Why Twelve Notes? The Economics of Musical Scale Selection

Working Paper

2026-02-01

Current Version: `https://eamonmcginn.com/papers/wp_scales.pdf`

Eamon McGinn

January 31, 2026

Abstract

Why do musical cultures converge on scales of 5–12 notes rather than using the full continuum of audible frequencies? This paper develops an economic model in which scale size emerges from joint optimisation over harmonic value, instrument construction cost, and performer cognitive cost. When harmonic returns diminish and cognitive costs grow superlinearly, a unique interior optimum exists. Calibration from acoustic fundamentals yields a harmonic value exponent of 1.86; the cognitive cost exponent is partially identified but plausibly exceeds 2.2. Empirical analysis of 468 classical compositions spanning 1485–1963 supports four predictions: harmonic complexity increases with economic prosperity ($p = 0.014$), rises over historical time (0.13 bits per century, $p < 0.001$), increases over composers' careers (0.33 bits per century, $p = 0.003$), and decelerates approaching a cognitive ceiling. The historical progression from pentatonic through diatonic to chromatic systems, and the persistence of 12-tone equal temperament in electronic music despite zero construction costs, emerge as natural consequences of declining construction costs against a fixed cognitive bound.

Keywords: Musical scales, cognitive constraints, cultural economics, pitch systems, optimisation

JEL Codes: Z11, D91, O33

1 Introduction

Humans can distinguish thousands of distinct pitches, yet the world's musical traditions converge on remarkably small subsets of the frequency spectrum. Pentatonic scales with five notes appear independently in Scottish folk songs, Chinese classical music, West African drumming, and Andean pan flute traditions. The seven-note diatonic scale dominates Western music from Gregorian chant to Taylor Swift. Twelve-note chromatic systems became standard with industrial-era piano manufacturing and persist in electronic music despite synthesisers' capacity to produce any frequency at negligible cost. This convergence is puzzling. If more notes enable richer harmony, why do musicians worldwide stop at 12? If cultural convention explains scale choice, why do unconnected traditions arrive at the same numbers?

This paper argues that scale size reflects economic optimisation. Musicians and instrument builders jointly maximise harmonic utility subject to two constraints: the physical cost of constructing instruments with more notes, and the cognitive cost of learning and performing larger pitch systems. The model is simple. A scale designer maximises $\pi(n) = \alpha n^{\beta} - c_1 n - c_2 n^{\gamma}$, where n is the number of notes, the first term captures harmonic value, the second is construction cost, and the third is cognitive cost. When harmonic returns diminish ($\beta < 2$) and cognitive costs eventually dominate ($\gamma > \beta$), a unique interior optimum exists. As construction costs fall through technological progress, optimal scale size expands until it hits a cognitive ceiling that binds even when construction becomes costless.

The mathematical properties of scales have attracted attention since Pythagoras. Modern music theory characterises which scales have desirable structural properties [Clough and Douthett, 1991, Carey and Clampitt, 1989, Tymoczko, 2011]. Recent work in computational musicology has modelled scale selection as a multi-constraint optimisation problem. Gill and Purves [2009] find that scales optimised for harmonic conformity to vocal spectra match those used cross-culturally. McBride et al. [2024] show that trade-offs between information rate constraints, motor constraints, and melody length can predict empirical scale distributions across 62 folk music corpora. These studies treat scale selection as a cognitive or biological optimisation problem, but not an economic one: they do not model construction costs, technological change, or the role of wealth in relaxing constraints.

Meanwhile, computational musicology has documented that harmonic complexity increased over the history of Western classical music. Serra-Peralta et al. [2021] analyse approximately 9,500 MIDI files spanning six centuries and find a clear increasing linear trend in harmonic vocabulary richness. Moss [2019] traces transitions of tonality across 2,000 pieces by 75 composers. Nakamura and Kaneko [2019] document a steady increase in dissonant interval frequencies over 400 years. Yet this literature describes trends without explaining them. Why did complexity increase? Why did it plateau? What economic or technological forces drove the trajectory?

Economic analyses of music focus on markets for performances and recordings [Throsby, 1994], intellectual property [Towse, 2001], or the cost disease afflicting orchestras [Baumol and Bowen, 1966], but treat musical content as exogenous. Scherer [2004] examines composer economics in detail but takes scale structure as given. To my knowledge, no prior work applies economic methodology—utility maximisation, cost-benefit analysis, comparative statics on prices and income—to explain why scales take the sizes they do.

This paper makes three contributions. First, it provides the first economic model of musical pitch system selection, treating scale size as an optimisation outcome responding to construction costs and cognitive costs. While information-theoretic models explain scale convergence through cognitive constraints alone [McBride et al., 2024, Gill and Purves, 2009], the economic framework uniquely incorporates supply-side factors: instrument manufacturing costs that vary with technology and wealth. This generates predictions about how scale complexity should change over time and across societies as economic conditions change.

Second, I show that the model's key parameters can be calibrated from acoustic fundamentals. The harmonic value exponent $\beta = 1.86$ is point-identified from Tenney height consonance measures with $R^2 = 0.996$. The cognitive cost exponent γ is partially identified:

the observed cognitive bound at 12 notes restricts (γ, c_2) to a curve in parameter space, with the curve's shape implying γ most likely exceeds 2.2. This partial identification suffices for the model's comparative static predictions, which depend only on $\gamma > \beta$.

Third, I test the model's predictions using a dataset of 468 classical compositions spanning 1485–1963, matched to composer biographies and historical GDP estimates. While prior computational musicology has documented increasing harmonic complexity over time [Serra-Peralta et al., 2021, Moss, 2019, Nakamura and Kaneko, 2019], no study has linked these trends to economic variables or tested theory-driven predictions about the determinants of complexity. Four of five predictions find statistically significant support: complexity rises with national income ($p = 0.014$), advances with technology (0.13 bits per century, $p < 0.001$), increases over composers' careers (0.33 bits per century, $p = 0.003$), and decelerates as it approaches a cognitive ceiling. The economic framework thus explains not just that complexity increased, but why it increased when it did.

The paper proceeds as follows. Section 2 introduces the acoustic and musical concepts necessary for readers unfamiliar with music theory. Section 3 reviews related literature. Section 4 presents the theoretical model and derives comparative statics. Section 5 calibrates the harmonic value function from consonance calculations and characterises the partially identified set for cognitive cost parameters. Section 6 tests five predictions using MIDI data on classical compositions matched to historical GDP estimates, supplemented by a case study of flute and clarinet development. Section 7 concludes.

2 Musical Preliminaries

This section introduces the acoustic and musical concepts necessary for the formal model. Readers familiar with music theory may proceed to the next section.

2.1 Pitch, Frequency, and the Octave

A musical note corresponds to a sound wave of a specific frequency, measured in Hertz (Hz). By international convention established in 1939, the note A above middle C is defined as exactly 440 Hz, with all other notes defined relative to this reference pitch. Higher frequencies sound higher in pitch; lower frequencies sound lower.

The octave is the most fundamental interval in music. When one frequency is exactly double another, the two pitches are said to be an octave apart. This relationship is so fundamental that it appears to be a human universal: listeners across cultures perceive notes an octave apart as essentially equivalent, differing only in register. This perceptual equivalence is reflected in musical notation, where notes an octave apart share the same letter name: A_3 at 220 Hz, A_4 at 440 Hz, and A_5 at 880 Hz are all called "A."

The octave equivalence phenomenon means that the problem of selecting musical pitches reduces to the problem of selecting frequencies within a single octave. Any selection made within one octave can be replicated in higher and lower octaves by successive doubling or halving. The question then becomes: which frequencies between f and $2f$ should be included in the musical system?

2.2 Scales and Consonance

A *scale* is a selection of a set of frequencies within the octave that serves as the raw material for musical composition and performance. The choice of which frequencies to include determines which melodies can be sung, which chords can be played, and which harmonic relationships are available to composers and performers.

Not all combinations of frequencies are equally pleasing to the ear. Since antiquity, theorists have observed that certain intervals sound consonant while others sound dissonant. The Greek mathematician Pythagoras is credited with discovering that consonant intervals correspond to simple integer frequency ratios. When two notes have frequencies in the ratio 3:2, the interval is a perfect fifth, widely regarded as one of the most consonant intervals after the octave itself. The ratio 4:3 produces a perfect fourth; 5:4 produces a major third; 6:5 produces a minor third. These intervals sound stable and harmonious.

The psychoacoustic basis for consonance is now well understood. Musical tones are not pure sine waves but complex waveforms containing a fundamental frequency plus integer multiples called overtones or harmonics. When two notes form a simple integer ratio, their overtone series overlap substantially, producing a coherent sound. When the ratio is complex, the overtones fall at different frequencies, creating beating and roughness that listeners perceive as dissonance [Plomp and Levelt, 1965].

2.3 Common Scale Systems

Most readers will be familiar with the diatonic scale, which comprises seven notes per octave: in the key of C major, these are C, D, E, F, G, A, and B. The diatonic scale has dominated Western music since ancient Greece and remains the foundation of most popular and classical music today. Its seven notes are not equally spaced but arranged in a pattern that creates a rich variety of consonant intervals. The diatonic scale contains perfect fifths, perfect fourths, major and minor thirds, and major and minor sixths, all approximating simple integer ratios.

The pentatonic scale is even more parsimonious, selecting only five notes per octave. In C major, a common pentatonic scale comprises C, D, E, G, and A, omitting the fourth and seventh degrees. The pentatonic scale appears independently in folk traditions across the world, from Scottish ballads to Chinese classical music to West African drumming to Native American chant. This near-universality suggests that the pentatonic scale may represent something close to a human default, requiring minimal cultural transmission to discover. The pentatonic scale's prevalence of consonant intervals is even higher than the diatonic: by omitting the two notes that create the most dissonant intervals with others, it achieves a simpler, more uniformly harmonious sound.

The chromatic scale extends to 12 notes per octave, adding five notes between the seven diatonic pitches. On a piano, the white keys play the diatonic scale in C major while the black keys supply the remaining chromatic notes. In *equal temperament*, the modern tuning system that has dominated Western music since the nineteenth century, each of these 12 notes is separated from its neighbours by exactly the same frequency ratio: the twelfth root of 2, or approximately 1.0595. This equal spacing simplifies transposition between keys at the cost of slightly compromising the pure integer ratios favoured by earlier tuning systems.

2.4 Why These Numbers?

There is no acoustic necessity that scales contain 5, 7, or 12 notes. The octave is a continuous frequency space; any number of divisions is mathematically possible. Microtonal systems with 19, 24, 31, or 53 notes per octave have been explored by theorists and composers. Indian classical music employs a system of 22 *shruti* (microtonal intervals) from which ragas are constructed. Arabic *maqam* and Turkish *makam* systems divide the octave more finely than Western practice, employing quarter-tones and other intervals absent from the 12-note chromatic scale.

Yet despite this diversity, the overwhelming majority of the world's music operates within scales of 5 to 12 notes. Microtonal systems remain the province of specialists and experimenters. Even in traditions that theoretically recognise finer divisions, practical performance typically centres on a subset comparable in size to Western scales. This convergence across cultures suggests that scale size is constrained by forces more fundamental than arbitrary convention.

2.5 The Trade-off

As the number of notes in a scale increases, so does the potential for harmonic richness. More notes mean more possible intervals, more possible chords, and more possible melodic contours. A composer working with 12 notes has access to harmonic resources unavailable to one working with 5. However, these benefits are subject to diminishing returns. The most consonant intervals, the perfect fifth and fourth, appear in even the smallest scales.

Additional notes provide access to thirds, sixths, and eventually more dissonant intervals, but each marginal note adds less harmonic value than the last.

Against these benefits stand two costs that the traditional musicological literature largely ignores but that become central once we adopt an economic perspective.

First, instruments must be constructed to produce the selected pitches. Each note in a scale typically requires dedicated physical infrastructure: a pipe in an organ, a string in a harp, a hole or key in a woodwind, a fret on a lute. More notes mean more components, more materials, more labour in construction, and more complexity in maintenance. Before industrial manufacturing, the cost of adding notes to an instrument was substantial. A simple wooden flute with six finger holes is far cheaper to produce than a Boehm-system flute with its elaborate keywork.

Second, performers must learn to navigate the scale. Musicians do not merely memorise individual notes; they internalise relationships between notes, including fingering patterns, interval recognition, chord voicings, and voice-leading conventions. This knowledge grows combinatorially with scale size: n notes generate $n(n-1)/2$ intervals, $\binom{n}{3}$ triads, and exponentially many possible progressions. Working memory constraints limit how much of this structure a performer can hold in mind during real-time performance.

The selection of a scale thus involves a trade-off. Larger scales offer richer harmonic possibilities but impose higher construction costs on instrument builders and higher cognitive costs on performers. The question of optimal scale size is not merely aesthetic but economic: it concerns the allocation of scarce resources, including materials, labour, and cognitive capacity, to competing ends.

