

Novel Covariance-Based Neutrality Test of Time-Series Data Reveals Asymmetries in Ecological and Economic Systems

Alex Washburne^{1,4,5}, Josh Burby² and Daniel Lacker³

February 11, 2016

Author Affiliations:

¹Princeton University, program for Quantitative and Computational Biology

²Princeton University, Physics Department, Plasma Physics Lab

³Princeton University, Department of Operations Research and Financial
Engineering

⁴Duke University, Department of Biology

⁵Corresponding Author, alex.d.washburne@gmail.com

1 Abstract

2 Systems as diverse as the interacting species in a community, alleles at a
3 genetic locus, and companies in a market are characterized by competi-
4 tion (over resources, space, capital, etc) and adaptation. Neutral theory,
5 built around the hypothesis that individual performance is independent
6 of group membership, has found utility across the disciplines of ecology,
7 population genetics, and economics, both because of the success of the
8 neutral hypothesis in predicting system properties and because deviations
9 from these predictions provide information about the underlying dynam-
10 ics. However, most tests of neutrality are weak, based on static system
11 properties such as species-abundance distributions or the number of sin-
12 gletons in a sample. Time-series data provide a window onto a system's
13 dynamics, and should furnish tests of the neutral hypothesis that are more
14 powerful to detect deviations from neutrality and more informative about
15 to the type of competitive asymmetry that drives the deviation.

16 Here, we present a neutrality test for time-series data. We apply this
17 test to several microbial time-series and financial time-series and find that
18 most of these systems are not neutral. Our test isolates the covariance
19 structure of neutral competition, thus facilitating further exploration of
20 the nature of asymmetry in the covariance structure of competitive sys-
21 tems. Much like neutrality tests from population genetics that use relative
22 abundance distributions have enabled researchers to scan entire genomes
23 for genes under selection, we anticipate our time-series test will be useful
24 for quick significance tests of neutrality across a range of ecological, eco-
25 nomic, and sociological systems for which time-series data are available.
26 Future work can use our test to categorize and compare the dynamic fin-
27 gerprints of particular competitive asymmetries (frequency dependence,
28 volatility smiles, etc) to improve forecasting and management of complex
29 adaptive systems.

30 Author Summary

31 From fisheries and forestries to game parks and gut microbes, managing a com-
32 munity of organisms is much like managing a portfolio. Managers care about
33 diversity, and calculations of risk - for extinction or financial ruin - require
34 accurate models of the covariance between the parts of the portfolio.

35 To model the covariances in portfolios or communities, it helps to start sim-
36 ple with a null model assuming the equivalence of species or companies relative
37 to one another (termed "neutrality") and letting the data suggest otherwise.
38 Researchers in biology and finance have independently entertained and tested
39 neutral models, but the existing tests have used snapshots of communities or the
40 variance of fluctuations of individual populations, whereas tests of the covari-
41 ances between species can better inform the development of alternative models.

42 We develop a covariance-based neutrality test for time-series data and use
43 it to show that the human microbiome, North American birds, and companies
44 in the S&P 500 all have a similar deviation from neutrality. Understanding
45 and incorporating this non-neutral covariance structure can yield more accurate
46 alternative models of community dynamics which can improve our management
47 of "portfolios" of multi-species systems.

48 Introduction

49 "*(A)s more individuals are produced than can possibly survive, there must in*
50 *every case be a struggle for existence, either one individual with another of the*
51 *same species, or with the individuals of distinct species, or with the physical*
52 *conditions of life.*" - Charles Darwin, Origin of Species [1]

53 Adaptive evolution requires that rivalrous goods are consumed by agents,
54 those agents have heritable variation in how they acquire and consume the

55 rivalrous goods, and the fitness of agents increases with the amount of goods
56 consumed [2]. By Lewontin's listing of the necessary conditions of evolution, a
57 variety of systems can be seen as evolving. Genes in populations, species in a
58 community, companies in a market, and political groups in a society all satisfy
59 Lewontin's axioms [2].

