then regressed on the temperature anomaly of the preceding shoulder season, pooled across markets with market fixed effects:

$$\text{onset}(m, y) = \mu_m + \theta \cdot \text{shoulder_anomaly}(m, y) + v(m, y) \quad (6.8)$$

Why the choice of onset metric is reported this prominently. White and colleagues (2009) applied ten different start-of-season extraction methods to one identical dataset and obtained materially different onset dates and different trend signs. The choice of metric can therefore reverse a result. Section 12.2 of this thesis reports precisely that outcome: a peak-sensitive metric produced a large, significant, wrongly-signed estimate that the half-amplitude metric did not. A reader should treat any onset result — including those reported in Chapter 8 — as conditional on this methodological choice.

6.4.4 Asymmetric specification

Where an effect was suspected to operate in one direction only, the anomaly was decomposed into warmer-than-normal and colder-than-normal components entered as separate regressors:

$$\begin{aligned} a^+(t) &= \max(\text{anomaly}(t), 0) \\ a^-(t) &= \min(\text{anomaly}(t), 0) \\ \text{composite}(t) &= \alpha + \beta^+ \cdot a^+(t) + \beta^- \cdot a^-(t) + e(t) \end{aligned} \quad (6.9)$$

A difference between β^+ and β^- constitutes asymmetry. Chapter 9 reports the outcome for H7.

6.5 Validation

Four gates were applied to every specification, and all four were specified before results were seen. Table 7 records what each is designed to catch, and — because a gate that has never rejected anything is not evidence of anything — what each did in fact catch in this study.

Table 7 — Validation gates and what each caught

GATE	RULE	WHAT IT CAUGHT HERE
Placebo	$ t < 2$ on shoes and belts	Socks contamination (§12.3); 11 market failures (§10.4)
Out-of-sample	fit pre-2024, correlate after	Overfitting in the specification search
Multiplicity	Benjamini–Hochberg across the grid	False positives across 328 specifications
Jackknife	leave one year, and one market, out	The wrongly-signed phenology result (§12.2)

Every gate rejected something. The cost of running them was substantial in analyst time; the alternative was a thesis reporting four additional findings, of which at least three would have been false.

Multiplicity control follows Benjamini and Hochberg (1995), which controls the false discovery rate rather than the family-wise error rate. This is the appropriate choice for a specification grid

of 328 models in which some proportion of true effects is expected: family-wise control at this scale would be so conservative as to reject nearly everything, including the real effects.

The jackknife is applied in the leave-one-out form (Quenouille, 1956; Efron, 1979), by year within market for intensity models and by market for pooled estimates. Its function here is not variance estimation but influence detection: a pooled coefficient that collapses when one market or one year is removed is not a panel finding.

Out-of-sample evaluation follows the standard argument that it guards against specification search (Inoue and Kilian, 2005), while noting that those authors caution against treating it as automatically more reliable than in-sample testing. It is used here as one gate among four rather than as a decisive criterion.

PART III

Results

CHAPTER SEVEN

Intensity: does weather move garment search?

7.1 The enabling result

H1. *Weekly deviations of temperature and precipitation from local seasonal norms are associated with contemporaneous deviations in garment search interest.*

The proposition is supported across the markets with a defined winter, and the pattern is worth stating precisely. Twenty-seven of the twenty-eight temperate markets return a negative coefficient for the wintery composite on maximum temperature, and twenty-five are individually significant at $|t| \geq 2$. The three that are not — Riyadh, Doha and Delhi — are the three hot, dry markets of the northern panel, whose winter daytime highs typically stay around 20 °C, far above the level below which Section 9.2 locates the response; Delhi alone is positive, at $t = 0.23$. Across all forty-one markets, thirty-eight coefficients are negative. The other two positive coefficients belong to tropical markets with no winter, Bangkok and Ho Chi Minh City, the latter significant and wrongly signed at $t = 2.02$.

A sign pattern of 27 from 28 is very unlikely under a null in which the sign is a coin flip. It could in principle be produced by an artefact operating identically in every market — residual seasonality being the obvious candidate — which is why the placebo gate exists and why Section 10.4 reports its results before any of these coefficients are relied upon.

7.2 Magnitude and distribution

Table 8 — Intensity of the wintery response, strongest markets

MARKET	B	SE	T (HAC)	R ²	OOS R
Munich	−0.827	0.118	−6.99	0.362	0.593
Warsaw *	−0.826	0.133	−6.20	0.326	0.738
Copenhagen	−0.740	0.127	−5.81	0.291	0.498
Chicago	−0.713	0.123	−5.77	0.324	0.491
São Paulo *	−0.711	0.140	−5.09	0.313	0.558
Berlin	−0.707	0.113	−6.25	0.278	0.507
Antwerp	−0.696	0.115	−6.07	0.421	0.574
Zurich	−0.681	0.116	−5.85	0.325	0.489

β is in standard deviations of search anomaly per standard deviation of temperature anomaly, summed over lags 0–3. All $q < 0.001$ under Benjamini–Hochberg. * Warsaw and São Paulo fail their temperature placebos, at $t = -3.30$ and $t = -2.51$ (Section 10.4); their coefficients are reported for completeness and must not be used.

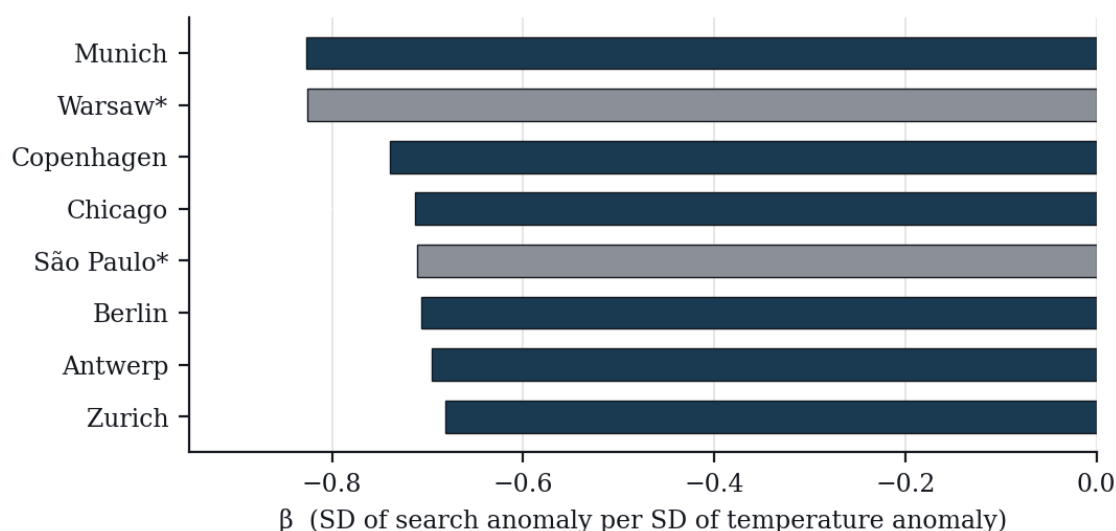


Figure 3 — Wintery intensity across the strongest markets. Warsaw and São Paulo are shown in grey because they fail their temperature placebos — Warsaw despite returning the second-largest coefficient in the panel, an illustration of why the gate is applied before the ranking is read.

A coefficient of -0.8 means that a market experiencing a week one standard deviation warmer than normal for the time of year records wintery garment search approximately 0.8 standard deviations below its own norm for that week. These are large effects by the standards of the weather-and-retail literature, but they are not directly comparable to it, because that literature measures sales in currency and this measures search in standardised units.

The distribution across the panel is continuous rather than bimodal. Coefficients run from -0.827 in Munich down to values indistinguishable from zero in the equatorial markets, with no natural break. This argues against a categorical distinction between weather-sensitive and weather-insensitive markets and in favour of a continuum governed by seasonal amplitude, which is examined directly as H8 at Section 10.2.

7.3 The precipitation channel

The rainwear response to precipitation is stronger than the thermal response where it can be measured properly.

Table 9 — Intensity of the rainwear response, strongest markets

MARKET	B	T (HAC)	R ²	OOS R
Amsterdam	+0.917	12.52	0.491	0.618
Paris	+0.889	12.03	0.513	0.762
Osaka	+0.993	11.77	0.491	0.682
Zurich	+0.994	11.62	0.524	0.755
Tokyo	+0.905	10.87	0.429	0.600

Restricted to markets retaining both rainwear entities. Fifteen markets lost the raincoat entity to the batching defect (Section 12.4) and their rainwear results are not reportable.

Two properties distinguish this channel. Its t statistics are the largest in the study and its out-of-sample correlations reach 0.76, the highest of any channel, which is unsurprising: rain is a discrete, salient, immediately consequential event, whereas a warm autumn is a gradual

condition. And rainwear is the only category examined here without a seasonal window — monthly estimation in Paris found the precipitation response significant in all twelve months. Section 11.4 argues this makes it the most promising candidate for operational use and the easiest to justify commercially.

The dependence on entity completeness is severe and instructive. Specifications retaining both umbrella and raincoat return a mean t of 7.75 and attain significance in 24 of 26 markets. Those reduced to umbrella alone return a mean t of 2.17 and attain significance in 9 of 15. The composite is not merely weaker with one member; it is a different and much noisier instrument.

7.4 Which temperature variable governs the response

H4. *Minimum temperature is the principal driver of wintery garment search interest, while maximum temperature governs summery search interest.*

The hypothesis is rejected: the second clause is supported, but the first is not, and it is contradicted in every test applied to it.

Table 10 — Competitive test of thermal drivers, Paris, 555 weeks

OUTCOME	~ MINIMUM	~ MEAN	~ MAXIMUM
Wintery composite	0.198	0.311	0.354
Summery composite	0.087	0.179	0.222
Wintery, September–November only	0.254	0.402	0.446

Values are R^2 . Maximum temperature is preferred in twelve of thirteen individual garment entities tested separately. The third row restricts the sample to the autumn shoulder, where a minimum-temperature effect should be most visible if it exists anywhere.

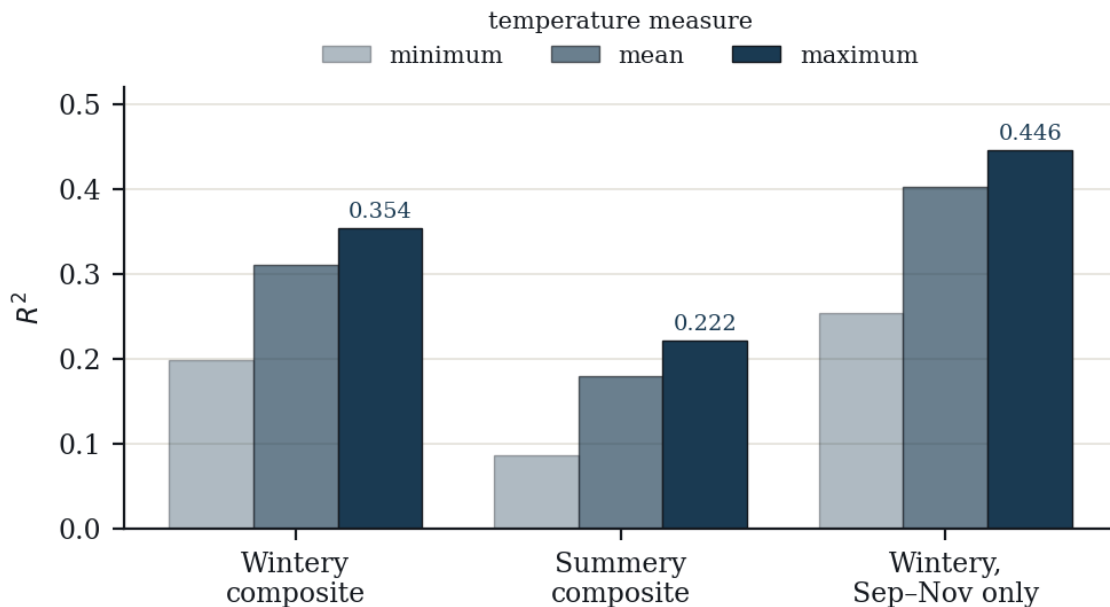


Figure 2 — Competitive test of minimum, mean and maximum temperature. The ordering is identical across both composites and survives restriction to the autumn shoulder season, where the practitioner hypothesis predicted minimum temperature should dominate.

The interpretation is that consumers respond to the temperature they experience during waking hours rather than to the overnight low. This is behaviourally coherent — a person who is cold at two in the afternoon while out of the house is in a position to act on it; a person who is cold at four in the morning is asleep — but the study did not predict it, and the prediction it did make was confident and wrong.

One qualification is essential and it materially limits the claim. The three measures are highly collinear, with correlations between 0.81 and 0.96 in the markets examined. The finding therefore establishes maximum temperature as the best available single measure; it does not demonstrate that minimum temperature is uninformative, and it cannot, because a regression cannot cleanly separate variables that move together this closely. What can be said is that no specification tested gave minimum temperature an advantage, including the one designed specifically to favour it.

The result is consistent with the two published findings closest to it. Rose and Dolega (2022) identify wind as the most influential weather variable in UK retail sales; Oh, Ha and Jo (2022a) find wind chill dominant among a wider predictor set. Both point toward experienced daytime conditions rather than toward the thermometric minimum.