3 Literature Review

3.1 Music Theory and Acoustics

The mathematical properties of musical scales have attracted scholarly attention since Pythagoras discovered that consonant intervals correspond to simple frequency ratios. Modern music theory has characterised the structural properties that distinguish commonly used scales from arbitrary pitch collections. Clough and Douthett [1991] introduce the concept of maximal evenness, showing that diatonic and pentatonic scales distribute their notes as uniformly as possible around the octave given the constraints of 12-tone equal temperament. Carey and Clampitt [1989] demonstrate that these same scales can be generated by stacking perfect fifths, connecting scale structure to the acoustics of the harmonic series. Tymoczko [2011] provides a comprehensive geometric framework for understanding voice leading and chord relationships, treating pitch space as a continuous manifold in which scales define discrete lattices.

Consonance perception has been studied extensively in psychoacoustics. Plomp and Levelt [1965] establish that consonance depends on the absence of beating between adjacent partials in the harmonic series. Tenney [1988] propose measuring interval complexity by the Tenney height $\log_2(p \times q)$ for a frequency ratio $p : q$ in lowest terms, with lower values indicating greater consonance. This measure has strong empirical support and provides the foundation for the harmonic value function calibrated in Section 5.

3.2 Optimisation Models of Scale Selection

Recent work models scale selection as multi-constraint optimisation, though with fixed rather than endogenous constraints. Gill and Purves [2009] computationally evaluate 40 million potential scales by harmonic conformity to vocal spectra, finding top-ranked scales match those used cross-culturally. Brown and Phillips [2025] find scales optimise for vocal production imprecision. McBride et al. [2023] document statistical convergence on 5–7 note equidistant scales across 46 countries. Most comprehensively, McBride et al. [2024] present a formal model where scale size emerges from trade-offs between information rate, motor constraints, and melody length, successfully predicting scale degree distributions across 62 folk music corpora.

These studies share the present paper's core insight: scale properties emerge from optimisation rather than arbitrary convention. The key distinction is that existing models treat constraints as fixed biological or cognitive parameters. This explains cross-sectional convergence but not historical dynamics. Why did Western scales expand from pentatonic to diatonic to chromatic over centuries? The economic framework developed here allows constraints to vary with technology and wealth, generating dynamic predictions testable against historical data.

3.3 Computational Musicology and Historical Trends

Computational analyses of large musical corpora have established an empirical regularity the present paper seeks to explain: harmonic complexity increased substantially over Western music history, then plateaued. Serra-Peralta et al. [2021] analyse approximately 9,500 MIDI files spanning six centuries and document a clear increasing linear trend in harmonic vocabulary richness. Buongiorno Nardelli et al. [2022] corroborate this using network entropy measures across 500 years of scores. Moss [2019], Moss et al. [2024] analyse roughly 3 million notes across 2,000 pieces, documenting increasingly complex tonal interval relations from 1360–1940, with particular focus on nineteenth-century transformations. Nakamura and Kaneko [2019] find steady increases in dissonant interval frequencies over 400 years.

This literature documents trends without explaining them. What economic or technological forces drove the trajectory? Why did complexity plateau rather than continue expanding? The present paper addresses these questions by deriving predictions from an optimisation model with endogenous constraints and testing whether complexity trends correlate with economic variables as the theory predicts.

3.4 Cultural Economics of Music

Economic analyses of music have focused on markets and institutions rather than musical structure itself. Throsby [1994] surveys performing arts economics; Towse [2001] examines intellectual property; Baumol and Bowen [1966] analyse orchestra economics; Scherer [2004] provides detailed analysis of composer careers and earnings but treats musical content as entirely exogenous. Work on cultural transmission [Bikhchandani et al., 1992, Salganik et al., 2006] studies how songs diffuse rather than why they take particular forms.

To my knowledge, no prior work applies economic methodology to pitch system selection. The present paper addresses this gap.

3.5 Cognitive Constraints in Music

Psychological research establishes that working memory limits constrain musical processing. Miller [1956] famously identified the "magical number seven, plus or minus two" as the capacity of short-term memory for distinct items. Subsequent work has refined this estimate and explored how chunking extends effective capacity. Cowan [2001] argues that the true limit is closer to four items when rehearsal is prevented, but that expertise enables increasingly sophisticated chunking.

In musical contexts, Snyder [2000] shows how memory constraints shape melodic structure, with phrases typically spanning seven notes or fewer. Krumhansl [1990] documents the "tonal hierarchy": listeners perceive pitch classes as more or less stable relative to an established key, suggesting that even within a 12-note chromatic system, cognitive processing imposes effective constraints on which notes are salient.

These findings motivate the cognitive cost function in the present model. The combinatorial structure of musical knowledge—encompassing intervals, triads, seventh chords, voice-leading conventions, and their interrelationships—grows superlinearly in the number of pitch classes. A performer working with n notes must internalise $n(n-1)/2$ intervals, $\binom{n}{3}$ possible triads, and exponentially many chord progressions. Memory limits ensure that cognitive costs eventually dominate harmonic benefits, preventing scale explosion even when construction costs vanish.

4 Theory

This section formalises the trade-off described above.

4.1 Model Setup

Let n denote the number of notes in a scale. I treat the choice of which n notes to include as solved at the intensive margin: the optimal n-note scale is well approximated by stacking $n-1$ perfect fifths and folding into a single octave, a procedure that generates the pentatonic scale at $n = 5$, the diatonic scale at $n = 7$, and the chromatic scale at $n = 12$ [Carey and Clampitt, 1989]. The extensive margin problem is to choose n.

A scale designer maximises net benefit:

$$\pi(n) = H(n) - C(n) - M(n) \tag{1}$$

where $H(n)$ is harmonic value, $C(n)$ is construction cost, and $M(n)$ is memorisation cost.

This specification raises the question of who or what this "scale designer" is. The optimisation framework should not be interpreted as literal maximisation by a single agent. Rather, scale selection emerges from decentralised interaction among three groups: instrument builders, who face construction costs and experiment with adding notes; composers, who exploit the harmonic possibilities of available instruments; and performers, who bear cognitive costs and gravitate toward systems they can master. No single agent solves the optimisation problem. Instead, the outcome resembles a market equilibrium where instrument builders who add notes that composers want and performers can handle see their designs adopted while those who do not see their instruments fall out of use. Composers who write for instruments that exist and performers who can play attract audiences; those who demand unavailable pitches or exceed cognitive limits do not. The profit function $\pi(n)$ is a reduced-form representation of this joint exploration process, analogous to how a social welfare function summarises decentralised market outcomes without implying a central planner. In this approach, comparative statics describe how the equilibrium shifts when underlying parameters change, not the decision rule of any particular agent.

4.1.1 Harmonic Value

The harmonic value of a scale depends on the number and quality of intervals it contains. With n notes, there are $n(n-1)/2$ distinct unordered intervals. Not all intervals are equally valuable: those approximating simple integer frequency ratios produce consonance through overlapping overtone series. The octave (2:1), perfect fifth (3:2), perfect fourth (4:3), and major and minor thirds (5:4 and 6:5) are highly consonant; the tritone and minor second are relatively dissonant.

I model harmonic value as a power function:

$$H(n) = \alpha n^{\beta} \tag{2}$$

where $\alpha > 0$ is a scaling parameter and $\beta > 0$ governs returns to scale.

This functional form is deliberately flexible. The value of β is an empirical question, not an assumption, and theoretical arguments exist for different parameter ranges:

- If $\beta > 2$, harmonic value exhibits increasing returns that outpace the growth in interval count. This would arise if intervals were complementary, with each new note making existing notes more valuable through network-like effects.

- If $\beta = 2$, harmonic value grows proportionally with the number of intervals. Each new interval would contribute equally to total value regardless of scale size.

- If $1 < \beta < 2$, harmonic value grows faster than linearly but slower than the number of intervals. More notes add value, but each marginal interval contributes less on average than previous intervals.

- If $\beta = 1$, harmonic value is linear in scale size. Each note contributes a fixed amount regardless of what other notes are present.

- If $\beta < 1$, harmonic value exhibits diminishing returns even in absolute terms. Additional notes add less and less value.

Conceptually it is likely that $1 < \beta < 2$. The case for $\beta < 2$ rests on interval quality degradation. The most consonant intervals, the perfect fifth and fourth, appear in even small scales generated by stacking fifths. As more notes are added, the marginal intervals approximate these ideal ratios less closely or introduce dissonance. The case for $\beta > 1$ rests on the value of harmonic variety. Composers benefit from having access to a richer palette of intervals, even if marginal intervals are individually less consonant than inframarginal ones.

The value of β will be covered in Section 6 which calibrates β from acoustic fundamentals by computing the total consonance of optimal n-note scales.

4.1.2 Construction Cost

Physical instruments require dedicated hardware for each note: pipes in organs, strings in harps and pianos, holes in woodwinds, frets on lutes, keys on brass instruments. I model construction cost as linear in scale size:

$$C(n) = c_1 n \tag{3}$$

where $c_1 \geq 0$ varies with manufacturing technology.

Intuitively, three historical eras correspond to distinct cost regimes. In the pre-industrial period, instrument makers carved wood by hand, forged metal components individually, and tuned each instrument as a one-off creation. Adding a note to a flute meant boring an additional hole and fitting a key mechanism by hand; adding a string to a harpsichord required skilled labour for the frame, soundboard, and tuning. Construction costs per note were substantial, limiting practical scale sizes for all but the wealthiest patrons.

The industrial revolution transformed instrument manufacturing. Precision boring machines could produce uniform woodwind tubes. Standardised metal production lowered the cost of keys, strings, and mechanical components. Factory organisation enabled division of labour and economies of scale. The Boehm flute (1847), with its complex system of keys and rings, would have been prohibitively expensive under craft production but became accessible to professional and amateur musicians alike. Construction costs per note fell dramatically, though they remained positive as each additional key still required materials, machining, and assembly.

The electronic synthesiser represents a discontinuous shift. A digital oscillator produces any frequency with equal computational cost. Adding a note to a software instrument requires no additional hardware, no materials, no labour. In contemporary economies, the marginal construction cost c_1 approaches zero.

The linearity assumption is a simplification. In practice, construction costs may exhibit economies or diseconomies of scale depending on the instrument and production technology. However, the key comparative static, that technological progress reduces c_1 over time, likely with discrete jumps corresponding to manufacturing revolutions, is robust to alternative functional forms.

4.1.3 Cognitive Cost

Performers and composers face cognitive burdens that increase with scale size. These costs extend beyond simple memorisation to cover learning, composing, practising, and performing.

I model cognitive cost as:

$$M(n) = c_2 n^{\gamma} \tag{4}$$

where $c_2 > 0$ and $\gamma > 1$.

Consider the sources of cognitive load. A performer must internalise not just n individual notes but the relationships between them. Fingering patterns must be learned for each note and each transition between notes. Interval recognition, essential for sight-reading and improvisation, requires familiarity with $n(n-1)/2$ distinct intervals. Chord voicings multiply further: $\binom{n}{3}$ triads, $\binom{n}{4}$ seventh chords, and so on. Voice-leading conventions govern how these chords connect, adding another combinatorial layer.

Composers face analogous challenges. Writing for a larger scale means navigating a more complex harmonic space, with more possible chord progressions, more potential melodic contours, and more opportunities for both consonance and dissonance. The cognitive effort of composing a coherent piece increases with the size of the palette.

Practice time scales with complexity. A pianist learning to play in all 12 major keys must master 12 distinct fingering patterns for scales, arpeggios, and common chord progressions. A violinist working in a microtonal system with 24 notes per octave faces roughly twice this burden. Performance risk also increases as more notes mean more opportunities for error, more possibilities that must be excluded in real time, and greater demands on working memory during performance.

The value of γ will be covered in Section 6 which calibrates γ from actual performance data.

The psychological literature provides some guidance. Miller [1956] established that working memory holds approximately seven items, later refined to approximately four without rehearsal [Cowan, 2001]. These limits apply to unstructured information; expertise enables chunking that expands effective capacity. However, the relational structure of musical knowledge, where items are defined by their connections to other items, imposes costs that grow faster than item counts. Empirical calibration in Section 6 yields $\gamma \approx 2.5$, consistent with cognitive costs growing faster than quadratically but slower than cubically.

The parameter c_2 varies across individuals and contexts. Professional musicians with years of deliberate practice will likely have lower effective c_2 than amateurs. Written notation reduces c_2 relative to oral traditions by offloading memory to external storage. Formal music theory education may reduce c_2 by providing systematic frameworks for understanding harmonic relationships. These sources of variation generate testable predictions about which musicians and traditions should employ more complex scales.

4.2 Model Solution

The optimisation problem is:

$$\max_n \pi(n) = \alpha n^\beta - c_1 n - c_2 n^\gamma \tag{5}$$

The first-order condition is:

$$\frac{d\pi}{dn} = \alpha\beta n^{\beta-1} - c_1 - \gamma c_2 n^{\gamma-1} = 0 \tag{6}$$

Rearranging:

$$\alpha\beta n^{\beta-1} = c_1 + \gamma c_2 n^{\gamma-1} \tag{7}$$

The left-hand side is marginal harmonic benefit. The right-hand side is marginal cost, comprising construction cost c_1 per note plus marginal cognitive cost $\gamma c_2 n^{\gamma-1}$. This is a transcendental equation with no closed-form solution for general parameter values. However, we can characterise when an interior solution exists and derive its key properties.