60 Canopy space is a rivalrous resource in the multi-species closed-canopy forests.
61 If nothing else intervenes, a competitively superior tree species will dominate
62 the canopy just like a competitively superior gene will become fixed in a pop-
63 ulation. In economic systems, companies compete over capital, customers, and
64 labor, and a company well-adapted to a market will increase its share of the
65 resources. In social systems, political groups compete over votes and occupied
66 positions of power, and political groups with superior recruitment compared to
67 other groups - either by persuasion, coercion, aggression, or reproduction with
68 vertical transmission of culture - will increase the votes it receives and/or its
69 representation in various positions of power.

70 These generalized competitive systems are examples of "complex adaptive
71 systems" [3, 4, 5] and understanding how they evolve can provide insight into the
72 drivers of adaptive evolution [6], diversity maintenance in human and natural
73 systems [7], portfolio construction in a market [8], and problems of recognition
74 and representation in multicultural societies [9]. Much literature has explored
75 the stochastic fluctuations of individual populations (e.g. [10, 11]) or asset prices
76 [12] in these systems, and accurate models of the stochastic time-evolution of
77 multi-species systems can enable calculations of the risk of extinction [13], the
78 dynamics of diversity (such as the entropy or evenness of a system), portfolio
79 analysis, and other features of interest.

80 A common stochastic model in which all groups are functionally equivalent,
81 termed "Neutral Theory" in ecology and population genetics, has been used

82 across many systems [14, 15, 16, 17, 18]. By “functionally equivalent”, we mean
83 that every agent’s performance in acquiring the rivalrous resource is indepen-
84 dent of their group membership. In other words, an organism’s species identity,
85 a company’s strategy or sector, a citizen’s political identity, or a political party’s
86 platform have no impact on their ability to hold or acquire new rivalrous re-
87 sources. Neutrality is a parsimonious starting point for community modeling
88 because it is based on first principles of random birth and death or acquisition
89 and release of resources that are appropriate for many competitive systems, and,
90 because neutrality does not assume particular traits that distinguish groups and
91 complex interactions between groups, it is invariant to grouping: the popula-
92 tions of neutral species can be aggregated into larger groups whose competition
93 is also neutral.

94 Neutrality is often posed as a null model for multi-species systems because
95 it can be parsimonious to assume, initially, that all species are equivalent. The
96 mathematical tractability of neutral systems has allowed for useful calculations
97 [19] that can sometimes accurately describe features of the system. However,
98 despite the mathematical ease, calculations for features such as extinction time
99 or the dynamics of portfolio diversity based on neutrality may be inaccurate
100 for systems with non-neutral dynamics such as positive or negative frequency-
101 dependent selection. Thus, there is a need for powerful and informative tests of
102 neutrality to assess whether or not the dynamics of the competitive system are
103 neutral.

104 In population genetics, tests of neutrality [20, 21] have facilitated rapid con-
105 ceptual and empirical advancements [22], allowing researchers to scan entire
106 genomes for neutral loci and identify loci that have been under selection. Neu-
107 trality tests developed in ecological and sociological systems test features of
108 rank-abundance and frequency-abundance distributions [16, 23, 24]. Many of

109 these existing neutrality tests utilize snapshots of a competitive system, but
110 time-series contain a tremendous amount of data and can enable stronger tests
111 of neutrality.

112 Some work has been done developing and utilizing tools to test whether
113 or not population dynamics are consistent with Neutral Theory [10, 25, 11].
114 These tests rely on a particular description of Neutral Theory as the one-step
115 process [26] posed by Kimura and Hubbell, but the proof [27] that the non-
116 zero-sum volatility-stabilized market models [17] converge to neutral drift in
117 relative abundances motivates a broader definition and more general tests of
118 neutrality. Since neutrality is the per capita equivalence between species, it is
119 necessarily relative, not absolute; a population is not neutral per se, but can
120 only be neutral relative to another population. To the best of our knowledge, the
121 existing time-series tests have all analyzed whether or not the variance in jumps
122 in abundance increase linearly or quadratically with the population size prior
123 to the jump, and none have examined the covariance structure of fluctuations
124 in relative abundance.