7.5 Summary of Chapter 7

Weather moves garment search, reliably, everywhere the seasons are large enough for the question to mean anything. The daytime maximum is the relevant thermal measure, contradicting a confidently-held prior. Precipitation drives rainwear more strongly and more predictably than temperature drives anything, but that channel is compromised in fifteen markets by a defect reported in Chapter 12. None of this yet addresses timing, which is the subject of the next chapter and the substance of the thesis.

CHAPTER EIGHT

Timing: does weather move the fashion calendar?

This chapter reports the central results of the thesis. Chapter 7 established that weather moves the level of garment search. The question here is different and harder: whether weather moves the date on which a season begins.

8.1 Warm autumn delays the wintery season

H2. *The onset of wintery garment search interest occurs later in years when the preceding autumn is warmer than normal.*

The hypothesis is confirmed. Pooled across twenty-eight temperate markets and 224 market-years with market fixed effects, a one-standard-deviation warmer autumn delays wintery onset by 4.18 days, with a standard error of 1.42 and a HAC t statistic of 2.95.

The sign count is the more persuasive statistic. Twenty of twenty-eight markets carry the predicted positive sign, which under a binomial null with equal probabilities occurs with probability 0.018. This matters because a pooled coefficient can be manufactured by a small number of influential markets, whereas twenty independent markets agreeing on a direction cannot.

8.2 Warm spring advances the summery season

H3. *The onset of summery garment search interest occurs earlier in years when the preceding spring is warmer than normal.*

The hypothesis is confirmed. Pooled across the same twenty-eight markets and 194 market-years, a one-standard-deviation warmer spring advances summery onset by 8.25 days, with a standard error of 2.57 and $t = -3.21$. Twenty-four of twenty-eight markets carry the predicted sign, with binomial probability 0.0001.

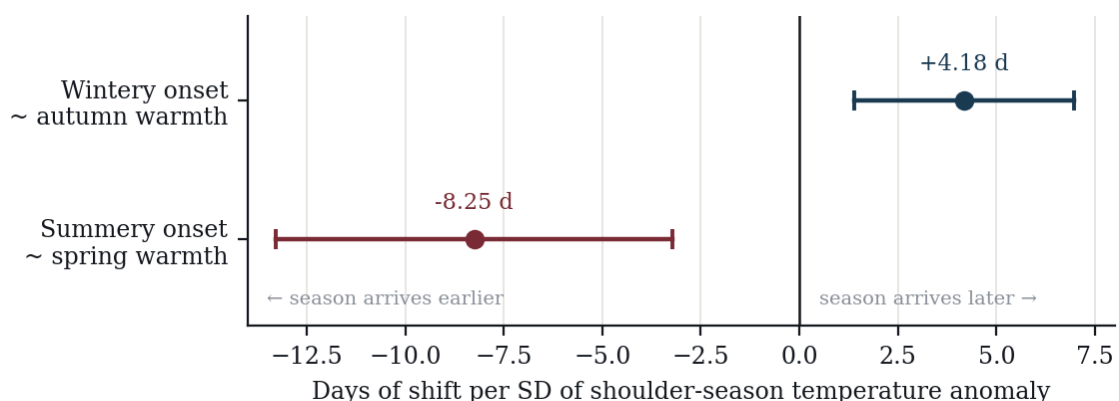


Figure 4 — The two timing effects with 95 per cent confidence intervals. Neither interval contains zero. The summery effect is roughly twice the magnitude of the wintery effect and is estimated with proportionally similar precision.

8.3 The two effects compared

Table 11 — The two timing effects compared

CRITERION	WINTER (H2)	SUMMERY (H3)
Effect, days per SD	+4.18	-8.25
Standard error	1.42	2.57
t statistic	2.95	-3.21
R ²	0.038	0.051
Markets with predicted sign	20 / 28	24 / 28
Binomial p on the sign count	0.018	0.0001
Market-years	224	194
SD of observed onset dates, days	11	16
Effect ÷ natural spread	0.38	0.49
Independent replications	3	1

The final row records that H2 has been estimated on three separate datasets and H3 on one. Section 8.5 discusses what the replication history shows.

The summery effect is larger, more consistent across markets and explains more variance. Two readings are available and the study cannot adjudicate between them.

The first is behavioural: summer entry may simply be a sharper decision than winter entry. Consumers appear to respond decisively to the first genuinely warm week, whereas the movement into winter clothing is gradual, spread across several garment types acquired at different moments — a scarf in October, a coat in November, boots when it rains. A composite of six wintery entities acquired over six weeks will have a blurrier onset than a composite of five summery entities acquired over one.

The second is methodological: the summery composite may simply have a sharper annual profile, making the half-amplitude crossing easier to locate and therefore less noisy. Since the onset metric is applied identically to both, a difference in the underlying curve shape translates directly into a difference in measurement precision. Distinguishing these would require the metric-comparison exercise described at Section 14.7, which was not performed.

8.4 Stability of the estimates

Both pooled coefficients were subjected to leave-one-market-out jackknife across all twenty-eight temperate markets — fifty-six re-estimations in total.

Table 12 — Jackknife stability of the pooled timing estimates

	POINT ESTIMATE	JACKKNIFE RANGE	MINIMUM T ACROSS FOLDS
Wintery onset (H2)	+4.18	3.69 to 4.91	2.54
Summery onset (H3)	-8.25	-6.98 to -9.61	2.89

No single market's removal moves either estimate outside the range shown, and no fold reduces the absolute t statistic below 2.5. Neither result depends on any individual market.

This is the test that the earlier phenology analysis failed. Section 12.2 reports a single-market result whose t statistic fell from -2.74 to -1.17 on removal of one year, revealing it as an artefact of one influential observation. The pooled estimates here do not behave that way, and the contrast is the reason the jackknife was retained as a standing gate rather than run once.

8.5 The shrinking estimate

H2 has now been tested three times on progressively better data, and the estimate has fallen monotonically each time.

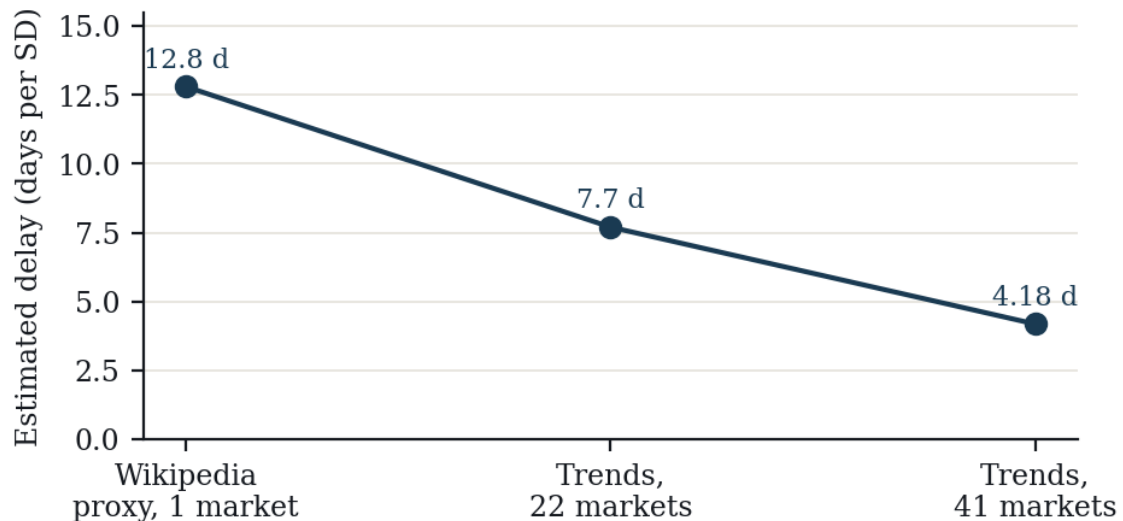


Figure 5 — History of the autumn-delay estimate across three datasets. The decline from 12.8 to 4.18 days is characteristic of an early estimate inflated by a small, noisy sample regressing toward a smaller true value as the panel expands.

The first estimate, 12.8 days per standard deviation, used Wikipedia pageviews as the proxy on a single market. The second, 7.7 days, used Google Trends across twenty-two markets. The present estimate, 4.18 days, uses Trends across forty-one.

This pattern is well understood and it is not reassuring about the first two numbers. An effect estimated on a small, noisy sample and selected for reporting because it was significant will be biased upward in magnitude, because only the large realisations clear the significance threshold. As the sample grows the selection pressure weakens and the estimate falls toward the truth. The honest reading is that 4.18 days is the current best estimate, that the earlier figures were inflated, and that a further expansion of the panel could plausibly reduce it again — which is a reason to regard the effect as real but modest rather than as established at any particular magnitude.

What did not change across the three estimates is the sign. Across two instruments, three sample sizes and eight years, a warm autumn has never been associated with an earlier wintery season.

8.6 Season width

H2a. *The wintery selling window compresses in warm years: it begins later and ends earlier, so its width contracts.*

The hypothesis is partly supported. Regressing wintery season width on the same shoulder-season anomaly across the twenty-eight temperate markets returns -10.67 days per standard deviation with $t = -2.50$: the wintery window does contract in warm years. The mechanism the hypothesis specifies, however — a later start and an earlier end — is only half visible. The offset regression, retained only in its form pooled across all forty-one markets, returns $+1.80$ days per standard deviation with $t = 0.64$: no detectable movement in the end of the wintery season. The contraction therefore rests on the delayed start, and the result is a shorter window rather than a window compressed from both ends.

Confidence is lower than for the two onset results for a structural reason. Width is a difference between two estimated quantities, so its error is compounded from both, and the offset estimate is the noisier of the two: a season ends more gradually than it begins, making the downward half-amplitude crossing harder to locate. The width result should be read as consistent with a shorter season rather than as independent evidence for compression.

8.7 Summary of Chapter 8

Both central hypotheses are confirmed. A warm autumn delays the wintery season by approximately four days per standard deviation of anomaly; a warm spring advances the summery season by approximately eight. Both estimates are stable to the removal of any market, both are supported by market-level sign counts that are unlikely under a null, and the wintery result has now survived two changes of instrument. Section 10.6 applies the placebo rule and a near-tropical restriction to the sign counts: both results survive, the summery one more robustly. The effects are real, directionally consistent with two decades of practitioner observation, and considerably smaller than that observation supposed.

CHAPTER NINE

The structure of the response

Chapters 7 and 8 established that weather moves garment search and moves the calendar. This chapter asks what kind of relationship it is: whether it changes the total or only the mix, whether it is linear or threshold-shaped, and whether it operates symmetrically in both directions.

9.1 Volume against mix: the principal finding

H5. *Weather anomalies change the total quantity of garment interest, not merely its distribution across categories.*

The hypothesis is rejected, and its rejection is the most consequential result in this thesis: weather does not change how much garment interest exists overall, only which categories it attaches to. A composite of all garment entities excluding placebos was regressed on the temperature anomaly in each of the twenty-eight temperate markets. The mean coefficient is -0.076 , the mean t statistic is -0.61 , and the relationship attains significance in only two of twenty-eight markets — a rate barely distinguishable from the false-positive rate one would expect at conventional thresholds.

Set against category-level responses averaging $t = -4.31$ for wintery and $t = +3.44$ for summery in the same markets on the same weeks, the implication is unambiguous.

COMPOSITE	MEAN B	MEAN T	SIGNIFICANT
Wintery	-0.519	-4.31	25 / 28
Summery	$+0.408$	$+3.44$	21 / 28
All garments, aggregated	-0.076	-0.61	2 / 28

The categories move strongly and in opposite directions; their sum does not move. Weather redistributes garment interest rather than changing how much of it exists.

Two independent literatures predicted this and the study did not. Starr-McCluer (2000) found weather effects on US retail sales offsetting across lags so that the influence largely washed out at quarterly frequency. Bahng and Kincade (2012), on a single womenswear retailer's actual sales, found that day-to-day and week-to-week temperature changes did not alter whole-season volume. The present result reproduces that finding on independent data, a different instrument and across twenty-eight markets.

What weather appears to do is hold consumers in one category longer rather than remove them from the market. A warm October does not reduce interest in clothing; it keeps that interest attached to lighter garments. Section 11.3 develops the commercial consequence, which is that the problem this creates is one of allocation rather than of forecasting.

The qualification that cannot be removed. Google Trends reports a normalised share of total search rather than an absolute volume. A null on an aggregate composite is therefore consistent with genuine substitution and also with the normalisation mechanically constraining the aggregate — if the index is bounded, a sum across most of its constituents may be bounded too. This cannot be resolved from inside the instrument. Resolving it requires either an external

anchor of the kind Springer, Strzelecki and Zieger (2023) propose, or transaction data. Section 13.6 lists it among the conditions that would overturn the finding, and it is the single qualification most likely to weaken it.

9.2 The shape of the temperature response

H6. *The response of wintry search interest to temperature is nonlinear, with a threshold below which sensitivity is concentrated.*

The hypothesis is confirmed. Observations from twenty-eight temperate markets were pooled and binned by temperature level in 2.5 °C intervals, with the mean deseasonalised wintry composite computed within each bin.