4.3 Theoretical Results

Proposition 1 (Existence and Uniqueness)**.** *For $\beta < \gamma$ and sufficiently small costs c_1 and c_2, there exists a unique interior maximum $n^* > 1$.*

For a finite interior optimum to exist, marginal cost must eventually exceed marginal benefit as n grows. Marginal harmonic benefit grows as $n^{\beta-1}$ whilst marginal cognitive cost grows as $n^{\gamma-1}$. If $\gamma > \beta$, cognitive cost growth dominates for large n, ensuring the curves cross at some finite n^*. If $\gamma \leq \beta$, marginal benefit grows at least as fast as marginal cost, and scale size would expand without bound as construction costs fall.

The condition $\gamma > \beta$ is thus necessary for a well-behaved model. Harmonic value derives from intervals, whose count grows combinatorially, but degradation in the harmonic quality of note pairings suggests $\beta < 2$. Cognitive costs derive from relational knowledge with combinatorial structure, suggesting γ exceeds 2. If $\beta < 2 < \gamma$, the condition for an interior solution is comfortably satisfied.

Proposition 2 (Construction Cost Comparative Static). $\partial n^*/\partial c_1 < 0$. *Higher construction cost reduces optimal scale size.*

Proof. By implicit differentiation of the first-order condition. See Appendix A. □

Proposition 3 (Cognitive Cost Comparative Static). $\partial n^*/\partial c_2 < 0$. *Higher cognitive cost reduces optimal scale size.*

Proof. By implicit differentiation of the first-order condition. See Appendix A. □

These comparative statics push towards parsimony in scale design: when building or learning additional notes is expensive, musicians economise on scale size.

Proposition 4 (Cognitive Bound). *As $c_1 \to 0$, n^* converges to a finite upper bound:*

$$\bar{n} = \left(\frac{\alpha\beta}{\gamma c_2}\right)^{1/(\gamma-\beta)} < \infty \tag{8}$$

Derivation. When construction costs vanish ($c_1 = 0$), the first-order condition simplifies to:

$$\alpha\beta n^{\beta-1} = \gamma c_2 n^{\gamma-1} \tag{9}$$

Dividing both sides by $n^{\gamma-1}$:

$$\alpha\beta n^{\beta-\gamma} = \gamma c_2 \tag{10}$$

Solving for n yields the cognitive bound. □

When construction costs approach zero, the optimal scale size converges to a finite bound. The economic mechanism is straightforward: marginal harmonic value grows at rate $n^{\beta-1}$ while marginal cognitive cost grows at rate $n^{\gamma-1}$. Since $\gamma > \beta$, cognitive costs eventually dominate, creating an interior optimum even when construction is costless. This bound depends only on harmonic returns (α, β) and cognitive parameters (c_2, γ), not on construction technology. This independence generates a testable prediction: electronic instruments, which face approximately zero construction costs, should employ scales no larger than $\bar{n}$. Calibration in Section 5 yields $\bar{n} \approx 10$–12, consistent with the persistence of 12-tone equal temperament in electronic music.

4.4 Testable Predictions

The propositions generate five empirically testable predictions.

Prediction 1 (Wealth Effects): Wealthier societies should exhibit larger scale sizes (from Proposition 2). Higher wealth likely reduces the relative construction cost c_1 through weaker budget constraints for consumers and increased capital stocks for producers. Both channels suggest that $\partial n^*/\partial\text{wealth} > 0$ operating through $\partial c_1/\partial\text{wealth} < 0$.

Prediction 2 (Technological Progress): Scale size should increase over time as manufacturing innovations reduce per-note construction costs (from Proposition 2). If $c_1(t)$ declines through technological progress, then $dn^*/dt > 0$ operating through $\partial n^*/\partial c_1 < 0$.

Prediction 3 (Instrument Heterogeneity): Other things constant, instruments with higher per-note construction costs should exhibit smaller scale size (from Proposition 2). This prediction relies on the fact that acoustic physics generates exogenous variation in c_1 across instruments. Alternatively, looking over time, instruments with higher c_1 should reach scale size thresholds later than instruments with inherently lower construction costs.

Prediction 4 (Cognitive Heterogeneity): Musicians with lower cognitive costs should use larger or more complex scale systems (from Proposition 3). If training and experience reduce c_2, then by $\partial n^*/\partial c_2 < 0$, we predict $\partial n^*/\partial \text{experience} > 0$ operating through $\partial c_2/\partial \text{experience} < 0$. Individual composers' harmonic complexity should increase over their careers as learning reduces the cognitive burden of navigating larger pitch systems.

Prediction 5 (Electronic Persistence): As electronic synthesis reduces $c_1 \to 0$, scale size should converge to a finite bound rather than expanding without limit (from Proposition 4). If the theory is correct, we should not observe an explosion in scale complexity when construction costs vanish. Additionally, the specific value at which electronic music stabilises reveals the cognitive bound $\bar{n}$ and provides information about model parameters. If music composed after the introduction of synthesis predominantly still uses 12-tone equal temperament, this implies $\bar{n} \approx 12$, which constrains the parameter combination $(\alpha\beta/\gamma c_2)^{1/(\gamma-\beta)}$.

5 Calibration

This section calibrates the model parameters from acoustic fundamentals and empirical observation. The harmonic value parameters (α, β) are point-identified from consonance calculations. The cognitive cost parameters (γ, c_2) are partially identified: the data determine a curve in parameter space rather than a unique point, but this suffices to establish the key theoretical condition $\gamma > \beta$.

5.1 Calibrating the Harmonic Value Function

The harmonic value function $H(n) = \alpha n^\beta$ can be empirically calibrated from acoustic fundamentals. Musical cultures worldwide have independently converged on scales generated by stacking perfect fifths. The procedure is simple: beginning from a root note, each successive note is placed a perfect fifth (frequency ratio 3:2, or 702 cents) above the previous, then folded back into a single octave. The resulting notes are sorted by pitch to yield the scale.

This method produces the familiar scales of Western music. Two stacked fifths yield the root and fifth (0, 702 cents). Five fifths produce the major pentatonic. Seven fifths generate the diatonic major scale. Twelve fifths yield the chromatic scale, at which point the cycle nearly closes: the thirteenth note falls just 24 cents (the Pythagorean comma) from the root.

Table 1 reports the scales generated for $n = 2$ to $n = 12$. The ubiquity of these structures across independent musical traditions suggests fifth-stacking represents a near-optimal solution to the problem of selecting n notes from the continuous pitch spectrum. I therefore take fifth-stacked scales as the empirical basis for calibrating $H(n)$.

Table 1: Fifth-Stacked Scales

n	Scale (cents)	Musical interpretation
2	0, 702	Root and fifth
3	0, 204, 702	Trichord
4	0, 204, 702, 906	Tetrachord
5	0, 204, 408, 702, 906	Pentatonic
6	0, 204, 408, 702, 906, 1110	Hexatonic
7	0, 204, 408, 612, 702, 906, 1110	Heptatonic
8	0, 114, 204, 408, 612, 702, 906, 1110	—
9	0, 114, 204, 408, 612, 702, 816, 906, 1110	—
10	0, 114, 204, 318, 408, 612, 702, 816, 906, 1110	—
11	0, 114, 204, 318, 408, 612, 702, 816, 906, 1020, 1110	—
12	0, 114, 204, 318, 408, 522, 612, 702, 816, 906, 1020, 1110	Chromatic

Note: Each scale is generated by stacking $n - 1$ perfect fifths (702 cents) from the root and folding into a single octave.

To compute $H(n)$ for each scale, I measure the consonance of every distinct interval pair and sum the results. Following Tenney [1988], interval consonance is based on the Tenney height: for a just ratio $p : q$ in lowest terms, the Tenney height is $\log_2(p \times q)$. Lower values indicate greater consonance. The octave (2:1) has Tenney height 1; the perfect fifth (3:2) has Tenney height 2.58; the tritone (45:32) has Tenney height 10.5.

The consonance of an interval at x cents is:

$$\text{consonance}(x) = \frac{\exp\left(-d^2/2\sigma^2\right)}{1 + T} \tag{11}$$

where d is the distance in cents to the nearest just interval, T is its Tenney height, and $\sigma = 10$ cents governs the tolerance for mistuning. The total harmonic value of an n-note scale $S = \{s_1, \ldots, s_n\}$ is then:

$$H(S) = \sum_{i<j} \text{consonance}(|s_j - s_i|) \tag{12}$$

With $H(n)$ computed for each scale size, I estimate the power law parameters by ordinary least squares on the log-linearised model:

$$\log H(n) = \log \alpha + \beta \log n + \varepsilon \tag{13}$$

Figure 1 plots the results. The fitted relationship is:

$$H(n) = 0.066 \times n^{1.861} \tag{14}$$

with $R^2 = 0.996$. The estimate $\hat{\beta} = 1.861$ has a standard error of 0.042, yielding a 95% confidence interval of $[1.77, 1.95]$. The estimate lies comfortably below 2, confirming diminishing returns to scale size. Fifth-stacking front-loads the most consonant intervals: the perfect fifth and fourth appear at $n = 2$, major and minor thirds enter at $n = 4$–5, and subsequent notes contribute progressively less consonant intervals.

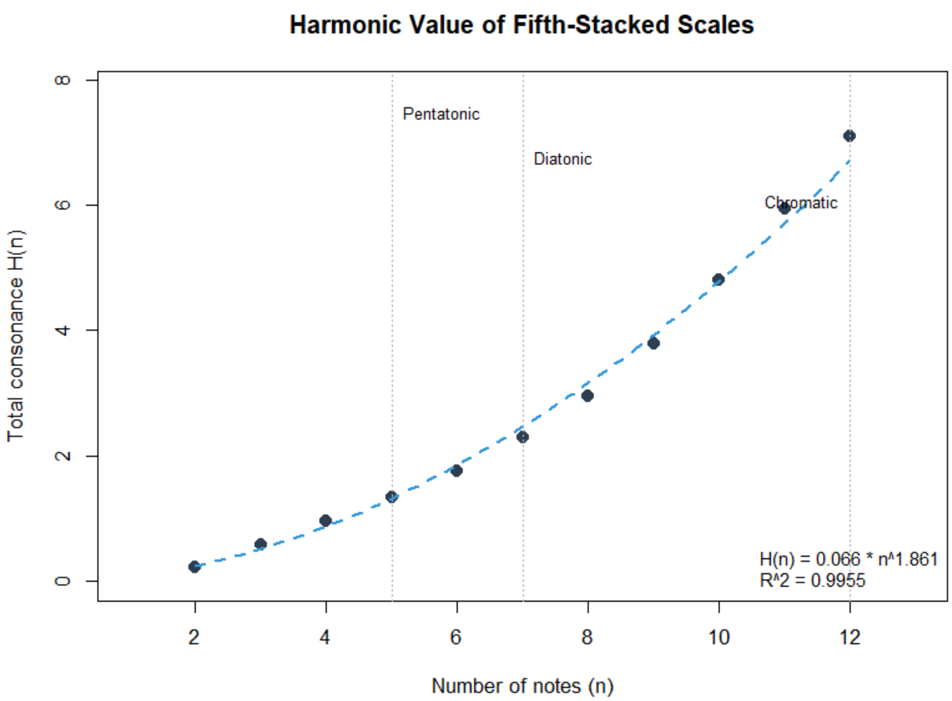

Figure 1: Harmonic Value of Fifth-Stacked Scales
Note: Points show total consonance $H(n)$ for scales generated by stacking perfect fifths. Dashed line shows fitted power law $H(n) = 0.066 \times n^{1.861}$ ($R^2 = 0.996$). Vertical lines mark pentatonic (5), diatonic (7), and chromatic (12) scales.

5.2 Calibrating the Cognitive Cost Function

The cognitive cost function $M(n) = c_2 n^\gamma$ contains two parameters. Unlike the harmonic value function, these cannot be separately identified from acoustic fundamentals. However, the observed apparent cognitive bound at $\bar{n} \approx 12$, explored further in Section 6, constrains the parameters to lie on a curve, and the shape of this curve provides insight into the feasible parameter range. Taking that the cognitive bound binds, Proposition 4 implies:

$$\bar{n} = \left(\frac{\alpha\beta}{\gamma c_2} \right)^{1/(\gamma - \beta)} \tag{15}$$

Rearranging to express c_2 as a function of γ:

$$c_2(\gamma) = \frac{\alpha\beta}{\gamma \cdot \bar{n}^{\gamma - \beta}} \tag{16}$$

With $\alpha = 0.066$, $\beta = 1.861$, and $\bar{n} = 12$, equation (16) defines a curve in (γ, c_2) space. Every point on this curve is consistent with the observed cognitive bound. Figure 2 plots the relationship.