125 Here, we present, to our knowledge, the first covariance-based neutrality
126 test for time-series data. Our test provides deeper insight into the nature of
127 non-neutrality than traditional tests of rank-abundance distributions and the
128 volatility of individual populations. Our test utilizes the grouping invariance of
129 neutral systems to isolate and test the covariance between changes in species'
130 relative abundances over small time intervals, allowing a rejection of neutral-
131 ity for the entire community considered. We apply our test to 6 metagenomic
132 time-series [28], a time-series of breeding birds across North America [29], and
133 a time-series of market capitalization of companies in the S&P 500 from 2000-
134 2005. We show that even some systems whose rank-abundance distributions ap-
135 pear neutral can exhibit significantly non-neutral covariances between species

Figure 1: Illustration of our method **(A)** The dynamics of a 15-species neutral community of 10,000 individuals and migration probability $m = 0.0002$ (shown here) can be approximated by a WFP with $\lambda = 20$. If the community is neutral, then the CVTs should yield homoskedastic plots of ν_t versus f_t . We test neutrality by randomly drawing from the 2^n possible CVTs, performing homoskedasticity tests on ν_t versus f_t , and then testing the uniformity of the resulting P-value distribution using a modified KS-test (see details in supplement section S3). **(B)** The relative abundances of 15 independent, mean-reverting geometric Brownian motions, $d \log X_t = \mu(b - \log X_t)dt + \sigma dW_t$ with $\mu = 15$, $\sigma = 30$, $b = 10$. Neutrality is rejected by the highly non-uniform distribution of P-values. The left-skewed P-value distribution indicates many CVTs had volatilities that depended on the state variable, f_t .

136 as detected by our test. Furthermore, our test, based on random groupings
137 of species, illustrates how to analyze the volatility of randomly formed groups
138 to reveal state-dependent volatility that differs from neutrality. The non-neutral
139 state-dependent covariance structure uncovered here can be incorporated to im-
140 prove our models of community dynamics and calculations of species' extinction
141 times, portfolio risk, and more.

142 Results

143 Our method is proven analytically in the Materials & Methods section, and a
144 demonstration of the method is provided in Figure 1.

145 We apply our neutrality test to 8 different datasets. Six of these datasets
146 are sequence-count data of microbial communities [28] from three body sites on
147 two individuals. One dataset is survey of breeding birds across North America
148 from 1966-2014 [29], and one dataset is financial data, obtained from the Cen-
149 ter for Research in Security Prices, of the day-end market shares and market
150 capitalization of 451 companies in the S&P 500 from January 1, 2000 to Jan-
151 uary 1, 2005. These datasets are long, time-series datasets, many of which have
152 rank-abundance distributions that are decently fit by neutral theory's expected
153 rank-abundance distribution (see supplement part S1 for a detailed description

Figure 2: Applying our test to time-series datasets reveal non-neutral competitive dynamics in microbial and financial systems. Goodness of fit P-values displayed are from a modified KS-test which accounts for dependence among the observations. Despite decent fits of neutral species-abundance distributions, our time-series test reveals that competitive asymmetries are important drivers in all systems except the female tongue bacteria.

154 of the datasets and fits of neutral theory’s rank-abundance distributions).

155 Our test relies on multiple groups of species. To group the species, we
156 randomly selected $a_i = \pm 1$, for 4,000 independent groups, we then calculated
157 ν_t as in equation (6), and used a White test [30] to test the homoskedasticity
158 of ν_t . The White test performed auxiliary regression with a generalized linear
159 model with a log-normal link function of the form

$$\hat{\epsilon}^2 = \beta_0 + \beta_1 f_t + \beta_2 f_t^2 + \gamma \quad (1)$$

160 where $\hat{\epsilon}^2$ are the squared residuals from similar quadratic regression of f_t on
161 ν_t . With a 0.005 significance threshold, neutrality was rejected in all but one
162 dataset (figure 2).