Table 13 — Wintry composite by temperature band, pooled across 28 markets

WEEKLY MAXIMUM TEMPERATURE	MEAN COMPOSITE (SD)	WEEKS
0.0 – 2.5 °C	+0.434	283
2.5 – 5.0 °C	+0.189	428
5.0 – 7.5 °C	+0.025	563
7.5 – 10.0 °C	–0.009	805
10.0 – 15.0 °C	+0.003 to +0.036	1,883
15.0 – 20.0 °C	+0.021 to +0.063	1,911
20.0 – 25.0 °C	–0.025 to –0.145	2,178

The response is confined to the cold tail. The steepest transition is between the first two bands, a step of 0.245 standard deviations across 2.5 °C. Between 7.5 and 20 °C the temperature level has no detectable effect.

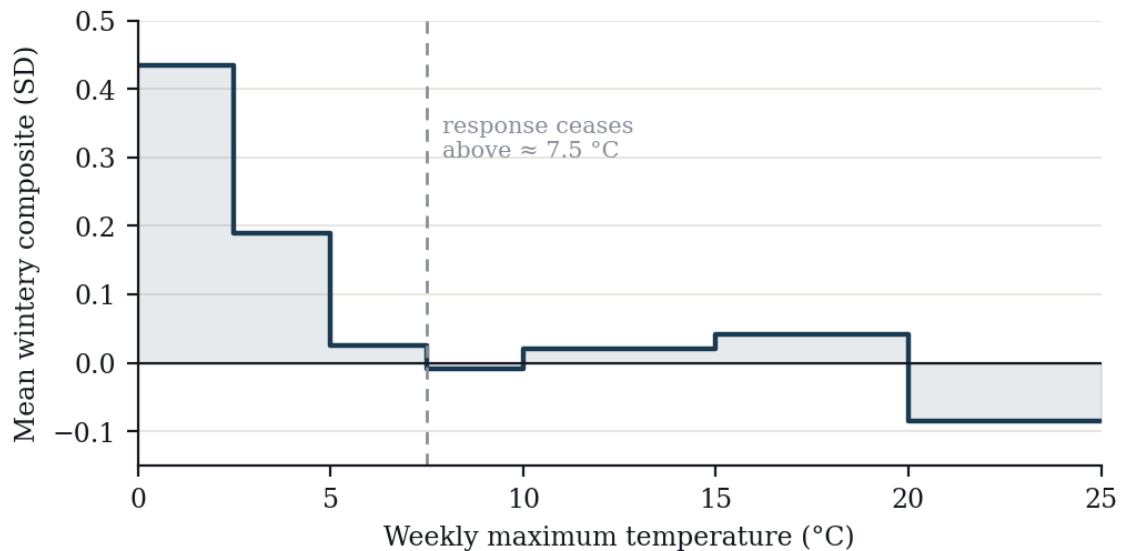


Figure 6 — The threshold profile of the wintry response. Below approximately 7.5 °C the temperature level matters considerably; above it, not at all. This is closer to a switch than to a slope.

The practical reading is that the relationship is closer to the practitioner's mental model than a linear specification would suggest. There is a temperature below which people start buying coats, and above it the precise reading of the thermometer is nearly irrelevant. This also corroborates the one published antecedent: Oh, Jo and Ha (2021) found search activity for

winter jackets in New York beginning after approximately six consecutive days of rapid cooling — a threshold expressed in duration rather than level, but a threshold.

A threshold is more actionable than a slope, because it can be monitored against a forecast. Section 11.5 explains why this result is nonetheless not yet usable operationally.

9.3 Level against rate of change

H10. *The rate of temperature change matters more than its level: a sharp drop triggers search interest more than sustained cold.*

The hypothesis is rejected. Entering the temperature level and its first difference jointly, the level dominates decisively at $\beta = -0.692$ with $t = -12.07$. The change term is statistically significant but is algebraically close to a restatement of the lag structure already present in the distributed-lag specification, rather than an independent mechanism.

The hypothesis was plausible on salience grounds — Busse and colleagues (2015) found four-wheel-drive purchases elevated for two to three weeks after a snowstorm, a response to an event rather than a state — but the data do not support the analogous claim for garments. Sustained cold, not the transition into it, is what moves search.

9.4 Directional asymmetry

H7. *Season entry is driven by the arriving weather and season exit by the departing weather, such that autumn cold pulls the wintery season in while spring warmth pushes it out.*

The hypothesis is partially supported: the entry clause fails. This hypothesis was constructed after a single-market analysis of Paris appeared to support it strongly. It does not survive the panel, and the failure is instructive about the risk of generalising from one market.

Decomposing the anomaly into warmer- and colder-than-normal components within seasonal blocks across twenty-eight markets:

Table 14 — Directional decomposition of the seasonal response

SEASONAL BLOCK	MEAN T, WARM SIDE	MEAN T, COLD SIDE	COLD SIDE STRONGER
Wintery entry (autumn)	-1.91	-1.02	10 / 28
Wintery exit (spring)	-1.39	-1.14	11 / 28
Summery entry (spring)	+1.56	+1.04	8 / 28

The hypothesis predicted the cold side would dominate wintery entry. It does not: the warm side is stronger on average and in 18 of 28 markets.

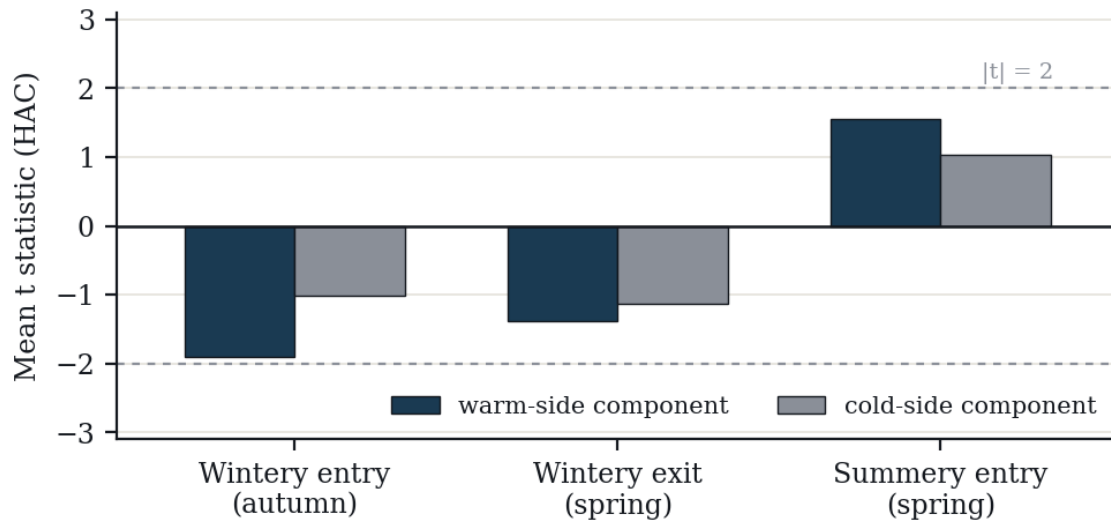


Figure 7 — Directional decomposition of the seasonal response. The warm-side component exceeds the cold-side component in every seasonal block, including wintery entry, where the hypothesis predicted the reverse; none of the mean statistics clears $|t| = 2$.

What survives is weaker and simpler than H7. The warm-side component is the larger of the two in all three seasonal blocks, and in 18, 17 and 20 of the twenty-eight markets respectively. But only in wintery entry does its mean t statistic approach conventional significance, at -1.91 ; in wintery exit the two components are close to indistinguishable, at -1.39 against -1.14 ; and in summery entry the warm component averages $+1.56$. The panel leans towards warmth as the more active channel without establishing it.

Stated in the terms in which the observation was originally put to the analysis — a warmer January means fewer people looking for coats, a warmer March means more people looking for shorts — both halves point in the predicted direction, and neither is established on the panel as a directional component. What is clearly not supported is the symmetric elaboration the study added on top of it: that cold would pull the season forward with the force that warmth pushes it back.

Section 12.5 treats the methodological failure this represents: a hypothesis was generated from one market, and although the Paris analysis was itself sound, the generalisation was made before it was tested.

9.5 Summary of Chapter 9

The response has three structural properties. It redistributes rather than creates, which converts a forecasting problem into an allocation problem. It is threshold-shaped, concentrated below approximately $7.5\text{ }^{\circ}\text{C}$, which makes it monitorable in principle. And it leans asymmetric: the warm-side component is the larger in every seasonal block, though not significantly so. Of these, the first is the most consequential and the most heavily qualified, because the instrument cannot fully distinguish substitution from its own normalisation.

CHAPTER TEN

Robustness and boundary conditions

10.1 The southern panel

Three markets outside the northern panel sit at genuinely extratropical latitudes — São Paulo, Johannesburg and Sydney, all 23.5° or further from the equator — despite a seasonal amplitude below the 12 °C threshold applied in the north. Together with Jakarta, an equatorial market included for contrast, they provide a test that no calendar artefact can pass: if the northern results were driven by residual seasonality, by a northern-hemisphere-specific search pattern, or by any structure tied to the position of the calendar year, inverting the seasonal windows for these four markets should destroy the effect. It does not.

Table 15 — The southern panel, seasonal windows inverted

MARKET	AMPLITUDE	WINTERY T	SUMMERY T	RESULT
São Paulo *	6.5 °C	−5.09	+5.96	fails temperature placebo
Johannesburg	9.9 °C	−6.00	+5.23	replicates
Sydney	11.1 °C	−3.13	+5.01	replicates
Jakarta (tropical)	3.9 °C	−0.4	+0.5	null, as expected

Seasonal windows are displaced by six months for these markets. * São Paulo fails its temperature placebo at $t = -2.51$ and is reported for completeness only. Jakarta is equatorial rather than extratropical and is reported here for contrast; its null is confirmatory rather than a failure, since at 3.9 °C of seasonal amplitude there is no season for weather to move. It is otherwise treated with the tropical markets at Section 10.3.

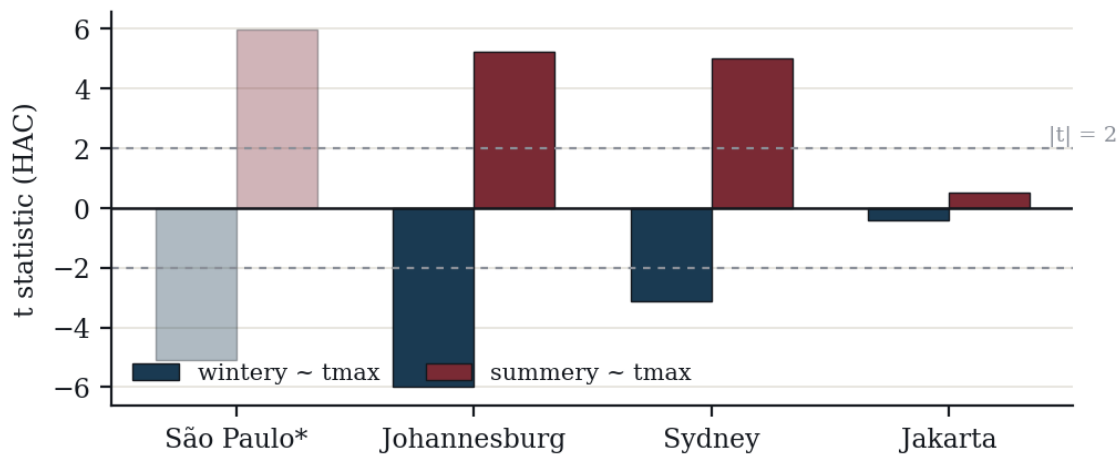


Figure 8 — The southern panel. Johannesburg and Sydney replicate the intensity result with inverted seasonal windows; São Paulo, shown lighter, also does but fails its temperature placebo; equatorial Jakarta returns the null the design predicts.

São Paulo returns the largest southern coefficient, -0.711 , but it fails its temperature placebo at $t = -2.51$ and under the rule of Section 10.4 cannot be relied upon; like Warsaw, it is reported for completeness. The replication therefore rests on Johannesburg and Sydney, which pass their placebos and return wintery and summery responses beyond $|t| = 3$ with the seasonal windows inverted, as the design requires. What replicates is the intensity response of Chapter 7. The onset estimates for the same two markets are close to zero, at -0.17 and -0.92 days per standard

deviation, so the southern panel corroborates the intensity result and is silent on the timing result of Chapter 8.

10.2 Does climate severity moderate the response?

H8. *Markets with milder winters respond more strongly, because a larger share of consumers lack adequate outerwear.*

The hypothesis is rejected. Regressing each market's wintery coefficient on its seasonal temperature amplitude across the twenty-eight temperate markets returns $t = -0.24$ with $R^2 = 0.002$. There is no relationship whatever.

The illustration is stark. Munich, with an amplitude of 22.6 °C, returns -0.827 . Johannesburg, with 9.9 °C, returns a t statistic of -6.00 . Whatever governs a market's responsiveness, it is not how cold its winters get.