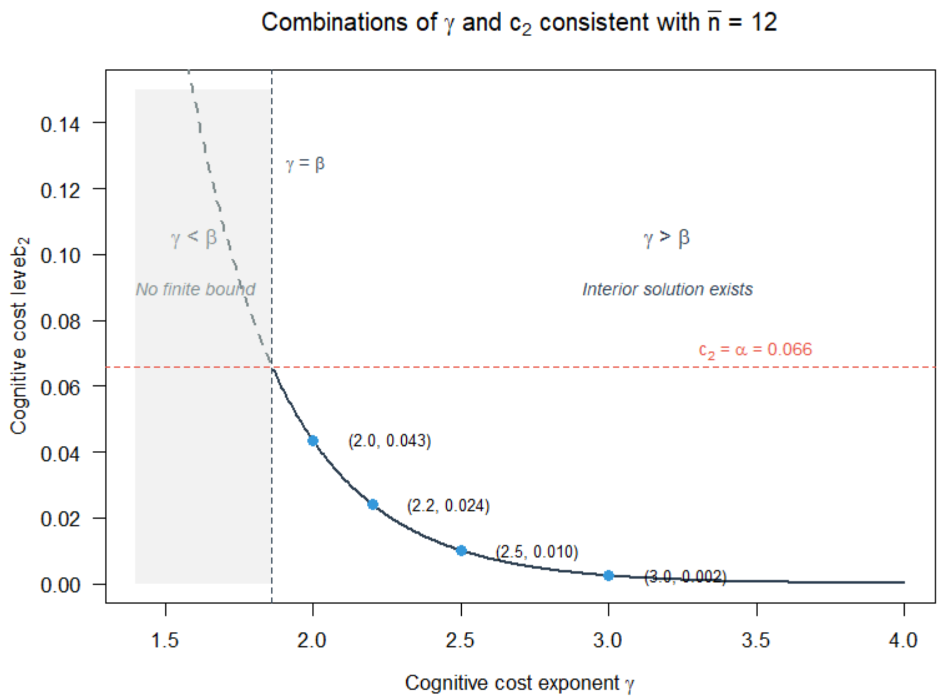

Figure 2: Partial Identification of Cognitive Cost Parameters
Note: Curve shows combinations of γ and c_2 consistent with cognitive bound $\bar{n} = 12$. The curve passes through $(\beta, \alpha) = (1.861, 0.066)$, which partitions the parameter space. The shaded region $\gamma < \beta$ admits no finite cognitive bound.

The curve has three notable properties.

First, it slopes downward in the valid region. Higher γ (cognitive costs that rise more steeply with scale size) requires lower c_2 (lower baseline cognitive cost) to generate the same ceiling at $\bar{n} = 12$. Intuitively, if cognitive costs accelerate rapidly, even a small baseline produces a binding constraint.

Second, the curve passes through the point $(\gamma, c_2) = (\beta, \alpha)$. Evaluating equation (16) at $\gamma = \beta$:

$$c_2(\beta) = \frac{\alpha\beta}{\beta \cdot 12^0} = \alpha \tag{17}$$

This crossing point partitions the parameter space. For $\gamma > \beta$, the exponent $(\gamma - \beta)$ is positive, so $12^{\gamma-\beta} > 1$, which pushes c_2 below α. For $\gamma < \beta$, the exponent is negative, so $12^{\gamma-\beta} < 1$, which pushes c_2 above α.

Third, the region $\gamma < \beta$ has no economic interpretation. When $\gamma < \beta$, harmonic benefits grow faster than cognitive costs as scale size increases. No finite cognitive bound exists and so, as construction costs fall, optimal scale size expands without limit. The shaded region in Figure 2 is therefore incompatible with a binding cognitive constraint.

The crossing point thus represents a boundary. Valid parameterisations must have $\gamma > \beta$, which implies $c_2 < \alpha$. The closer γ is to β, the closer c_2 approaches α from below. Table 2 reports selected points on the curve in the valid region.

Table 2: Cognitive Cost Parameters: Points on the Identified Set

γ	$\gamma - \beta$	c_2	c_2/α
2.0	0.14	0.043	65%
2.2	0.34	0.024	36%
2.5	0.64	0.010	15%
3.0	1.14	0.002	3%

Note: All combinations yield $\bar{n} = 12$ given $\alpha = 0.066$ and $\beta = 1.861$.

If γ is only slightly above β (say, $\gamma = 2.0$), the model requires baseline cognitive costs equal to 65% of baseline harmonic value. This would imply that even a two-note scale imposes substantial cognitive burden relative to its harmonic benefit. If instead γ is comfortably above β (say, $\gamma = 2.5$ or higher), baseline cognitive costs fall to 15% of harmonic value or less. Small scales would then be cognitively easy, with costs only dominating at larger n.

The latter scenario aligns better with musical reality. Simple scales like the pentatonic are learned readily by children and appear spontaneously across cultures. Cognitive costs seem to become salient only as musicians attempt to master larger pitch systems. This intuition suggests γ lies comfortably above β, perhaps $\gamma \geq 2.2$, without requiring an exact point estimate.

In summary, while γ and c_2 are not separately point-identified, the analysis establishes that γ most likely exceeds 2.2. Values closer to β would require implausibly high baseline cognitive costs relative to harmonic value. This partial identification does not constrain the empirical analysis that follows, which focuses on comparative statics. The predictions that complexity increases with wealth, advances with technology, and rises over composers' careers depend only on the signs of $\partial n^*/\partial c_1 < 0$ and $\partial n^*/\partial c_2 < 0$, which hold for any (γ, c_2) pair satisfying $\gamma > \beta$. The qualitative patterns are robust across the entire identified set.

6 Empirical Tests

Long-run trends in musical composition and instrument design provide direct tests of the model's predictions. The primary empirical strategy analyses a large corpus of MIDI-encoded compositions spanning 1485 to 1963, measuring harmonic complexity across composers, countries, and eras. This computational approach enables systematic testing of Predictions 1, 2, 4, and 5.

Prediction 3, concerning instrument heterogeneity, is better addressed using different evidence. Acoustic physics generates exogenous variation in per-note construction costs across instruments, but this variation does not appear in compositional data. I therefore employ a detailed case study of instrument development, examining how manufacturing constraints shaped chromatic capability over time. This approach follows the tradition of classical political economy, where Smith's analysis of pin manufacturing and Marx's detailed examination of machinery in *Capital* grounded economic theory in the material realities of production technology. Here, the physical constraints of flute versus clarinet construction provide comparable insight into how cost parameters shape optimal scale selection.

The remainder of this section presents tests of each prediction in turn.

6.1 Data

6.1.1 MIDI Corpus

The primary data source comprises 899 MIDI files from kunstderfuge.com, the largest repository of classical music in MIDI format. MIDI (Musical Instrument Digital Interface) encodes music as sequences of discrete note events, each specifying pitch, timing, velocity, and channel. This format enables systematic measurement of harmonic complexity across a large corpus.

I focus on 48 composers for whom comprehensive biographical information (birth year, death year, country) is readily available, covering 638 files representing 71% of the original corpus. Fuzzy matching of piece titles to musicological databases to obtain composition dates, supplemented with manual verification for ambiguous cases, yields 620 compositions with reliable dating. Finally, excluding files with MIDI parsing errors produces the final corpus of 468 compositions by 45 composers from 12 countries, spanning 1485 to 1963.

Table 3 summarises the sample by historical era.

Table 3: Sample Summary by Era

Era	Years	Compositions	Composers	Representatives
Renaissance	1485–1600	38	5	Palestrina, Josquin
Baroque	1600–1750	100	11	Vivaldi, Bach
Classical	1750–1820	96	9	Beethoven, Mendelssohn
Romantic/Modern	1820–1963	234	21	Stravinsky, Tchaikovsky
Total	1485–1963	468	45	

A potential concern is selection into the corpus. Compositions surviving to the present, transcribed to MIDI, and included in digital repositories may not be representative of their eras. If harmonically sophisticated pieces were more likely to be preserved and digitised, time trends could be biased upward. Early-period compositions in the sample may be an unrepresentatively sophisticated subset of what was actually performed and heard. This concern is partially mitigated by two features of the empirical strategy. The within-composer analysis (Prediction 4) examines career trajectories conditional on inclusion in the corpus, so survivorship bias affects only the level of complexity, not its slope over a composer's career. The cross-sectional wealth results (Prediction 1) compare contemporaneous compositions from different countries, where selection pressures are plausibly similar across nations in the same era. Nevertheless, the absolute levels of early-period complexity should be interpreted cautiously, and the secular time trend may overstate the true rate of increase if preservation rates correlate with harmonic sophistication.

6.1.2 Harmonic Complexity Measure

The primary dependent variable is *pitch class entropy*, a measure of how uniformly a composition distributes notes across the available pitch classes. This metric derives from Shannon's information theory [Shannon, 1948] and has been widely applied in computational musicology to quantify tonal diversity.

For a composition with pitch class distribution $(p_0, p_1, \ldots, p_{k-1})$ where k is the number of distinct pitch classes in the chromatic system (typically 12 in Western music), entropy is:

$$\text{Entropy} = -\sum_{i=0}^{k-1} p_i \log_2 p_i \tag{18}$$

The measure is expressed in bits, the standard unit of information content. Intuitively, entropy measures the number of binary questions (yes/no) required to identify which pitch class will appear next in a composition.

To illustrate: a pentatonic composition using 5 pitch classes with equal frequency has entropy $\log_2(5) = 2.32$ bits. A diatonic composition using 7 pitch classes equally has entropy $\log_2(7) = 2.81$ bits. A fully chromatic composition using all 12 pitch classes equally has entropy $\log_2(12) = 3.58$ bits. The minimum is 0 bits (only one pitch class used); the maximum for Western music is 3.58 bits. Non-Western systems employing finer pitch divisions could reach higher values. For example, Indian raga using all 22 shrutis would yield up to $\log_2(22) = 4.46$ bits, whilst Arabic maqam with 24 quarter-tones could reach $\log_2(24) = 4.58$ bits. Higher entropy indicates greater harmonic complexity and fuller use of the available pitch system.

In the sample, pitch class entropy ranges from 2.05 to 3.56 bits with mean 3.17 bits (median 3.21 bits, standard deviation 0.25 bits). Approximately 92% of compositions exceed the diatonic baseline of 2.81 bits, while 37% approach the chromatic ceiling (within 0.3 bits of 3.58). To contextualise these values: Josquin des Prez's *Vox Regis* (1495) achieves 2.17 bits, representing truly diatonic writing characteristic of Renaissance polyphony, well below the theoretical maximum of 2.81 bits for equal use of seven pitch classes. J.S. Bach's *Fughetta* (1680) at 2.81 bits exploits the full diatonic palette more uniformly. By contrast, works by Messiaen, Bartók, and Ravel from the 1920s-1930s reach 3.56 bits, effectively exhausting the chromatic scale. The distribution exhibits clear historical progression as Renaissance compositions average 2.85 bits, Baroque 3.01 bits, Classical 3.14 bits, and Romantic/Modern 3.30 bits.

The theoretical model predicts optimal scale size n^*, but the empirical measure is pitch class entropy, which captures usage intensity within available scales. These are related but not identical as a composer working with 12-TET instruments can still write purely diatonic music. However, usage provides a reasonable proxy for effective scale selection. Performers cannot use notes excluded from the physical instrument or absent from the composition, so usage is bounded above by technological availability. When instrument builders add a note, composers do not exploit it immediately. Instead, they gradually explore the expanded harmonic palette as they learn its possibilities. Rising entropy therefore indicates that composers are drawing on a larger effective scale, whether because new notes have become available or because existing availability is being more fully exploited. Both channels are consistent with the model as declining construction costs first expand available scales, then composers progressively incorporate the new possibilities into practice. What we measure is the outcome of this joint process.

Pitch class entropy is the primary measure because it directly captures the theoretical object of interest (how fully composers exploit available pitch classes). Alternative measures which capture related but distinct dimensions of complexity are explored in Appendix B. Dissonance ratio measures harmonic tension rather than chromatic breadth. Unique pitch classes counts scale size but ignores usage intensity. Maximum simultaneous notes captures textural density orthogonal to pitch selection. Appendix B tests all four predictions across these alternative measures. The time trend (Prediction 2) is unanimous across measures; the wealth effect (Prediction 1) and ceiling effect (Prediction 5) hold for three of four while the career effect (Prediction 4) is unanimous in sign though significant only for entropy. The primary measure thus performs well across all predictions, with alternative measures providing broadly consistent support.

6.1.3 Economic and Biographical Variables

Composer biographical data (birth year, death year, birth country) were manually compiled from musicological sources including Grove Music Online and composer biographies. I construct *career year* as the years elapsed since a composer's first composition in the sample, measuring accumulated experience and training effects.

Composition dates were obtained through fuzzy string matching of piece titles against musicological databases including the International Music Score Library Project (IMSLP) and Grove Music Online. Manual verification resolved ambiguous cases.

I match compositions to historical GDP per capita estimates using the Maddison Project

Database [Bolt and van Zanden, 2020]. GDP is interpolated to annual frequency and matched on the composer's birth country at composition date. This provides a proxy for economic conditions affecting instrument availability, manufacturing sophistication, and training opportunities.