163 Rejecting neutrality for these competitive systems motivates further investi-
164 gation on whether the rejection of neutrality stems from sampling error or from
165 true competitive asymmetries in the system. For the financial datasets, there
166 is no sampling error - the reported values of day-end prices are the true values.
167 For the metagenomic datasets, sampling error for sequence count data could be
168 driving apparent non-neutrality.

169 Figure 3 examines the competitive asymmetry in the male tongue’s bacterial
170 community. Scatter plots of ν_t vs. f_t reveal a downward trend indicative of mean
171 reversion - jumps in f_t are positive when f_t is above its mean and negative when
172 f_t is above its mean (figure 3A). The mean reversion is regressed out prior to
173 our auxiliary regression, but such strong mean reversion is not apparent in the
174 simulated neutral community of figure 1. The discrepancy between the mean

Figure 3: (A-D) Analyzing the non-neutrality of competitive systems (A) The negative relationship between ν_t vs. f_t indicates mean reversion. Overlaying ν_t vs. f_t scatter plot from a particular CVT from the male tongue data onto the results from 4,000 WFP trajectories with long sampling intervals, Δt , shows that mean reversion can be accounted for by sparse time-sampling of the data. (B) However, even when correcting for sparse time-sampling, the left-skewed P-value distribution in the male tongue indicates stronger signal of non-neutral volatility than 16,000 surrogate WFPs. (C) The parameter β_2 from significantly ($P < 0.001$ for male tongue, $P < 0.01$ for surrogate data) heteroskedastic auxiliary regressions in equation 7 reveals significantly more $\beta_2 > 0$ than $\beta_2 < 0$ in the data. The different P-value cutoffs are for visualization - the same bias for $\beta_2 > 0$ holds for a standard cutoff of $P < 0.01$ (D) Overlaying scatterplots of the residuals, $\hat{\epsilon}^2$, from all heteroskedastic cases ($P < 0.01$) of the male tongue data, reveals the empirical pattern of heteroskedasticity. Compared to surrogate neutral data, the male tongue is more volatile when the groupings are uneven, suggesting that either rare or abundant groups are more volatile - or equal groupings are relatively less volatile - than neutrality would predict. (E) All datasets have the same over-abundance of $\beta_2 > 0$ for heteroskedastic ($P < 0.05$) CVTs.

175 reversion in the data and the simulated neutral community may be due to a
176 stronger and/or non-linear mean reversion in the microbial system, or it may
177 be due to the relatively long time between time points in the data relative to
178 the turnover rate of the community. Such sparse time-sampling could affect the
179 accuracy of our test, which relies on the assumption that Δt in equation 3 is
180 small. A neutral community may still have mean reversion due to migration
181 from a metacommunity or mutation/conversion rates between classes of agents,
182 and sparsely sampling a time-series of such a neutral community may yield the
183 same downward trend on plots of ν_t versus f_t .

184 To examine if the long time between time points accounts for the perceived
185 non-neutrality from our analysis of figure 2, we produced surrogate data by
186 simulating neutral communities with similarly sparse time points. Parameter
187 estimation of λ , ρ and WFP simulation is described in the supplement section S4.
188 For one particular CVT, we simulated 4,000 independent trajectories to allow
189 the superposition of the ν_t vs f_t scatter plots from the male tongue data over
190 the points from the surrogate data. Much of the strong mean reversion in the
191 data can in fact be accounted for by the sparsity of time points (figure 4A), but

192 the P-value distribution from constant-volatility tests of 16,000 randomly drawn
193 surrogate CVT simulations is much more uniform than the same distribution
194 from the male tongue, which has many small P-values indicative of consistent
195 state-dependent volatility of f_t (figure 3b). Thus, the non-neutrality of the male
196 tongue dataset is not due to long sampling intervals.