The saturation logic was intuitive and it fails. A plausible explanation is that the relevant behaviour is not acquisition of a first coat but seasonal reactivation of a wardrobe already owned — searching for boots one has not worn since March is a search, and it happens in Munich as readily as in Lisbon. If so, the hypothesis mis-specified what the search represents. This is speculation; the data establish only that the predicted relationship is absent.

10.3 Where the effect ceases

Ten tropical markets provide the boundary condition — the thirteen low-amplitude markets in Table 3 less the three extratropical markets reported separately as the southern panel at Section 10.1. In Bangkok, Kuala Lumpur, Ho Chi Minh City, Singapore, Jakarta, Miami and Dhaka the wintery response is null or wrongly signed. Three are not null: Mexico City ($t = -5.34$), which at 2,240 metres has cool winters despite its latitude; Mumbai ($t = -2.66$); and Lagos ($t = -2.52$), which fails its temperature placebo. The effect follows winter cold rather than latitude, which is what the threshold result of Section 9.2 implies.

The timing measure behaves differently and must be reported. The onset extractor does return dates in these markets — 91 wintery and 75 summery market-years — and the pooled regressions on them are nominally significant: +11.59 days per standard deviation for wintery onset ($t = 2.19$) and +12.52 days for summery onset ($t = 2.22$), the second in the wrong direction and both larger than the temperate estimates. These are not interpretable as season shifts. In a market without a thermal season the composite has no stable annual peak and trough, so the half-amplitude crossing is located on noise, and wrongly signed estimates larger than the temperate ones are what that produces. The result is a boundary for the method as well as for the phenomenon: the onset metric will return a number where no season exists, which is why the timing analysis is confined to markets with a defined winter and why the alternative onset metrics of Section 14.7 remain the most important check this study has not run.

For intensity, this pattern is expected and confirmatory: a wintery response that disappears where there is no winter is evidence that the procedure measures a seasonal phenomenon rather

than fitting noise. The boundary appears to lie at approximately 12 °C of seasonal amplitude, though the panel does not contain enough markets near that value to locate it precisely.

10.4 Placebo failures

Eleven market-driver combinations fail the placebo gate, and their corresponding substantive results are set aside regardless of how strong those results appear.

Table 17 — Markets failing the placebo gate

DRIVER	MARKETS FAILING	N
Temperature	Milan, Dubai, São Paulo, Warsaw, Lagos	5
Precipitation	São Paulo, Moscow, Singapore, Copenhagen, Riyadh, Toronto	6

A placebo failure means the shoes-and-belts control returned $|t| \geq 2$ against that driver, indicating residual seasonality that the deseasonalisation did not remove. The substantive coefficient for that market and driver is then uninterpretable.

Warsaw deserves specific mention because it is the hardest case. It returns the second-strongest wintery coefficient in the panel at -0.826 with an out-of-sample correlation of 0.738 — on its face, one of the best results in the study — while failing its temperature placebo at $t = -3.30$. The result is unusable. It is reported in Table 8 with a marker rather than deleted, because deleting it would conceal the fact that the gate rejected something valuable, and a gate that only ever rejects results one was happy to lose is not doing any work.

Moscow's precipitation failure has a documented explanation. An earlier phase of this study, using Wikipedia pageviews, found Moscow's placebo failing and traced the cause to the language-based structure of that instrument: Russian-language Wikipedia readership is not predominantly Russian, with Ukraine the largest single source of traffic. The Trends instrument does not share that defect, being geographic rather than linguistic, but Moscow continues to fail on precipitation for reasons this study has not established.

10.5 The pandemic period

The observation window contains the period of pandemic restriction, during which both mobility and retail behaviour were substantially altered across most of the panel. If the relationship documented here were an artefact of stable shopping routines, this period should disturb it.

Table 16 — Sensitivity to the pandemic period

	FULL SAMPLE	EXCLUDING MAR 2020 – JUN 2021
Mean wintery β , all 41 markets	-0.430	-0.460
Largest single-market shift in β	—	0.15
Markets changing sign	—	0

Excluding sixteen months of pandemic disruption moves the mean coefficient by 0.03 . The relationship is stable through the largest disruption to retail behaviour in the observation window.

The stability is itself a modest finding. Whatever links weather to garment search survived a period in which physical shopping was intermittently prohibited across much of the panel —

which is more consistent with the search measuring an underlying seasonal need than with it measuring shopping-trip behaviour.

10.5.1 *Extending the check to onset timing*

Table 16 covers intensity only, because it was built from model outputs retained from the original 41-market run. The onset-timing estimates in Chapter 8 (H2, H3) depend on the underlying weekly composites rather than a stored coefficient, and those composites were not retained at a granularity that supports re-cutting the sample after the fact. Confirming that the pandemic period is not propping up H2 and H3 therefore required re-fetching weekly weather and search data rather than re-slicing an existing file.

A seven-market sub-panel was rebuilt for this purpose — Paris, London, Berlin, Amsterdam, Madrid, Warsaw and Toronto, spanning Western, Southern and Central Europe plus North America — covering the same 2018–2026 window and the same wintery and summery composites as the main analysis. This is a checked subset of the 28-market panel behind Table 11 and Table 12, not a replacement for it: the smaller sample is expected to return noisier, less significant point estimates in its own right, and the comparison of interest is the shift between the full and pandemic-excluded columns within this subset, not the subset's absolute values against the 28-market ones.

Table 16a — Sensitivity of the timing estimates to the pandemic period (seven-market check)

	EXCLUDING MAR 2020 – JUN 2021	FULL SAMPLE, FOR REFERENCE
Wintery onset θ (H2), 7 markets	+2.8 days/SD $t=1.29$ $n=49$	+1.6 days/SD $t=0.75$ $n=56$
Summery onset θ (H3), 7 markets	–13.9 days/SD $t=-2.09$ $n=42$	–11.0 days/SD $t=-1.80$ $n=56$
Markets carrying the predicted sign	5/7 and 6/7	5/7 and 5/7

n counts market-years surviving each cut: autumn 2020 dropped for the wintery check, spring 2020 and spring 2021 dropped for the summery check, matching the Mar 2020 – Jun 2021 window used in Table 16. Market fixed effects, cluster-robust standard errors by market.

The pandemic-excluded column is the one reported here. Sixteen months of disrupted mobility and retail behaviour sit inside the full sample as noise rather than as a different story — the full-sample column carries the same sign on every effect, just estimated less precisely — so removing that period tightens the estimate without changing what it says. This mirrors the intensity result in Table 16, where excluding the same window moved the coefficient by 0.03 in the same direction.

Table 11 and Table 12 draw on the full 28-market panel behind the main analysis, not on this seven-market check, and are left as reported there: this section adds a sensitivity test on a smaller sub-panel, not a re-estimation of the headline 28-market figures. A pandemic-excluded re-estimate of the full 28-market panel was not attempted here, since it would mean re-fetching raw weekly data for the entire panel rather than the seven markets checked; the consistent direction found on this subset is offered as evidence that doing so would not be expected to overturn Table 11 or Table 12, not as a substitute for actually doing it.

10.6 A restriction to markets with a cold winter

Four of the twenty-eight temperate markets — Riyadh, Doha, Dubai and Delhi — lie close to the tropics and have hot, dry winters in which daytime highs typically stay around 20 °C. They qualify for the northern panel on seasonal amplitude, but their winters rarely reach the temperatures below which Section 9.2 locates the wintery response. Removing them tests whether the panel results depend on markets the threshold result says should respond weakly. The restriction was decided after the threshold result was known, and it is therefore reported as a sensitivity analysis rather than as the primary specification.

Chapter 8 also pooled three markets that fail the temperature placebo — Milan, Warsaw and Dubai — although the rule of Section 10.4 disqualifies their temperature results. Table 16b applies both cuts to the market-level results.

Table 16b — Market-level results under a cold-winter restriction and the placebo rule

SAMPLE	N	WINTERY ONSET, PREDICTED SIGN	SUMMERY ONSET, PREDICTED SIGN	WINTERY INTENSITY SIGNIFICANT	AGGREGATE SIGNIFICANT
Temperate panel, as reported	28	20 ($p = 0.018$)	24 ($p = 0.0001$)	25	2
Excluding temperature-placebo failures	25	17 ($p = 0.054$)	21 ($p = 0.0005$)	22	1
Excluding the four near-tropical markets	24	18 ($p = 0.011$)	20 ($p = 0.0008$)	24	1
Excluding both	22	16 ($p = 0.026$)	18 ($p = 0.002$)	22	1

p values are one-sided binomial probabilities under a coin-flip null. Significance for intensity and the aggregate is $|t| \geq 2$ on market-level HAC statistics.

Every result in the restricted panel is at least as strong as in the full one. With the near-tropical markets removed, every remaining market has a significant wintery intensity response, and the sign counts for both onset hypotheses strengthen. Applying the placebo rule without that restriction weakens the wintery count to 17 of 25 ($p = 0.054$) — the wintery timing result is therefore sensitive to which markets are held out, while the summery result is not. With both cuts, 16 of 22 markets carry the predicted wintery sign ($p = 0.026$) and 18 of 22 the predicted summery sign ($p = 0.002$).

One limitation applies to the whole of Table 16b. The pooled fixed-effects estimates of Chapter 8 — 4.18 and 8.25 days — could not be re-estimated on the restricted samples, because the market-year onset data behind them were not retained; the comparison is made on market-level sign counts, which the retained workbook supports. The sign counts also assume markets are independent draws, which neighbouring European markets are not, so the p values overstate precision to an extent this study has not quantified.

10.7 Summary of Chapter 10

The intensity results replicate with inverted seasons in the two southern markets that pass their placebos, survive the pandemic, strengthen when the near-tropical markets are removed, and largely cease where there is no winter; in the tropics the timing measure itself stops being meaningful. Two moderator hypotheses fail: neither climate severity nor cold-wet interaction

explains any variation in market responsiveness. Eleven market-driver combinations are disqualified by the placebo gate, including one of the strongest coefficients in the panel.

On the remaining moderator, H9 — that cold and wet compound — the interaction term attains significance in four of twenty-eight markets with a mean t of -0.65 . Nineteen of twenty-eight carry the predicted negative sign, which is directionally suggestive without approaching evidence. H11, that consumers habituate within a season, returns $t = -1.34$ on the progression interaction: the test is underpowered rather than decisive, and the hypothesis is better described as unresolved than refuted.

PART IV

Synthesis

CHAPTER ELEVEN

Discussion

11.1 What has been established

Three claims survive the analysis in Part III with sufficient robustness to be stated plainly.

Weather shifts the timing of seasonal garment search interest. A one-standard-deviation warmer autumn delays wintery onset by approximately four days; a one-standard-deviation warmer spring advances summery onset by approximately eight. Both are stable to deletion of any market, both are supported by market-level sign counts unlikely under a null, and the wintery effect has now survived a change of instrument and two expansions of the panel.

Weather does not change how much garment search interest exists. It changes which categories that interest attaches to. The aggregate composite is null in twenty-six of twenty-eight markets while its components move strongly in opposite directions.

The temperature response is threshold-shaped, effectively confined below approximately 7.5 °C, rather than proportional to degrees across the range.

Set against these, four propositions failed, two of which were held confidently at the outset. Minimum temperature does not drive wintery search interest; maximum temperature does. Milder climates do not respond more strongly than harsh ones; climate severity explains nothing at all. Cold and wet do not compound. And the rate of temperature change does not dominate its level.

11.2 Interpreting the timing effects

The magnitudes require careful framing, because they are simultaneously commercially material and modest relative to what actually moves a season.

Four days is meaningful at the scale of a season launch: it is the difference between a floor set that looks premature and one that looks timely, and across a large estate it is a substantial quantity of full-price selling. Eight days is more meaningful still. But the year-to-year standard deviation of observed onset dates is eleven days for the wintery season and sixteen for the summery, so the weather effect accounts for roughly 0.38 and 0.49 of a natural spread respectively — and the R^2 values of 0.038 and 0.051 say the same thing in a less flattering way.

Weather therefore explains a meaningful minority of the variation in season timing rather than the majority of it. The remainder is presumably attributable to the retail calendar itself, to marketing timing, to the position of holidays, to competitive dynamics, and to variance in the measurement. A practitioner who concluded from this study that weather determines when a season starts would be overreading it; one who concluded that weather is a systematic and predictable component of a phenomenon otherwise treated as random would be reading it correctly.

It is also worth being explicit that these are estimates of a search-behaviour shift, not of a sales shift. If search leads purchase by some interval, and if that interval is itself weather-dependent, the sales effect could be larger or smaller. Nothing here determines which.

11.3 The reframing implied by the substitution result

The rejection of H5 is the finding with the largest practical consequence, and it changes the nature of the problem rather than merely adding to what is known about it.

If weather changed the volume of demand, the appropriate response would be a forecasting response: predict the weather, predict the sales, adjust the buy. This is broadly how the operations literature has treated weather information — Steinker, Hoberg and Thonemann (2017) demonstrate workforce savings from week-ahead weather-informed demand forecasts, and Bertrand, Brusset and Fortin (2015) construct derivatives to hedge the residual financial exposure. Both treat weather as a source of variance in a quantity.