6.2 Prediction 1: Wealth Effect

The model predicts that higher GDP per capita reduces effective construction costs, enabling more complex scale systems. I test this using the specification:

$$\text{Entropy}_{ict} = \beta_1 \log(\text{GDP}_{ct}) + \beta_2 \text{Year}_t + \mu_c + \varepsilon_{ict} \tag{19}$$

where i indexes compositions, c indexes countries, t indexes years, and μ_c are country fixed effects.

Identification. The country fixed effects absorb time-invariant cross-sectional differences between nations such as musical traditions, institutional structures, and baseline cultural orientations toward harmony. The year control absorbs common shocks affecting all countries simultaneously. Identification therefore comes from within-country deviations from trend and so the coefficient β_1 measures whether, in periods when a country is richer than its own historical trend would predict, its composers write more harmonically complex music.

This identification strategy requires that within-country GDP variation, conditional on secular time trends, is uncorrelated with other determinants of harmonic complexity. Several factors could violate this assumption. Musical institutions such as conservatories and opera houses may expand during economic booms, confounding wealth with training infrastructure. Patronage opportunities may fluctuate with business cycles, affecting which composers can work and what risks they can take. Cultural confidence or openness to experimentation might correlate with economic conditions independently of the cost channels the model emphasises.

These concerns are partially mitigated by the long time horizon. Year-to-year business cycle fluctuations are unlikely to drive compositional choices that reflect years of training and artistic development. The variation identifying β_1 is essentially sustained differences in prosperity across decades rather than transitory shocks. Nevertheless, the coefficient should be interpreted as a conditional correlation with strong theoretical motivation rather than a cleanly identified causal effect.

Results. Table 4 reports the results. The coefficient on log GDP is 0.089 ($p = 0.014$). Economically, doubling GDP per capita is associated with 0.062 bits higher entropy, controlling for secular time trends and country-specific factors. To contextualise this effect, the shift from a purely diatonic scale (entropy 2.81 bits) to full chromaticism (entropy 3.58 bits) spans 0.77 bits. The estimated wealth effect of doubling GDP corresponds to moving roughly 8% of the way along this trajectory, equivalent to expanding from strict diatonic writing to occasional use of accidentals.

Table 4: Prediction 1: Wealth Effect on Harmonic Complexity

	Coefficient	Std. Error	t-statistic	p-value
log(GDP per capita)	0.089	0.036	2.47	0.014
Composition year	0.00104	0.00014	7.28	$<$0.001
Country fixed effects	Yes (12 countries)			
R^2 (within)	0.345			
Observations	468			

Note: Dependent variable is pitch class entropy (bits). Standard errors clustered by composer. Identification from within-country variation over time.

Figure 3 displays the partial relationship between GDP and complexity. The plot shows residualised entropy against residualised log GDP, both after removing time trends and country effects. The positive slope is evident despite substantial composition-level noise.

Interpretation. The wealth coefficient should be interpreted as a conditional correlation with strong theoretical motivation rather than a cleanly identified causal effect. The identification strategy removes time-invariant country differences and common time trends, isolating

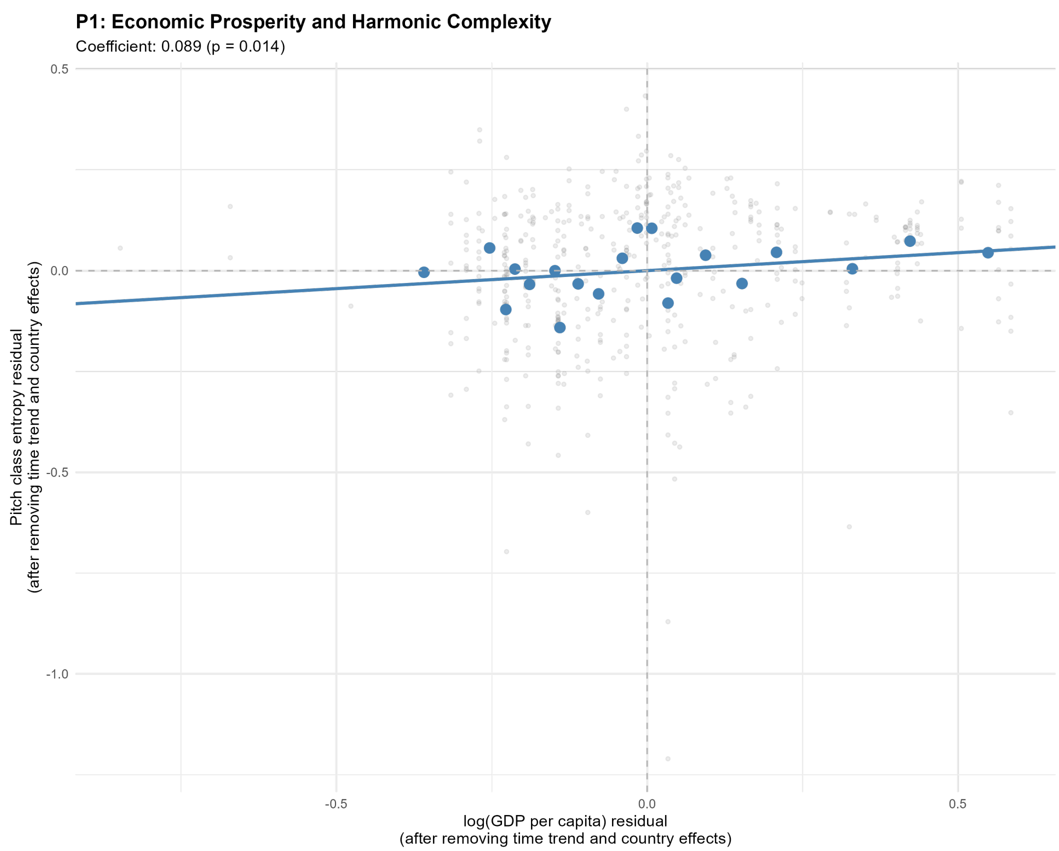

Figure 3: Prediction 1: Economic Prosperity and Harmonic Complexity
Note: Scatter plot shows entropy residuals against log GDP residuals after partialling out composition year and country fixed effects. Blue points are binned means; grey points are individual compositions. Line shows OLS fit.

within-country variation in GDP. However, this variation is not exogenous. I cannot rule out that omitted variables drive both GDP and compositional sophistication.

What the coefficient does establish is that the wealth-complexity relationship predicted by the model appears in the data with the correct sign and survives controls for secular technological progress and national musical traditions. Further robustness checks below also show that it holds across every well-sampled country in the corpus. The theoretical framework provides a mechanism linking prosperity to complexity through reduced effective construction costs and expanded access to training. While alternative mechanisms could generate similar correlations, the consistency of the pattern across diverse national contexts.

Reverse causality. One might ask whether the relationship runs in the opposite direction: could harmonic complexity somehow drive GDP rather than vice versa? This concern is implausible for several reasons. First, music represents a negligible share of economic activity, so any productivity spillovers from musical innovation would be too small to detectably affect national income. Second, the mechanisms by which GDP affects musical complexity are plausible. For example, wealth enables patronage systems, subsidises instrument construction and acquisition, and supports professional training institutions. The reverse channel lacks any credible mechanism. A potential way to address this concern more formally would be by looking for an instrumental variable. However, the relationship we are interested in operates over decades within a country. An instrumental variables approach would require an instrument affecting long-run economic development but not musical practice directly, which is difficult to conceive. However, the relevant counterfactual for IV would be identifying short-run causal effects, not the long-run structural relationship the model describes. The theoretical prediction concerns how equilibrium scale complexity responds to construction costs over extended periods, not whether a shock to GDP causes immediate changes in composition style. Overall, it's therefore reasonable to interpret the relationship as a conditional correlation with strong theoretical motivation whereby the causal channel runs from wealth to musical practice rather than a cleanly identified causal effect.

Country-level robustness. As a simple robustness check, I estimate bivariate regressions of complexity on log GDP within each country. Table 5 reports the results. All ten countries with sufficient sample size (more than twenty observations) exhibit positive coefficients, with eight significant at the 10% level. These correlations do not control for secular time trends and thus conflate GDP with other sources of historical change, but they confirm that the wealth-complexity relationship appears in every national subsample rather than being driven by a few influential countries.

Table 5: Prediction 1: Country-Level Robustness

Country	N	Coefficient	Std. Error	p-value
Belgium	32	0.793	0.071	<0.001
Italy	81	0.413	0.095	<0.001
France	61	0.391	0.045	<0.001
Spain	23	0.317	0.302	0.305
Germany	75	0.278	0.049	<0.001
Russia	69	0.278	0.158	0.082
Austria	26	0.224	0.060	0.001
England	23	0.215	0.033	<0.001
Hungary	33	0.199	0.076	0.014
Norway	20	0.124	0.188	0.517

Note: Each row reports a bivariate OLS regression of pitch class entropy on log GDP per capita within that country, without time controls. Countries ordered by coefficient magnitude.

Structural break at industrialisation. One might expect the wealth-complexity relationship to strengthen after industrialisation if the primary channel runs through manufacturing costs. I test for a structural break around 1850 by interacting log GDP with a post-1850 indicator:

$$\text{Entropy}_{ict} = \beta_1 \log(\text{GDP}_{ct}) + \beta_2 \left[\log(\text{GDP}_{ct}) \times \mathbf{1}_{t\geq 1850}\right] + \beta_3 \text{Year}_t + \mu_c + \varepsilon_{ict} \qquad (20)$$

The coefficient β_1 captures the pre-1850 relationship, while $\beta_1 + \beta_2$ captures the post-1850 relationship.

The interaction term is not statistically significant ($\hat{\beta}_2 = -0.004$, $p = 0.489$), with pre-1850 and post-1850 slopes of 0.106 and 0.102 respectively. A formal test fails to reject the null hypothesis that the slopes are equal. This pattern is robust to alternative break years as testing breaks at 1825, 1875, and 1900 yield no significant slope changes.

This stability is consistent with the history of instrument development. Chromatic keyboard instruments existed by the early eighteenth century. The Hotteterre flute achieved practical chromaticism in the seventeenth century. Chromatic brass instruments developed through the Baroque period. The construction cost constraint for chromatic capability therefore relaxed well before industrialisation reduced manufacturing costs more broadly.

If chromatic instruments were already available before 1850, the GDP coefficient is unlikely to capture manufacturing cost reductions specifically. Instead, it likely reflects channels that respond to economic growth throughout the sample period such as the affordability of existing instruments to a wider population of musicians, the funding of conservatories and training institutions, and the depth of patronage systems that support compositional experimentation. These channels operate through both construction costs c_1 (via purchasing power) and cognitive costs c_2 (via access to training), and none depends on a discrete manufacturing revolution.

6.3 Prediction 2: Technological Progress

The model predicts that, as technological progress reduces construction costs over time, we should see increases in scale complexity. I test this with country fixed effects:

$$\text{Entropy}_{ict} = \beta \text{Year}_t + \mu_c + \varepsilon_{ict} \tag{21}$$

Table 6 reports the results. The coefficient is 0.00132 ($p < 0.001$), equivalent to 0.13 bits per century. To contextualise this rate of change: the full progression from diatonic writing (2.81 bits) to complete chromaticism (3.58 bits) spans 0.77 bits, roughly equivalent to adding the five black keys on a piano. The estimated time trend corresponds to traversing this entire journey in approximately six centuries. Over the actual 478-year sample period, the model predicts an increase of 0.63 bits, closely matching the observed difference of 0.45 bits between Renaissance (mean 2.85 bits) and Romantic/Modern (mean 3.30 bits) compositions.

Figure 4 displays entropy distributions by era. The progression from Renaissance through Baroque, Classical, and Romantic/Modern periods shows steady upward shift, with the distribution approaching but not exceeding the chromatic ceiling at 3.58 bits.

Table 6: Prediction 2: Time Trend in Harmonic Complexity

	Coefficient	Std. Error	t-statistic	p-value
Composition year	0.00132	0.00009	15.18	<0.001
Country fixed effects	Yes (12 countries)			
R^2 (within)	0.336			
Observations	468			

Note: Dependent variable is pitch class entropy (bits). Standard errors clustered by composer.