197 The male tongue microbiome deviates from neutrality by having a signifi-
198 cantly more $\beta_2 > 0$ than $\beta_2 < 0$ for those auxiliary regressions yielding signifi-
199 cant heteroskedasticity (figure 3C). $\beta_2 > 0$ indicates that the volatility increases
200 farther away from the mean and plotting the residuals, $\hat{\epsilon}^2$ against the state vari-
201 able, f_t , reveals the form of heteroskedasticity (figure 3D). A similar significant
202 hyper-abundance of $\beta_2 > 0$ exists for all datasets considered here (figure 3E).

203 Discussion

204 Evolution, driven by competition over rivalrous goods or limiting resources,
205 is a phenomenon common to ecology, economics and sociology, and accurate
206 statistical models of how competitive systems evolve can allow us to forecast,
207 manage, and invest in them [2, 3, 8, 1, 31]. Neutral Theory is a null model of
208 competition which assumes that all players are equal - that a canopy tree fills a
209 gap in the canopy independent of its species' identity, a dollar finds its way to
210 another dollar independent of who owns the dollar, and a seat in congress is filled
211 by someone independent of the racial, cultural or political traits of the successor
212 or predecessor. It's been hypothesized that neutrality could arise naturally as
213 a result of competitively inferior species going extinct [32], and thus systems
214 would tend towards neutrality over long periods of time, but the accuracy and
215 generality of Neutral Theory as a dynamical model for a range of competitive
216 systems was unclear.

217 We have provided a time-series test of neutral covariance structure that

218 reveals a common feature of non-neutrality across a range of ecological and
219 economic systems. Our test is based on the grouping invariance of neutral
220 communities, and this grouping invariance is maintained by a particular co-
221 variance structure of volatilities, namely where the volatility of a group, X_t^i ,
222 is $v(X_t^i) = cX_t^i(1 - X_t^i)$ for some constant, c , which can only be invariant to
223 grouping if the covariance between groups i , and j is $\Sigma_{i,j} = -cX_t^iX_t^j$. We test
224 neutrality by randomly grouping species and testing if the volatility of those
225 random groups is of the form of $v(x)$. A deviation from v can indicate non-
226 neutral covariance structure. If, for instance, instead of a quadratic curve of
227 v , the volatility of random groups in the data follows a bell-shaped curve with
228 positive curvature at the end-points, it could indicate that there is a positive
229 covariance between rare species, possibly due to kill-the-winner effects or due
230 to the relative abundances being driven by fluctuations in the largest popula-
231 tions. Conversely, excessive negative curvature could indicate strong negative
232 covariances between rare species, possibly due to relatively constant populations
233 of abundant species. Future work analyzing the volatility surfaces of random
234 groups can improve our test and allow researchers to quickly isolate particular
235 forms and fingerprints of common, non-neutral competitive asymmetries.

236 The results presented here are limited to the particular choice of species (the
237 taxonomic scale), resource (trophic scale), and time scale. However, an analysis
238 restricted to a particular scale means that our test can also allow researchers to
239 probe multiple taxonomic, trophic and time-scales to see if there are patterns in
240 which scales are most/least neutral in their dynamics. An alternative grouping
241 of these original OTUs by genera may reveal different results by conditioning the
242 groupings of species on a particular sub-set of possible groups, namely by group-
243 ing species with recent ancestors and shared traits together, and consequently
244 this test could serve as a tool for evaluating competition at multiple taxonomic

245 scales. The results are also limited to the choice of resource: companies within
246 a sector governed by trust-busting policies which break the neutral symmetry in
247 their market capitalization dynamics may still be neutral in their competition
248 over the ethnic or cultural composition of their labor force. It's possible that
249 microbes in the gut are not neutral in their short-term fluctuations over the
250 course of a year, but perhaps are neutral over longer time-scales that average
251 out short-term fluctuations in diet and physiological state that are known to
252 have predictable effects on microbial communities [33, 34].