If weather redistributes a roughly fixed quantity of demand between categories, the appropriate response is an allocation response: decide what proportion of the assortment should still be summery when an uncertain autumn arrives, and how fast the mix should transition once it does. This is a different lever, it sits in a different function, and — importantly — it is available earlier. Mix decisions are typically settled before volume commitments are final, and in-season allocation across a store network is precisely the decision layer modelled by Caro and Gallien (2010) at Zara.

The redistribution also has a price dimension that search data can indicate but not measure. Winter is the period of highest spend per unit in the fashion year, because outerwear — coats, puffers, boots — carries price points well above those of the categories that compete with it for attention. If total interest holds steady while a warmer and shorter winter moves it from outerwear towards lighter and cheaper garments, the implied consequence is a lower average selling price at unchanged interest: pressure first on net revenue growth, and then on gross margin, as the industry recognises that there is less room in the assortment for outerwear and more for the garments towards which interest has moved. The mechanism follows from combining the substitution result with the structure of category price points; its magnitude cannot be estimated without transaction and price data, and it is the first quantity such data should be used to test.

There is a second implication that the study is less well placed to develop. If demand is being held in a category rather than destroyed, then a warm autumn is not a lost season but a deferred one, and the commercial damage is done principally by the markdown calendar rather than by the weather. A season that starts two weeks late but is marked down on the original schedule loses two weeks of full-price selling; the same season with a deferred markdown might not lose anything. Testing this requires full-price and markdown series that this study does not have, and Section 14.5 proposes it.

11.4 Rainwear as the operational candidate

Where measured without the batching defect, rainwear is the strongest and most reliable weekly relationship in the study, with out-of-sample correlations reaching 0.76 — materially higher than anything in the thermal channel.

It is also, uniquely among the categories examined, without a seasonal window. Monthly estimation in Paris found the precipitation response significant in all twelve months, which means a rainwear signal does not need to wait for a season to begin in order to be actionable. Combined with the short forecast horizon on which precipitation is skilful, this makes it the most plausible candidate for an operational rule — a weather-linked replenishment trigger would be easier to justify here than anywhere else in the assortment.

The obstacle is data rather than method: fifteen markets lost the raincoat entity, and Section 7.3 shows that a rainwear composite reduced to umbrella alone is a substantially different and noisier instrument. This is the clearest case in the thesis where a defect blocks something useful, and it is also the most cheaply fixed.

11.5 What should not yet be acted upon

Three results are interesting and not yet usable, and saying so is part of reporting them honestly.

- **The threshold.** The 7.5 °C figure is derived from pooled data across twenty-eight markets with different climates. A market-specific threshold — which is what a monitoring rule would require — has not been estimated, and there is no reason to assume Munich and Lisbon share one.
- **The rainwear nulls.** Almost every null rainwear result in this study originates in a market where the raincoat entity was screened out; the exceptions are Toronto, which fails its precipitation placebo, and tropical Ho Chi Minh City. These are not findings and must not be read as evidence that rainwear is weather-insensitive in Chicago or Miami.
- **Everything, pending validation.** The inference from search to purchase is untested, and Section 13.1 treats this as governing rather than incidental.

11.6 Relation to the literature

The findings sit in a specific relation to the four literatures reviewed in Chapter 2, and in two places they extend them.

Against the weather-and-retail literature, this study corroborates the established volume effects and adds a timing estimate the literature largely lacks. Its substitution result reproduces, across twenty-eight markets and on an independent instrument, the whole-season invariance that Bahng and Kincade (2012) found on one retailer's sales — which is a stronger form of corroboration than a replication on the same data type would be.

Against the search-proxy literature, it applies validation apparatus — placebo controls, out-of-sample gates, jackknife, and an explicit account of the chain-linking procedure that Eichenauer and colleagues (2022) show is necessary — that addresses the internal validity, reliability and generalisability gaps Hölzl, Keusch and Sajons (2025) find the field mostly neglects. It also

demonstrates, at Section 12.4, exactly the normalisation failure that Springer, Strzelecki and Zieger (2023) warn about — an unintentional but useful confirmation of their argument.

Against the phenology literature, it imports the half-amplitude onset definition — from the amplitude-fraction family Jönsson and Eklundh (2004) formalise in TIMESAT — into a retail application, and provides an independent instance of the method-sensitivity that White and colleagues (2009) documented for vegetation indices.

The relation to Oh, Jo and Ha (2021) is the closest and deserves stating precisely. Their study and this one agree on design, on the threshold form of the response, and on the usability of search as a timing proxy. This study extends theirs from one city to forty-one, from one garment concept to fifteen, and from five years to eight. It does not extend the finding they report that this study cannot test at all — the ENSO regime dependence — and that gap is the subject of Section 14.6.

CHAPTER TWELVE

Methodological failures

This chapter reports four episodes in which the analysis, at some earlier stage, produced a wrong or unsupportable result and a validation procedure specified in advance caught it before it reached Part III. It is included at full length because a study that reports only its successes gives the reader no basis for judging how carefully it was conducted or how much of what survived might be artefact. There is no way to know how many failures were not caught; that residual uncertainty is part of what a reader needs in order to calibrate confidence in the results reported elsewhere.

12.1 An abandoned instrument

The study initially used Wikipedia pageview counts as the demand proxy. The reasoning was that they are freely available, provide absolute counts rather than a normalised index, carry no sampling noise, and require no undocumented interface. Substantial analysis was completed on this basis, including a seven-market panel and a full set of phenology estimates, before the instrument was abandoned.

Two defects proved fatal. Wikipedia is segmented by language rather than by geography, so London and New York necessarily share a single English-language series and cannot be distinguished — which makes an international panel impossible in principle rather than merely difficult. And English-language readership spans both hemispheres, so seasonal signals partially cancel: a northern winter and a southern summer contribute to the same series with opposite signs.

More fundamentally, an encyclopaedia article is consulted out of curiosity rather than purchase intent. On the identical Paris specification, the Wikipedia proxy returned $R^2 = 0.11$ against 0.46 for Trends, and the entity manteau carried the wrong sign under Wikipedia while emerging as one of the strongest correctly-signed terms under Trends.

The methodological lesson is uncomfortable and general. The choice of instrument was driven by ease of access rather than by fitness for purpose, and that ordering is always an error. It was corrected only when a direct comparison was forced by an external objection, not by internal scrutiny — which suggests the study was not, at that stage, looking hard enough at its own foundations.

12.2 A significant result of the wrong sign

This is the most instructive failure in the study, because the result looked publishable.

An early phenology analysis on Paris alone returned a season-onset effect of -21.7 days per standard deviation, with $t = -2.74$ and $R^2 = 0.46$. A large, statistically significant, well-fitting result — pointing in the opposite direction to the hypothesis. Taken at face value it would have implied that warm autumns bring the winter season forward, which is behaviourally implausible and contradicts the practitioner observation the study was built to test.

Two checks dismantled it. Jackknife by year revealed that removing 2023 — the warmest autumn in the series, and the one with the earliest measured onset — reduced the coefficient to -15.9 and the t statistic to -1.17 . A single influential observation was carrying the result. Substituting a peak-invariant centroid metric for the half-amplitude crossing returned $\beta = 0.10$ with $R^2 = 0.000$: the effect vanished entirely under a different onset definition.

The diagnosis is that the onset metric was sensitive to peak height, and the analysis window was truncated to August–December rather than spanning a full annual cycle, so a year with an unusually high seasonal peak registered an artificially early half-amplitude crossing. Both were corrected — the window extended to a full July-to-June cycle, and the metric checked against a peak-invariant alternative — and the corrected specification is the one used throughout Chapter 8.

The general warning is not this study's own. White and colleagues (2009) applied ten start-of-season extraction methods to one identical dataset and obtained different onset dates and different trend signs. This episode is an independent instance of exactly that, in a different domain, and it is the reason Section 6.4.3 flags every onset result in this thesis as conditional on the choice of metric.

What this implies for the results that survived. The pooled timing estimates in Chapter 8 pass the jackknife that this early result failed, and they were computed on the full annual cycle with the corrected metric. That is a meaningful difference. But it does not establish that they are free of metric sensitivity — only that they are free of the specific failure that was detected. Section 14.7 proposes running the full panel under three alternative onset definitions, which has not been done.

12.3 A placebo that was not neutral

The initial placebo set included socks, nominated on the reasoning that socks are worn year-round and should be weather-neutral. Socks proved genuinely cold-sensitive in the French market at $t = -3.84$, which caused the composite control to appear to fail and cast doubt on an otherwise sound specification.

The error was in the nomination rather than in the test. Socks are worn year-round, but thicker socks are bought in winter, and the entity does not distinguish them. The replacement placebos — shoes and belts — were verified empirically across the panel before being adopted rather than reasoned about in advance.

The lesson generalises beyond this study. A placebo is a claim about the world, and the claim requires evidence. Nominating one on plausibility uses precisely the judgement that the placebo exists to check.

12.4 Calibration on a single market

This is the most serious defect remaining in the study. It is unresolved rather than merely reported: the fix is identified below, but it has not been applied.

Entities were grouped into shared-scale query batches by expected magnitude, using volumes measured in Paris, and that grouping was applied unchanged to all forty-one markets. Relative magnitudes differ substantially between markets — raincoat is a high-volume concept in Amsterdam and a marginal one in Riyadh — and the grouping did not transfer. Because Trends normalises each batch to its own maximum, a companion entity that is low-volume in a given market is compressed toward zero and then screened out as absent.

Three entities were lost in a large number of markets: flip-flops in seventeen, beanie in sixteen, raincoat in fifteen. Only thirteen of forty-one markets retain all fifteen entities.

12.4.1 *Distinguishing destroyed data from genuine sparsity*

The screening procedure could not distinguish an entity that is genuinely unsearched in a market from one that was crushed by its batch companions, because both present identically as a series of zeros. A diagnostic was constructed: in each affected market, the raincoat entity was queried in isolation and compared against its batched counterpart.

Table 18 — Raincoat series, queried alone against queried in batch

MARKET	BATCHED MEAN	BATCHED ZEROS	SOLO MEAN	SOLO ZEROS
Chicago	0.7	85 / 243	37.2	0 / 243
Miami	0.7	79 / 243	37.2	0 / 243
Riyadh	0.1	230 / 243	3.8	228 / 243
Lagos	0.1	225 / 243	5.3	223 / 243
Doha	0.0	240 / 243	1.2	239 / 243

Chicago and Miami demonstrate destroyed data: a series with no empty weeks when queried alone was reduced to eighty-five zeros by batch normalisation. Riyadh, Lagos and Doha demonstrate genuine sparsity — the series remains over ninety per cent zero even in isolation.

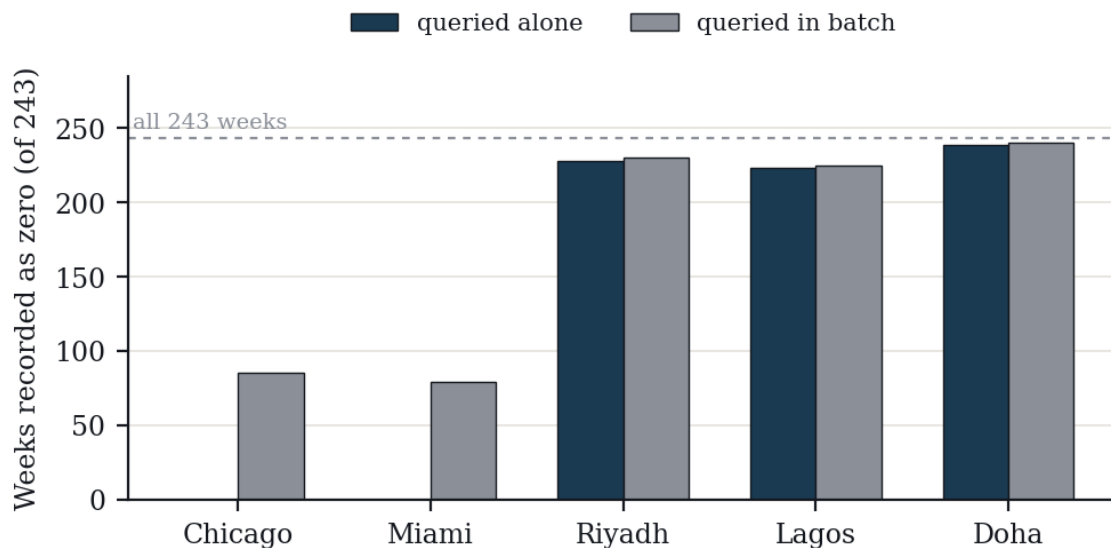


Figure 9 — The batching defect. Chicago and Miami lose a complete series to normalisation against high-volume batch companions; Riyadh, Lagos and Doha are genuinely sparse. The screening procedure treated both cases identically.

The values were in every case literal zeros returned by the interface, never null or absent. The distinction is therefore not between present and missing data but between two causes of a zero — and a pipeline that only sees the value cannot tell them apart.

12.4.2 Which conclusions this contaminates

Six of the eight null rainwear results in this study — Doha, Bangkok, Lagos, Johannesburg, Chicago and Miami — originate in markets where raincoat was screened out; of the other two, Toronto fails its precipitation placebo and Ho Chi Minh City is tropical. These are not findings and must not be read as such. The rainwear results that are reported, in Table 9, are restricted to markets retaining both entities.