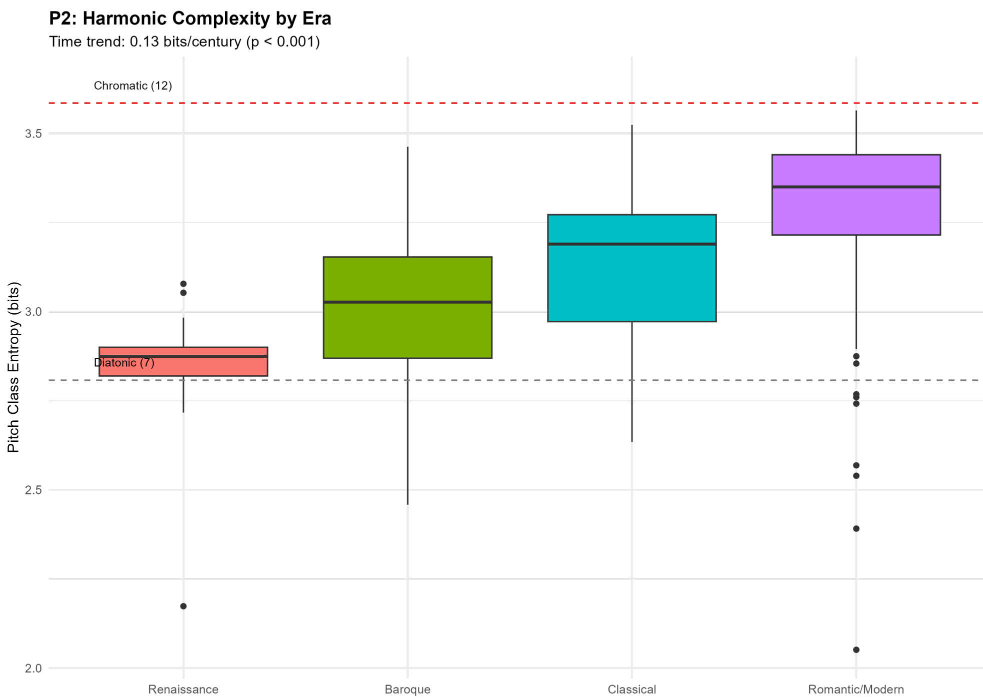

Figure 4: Prediction 2: Harmonic Complexity by Era
Note: Box plots show pitch class entropy distribution by historical era. Dashed grey line indicates diatonic baseline ($\log_2 7 = 2.81$ bits); dashed red line indicates chromatic ceiling ($\log_2 12 = 3.58$ bits). Time trend coefficient: 0.13 bits/century ($p < 0.001$).

6.4 Prediction 3: Instrument Heterogeneity

The model predicts that instruments with higher per-note construction costs should achieve chromatic capability later. The MIDI corpus does not contain instrument-level variation suitable for quantitative testing, so I establish this result through a detailed historical case study of the flute and clarinet.

The flute and clarinet provide a natural experiment in construction cost variation arising from acoustic physics. Both instruments developed within the same Western European musical culture, served the same repertoire, and were produced by overlapping communities of instrument makers. Yet they followed dramatically different trajectories toward chromatic completion.

The difference stems from how the instruments produce their upper register. The flute, with its quasi-cylindrical bore, overblows at the octave (frequency ratio 2:1). When a player increases air pressure, the pitch jumps exactly one octave. This means the fingering pattern for the lower register repeats identically in the upper register. Filling chromatic gaps in one octave automatically provides them in the other.

The clarinet, also cylindrical-bored, behaves differently due to its closed tube physics (the reed seals one end). It overblows at the twelfth (frequency ratio 3:1). When a player increases air pressure, the pitch jumps a twelfth (spanning an octave plus a fifth). This creates systematic gaps in the chromatic scale that cannot be filled by simple fingering repetition. For the clarinet, each missing chromatic note requires dedicated keywork which dramatically increases the per-note construction cost relative to the flute. Importantly, this difference arises from the physics of the instrument rather than from construction technologies.

Chromatic timeline. Table 7 documents key innovations for both the flute and clarinet. Overall, after normalising to the year each instrument's modern form first appeared (the single-key flute circa 1670 and Denner's two-key clarinet circa 1700) the flute achieved practical chromaticism immediately while the clarinet required 112 years of incremental development. Even comparing from the clarinet's invention, the lag is approximately 140 years (1670s versus 1812). That is, the flute achieved full chromatic capability during the construction of Versailles, while the clarinet took until Bonaparte's invasion of Russia.

Table 7: Chromatic Development: Flute versus Clarinet

Year	Instrument	Innovation
Flute		
1670s	Flute	Hotteterre adds single D♯ key, achieving practical chromaticism
1720s	Flute	Additional keys for G♯ and B♭ (luxury variants)
1760s	Flute	4-key standard emerges (D♯, E♭, F, G♯)
1847	Flute	Boehm system (15+ keys) perfects ergonomics
Clarinet		
1700	Clarinet	Denner invents two-key clarinet (severely limited chromatic capability)
1740s	Clarinet	5-key system emerges, still missing several chromatics
1780s	Clarinet	6-key standard (Mozart era), chromatic but awkward
1812	Clarinet	Müller 13-key system achieves full practical chromaticism
1839	Clarinet	Klosé-Boehm system (17+ keys)

This difference is not explained by demand conditions (both instruments served the same orchestral and chamber repertoire), technological capability (the same artisans made both), or cultural factors (development occurred in Germany, France, and Austria for both instruments). The exogenous variation comes purely from acoustic physics determining per-note construction cost.

Period sources confirm these limitations. Writing 82 years after the single-key flute's introduction, Quantz's 1752 *Versuch einer Anweisung die Flöte traversiere zu spielen* (Essay on Playing the Transverse Flute) describes the four-key flute as fully chromatic, while acknowledging certain fingerings as "impure" but functional [Quantz, 1752]. By contrast, Koch's 1802 *Musikalisches Lexikon*, written 102 years after the clarinet's invention, describes the six-key clarinet as still having "defective" chromatic notes that are "difficult to produce" and "impure in intonation" [Koch, 1802]. A full decade later, Müller's 1812 patent application (112 years post-invention) explicitly advertises his 13-key design as finally solving the "incomplete chromatic scale" problem [Rice, 1992]. This evolution demonstrates that, even among contemporaries, the flute was declared functionally chromatic within a century of its modern form, while the clarinet required over a century merely to reach adequacy, with contemporaneous sources documenting its chromatic deficiencies.

Corroborating case in brass instruments. The brass family exhibits a similar pattern. The trombone achieved full chromaticism by the late fifteenth century through a mechanically simple slide mechanism involving two nested tubes that extend continuously to produce any pitch. The horn and trumpet, by contrast, were confined to the harmonic series until the development of valve mechanisms in the early 1800s (*i.e.* the sparse set of natural overtones producible from a fixed tube length). Valves require precise tubing lengths, airtight pistons or rotary mechanisms, and springs, all of which increase per-note construction costs substantially relative to the slide. The result was a lag of approximately 350 years, even more pronounced than the flute-clarinet gap. That two independent instrument families exhibit the same pattern, with simpler mechanisms achieving chromaticism centuries earlier, strengthens the case that construction costs play a role in the timing of chromatic capability.

6.5 Prediction 4: Career Trajectories

If training reduces cognitive costs then the model predicts that individual composers should increase harmonic complexity over their careers. I test this using composer fixed effects:

$$\text{Entropy}_{it} = \beta \text{CareerYear}_{it} + \mu_i + \varepsilon_{it} \tag{22}$$

where μ_i absorbs all time-invariant composer characteristics including era, nationality, and baseline ability. Note that, with composer fixed effects, career year and biological age are equivalent specifications since they differ only by a composer-specific constant absorbed by the fixed effects.

Table 8 reports the results. The coefficient is 0.00331 ($p = 0.003$), equivalent to 0.33 bits per century of career time or approximately 0.1 bits over a 30-year career. To contextualise this effect, recall that the shift from diatonic to chromatic writing spans 0.77 bits. The career learning effect implies that a composer gains roughly one chromatic note's worth of complexity per decade of experience.

This within-composer effect (0.33 bits/century) is 2.5 times larger than the across-composer time trend from Prediction 2 (0.13 bits/century), consistent with individual learning effects operating more strongly than technological change at the aggregate level. Put differently, a composer's personal mastery increases faster than the technological frontier.

The model's estimates reveal how national prosperity advantages composers working in the same time period. Consider César Franck working in Belgium (1863-1876) and Pyotr Ilyich Tchaikovsky working in Russia during the same period. Both composers were in their early careers yet Franck recorded mean entropy of 3.42 bits, while Tchaikovsky recorded 3.23 bits, a gap of 0.19 bits.

The model attributes 26% of this gap to Belgium's economic advantage alone. Belgium's GDP ($3,086) was 1.7 times Russia's ($1,830). If Tchaikovsky had worked in Belgium rather than Russia during this period, the model predicts his complexity would have increased from 3.23 to 3.28 bits purely from access to Belgian economic conditions such as better instruments, training infrastructure, and compositional resources. Conversely, Franck transplanted to 1870s Russia would have been constrained to approximately 3.37 bits.

This illustrates how national prosperity creates compositional advantages between contemporaries. Two composers of comparable career stage, working simultaneously in different countries, achieve measurably different harmonic complexity due to economic geography alone.

Figure 5 displays individual composer trajectories by era. Each line traces one composer's entropy over their career. The dashed line shows the pooled within-composer slope from the fixed effects regression. The pattern is consistent across eras, though noisier in the Renaissance and Baroque periods where career coverage is sparser and sample sizes smaller.

Table 8: Prediction 4: Career Effect on Harmonic Complexity

	Coefficient	Std. Error	t-statistic	p-value
Career year	0.00331	0.00109	3.03	0.003
Composer fixed effects	Yes (45 composers)			
R^2 (within)	0.021			
Observations	468			

Note: Dependent variable is pitch class entropy (bits). Career year measured as years since composer's first known composition in sample. Standard errors clustered by composer.

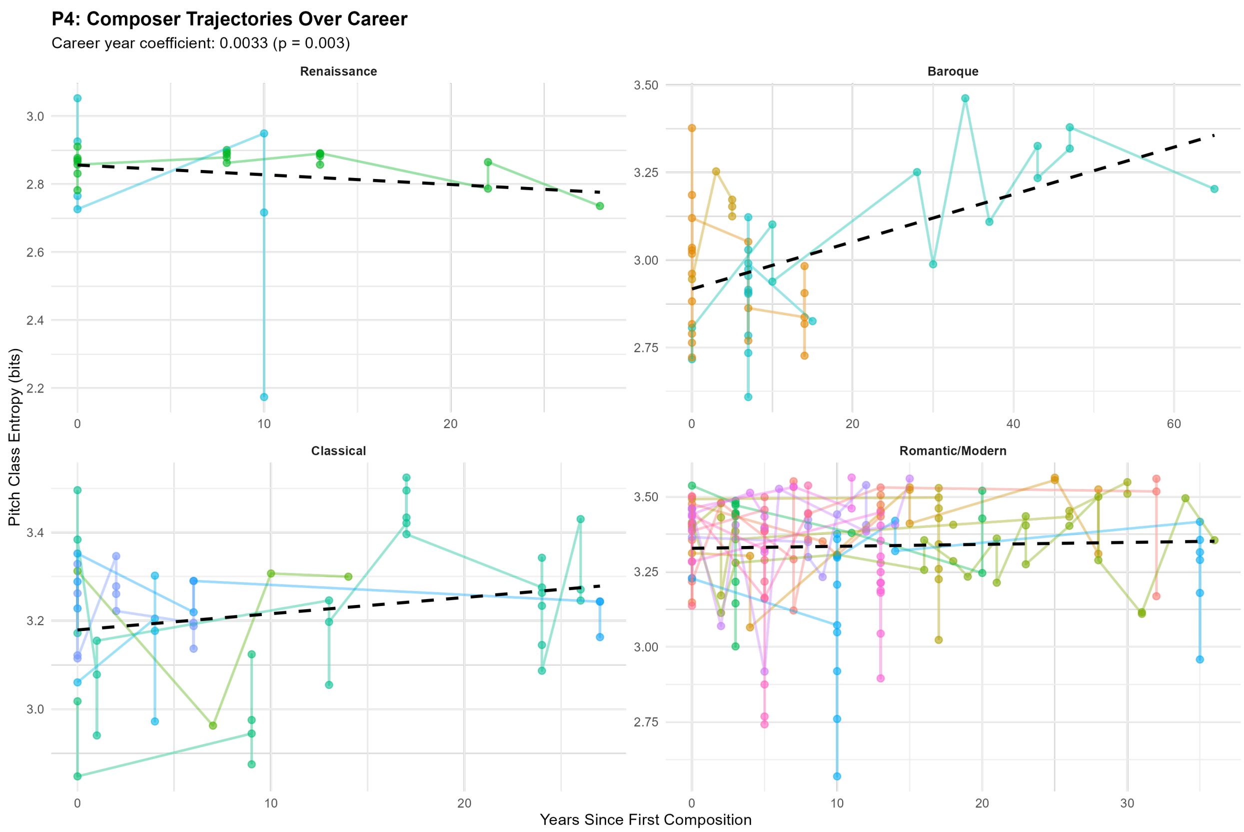

Figure 5: Prediction 4: Composer Trajectories Over Career
Note: Each coloured line traces one composer's entropy over their career (years since first composition). Dashed black line shows pooled within-composer slope from fixed effects regression. Panels separate eras. Career coefficient: 0.0033 bits/year ($p = 0.003$).