253 Our test can be applied to any community of competing agents classified into
254 discrete groups for which time-series of relative abundances (or market share,
255 etc) of the groups are available. This test is most effective when the time-series
256 is long and the spacing between samples is short relative to the turnover rate of
257 the underlying resources (trees, dollars, congressional seats). There are many
258 ways to build on our method. Explicit calculations of the drift and volatility
259 over long time-intervals can improve our method for datasets with sparse time
260 points. The dependence of the CVTs may be calculated as copulas allowing the
261 implementation of a more exact goodness of fit test [35]. There may be more
262 CVTs that solve the Hamilton-Jacobi equation, and different CVTs might be
263 more specialized at detecting different asymmetries in competitive systems.

264 Neutrality tests of time-series data can help us understand the stochastic
265 time-evolution of competitive systems and facilitate better prediction and man-
266 agement. For bacteria in the gut, for instance, understanding the important
267 non-neutral forces governing the dynamics could allow progress towards the
268 stochastic pharmacokinetics of probiotics [36]. Understanding the predictability
269 of invasions at different taxonomic scales can tell us whether to evaluate metage-
270 nomic or ecological datasets based on the species, or whether other taxonomic
271 levels will yield a more informative analysis - perhaps grouping tropical trees

272 into the family Fabaceae, the family Melastomataceae, the genus *Cecropia*, and
273 all other trees reveals trophic structure of tropical forests and competitive asym-
274 metries that are drowned out by analyses at the species level. Demonstrating
275 that the non-neutrality of portfolios is consistent with rare-species advantages
276 in the Atlas model [8] would have major implications for portfolio design. In all
277 cases, the first step of empirically demonstrating the existence of competitive
278 asymmetries in time-series data can now be done with the test provided here.

279 **Materials and Methods**

280 *Neutral Theory and the Wright-Fisher Process*

281 Large neutral communities are well-approximated by a Wright-Fisher Process
282 (WFP) [37, 38]. The convergence of discrete neutral communities to the contin-
283 uous diffusion model of the WFP is covered in [39], and some numerical methods
284 used for parameter estimation and simulation have been produced by [31]. The
285 WFP is a continuous-state, continuous-time approximation of Kimura’s theory
286 [14], it is an approximation of Hubbell’s neutral theory [15] when speciation
287 rates are negligible over the timescale of interest, and it describes the dynam-
288 ics of relative abundances of non-zero-sum volatility-stabilized market models
289 [17, 27]. Using the WFP as a continuous approximation of large, finite communi-
290 ties simplifies the covariance in the jumps between species’ relative abundances,
291 thereby permitting the analysis below.

292 The WFP models the stochastic time-evolution of relative abundances of n
293 species. Let $\mathbf{X}_t = (X_t^1, \dots, X_t^n)$ be the vector of relative abundances, the WFP
294 is defined by the Itô SDE

$$d\mathbf{X}_t = \lambda(\rho - \mathbf{X}_t) dt + \sigma(\mathbf{X}_t) d\mathbf{W}_t \quad (2)$$

295 where $\lambda > 0$ and $\rho > 0$. The covariation between relative abundances of different
296 species is given by the elements of $\Sigma = \sigma\sigma^T/2$, where

$$\Sigma_{i,j}(x_1, \dots, x_n) = \begin{cases} x_i(1-x_i) & i = j \\ x_i x_j & i \neq j \end{cases}. \quad (3)$$

297 The deterministic motion, or drift, of the WFP - $\lambda(\rho - \mathbf{X}_t)$ - yields exponen-
298 tial mean reversion like many dynamical systems reverting to a nearby stable
299 equilibrium. The quadratic covariation of the WFP, Σ , captures the stochastic
300 fingerprint of neutrality; it arises from randomly drawing a resource to be freed
301 from its agent followed by randomly drawing one of the remaining agents to
302 acquire that resource with a probability proportional to the agent's current re-
303 source ownership. The family of Wright-Fisher Processes is closed to grouping,
304 meaning that if a multi-species community's dynamics are governed by a WFP,
305 species can be grouped (e.g. collecting species into genera or higher taxonomic
306 levels) and the dynamics of the resulting, re-grouped community will also be
307 governed by a WFP.