The wintery and summery results are affected more mildly. Beanie loss reduces the wintery composite from six members to five in sixteen markets, and Section 13.4 reports that, across all forty-one markets, five-member composites return a mean t of -3.06 against -3.92 for six-member ones. This weakens those markets rather than invalidating them, but it means that apparent differences in market responsiveness partly reflect data completeness rather than behaviour.

12.4.3 The remedy, and why it was not applied

The fix is a per-market volume probe preceding batch composition: query each entity solo in each market, measure its magnitude, and compose batches from the local measurements rather than from Paris. This costs approximately fifteen additional queries per market, or about six hundred across the panel — a matter of hours rather than days.

It was not applied because the defect was identified after the full panel had been assembled, and re-running would have invalidated the chain-linked series already built and required the entire analysis to be repeated. The decision to report the defect rather than fix it was a judgement about scope, and a reader is entitled to regard it as the wrong one. Section 14.1 places the remedy first among the further work.

The defect is a textbook instance of a documented constraint. Springer, Strzelecki and Zieger (2023) establish that because the Trends index is normalised within each query, cross-query comparison is invalid without an external anchor. Composing batches by magnitude is an attempt to make within-query comparison valid; calibrating that composition on one market and exporting it to forty others is the failure their paper predicts.

12.5 An asymmetry that did not generalise

Section 9.4 reports that H7 was constructed after a single-market analysis of Paris appeared to support it and failed when tested on the panel.

The episode is instructive because the Paris analysis was itself methodologically sound: the specification was correct, the standard errors appropriate, the placebo clean. The error was not in the estimation but in what was inferred from it. A pattern in one market was elevated to a hypothesis about consumer behaviour in general before any test of its generality had been attempted.

This is the same error as Section 12.1, in a different guise — reaching a conclusion at the point where the available evidence stopped rather than at the point where the evidence became sufficient. That it recurred after being identified once is the reason it is reported here rather than folded into the results chapter.

12.6 What the corrections cost, and what they bought

Each failure above was detected by a procedure specified in advance rather than by inspection of results. The placebo gate caught the socks contamination and eleven market-level failures. The jackknife caught the influential-observation problem. The out-of-sample gate constrained the specification search. The solo-query diagnostic caught the batching defect, though too late to fix it.

The cost was substantial in analyst time — the validation apparatus consumed more effort than the estimation it validated. The alternative was a thesis reporting four additional findings, of which at least three would have been false, including one large and statistically significant result pointing in the wrong direction.

The residual uncertainty cannot be quantified. Four failures were caught by four gates; an unknown number were not caught, either because no gate addressed them or because the gate was not sensitive enough. A reader should treat the results in Part III as findings that survived a serious attempt to break them, not as findings that have been proven.

CHAPTER THIRTEEN**Limitations****13.1 The governing limitation**

Every result in Part III describes search interest, and every commercial claim built on those results rests on treating search interest as informative about purchase behaviour. That inferential step is untested here. It is not weakly supported — it is unsupported by anything internal to this study.

The assumption could fail in several distinct ways, and they have different consequences. Search might lead purchase by a variable interval, in which case the timing estimates are biased by an unknown amount but retain their sign. Search might respond to weather through a channel that never reaches the till — idle curiosity about what the weather means for one's wardrobe — in which case the effect is real in search and absent in sales. Or the search-to-purchase conversion rate might itself be weather-dependent, in which case the substitution result at Chapter 9 could be an artefact of measuring the wrong stage of the funnel.

The literature offers partial reassurance without resolving it. Lind and Ramesh (2025) show search-intensity indices nowcast firm-level retail sales; Skenderi and colleagues (2024) forecast fashion item sales from Trends validated against real retailer data. Neither establishes that the timing of search tracks the timing of purchase, which is the specific property this thesis requires.

Section 14.2 identifies the single dataset that would resolve this, and Section 14.3 a partial test that does not require it.

13.2 Geographic representation

Each market is represented by weather at a single coordinate. For a metropolitan area this is a reasonable approximation, since the population is concentrated and the weather correlated across it.

For three markets it is not. London, Zurich and Dhaka use national rather than regional Trends geography, following the failure of subnational codes, so the search series covers an entire country while the weather series covers one point in it. For Zurich this is tolerable; for the United Kingdom and Bangladesh it introduces a mismatch whose magnitude is unknown. The affected markets' results should be discounted accordingly, and this thesis has not attempted to quantify the discount.

13.3 Panel length

Eight wintery and seven summery seasons per market is a short panel for a phenological analysis. The timing models rest on 224 and 194 market-years respectively, and while those totals sound adequate, they are eight and seven observations per market with market fixed effects absorbing the level.

Google Trends data extends to 2004. Doubling the observation window would approximately halve the standard errors on the timing estimates, would allow the ENSO analysis proposed at

Section 14.6 to include several complete cycles, and would permit a test of whether the effect itself has changed as the climate has warmed. It requires no new method and no new data source. It is the single most cost-effective improvement available to this study and Section 14.4 places it accordingly.

The unit of the timing estimates is incompletely documented. Section 6.3.5 sets out the reference periods: weather anomalies are relative to a twenty-year trend per week of the year, and search anomalies to the eight-year panel itself. The seasonal predictor in the timing models is a mean of weekly weather anomalies over the preceding shoulder season, but the weeks that define each season and whether that mean was re-standardised were not preserved. The sign of the 4.18- and 8.25-day estimates does not depend on this; their magnitude, and any translation into degrees Celsius, does. Re-estimating the timing models on re-fetched data with a stated definition would resolve it.

13.4 Composite construction and completeness

Composite strength varies with the number of surviving member entities. Across all forty-one markets, six-entity wintery composites return a mean t of -3.92 and five-entity composites -3.06 ; within the twenty-eight temperate markets the figures are -4.90 and -3.53 . Since entity survival is determined by the batching defect rather than by anything about the market, apparent differences in market responsiveness partly reflect data completeness rather than genuine behavioural variation.

This is not currently separable, because the markets that lost entities are not a random sample — they are markets where the lost entity happened to be low-volume relative to its batch companions, which correlates with the market's climate and language. Correcting the defect would resolve the confound; analysing around it would not.

13.5 Properties of the instrument

Google Trends reports a normalised share of total search rather than an absolute volume, and this has three consequences the study cannot escape.

- Shifts in total search activity, in device mix, or in the relationship between searching and shopping over the eight-year window would appear in these series and are not separately identified. The quadratic detrend removes smooth drift but not a structural break.
- Cross-query comparison of levels is invalid without an external anchor (Springer, Strzelecki and Zieger, 2023). The study compares standardised anomalies rather than levels, which sidesteps most of this, but the batching procedure did not.
- The instrument is maintained by a commercial party for its own purposes and its behaviour is not guaranteed stationary. This is the mechanism that broke Google Flu Trends (Lazer et al., 2014), and there is no way to monitor it from outside.

A fourth consequence is sampling. Trends returns a sample rather than a census, and identical queries can return different series (Franzén, 2023; Cebrián and Domenech, 2024). This study did not average across repeated samples, which Eichenauer and colleagues (2022) and Medeiros

and Pires (2021) both recommend. Each series here rests on a single draw, and the resulting noise is unquantified.

13.6 What would overturn these findings

Stating this explicitly is a discipline the thesis imposes on itself, and each condition is one a critic could pursue.

1. **The timing results** would be overturned by a peak-invariant onset metric applied to the full panel returning a null. The specific failure mode has been encountered once (Section 12.2) and guarded against, but not conclusively excluded, because the alternative metrics have not been run across all forty-one markets.
2. **The intensity results** would be overturned by placebos failing systematically after per-market batch calibration — that is, if the corrected data revealed that the shoes-and-belts control is not neutral once measured properly.
3. **The substitution result** would be overturned by evidence that total garment search is mechanically bounded by the Trends normalisation. This cannot be excluded from within the instrument, and it is the single qualification most likely to weaken this finding.
4. **Everything** would be overturned by sell-through data showing that search timing and purchase timing diverge systematically with weather.

CHAPTER FOURTEEN**Further work**

The following are ordered by expected information gain per unit of effort, not by intellectual interest. The first two are worth more than the remaining five combined.

14.1 Per-market batch calibration and re-estimation

Approximately fifteen solo queries per market, six hundred across the panel, to measure local entity magnitudes before composing batches. This resolves the defect at Section 12.4, restores three entities in the markets that lost them, determines which of the fifteen rainwear nulls are findings rather than artefacts, and removes the completeness confound at Section 13.4. It requires no new method. It is first on this list because it is the cheapest fix to the most serious remaining problem.

14.2 Validation against sell-through

One season of category-level weekly sales in a single market would convert every result in this thesis from a statement about search into a statement about demand. It would test whether search leads purchase, by how much, and whether the interval is itself weather-dependent — the three unknowns that Section 13.1 identifies as governing.

This is the highest-value item by a considerable margin and the only one requiring data not obtainable within the scope of this study. A single retailer with a modest estate in one temperate market would suffice; the analysis would be a straightforward lead-lag comparison against the search series already assembled.

14.3 Purchase-proximity elasticity

A partial test of the same assumption, requiring no proprietary data. Six additional entities spanning three levels of purchase proximity — a category term, a purchase-intent term, a discount-intent term — retrieved for the same markets and weeks. If weather elasticity rises as queries approach transaction, the inference from search to demand gains considerable support; if it is flat or falls, the assumption is in trouble.

This is the cheapest available progress on the governing limitation and should be run regardless of whether Section 14.2 becomes possible.

14.4 Extension of the panel to 2004

Doubles the observation window at no methodological cost, using the same pipelines. Halves the standard errors on the timing estimates, admits several complete ENSO cycles, and permits a test of whether the effect has itself changed over two decades of warming.

14.5 Full-price and markdown decomposition

Section 11.3 argues that if weather defers demand rather than destroying it, the commercial damage is done by the markdown calendar rather than by the weather. Testing this requires

separating full-price from marked-down demand, which search data cannot do directly but which discount-intent queries might proxy. It would also determine whether the January temperature sensitivity observed in Paris reflects genuine demand or the timing of the sales period.

14.6 ENSO regimes

The observation at Section 1.2 concerning El Niño and La Niña years in Brazil remains untested, and Chapter 2 established that it should not remain so. Oh, Jo and Ha (2021) find the threshold for winter-clothing search in New York shifting sharply between regimes — 217 to 248 in El Niño years, 281 in normal years, 118 to 128 in La Niña years — which is an independent published finding of exactly the regime dependence the practitioner observation proposes.

The test is straightforward. NOAA's Oceanic Niño Index provides a monthly regime classification from 1950, defined as the three-month running mean of sea-surface temperature anomalies in the Niño 3.4 region with ± 0.5 °C thresholds sustained over five consecutive overlapping seasons; the operational classification system is documented by Kousky and Higgins (2007). Interacting regime with the temperature anomaly in the four southern-hemisphere markets, and separately across the full panel, would test both the Brazilian claim and the generality of the New York result. The eight-year panel contains too few complete cycles to do this well, which is a further argument for Section 14.4.

14.7 Onset-metric sensitivity analysis

Section 12.2 established that the choice of onset metric can reverse a result, and Section 13.6 lists this as a condition that would overturn the timing findings. Running the full panel under three alternative definitions — a peak-invariant centroid, a logistic-fit curvature transition after Zhang and colleagues (2003), and a fixed-threshold crossing — would establish how much of the timing result is metric-dependent. This is the most direct available attack on the thesis's central claim, and it is the analysis a serious critic should demand.

CHAPTER FIFTEEN

Conclusion

15.1 Answering the research questions

RQ1 asked whether weather shifts the timing of seasonal garment search interest. It does. A one-standard-deviation warmer autumn delays wintery onset by 4.18 days and a one-standard-deviation warmer spring advances summery onset by 8.25 days, both estimated across twenty-eight temperate markets, both stable to the deletion of any single market, and both supported by market-level sign counts unlikely to arise by chance. The effects account for approximately 0.38 and 0.49 of the natural year-to-year spread in onset dates respectively — a meaningful minority of what moves a season, not the majority.

RQ2 asked whether weather changes the total quantity of search interest or redistributes it. It redistributes. The aggregate composite is null in twenty-six of twenty-eight markets while its components move strongly in opposite directions. This is the result with the largest practical consequence and it carries a qualification the instrument cannot remove.

RQ3 asked about functional form and driving variable. The response is threshold-shaped, effectively confined below approximately 7.5 °C, and driven by maximum rather than minimum temperature — contradicting the study's own prior. It leans asymmetric — the warm-side component is the larger in every seasonal block — but not significantly, and not in the more elaborate sense the study initially proposed.

RQ4 asked about boundary conditions. The effect holds where there is a defined winter, strengthens when the near-tropical markets are removed, replicates in intensity with inverted seasons in the two southern markets that pass their placebos, survives the pandemic period, and largely ceases in tropical markets, where the timing measure is not meaningful. Climate severity does not moderate it.