Robustness. Appendix B tests the career effect across alternative complexity measures. All four measures show positive coefficients, but only pitch class entropy reaches statistical significance. This fragility warrants acknowledgment. One interpretation is that career learning operates specifically through more uniform exploitation of available pitch classes rather than through dissonance, raw pitch count, or polyphonic density. Training may reduce the cognitive burden of navigating the full chromatic palette, enabling composers to draw on more pitch classes more evenly, without necessarily increasing the proportion of dissonant intervals or the number of simultaneous voices. I report the entropy result as the primary finding given its theoretical alignment with the model's focus on pitch class navigation, but readers should treat the career effect as more tentative than the wealth and time trend results, which are robust across measurement approaches.

6.6 Prediction 5: Cognitive Ceiling

The model predicts that complexity growth should decelerate as it approaches the cognitive bound $\bar{n}$. I test this using a piecewise linear specification with a break at 1900:

$$\text{Entropy}_{it} = \beta_{\text{pre}} \cdot \mathbf{1}[t < 1900] \cdot (t - 1900) + \beta_{\text{post}} \cdot \mathbf{1}[t \geq 1900] \cdot (t - 1900) + \varepsilon_{it} \tag{23}$$

Table 9 reports the results. The pre-1900 slope is 0.00138 bits per year (0.14 bits per century, $p < 0.001$). The post-1900 slope is 0.00038 bits per year (0.04 bits per century, $p = 0.64$), statistically indistinguishable from zero. The pre-1900 rate of 0.14 bits per century would traverse the full diatonic-to-chromatic journey (0.77 bits) in approximately 550 years. The post-1900 rate of 0.04 bits per century suggests music has effectively reached a plateau. Mean post-1900 entropy of 3.31 bits represents 92% of the chromatic ceiling at 3.58 bits, consistent with the cognitive bound becoming binding in the twentieth century.

Table 9: Prediction 5: Deceleration Toward Cognitive Ceiling

	Coefficient	Std. Error	t-statistic	p-value
Pre-1900 trend (per year)	0.00138	0.00009	15.66	<0.001
Post-1900 trend (per year)	0.00038	0.00080	0.47	0.636
R^2		0.397		
Observations		468		

Note: Dependent variable is pitch class entropy (bits). Piecewise linear specification with break at 1900. Pre-1900 slope is 3.7× larger than post-1900 slope.

Figure 6 visualises this deceleration. The flattening of the trend line as it approaches the chromatic ceiling is evident, with the distribution bunching near 3.58 bits in the post-1900 period.

Table 10 reports piecewise regressions with alternative break years. The pre-break slope remains stable at 0.13–0.14 bits per century ($p < 0.001$) across all specifications, while the post-break slope is never statistically distinguishable from zero. The ratio of pre- to post-break slopes ranges from 1.7 to 31.6, confirming that the deceleration finding is not sensitive

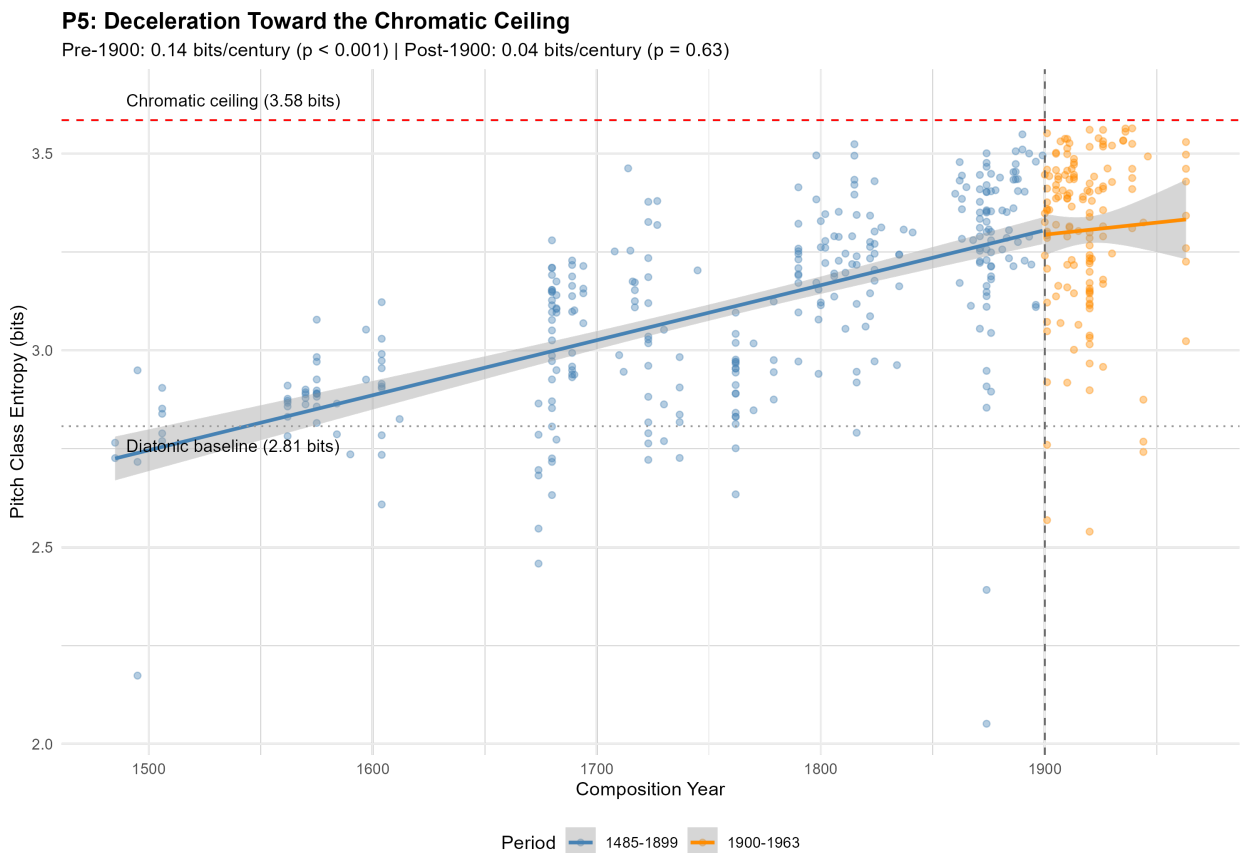

Figure 6: Prediction 5: Deceleration Toward the Chromatic Ceiling
Note: Scatter plot shows entropy by composition year, coloured by period (pre/post 1900). Lines show period-specific linear trends with 95% confidence bands. Dashed red line indicates chromatic ceiling ($\log_2 12 = 3.58$ bits); dotted grey line indicates diatonic baseline ($\log_2 7 = 2.81$ bits). Pre-1900 trend: 0.14 bits/century ($p < 0.001$); post-1900 trend: 0.04 bits/century ($p = 0.64$).

to the choice of break point.

Table 10: Robustness of Deceleration Finding to Alternative Break Years

Break Year	Pre-Break		Post-Break		Ratio
	Slope	p	Slope	p	
1880	0.14	<0.001	0.08	0.119	1.7
1890	0.14	<0.001	0.05	0.392	2.6
1900	0.14	<0.001	0.04	0.636	3.7
1910	0.14	<0.001	0.00	0.966	31.6
1920	0.13	<0.001	0.06	0.690	2.4

Notes: Piecewise linear regressions of pitch class entropy on composition year with slopes allowed to differ before and after the indicated break year. Slopes expressed as bits per century. Ratio is pre-break slope divided by post-break slope.

Electronic Music Robustness Check. The sample ends in 1963, before widespread electronic music production. The subsequent six decades offer a natural setting to examine whether the cognitive bound operates when construction costs vanish. Electronic synthesisers can produce any frequency with equal ease ($c_1 \approx 0$), yet the vast majority of electronic music employs 12-TET. This pattern holds from pioneers like Kraftwerk and Brian Eno through dance music (Daft Punk, The Chemical Brothers) to modern EDM (deadmau5, Skrillex, Avicii). Microtonal electronic music exists but remains niche despite zero technical barriers.

As an initial probe, I examine 500 tracks from the GiantSteps+ EDM Key Dataset [Faraldo, 2017], which provides expert-annotated pitch class sets across 26 subgenres including house, techno, trance and drum & bass. This dataset captures commercially dominant electronic dance music rather than experimental or art-music traditions, so the analysis tests whether cognitive constraints bind in mainstream production contexts rather than at the frontier of harmonic experimentation.

For each track, I compute maximum entropy as $H = \log_2(n)$ where n is the number of distinct pitch classes employed. Mean pitch class entropy is 2.69 bits (SD = 0.38), with a median of 2.81 bits corresponding exactly to diatonic usage ($\log_2 7 = 2.81$). Seventy-nine percent of tracks employ seven or fewer pitch classes, and the 95th percentile (3.17 bits) remains well below the chromatic maximum of $\log_2 12 = 3.58$ bits. Only a single track in the sample uses all twelve pitch classes.

These findings are consistent with cognitive constraints binding at roughly diatonic levels even when construction costs vanish. However, commercial EDM may face additional constraints beyond cognition, including audience expectations, genre conventions, and optimisation for danceability, that independently favour simpler harmonic structures. A stronger test would examine experimental electronic genres where such commercial pressures are weaker.

7 Conclusion

This paper develops an economic model of musical scale selection. The central insight is that scale size reflects joint optimisation over harmonic value, instrument construction cost, and performer cognitive cost. The objective function $\pi(n) = \alpha n^\beta - c_1 n - c_2 n^\gamma$ generates a unique interior optimum when harmonic returns diminish ($\beta < 2$) and cognitive costs eventually dominate ($\gamma > \beta$). Both conditions have empirical foundations. The harmonic value exponent $\beta = 1.86$ is point-identified from Tenney height consonance measures with $R^2 = 0.996$, comfortably below the threshold of 2. The cognitive cost exponent γ is partially identified. The binding cognitive bound at 12 notes in electronic music establishes $\gamma > \beta$, with the parameter curve implying γ most likely exceeds 2.2.

The model explains several patterns that appear unrelated under conventional accounts. The worldwide prevalence of pentatonic scales reflects high construction costs in pre-industrial settings. The historical progression from pentatonic to diatonic to chromatic systems reflects declining construction costs through manufacturing improvements. The 140-year lag of clarinet chromaticism behind the flute reflects exogenous differences in per-note construction costs arising from acoustic physics. The tendency of composers to increase harmonic complexity over their careers reflects declining cognitive costs through training. The persistence of 12-note keyboards in electronic music despite zero construction costs reflects the cognitive bound that prevents scale explosion.

Empirical analysis of 468 classical compositions spanning 1485–1963 supports four of five predictions. Harmonic complexity increases with economic prosperity ($p = 0.014$), advances secularly with technology (0.13 bits per century, $p < 0.001$), rises over composers' careers (0.33 bits per century, $p = 0.003$), and decelerates as it approaches the chromatic ceiling. The fifth prediction, regarding instrument heterogeneity, receives qualitative support from the flute-clarinet case study but awaits systematic quantitative testing.

Several limitations warrant acknowledgment. The MIDI corpus is restricted to Western classical music, leaving cross-cultural predictions untested, and survivorship bias in the corpus may inflate absolute complexity levels in earlier periods, though within-composer and cross-country results are less susceptible to this concern. The cognitive cost parameters are partially rather than point-identified, though this does not affect the model's comparative static predictions. The model treats scale selection as static optimisation, abstracting from network effects, path dependence, and coordination failures that may sustain suboptimal equilibria. Future work might extend the model to incorporate these dynamics, test predictions using non-Western musical traditions, and develop direct experimental measures of cognitive cost at different scale sizes.

Despite these limitations, the framework demonstrates that economic analysis can illuminate fundamental questions about cultural form. Musical scales are not arbitrary conventions or purely artistic choices but optimal responses to technological and cognitive constraints. The tools of price theory, applied to the production function of harmony itself, reveal the hidden economics of aesthetic choice.

References

William J. Baumol and William G. Bowen. *Performing Arts: The Economic Dilemma*. Twentieth Century Fund, New York, 1966.

Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5): 992–1026, 1992.

Jutta Bolt and Jan Luiten van Zanden. Maddison style estimates of the evolution of the world economy: A new 2020 update. *Maddison Project Working Paper*, 15, 2020.

Steven Brown and Emily Phillips. Musical scales optimize pitch spacing. *Humanities and Social Sciences Communications*, 12:89, 2025.

Marco Buongiorno Nardelli, Garland Culbreth, and Miguel A. Fuentes. Towards a measure of harmonic complexity in western classical music. *Advances in Complex Systems*, 25(5–6): 2250009, 2022.

Norman Carey and David Clampitt. Aspects of well-formed scales. *Music Theory Spectrum*, 11(2):187–206, 1989.

John Clough and Jack Douthett. Maximally even sets. *Journal of Music Theory*, 35(1/2): 93–173, 1991.

Nelson Cowan. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1):87–114, 2001.

Ángel Faraldo. GiantSteps+ EDM Key Dataset, 2017. URL `https://doi.org/10.5281/zenodo.1095691`.

Kamraan Z. Gill and Dale Purves. A biological rationale for musical scales. *PLOS ONE*, 4 (12):e8144, 2009.

Heinrich Christoph Koch. *Musikalisches Lexikon*. August Hermann der Jüngere, Frankfurt am Main, 1802.