308 *Testing Neutral Covariance Structure*

309 We developed a test that is intentionally sensitive to the state-dependent noise
310 of the WFP, allowing researchers to test the underlying stochastic model of the
311 random turnover of resources at the heart of neutral competition. Developing
312 a strong test of the state-dependent covariance is not trivial, though, because
313 direct measurement of the covariance of jumps conditioned on the state of the
314 system, $\text{Cov}[\Delta\mathbf{X}_t|\mathbf{X}_t]$, would require many replicate time points each with the
315 same initial state, \mathbf{X}_t , and, even with multiple time points at the same state,
316 the sparsity of the high-dimensional data challenges the accurate estimation
317 and significance testing of the covariance matrix. To circumvent these problems

318 of replicate time points and high dimensionality and develop a strong test the
319 state-dependent noise of the WFP, we find a variance-stabilizing transformation
320 for the WFP that allows a regression-based heteroskedasticity test [40].

321 To be precise, we are looking for a real-valued function $f(\mathbf{X}_t)$ such that for
322 \mathbf{X}_t obeying the WFP law in equation (1),

$$\lim_{\Delta t \rightarrow 0} \text{Var} \left[\frac{f(\mathbf{X}_{t+\Delta t}) - f(\mathbf{X}_t)}{\sqrt{\Delta t}} \right] = \text{const.} \quad (4)$$

323 This approach is conceptually similar to the variance-stabilizing tools used for
324 population fluctuation analyses [10, 25, 11] which stabilize the variance in jumps
325 of a population size, except our function must stabilize the covariance of jumps
326 between populations, not just the variance. In particular, to have a constant
327 volatility, our function f must satisfy the Hamilton-Jacobi equation,

$$\nabla f^T \Sigma \nabla f = \text{const.} \quad (5)$$

328 (see supplement part S2 for more details). The grouping invariance of the WFP
329 can be used to intuit and show that there are at least 2^n different variance-
330 stabilizing transformations of the WFP, parametrized by a vector \mathbf{a} :

$$f_{\mathbf{a}}(\mathbf{X}_t) = \arcsin(a_1 X_t^1 + \dots + a_n X_t^n), \quad (6)$$

331 where $a_i = \pm 1$ for all i .

332 After transforming the data with $f_{\mathbf{a}}$, we need to perform a constant-volatility
333 test. To test the constant volatility of f , we test the homoskedasticity of stan-
334 dardized jumps,

$$\nu_t = (f_{t+\Delta t} - f_t) / \sqrt{\Delta t}, \quad (7)$$

335 following regression on f_t to eliminate the state-dependent drift.

336 A homoskedasticity test of ν_t for a single CVT is a test of neutrality for
337 time-series data. However, with 2^n different CVTs, one can perform multiple
338 hypothesis tests. For any multiple-hypothesis tests, if the null hypothesis is true,
339 the distribution of P-values is uniform. In this paper, we test the uniformity
340 of the distribution of resultant P-values from homoskedasticity tests of ν_t for
341 a number of randomly drawn CVTs. Figure 1 illustrates this test for 2,000
342 randomly drawn CVTs and shows the successful rejection of the WFP for the
343 relative abundances of a mean-reverting geometric Brownian motion.

344 The P-values arising from homoskedasticity tests of different CVTs are not
345 independent. Consequently, a Kolmogorov-Smirnov test of the P-value distri-
346 bution against a uniform distribution would have a high false-positive rate. To
347 reduce the false-positive rate, we perform a perturbation analysis to generate
348 conservative estimates of cutoffs for the KS statistic at 0.05 and 0.005 signifi-
349 cance levels. Details of the sensitivity analysis are provided in the supplement
350 part S3.