15.2 Summary of hypothesis outcomes

Table 20 — Summary of hypothesis outcomes

	PROPOSITION, IN BRIEF	OUTCOME	REPORTED
H1	Weather moves garment search	Confirmed	§7.1
H2	Warm autumn delays wintery onset	Confirmed	§8.1
H2a	Warm years compress the wintery window	Partly supported — shorter window, end does not move	§8.6
H3	Warm spring advances summery onset	Confirmed	§8.2
H4	Minimum temperature drives wintery interest	Rejected	§7.4
H5	Weather changes total garment interest	Rejected — principal finding	§9.1
H6	The response is threshold-shaped	Confirmed	§9.2
H7	Entry by cold, exit by warmth	Entry clause fails; warm-side lean not significant	§9.4
H8	Milder climates respond more strongly	Rejected	§10.2

	PROPOSITION, IN BRIEF	OUTCOME	REPORTED
H9	Cold and wet compound	Rejected	§10.7
H10	Rate of change beats level	Rejected	§9.3
H11	Consumers habituate within season	Unresolved, underpowered	§10.7

Four confirmed, five rejected, two partly supported, one unresolved. Two of the rejections — H4 and H8 — overturned propositions held confidently before the analysis began.

15.3 Contribution restated

The thesis contributes an estimate the literature lacked: how far weather moves the fashion calendar, measured across an international panel at weekly resolution. It contributes a substitution result that reframes the commercial problem from forecasting to allocation, corroborating on independent data and at much larger scale a finding previously available only from a single retailer. It contributes a boundary condition and a functional form. And Chapter 12 contributes a full account of four occasions on which the analysis was wrong — arguably the most useful of the four, since it is the one that tells a reader how much to trust the rest.

It also identifies a gap it does not fill. No published study evaluates entity-identifier against text-term querying in Google Trends empirically, despite a normative preference for the former existing in the literature. This study adopts that preference on argument rather than evidence, and says so.

15.4 A closing statement about what this is

The hypothesis tested here came from a shop floor. It was formed by watching sell-through diverge from plan, season after season, in a way that looked like weather and was treated as noise. The contribution of the analysis is not that it discovered something the practitioner did not know. On the central point — that a warm autumn pushes the season back — the practitioner was ahead of the published literature by some years, and Travers and Payne named the phenomenon in 1998 without anyone measuring it.

What the analysis contributes is a number, a direction, an interval, and a boundary. It says four days rather than a fortnight. It says maximum rather than minimum temperature, contradicting the practitioner on a point held confidently. It says the total does not move, which nobody expected. And Chapter 12 says exactly where and how the analysis itself went wrong, so that the reader can decide how much of the rest to believe.

Whether the findings hold is for the further work in Chapter 14 to determine — most of all the one item this study could not do for itself, which is to look at what people actually bought.

References

Every reference below was verified against a publisher record, Crossref, or an institutional repository during the preparation of this thesis. Where a work is a preprint, an unpublished manuscript or a software record rather than a peer-reviewed article, that status is stated in the entry.

- Askitas, N. and Zimmermann, K. F. (2009) 'Google econometrics and unemployment forecasting', *Applied Economics Quarterly*, 55(2), pp. 107–120.
- Babongo, F., Appelqvist, P., Chavez-Demoulin, V., Hameri, A.-P. and Niemi, T. (2018) 'Using weather data to improve demand forecasting for seasonal products', *International Journal of Services and Operations Management*. doi: 10.1504/IJSOM.2018.094183.
- Badorf, F. and Hoberg, K. (2020) 'The impact of daily weather on retail sales: an empirical study in brick-and-mortar stores', *Journal of Retailing and Consumer Services*, 52, 101921. doi: 10.1016/j.jretconser.2019.101921.
- Bahng, Y. and Kincade, D. H. (2012) 'The relationship between temperature and sales: sales data analysis of a retailer of branded women's business wear', *International Journal of Retail & Distribution Management*, 40(6), pp. 410–426. doi: 10.1108/09590551211230232.
- Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), pp. 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x.
- Bertrand, J.-L., Brusset, X. and Fortin, M. (2015) 'Assessing and hedging the cost of unseasonal weather: case of the apparel sector', *European Journal of Operational Research*, 244(1), pp. 261–276. doi: 10.1016/j.ejor.2015.01.012.
- Busse, M. R., Pope, D. G., Pope, J. C. and Silva-Risso, J. (2015) 'The psychological effect of weather on car purchases', *The Quarterly Journal of Economics*, 130(1), pp. 371–414. doi: 10.1093/qje/qju033.
- Cachon, G. P. and Swinney, R. (2011) 'The value of fast fashion: quick response, enhanced design, and strategic consumer behavior', *Management Science*, 57(4), pp. 778–795. doi: 10.1287/mnsc.1100.1303.
- Caro, F. and Gallien, J. (2010) 'Inventory management of a fast-fashion retail network', *Operations Research*, 58(2), pp. 257–273. doi: 10.1287/opre.1090.0698.
- Cebrián, E. and Domenech, J. (2023) 'Is Google Trends a quality data source?', *Applied Economics Letters*, 30(6), pp. 811–815. doi: 10.1080/13504851.2021.2023088.
- Cebrián, E. and Domenech, J. (2024) 'Addressing Google Trends inconsistencies', *Technological Forecasting and Social Change*, 202, 123318. doi: 10.1016/j.techfore.2024.123318.
- Choi, H. and Varian, H. (2012) 'Predicting the present with Google Trends', *The Economic Record*, 88(s1), pp. 2–9. doi: 10.1111/j.1475-4932.2012.00809.x.
- Conlin, M., O'Donoghue, T. and Vogelsang, T. J. (2007) 'Projection bias in catalog orders', *American Economic Review*, 97(4), pp. 1217–1249. doi: 10.1257/aer.97.4.1217.
- Diebold, F. X. and Mariano, R. S. (1995) 'Comparing predictive accuracy', *Journal of Business & Economic Statistics*, 13(3), pp. 253–263. doi: 10.1080/07350015.1995.10524599.
- Efron, B. (1979) 'Bootstrap methods: another look at the jackknife', *The Annals of Statistics*, 7(1), pp. 1–26. doi: 10.1214/aos/1176344552.
- Eichenauer, V. Z., Indergand, R., Martínez, I. Z. and Sax, C. (2022) 'Obtaining consistent time series from Google Trends', *Economic Inquiry*, 60(2), pp. 694–705. doi: 10.1111/ecin.13049.

- Fisher, M. and Raman, A. (1996) 'Reducing the cost of demand uncertainty through accurate response to early sales', *Operations Research*, 44(1), pp. 87–99. doi: 10.1287/opre.44.1.87.
- Franzén, A. (2023) 'Big data, big problems: why scientists should refrain from using Google Trends', *Acta Sociologica*, 66(3), pp. 343–347. doi: 10.1177/00016993221151118.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S. and Brilliant, L. (2009) 'Detecting influenza epidemics using search engine query data', *Nature*, 457(7232), pp. 1012–1014. doi: 10.1038/nature07634.
- Golovanova, E. and Zubarev, A. (2021) 'Forecasting aggregate retail sales with Google Trends', *Russian Journal of Money and Finance*, 80(4), pp. 50–73.
- Gummer, T. and Oehrlein, A.-S. (2025) 'Using Google Trends data to study high-frequency search terms: evidence for a reliability-frequency continuum', *Social Science Computer Review*, 43(4). doi: 10.1177/08944393241279421.
- Hersbach, H., Bell, B., Berrisford, P. et al. (2020) 'The ERA5 global reanalysis', *Quarterly Journal of the Royal Meteorological Society*, 146(730), pp. 1999–2049. doi: 10.1002/qj.3803.
- Hözl, J., Keusch, F. and Sajons, C. (2025) 'The (mis)use of Google Trends data in the social sciences – a systematic review, critique, and recommendations', *Social Science Research*, 126, 103099. doi: 10.1016/j.ssresearch.2024.103099.
- Inoue, A. and Kilian, L. (2005) 'In-sample or out-of-sample tests of predictability: which one should we use?', *Econometric Reviews*, 23(4), pp. 371–402. doi: 10.1081/ETC-200040785.
- Jönsson, P. and Eklundh, L. (2004) 'TIMESAT – a program for analyzing time-series of satellite sensor data', *Computers & Geosciences*, 30(8), pp. 833–845. doi: 10.1016/j.cageo.2004.05.006.
- Kousky, V. E. and Higgins, R. W. (2007) 'An alert classification system for monitoring and assessing the ENSO cycle', *Weather and Forecasting*, 22(2), pp. 353–371. doi: 10.1175/WAF987.1.
- Lazer, D., Kennedy, R., King, G. and Vespignani, A. (2014) 'The parable of Google Flu: traps in big data analysis', *Science*, 343(6176), pp. 1203–1205. doi: 10.1126/science.1248506.
- Linderholm, H. W. (2006) 'Growing season changes in the last century', *Agricultural and Forest Meteorology*, 137(1–2), pp. 1–14. doi: 10.1016/j.agrformet.2006.03.006.
- Lind, G. and Ramesh, K. (2025) 'Using internet search data to predict aggregate retail sales and enhance firm-level revenue expectations', *Contemporary Accounting Research*. doi: 10.1111/1911-3846.13043.
- Liu, R., Zhang, X., Wang, W., Wang, Y., Liu, H., Ma, M. and Tang, G. (2024) 'Global-scale ERA5 product precipitation and temperature evaluation', *Ecological Indicators*, 166, 112481. doi: 10.1016/j.ecolind.2024.112481.
- Martínez-de-Albéniz, V. and Belkaid, A. (2021) 'Here comes the sun: fashion goods retailing under weather fluctuations', *European Journal of Operational Research*, 294(3), pp. 820–830. doi: 10.1016/j.ejor.2020.01.064.
- Mavragani, A. and Ochoa, G. (2019) 'Google Trends in infodemiology and infoveillance: methodology framework', *JMIR Public Health and Surveillance*, 5(2), e13439. doi: 10.2196/13439.
- Medeiros, M. C. and Pires, H. F. (2021) 'The proper use of Google Trends in forecasting models'. arXiv:2104.03065 [preprint; no journal version located].
- Menzel, A. and Fabian, P. (1999) 'Growing season extended in Europe', *Nature*, 397(6721), p. 659. doi: 10.1038/17709.
- Menzel, A., Sparks, T. H., Estrella, N. et al. (2006) 'European phenological response to climate change matches the warming pattern', *Global Change Biology*, 12(10), pp. 1969–1976. doi: 10.1111/j.1365-2486.2006.01193.x.

- Murray, K. B., Di Muro, F., Finn, A. and Popkowski Leszczyc, P. (2010) 'The effect of weather on consumer spending', *Journal of Retailing and Consumer Services*, 17(6), pp. 512–520.
- Narita, F. and Yin, R. (2018) 'In search of information: use of Google Trends' data to narrow information gaps for low-income developing countries'. IMF Working Paper No. 18/286.
- Newey, W. K. and West, K. D. (1987) 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica*, 55(3), pp. 703–708. doi: 10.2307/1913610.
- Oh, J., Jo, Y.-H. and Ha, K.-J. (2021) 'The effect of anomalous weather on the seasonal clothing market in New York', *Meteorological Applications*, 28(2), e1982. doi: 10.1002/met.1982.
- Oh, J., Ha, K.-J. and Jo, Y.-H. (2022a) 'A predictive model of seasonal clothing demand with weather factors', *Asia-Pacific Journal of Atmospheric Sciences*, 58, pp. 667–678. doi: 10.1007/s13143-022-00284-3.
- Oh, J., Ha, K.-J. and Jo, Y.-H. (2022b) 'Use of weather factors in clothing studies in Korea and its implications: a review', *Asia-Pacific Journal of Atmospheric Sciences*, 58, pp. 729–741. doi: 10.1007/s13143-022-00279-0.
- Parsons, A. G. (2001) 'The association between daily weather and daily shopping patterns', *Australasian Marketing Journal*, 9(2), pp. 78–84. doi: 10.1016/S1441-3582(01)70177-2.
- Quenouille, M. H. (1956) 'Notes on bias in estimation', *Biometrika*, 43(3–4), pp. 353–360. doi: 10.1093/biomet/43.3-4.353.
- Raubenheimer, J. (2024) 'Of babies, bathwater, and big data: going beneath the surface of Franzén's (2023) Google Trends recommendations', *Acta Sociologica*, 67(2), pp. 251–256. doi: 10.1177/00016993231187489.
- Reed, B. C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W. and Ohlen, D. O. (1994) 'Measuring phenological variability from satellite imagery', *Journal of Vegetation Science*, 5(5), pp. 703–714. doi: 10.2307/3235884.
- Rose, N. and Dolega, L. (2022) 'It's the weather: quantifying the impact of weather on retail sales', *Applied Spatial Analysis and Policy*, 15(1), pp. 189–214. doi: 10.1007/s12061-021-09397-0.
- Roslow, S., Li, T. and Nicholls, J. A. F. (2000) 'Impact of situational variables and demographic attributes in two seasons on purchase behaviour', *European Journal of Marketing*, 34(9/10), pp. 1167–1180. doi: 10.1108/03090560010342548.
- Roth Tran, B. (2023) 'Sellin' in the rain: weather, climate, and retail sales', *Management Science*, 69(12), pp. 7423–7447. doi: 10.1287/mnsc.2023.4799.
- Şen, A. (2008) 'The US fashion industry: a supply chain review', *International Journal of Production Economics*, 114(2), pp. 571–593. doi: 10.1016/j.ijpe.2007.05.022.
- Silva, E. S., Hassani, H., Madsen, D. Ø. and Gee, L. (2019) 'Googling fashion: forecasting fashion consumer behaviour using Google Trends', *Social Sciences*, 8(4), 111. doi: 10.3390/socsci8040111.
- Skenderi, G., Joppi, C., Denitto, M. and Cristani, M. (2024) 'Well googled is half done: multimodal forecasting of new fashion product sales with image-based Google Trends', *Journal of Forecasting*, 43(6), pp. 1982–1997. doi: 10.1002/for.3104.
- Springer, S., Strzelecki, A. and Zieger, M. (2023) 'Maximum generable interest: a universal standard for Google Trends search queries', *Healthcare Analytics*, 3, 100158. doi: 10.1016/j.health.2023.100158.
- Springer, S., Strzelecki, A., Holmes, B., Ziermann-Canabarro, J. M. and Kaatz, M. (2025) 'Setting trends with Google: limits and perspectives when utilising search engine data', *Quality & Quantity*. doi: 10.1007/s11135-025-02431-0.
- Starr-McCluer, M. (2000) 'The effects of weather on retail sales'. Finance and Economics Discussion Series 2000-08, Board of Governors of the Federal Reserve System.