Carol L. Krumhansl. *Cognitive Foundations of Musical Pitch*. Oxford University Press, New York, 1990.

John M. McBride, Sam Passmore, and Tsvi Tlusty. Convergent evolution in a large cross-cultural database of musical scales. *PLOS ONE*, 18(12):e0296291, 2023.

John M. McBride, Nahie Kim, Joren Nishikawa, Mekhmed Saadakeev, Marcus Pearce, and Tsvi Tlusty. Information and motor constraints shape melodic diversity across cultures. *arXiv preprint arXiv:2408.12635*, 2024.

George A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2):81–97, 1956.

Fabian C. Moss. *Transitions of Tonality: A Model-Based Corpus Study*. PhD thesis, École Polytechnique Fédérale de Lausanne, 2019.

Fabian C. Moss, Robert Lieck, and Martin Rohrmeier. Computational modeling of interval distributions in tonal space. *Humanities and Social Sciences Communications*, 11:215, 2024.

Eita Nakamura and Kunihiko Kaneko. Statistical evolutionary laws in music styles. *Scientific Reports*, 9:15993, 2019.

R. Plomp and W. J. M. Levelt. Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, 38(4):548–560, 1965.

Johann Joachim Quantz. *Versuch einer Anweisung die Flöte traversiere zu spielen*. Johann Friedrich Voß, Berlin, 1752.

Albert R. Rice. *The Baroque Clarinet*. Clarendon Press, Oxford, 1992.

Matthew J. Salganik, Peter Sheridan Dodds, and Duncan J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762): 854–856, 2006.

F. M. Scherer. *Quarter Notes and Bank Notes: The Economics of Music Composition in the Eighteenth and Nineteenth Centuries*. Princeton University Press, Princeton, 2004.

Marc Serra-Peralta, Joan Serrà, and Álvaro Corral. Heaps' law and vocabulary richness in the history of classical music harmony. *EPJ Data Science*, 10:40, 2021.

Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

Bob Snyder. *Music and Memory: An Introduction*. MIT Press, Cambridge, MA, 2000.

James Tenney. *A History of 'Consonance' and 'Dissonance'*. Excelsior Music Publishing, New York, 1988.

David Throsby. The production and consumption of the arts: A view of cultural economics. *Journal of Economic Literature*, 32(1):1–29, 1994.

Ruth Towse. *Creativity, Incentive and Reward: An Economic Analysis of Copyright and Culture in the Information Age*. Edward Elgar, Cheltenham, 2001.

Dmitri Tymoczko. *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*. Oxford University Press, New York, 2011.

A Theory Derivations

A.1 Proof of Proposition 1: Existence and Uniqueness

The optimisation problem is:

$$\max_{n} \pi(n) = \alpha n^{\beta} - c_1 n - c_2 n^{\gamma} \tag{24}$$

The first-order condition is:

$$\frac{d\pi}{dn} = \alpha\beta n^{\beta-1} - c_1 - \gamma c_2 n^{\gamma-1} = 0 \tag{25}$$

Define $\text{MHB}(n) = \alpha\beta n^{\beta-1}$ (marginal harmonic benefit) and $\text{MC}(n) = c_1 + \gamma c_2 n^{\gamma-1}$ (marginal cost). We establish existence and uniqueness of a crossing for $n > 0$.

Case 1: $\beta < 1$. As $n \to 0^+$, $\text{MHB}(n) \to \infty$ while $\text{MC}(n) \to c_1$ (finite). As $n \to \infty$, $\text{MHB}(n) \to 0$ while $\text{MC}(n) \to \infty$ (since $\gamma > 1$). By the intermediate value theorem, a crossing exists. Uniqueness follows because MHB is strictly decreasing and MC is strictly increasing.

Case 2: $\beta \geq 1$ (including the empirically relevant case $\beta \approx 1.86$). As $n \to 0^+$, $\text{MHB}(n) \to 0$ and $\text{MC}(n) \to c_1$. As $n \to \infty$, both functions diverge, but MC grows faster since $\gamma > \beta$. The existence of a crossing depends on whether MHB exceeds MC for some $n > 0$.

At $n = 1$: $\text{MHB}(1) = \alpha\beta$ and $\text{MC}(1) = c_1 + \gamma c_2$. A sufficient condition for an interior optimum is $\alpha\beta > c_1 + \gamma c_2$, which holds for sufficiently small c_1 and c_2 given positive α and β. When construction costs are prohibitively high, the corner solution $n^* = 0$ (or the minimal musically meaningful scale) may be optimal, consistent with the model's economic interpretation.

For uniqueness when $1 < \beta < 2$ and $\gamma > 2$: MHB is increasing and strictly concave (since $\frac{d^2\text{MHB}}{dn^2} = \alpha\beta(\beta-1)(\beta-2)n^{\beta-3} < 0$), while MC is increasing and strictly convex (since $\frac{d^2\text{MC}}{dn^2} = \gamma(\gamma-1)(\gamma-2)c_2 n^{\gamma-3} > 0$). A strictly concave function and a strictly convex function, both increasing, can cross at most once. Combined with the boundary behaviour (MC dominates for large n), there is exactly one crossing when an interior solution exists.

Second-order condition: At any interior critical point n^*:

$$\frac{d^2\pi}{dn^2} = \alpha\beta(\beta-1)n^{\beta-2} - \gamma(\gamma-1)c_2 n^{\gamma-2} \tag{26}$$

For $1 < \beta < 2$, the first term is positive (since $\beta - 1 > 0$) but small relative to the second term at the optimum. For the second term, $\gamma > 2$ implies $\gamma - 1 > 0$ and $\gamma - 2 > 0$, making the term negative. The second-order condition is satisfied, confirming a maximum. □

A.2 Proof of Proposition 2: Construction Cost Comparative Static

The first-order condition implicitly defines $n^*(c_1)$:

$$F(n, c_1) \equiv \alpha\beta n^{\beta-1} - c_1 - \gamma c_2 n^{\gamma-1} = 0 \tag{27}$$

By the implicit function theorem:

$$\frac{\partial n^*}{\partial c_1} = -\frac{\partial F/\partial c_1}{\partial F/\partial n} = -\frac{-1}{\alpha\beta(\beta-1)n^{\beta-2} - \gamma(\gamma-1)c_2 n^{\gamma-2}} \tag{28}$$

The denominator is negative (from the second-order condition), so $\partial n^*/\partial c_1 < 0$. □

A.3 Proof of Proposition 3: Cognitive Cost Comparative Static

Similarly:

$$\frac{\partial n^*}{\partial c_2} = -\frac{\partial F/\partial c_2}{\partial F/\partial n} = -\frac{-\gamma n^{\gamma-1}}{\alpha\beta(\beta-1)n^{\beta-2} - \gamma(\gamma-1)c_2 n^{\gamma-2}} \tag{29}$$

The numerator after the minus sign is positive ($\gamma n^{\gamma-1} > 0$), and the denominator is negative, so $\partial n^*/\partial c_2 < 0$. □

A.4 Proof of Proposition 4: Cognitive Bound

Setting $c_1 = 0$ in the first-order condition:

$$\alpha\beta n^{\beta-1} = \gamma c_2 n^{\gamma-1} \tag{30}$$

Rearranging:

$$n^{\gamma-\beta} = \frac{\alpha\beta}{\gamma c_2} \tag{31}$$

Solving for n:

$$\bar{n} = \left(\frac{\alpha\beta}{\gamma c_2}\right)^{1/(\gamma-\beta)} \tag{32}$$

Since $\gamma > \beta$, the exponent $1/(\gamma - \beta) > 0$, and the base is positive and finite, $\bar{n}$ is positive and finite. $\square$

B Robustness: Alternative Complexity Measures

The main text uses pitch class entropy as the primary measure of harmonic complexity. This appendix tests whether the key findings hold across alternative measures capturing different dimensions of compositional complexity.

B.1 Alternative Measures

I examine three additional measures from the MIDI data:

- *Unique pitch classes*: Count of distinct pitch classes used (0–12). A direct measure of effective scale size, though insensitive to usage intensity. Where entropy rewards uniform distribution across pitch classes, this measure treats any appearance equally.
- *Maximum simultaneous notes*: Peak polyphonic density within a composition. This captures textural complexity orthogonal to pitch selection, measuring the vertical dimension of compositional ambition.
- *Dissonance ratio*: Proportion of chords containing intervals traditionally classified as dissonant (minor seconds, major sevenths, tritones). This captures harmonic tension directly, complementing entropy's focus on chromatic breadth.

Prediction 1: Wealth Effect Table 11 reports the GDP coefficient from the main specification across all measures.

Table 11: Prediction 1 Robustness: Wealth Effect Across Measures

Measure	Coefficient	Std. Error	p-value
Pitch class entropy	0.089	0.036	0.014 *
Dissonance ratio	0.066	0.027	0.015 *
Unique pitch classes	−0.115	0.187	0.537
Max simultaneous	3.588	1.477	0.015 *

Notes: Each row reports the coefficient on log GDP per capita from the main specification with country fixed effects and year controls. * $p < 0.05$; ** $p < 0.01$.

Three of four measures show positive coefficients, with all three reaching statistical significance. The lone exception, unique pitch classes, is negative but far from significant ($p = 0.537$). The wealth effect operates through how intensively and dissonantly composers use available pitch classes, and through textural richness, rather than through expanding the raw count of pitches employed.

Prediction 2: Time Trend Table 12 reports the time trend across all measures.

Table 12: Prediction 2 Robustness: Time Trend Across Measures

Measure	Coefficient	Std. Error	p-value
Pitch class entropy	0.00132	0.00009	<0.001 **
Dissonance ratio	0.00086	0.00006	<0.001 **
Unique pitch classes	0.00659	0.00045	<0.001 **
Max simultaneous	0.01252	0.00355	<0.001 **

Notes: Each row reports the coefficient on composition year from the specification with country fixed effects. Coefficients expressed per year. * $p < 0.05$; ** $p < 0.01$.

All four measures show positive and highly significant time trends. The secular increase in harmonic complexity documented in the main text is unanimous across measurement approaches.

Prediction 4: Career Effect Table 13 reports the career year coefficient across all measures.

Table 13: Prediction 4 Robustness: Career Effect Across Measures

Measure	Coefficient	Std. Error	p-value
Pitch class entropy	0.00331	0.00109	0.003 **
Dissonance ratio	0.00027	0.00083	0.749
Unique pitch classes	0.00518	0.00519	0.319
Max simultaneous	0.00535	0.04129	0.897

Notes: Each row reports the coefficient on career year from the specification with composer fixed effects. * $p < 0.05$; ** $p < 0.01$.

All four measures show positive coefficients, consistent with the prediction that composers increase complexity over their careers. However, only pitch class entropy reaches statistical significance. The career learning effect appears to operate specifically through more uniform exploitation of the chromatic scale rather than through dissonance, pitch count, or polyphonic density. This specificity is consistent with the theoretical model's focus on cognitive costs of navigating pitch systems. Training reduces the burden of using more pitch classes uniformly, while dissonance and texture may depend more on stylistic choice than accumulated skill.

Prediction 5: Cognitive Ceiling Table 14 reports the piecewise regression results across all measures.

Table 14: Prediction 5 Robustness: Deceleration Across Measures

Measure	Pre-1900		Post-1900		Equality
	Slope	p	Slope	p	p
Pitch class entropy	0.14	<.001	0.04	.636	.231
Dissonance ratio	0.08	<.001	0.26	<.001	.003
Unique pitch classes	0.75	<.001	−0.73	.061	<.001
Max simultaneous	2.25	<.001	−3.60	.261	.082

Notes: Slopes expressed per century. "Equality p" tests the null hypothesis that pre-1900 and post-1900 slopes are equal.

Three of four measures show deceleration, with pre-1900 slopes exceeding post-1900 slopes. For pitch class entropy, the pre-1900 trend of 0.14 bits per century falls to an insignificant 0.04 bits per century after 1900. Unique pitch classes and maximum simultaneous notes actually show negative post-1900 slopes, suggesting reversal rather than mere deceleration.

The exception is dissonance ratio, which accelerates from 0.08 to 0.26 per century, with the slope difference highly significant ($p = 0.003$). This finding supports rather than undermines the theoretical interpretation. The cognitive bound constrains how many pitch classes composers can effectively manage, not how dissonantly they deploy them. Once chromatic saturation was approached around 1900, composers could not add more notes but could push harder on dissonance. The works of Schoenberg, Bartók, and Stravinsky exemplify this pattern. They are operating within 12-tone systems but deploy those tones with unprecedented harmonic tension. The ceiling is specifically about scale size, not about harmonic adventurousness more broadly.

In summary, the time trend (P2) receives unanimous support. The wealth effect (P1) and ceiling effect (P5) hold for three of four measures, with the P5 exception (dissonance acceleration) theoretically interpretable. The career effect (P4) is unanimous in sign but statistically significant only for pitch class entropy. Overall, the primary measure performs well across all predictions, and the alternative measures provide consistent support for the model's core implications.