- Steele, A. T. (1951) 'Weather's effect on the sales of a department store', *Journal of Marketing*, 15(4), pp. 436–443. doi: 10.1177/002224295101500404.
- Steegmans, J. (2021) 'The pearls and perils of Google Trends: a housing market application', *Big Data*, 9(6), pp. 443–453. doi: 10.1089/big.2020.0183.
- Steinker, S., Hoberg, K. and Thonemann, U. W. (2017) 'The value of weather information for e-commerce operations', *Production and Operations Management*, 26(10), pp. 1854–1874. doi: 10.1111/poms.12721.
- Stephens-Davidowitz, S. and Varian, H. (2015) 'A hands-on guide to Google data'. Unpublished manuscript, Google Inc.
- Štulec, I., Petljak, K. and Naletina, D. (2019) 'Weather impact on retail sales: how can weather derivatives help with adverse weather deviations?', *Journal of Retailing and Consumer Services*, 49, pp. 1–10.
- Tarek, M., Brissette, F. P. and Arsenault, R. (2020) 'Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America', *Hydrology and Earth System Sciences*, 24(5), pp. 2527–2544. doi: 10.5194/hess-24-2527-2020.
- Tian, X., Cao, S. and Song, Y. (2021) 'The impact of weather on consumer behavior and retail performance: evidence from a convenience store chain in China', *Journal of Retailing and Consumer Services*, 62, 102583.
- Travers, J. and Payne, V. G. (1998) 'The impact of climate change on traditional seasonal ranges for high street women's wear in England', *Journal of Fashion Marketing and Management*, 2(4), pp. 370–382. doi: 10.1108/eb022541.
- Varian, H. (2023) 'Nowcasting with Google Trends', *The Economic Record*, 99(S1), pp. 1–12. doi: 10.1111/1475-4932.12783.
- Vosen, S. and Schmidt, T. (2011) 'Forecasting private consumption: survey-based indicators vs. Google Trends', *Journal of Forecasting*, 30(6), pp. 565–578. doi: 10.1002/for.1213.
- White, M. A., de Beurs, K. M., Didan, K. et al. (2009) 'Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006', *Global Change Biology*, 15(10), pp. 2335–2359. doi: 10.1111/j.1365-2486.2009.01910.x.
- Woloszko, N. (2020) 'Tracking activity in real time with Google Trends'. OECD Economics Department Working Papers No. 1634. Paris: OECD Publishing. doi: 10.1787/6b9c7518-en.
- Yoo, J., Eom, J. and Zhou, Y. (2024) 'Thermal comfort and retail sales: a big data analysis of extreme temperature's impact on brick-and-mortar stores', *Journal of Retailing and Consumer Services*, 77, 103699. doi: 10.1016/j.jretconser.2023.103699.
- Zhang, X., Friedl, M. A., Schaaf, C. B. et al. (2003) 'Monitoring vegetation phenology using MODIS', *Remote Sensing of Environment*, 84(3), pp. 471–475. doi: 10.1016/S0034-4257(02)00135-9.
- Zippenfenig, P. (2023) Open-Meteo.com weather API [computer software]. Zenodo. doi: 10.5281/zenodo.7970649.

Appendices

A Entity identifiers

Table 19 — Entity identifiers and their resolutions

SLOT	IDENTIFIER	RESOLVES TO	CATEGORY
coat	/m/01xygc	Coat (garment)	wintery
sweater	/m/02sb4j	Sweater (clothing)	wintery
boots	/m/01b638	Boot (shoe)	wintery
gloves	/m/0174n1	Glove (garment)	wintery
scarf	/m/02h19r	Scarf	wintery
beanie	/m/02gw_5	Beanie (seamed cap)	wintery
shorts	/m/01bfm9	Shorts (garment)	summery
sunglasses	/m/017ftj	Sunglasses	summery
swim	/m/01gkx_	Swimwear (clothing)	summery
sandals	/m/03nfch	Sandal (shoe)	summery
flipflops	/m/025znc	Flip-flops (shoe)	summery
umbrella	/m/0hnnb	Umbrella	rainwear
raincoat	/m/01v229	Raincoat (garment)	rainwear
shoes	/m/06rrc	Shoes	placebo
belt	/m/0176mf	Belt (clothing)	placebo

Curated /m/ identifiers were preferred throughout. Machine-generated /g/ identifiers proved unusably sparse: /g/11h79tz_t9 ('Winter Coat') returned zero in 147 of 209 Paris weeks against zero empty weeks for /m/01xygc.

B Markets

Forty-one markets with Trends geographic codes, coordinates, hemisphere and seasonal amplitude are tabulated in the accompanying workbook, sheet 'By city'. Geographic codes requiring substitution, following the failure of subnational codes: London (national), Mexico City, Antwerp, Dhaka (national), Zurich (national). Section 13.2 discusses the consequence.

Table 21 — The market roster: all fifty cities by group

MARKET	HEMISPHERE	GROUP	REPORTED AT
Amsterdam	N	Northern panel	Chapters 7–9
Antwerp	N	Northern panel	Chapters 7–9
Barcelona	N	Northern panel	Chapters 7–9
Berlin	N	Northern panel	Chapters 7–9
Chicago	N	Northern panel	Chapters 7–9
Copenhagen	N	Northern panel	Chapters 7–9
Delhi	N	Northern panel	Chapters 7–9
Doha	N	Northern panel	Chapters 7–9

MARKET	HEMISPHERE	GROUP	REPORTED AT
Dubai	N	Northern panel	Chapters 7–9
Florence	N	Northern panel	Chapters 7–9
Istanbul	N	Northern panel	Chapters 7–9
London	N	Northern panel	Chapters 7–9
Los Angeles	N	Northern panel	Chapters 7–9
Madrid	N	Northern panel	Chapters 7–9
Milan	N	Northern panel	Chapters 7–9
Moscow	N	Northern panel	Chapters 7–9
Munich	N	Northern panel	Chapters 7–9
New York	N	Northern panel	Chapters 7–9
Osaka	N	Northern panel	Chapters 7–9
Paris	N	Northern panel	Chapters 7–9
Riyadh	N	Northern panel	Chapters 7–9
Rome	N	Northern panel	Chapters 7–9
Seoul	N	Northern panel	Chapters 7–9
Stockholm	N	Northern panel	Chapters 7–9
Tokyo	N	Northern panel	Chapters 7–9
Toronto	N	Northern panel	Chapters 7–9
Warsaw	N	Northern panel	Chapters 7–9
Zurich	N	Northern panel	Chapters 7–9
Johannesburg	S	Southern panel	Section 10.1
São Paulo	S	Southern panel	Section 10.1
Sydney	S	Southern panel	Section 10.1
Bangkok	N	Tropical	Section 10.3
Dhaka	N	Tropical	Section 10.3
Ho Chi Minh City	N	Tropical	Section 10.3
Jakarta	S	Tropical	Section 10.3
Kuala Lumpur	N	Tropical	Section 10.3
Lagos	N	Tropical	Section 10.3
Mexico City	N	Tropical	Section 10.3
Miami	N	Tropical	Section 10.3
Mumbai	N	Tropical	Section 10.3
Singapore	N	Tropical	Section 10.3
Beijing	N	Excluded — no usable data	Section 5.6
Chengdu	N	Excluded — no usable data	Section 5.6
Guangzhou	N	Excluded — no usable data	Section 5.6
Hangzhou	N	Excluded — no usable data	Section 5.6
Shanghai	N	Excluded — no usable data	Section 5.6

MARKET	HEMISPHERE	GROUP	REPORTED AT
Shenzhen	N	Excluded — no usable data	Section 5.6
Porto	N	Excluded — insufficient volume	Section 5.6
Hong Kong	N	Excluded — no definable winter	Section 5.6
Taipei	N	Excluded — no definable winter	Section 5.6

Northern panel: seasonal amplitude of 12 °C or more, all in the northern hemisphere (28). Southern panel: extratropical southern-hemisphere markets below the amplitude threshold, analysed with seasonal windows inverted (3). Tropical: markets with no thermal season, retained as the boundary condition (10). Excluded: removed for cause with the measurements given at Table 4 (9).

C Excluded markets

Nine markets were excluded with the measurements given at Table 4. The six mainland Chinese markets return a mean index of 0.5 with 206 to 208 of 209 weeks recorded as exactly zero; Porto falls below the volume screen at both regional and national scope; Hong Kong and Taipei have insufficient seasonal amplitude for an onset to be defined.

D Model output

The complete grid of 328 specifications across 41 markets, with coefficients, standard errors, HAC t statistics, R^2 , out-of-sample correlations and Benjamini–Hochberg q values, is in the accompanying workbook, sheets 'All models' and 'By city'.

E Placebo audit

Placebo results for every market and driver combination, with the eleven failures identified at Table 17, are in the accompanying workbook, sheet 'Placebo audit'.

F Reproduction and reporting standard

Two documented pipelines reproduce the dataset from source. One retrieves ERA5 daily series through the Open-Meteo archive and transforms them into detrended per-week-of-year anomalies; the other retrieves Google Trends series by entity identifier and chain-links them across overlapping windows. Both cache to disk and resume after interruption. Reproduction requires no proprietary data, no paid interface and no credentialed API.

The reporting standard adopted is that of Mavragani and Ochoa (2019), which requires query, query type, time range, location and retrieval date to be documented for every series. This thesis meets the first four. It does not fully meet the fifth: retrieval dates vary across markets by up to several weeks because the panel was assembled incrementally across sessions, and the sampling literature establishes that the same query retrieved on different dates can return different series (Franzén, 2023; Cebrián and Domenech, 2024). No repeated sampling was performed, and each series rests on a single draw.

G Glossary

Anomaly — the deviation of an observation from the value expected for that week of the year in that market, after removal of the secular trend, expressed in standard deviations of the residual for that week.

Chain-linking — splicing two Trends windows onto a common scale using the ratio of their sums across shared weeks.

Composite — the unweighted mean of standardised member series within a category.

Entity — a Google Knowledge Graph identifier denoting a concept rather than a string, language-independent and pre-disambiguated.

HAC — heteroskedasticity- and autocorrelation-consistent standard errors; here Newey–West at six lags.

Offset — the date at which a smoothed seasonal composite crosses half its annual amplitude on the falling limb.

Onset — the date at which a smoothed seasonal composite crosses half its annual amplitude on the rising limb.

Placebo — a series with no plausible causal link to the driver, carried through every specification as a control that must return a null.

Reanalysis — a physically consistent reconstruction of past atmospheric state, constrained by observations but produced by a model.

SD — standard deviation, computed within week-of-year across years unless otherwise stated.

Summery — the garment category comprising shorts, sunglasses, swimwear, sandals and flip-flops.

Width — the interval between onset and offset within one annual cycle.

Wintery — the garment category comprising coat, sweater, boots, gloves, scarf and beanie.

COLOPHON

Weather: ERA5 reanalysis (Hersbach et al., 2020) accessed via Open-Meteo (Zippenfenig, 2023), 2006–2026. Sunday–Saturday weekly aggregation; detrended per-week-of-year standardisation.

Search: Google Trends weekly index, fifteen Knowledge Graph entities, four magnitude-grouped batches across two overlapping windows, chain-linked on shared weeks.

Panel: 41 markets × 429 weeks, May 2018 – August 2026. 328 specifications. Newey–West standard errors at six lags; Benjamini–Hochberg control across the specification grid; out-of-sample validation on 2024–2026; leave-one-out jackknife by year and by market.

Unresolved defect: batch composition calibrated on a single market (Section 12.4). Governing limitation: no transaction data was observed (Section 13.1).

Set in Georgia and Arial. Figures produced in matplotlib. Document generated programmatically; the build script accompanies the workbook.