351 Acknowledgments

352 The authors would like to thank R. Chisholm, S. Levin, S. Pacala, J. O'dwyer,
353 and J. B. Socolar for their discussions and feedback. In particular, A.D.W.
354 would like to thank J. B. Socolar for numerous excellent recommendations that
355 have greatly improved this manuscript, and both A. Gammie and D. Nemergut
356 for their encouragement and support.

357 References

- 358 [1] Darwin C, Bynum WF. The origin of species by means of natural selection:
359 or, the preservation of favored races in the struggle for life. AL Burt; 2009.

- 360 [2] Lewontin RC. The units of selection. Annual review of ecology and sys-
361 tematics. 1970;p. 1–18.
- 362 [3] Holland JH. Complex adaptive systems. Daedalus. 1992;p. 17–30.
- 363 [4] Levin SA. Ecosystems and the biosphere as complex adaptive systems.
364 Ecosystems. 1998;1(5):431–436.
- 365 [5] Arthur WB. Complexity and the economy. science. 1999;284(5411):107–
366 109.
- 367 [6] Grant PR, Grant BR. Evolution of character displacement in Darwin’s
368 finches. science. 2006;313(5784):224–226.
- 369 [7] Bengtsson J. Interspecific competition increases local extinction rate in a
370 metapopulation system. Nature. 1989;340(6236):713–715.
- 371 [8] Fernholz ER. Stochastic portfolio theory. Springer; 2002.
- 372 [9] Taylor C, et al.. Multiculturalism: Examining the Politics of Recognition,
373 ed. Amy Gutmann. Princeton: Princeton University Press; 1994.
- 374 [10] Lande R, Engen S, Saether BE. Stochastic population dynamics in ecology
375 and conservation. Oxford University Press Oxford; 2003.
- 376 [11] Kalyuzhny M, Seri E, Chocron R, Flather CH, Kadmon R, Shnerb NM.
377 Niche versus Neutrality: A Dynamical Analysis. The American Naturalist.
378 2014;184(4):439–446.
- 379 [12] Black F, Scholes M. The pricing of options and corporate liabilities. The
380 journal of political economy. 1973;p. 637–654.
- 381 [13] Halley JM, Iwasa Y. Neutral theory as a predictor of avifaunal extinc-
382 tions after habitat loss. Proceedings of the National Academy of Sciences.
383 2011;108(6):2316–2321.

- 384 [14] Kimura M. The neutral theory of molecular evolution. Cambridge Univer-
385 sity Press; 1985.
- 386 [15] Hubbell SP. The unified neutral theory of biodiversity and biogeography
387 (MPB-32). vol. 32. Princeton University Press; 2001.
- 388 [16] Bentley RA, Hahn MW, Shennan SJ. Random drift and culture change.
389 Proceedings of the Royal Society of London B: Biological Sciences.
390 2004;271(1547):1443–1450.
- 391 [17] Fernholz R, Karatzas I. Relative arbitrage in volatility-stabilized markets.
392 Annals of Finance. 2005;1(2):149–177.
- 393 [18] Dangerfield C, Kay D, Burrage K. Stochastic models and simulation of ion
394 channel dynamics. Procedia Computer Science. 2010;1(1):1587–1596.
- 395 [19] Volkov I, Banavar JR, Hubbell SP, Maritan A. Neutral theory and relative
396 species abundance in ecology. Nature. 2003;424(6952):1035–1037.
- 397 [20] Tajima F. Statistical method for testing the neutral mutation hypothesis
398 by DNA polymorphism. Genetics. 1989;123(3):585–595.
- 399 [21] Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. Genetics.
400 2000;155(3):1405–1413.
- 401 [22] Leigh EG. Neutral theory: a historical perspective. Journal of evolutionary
402 biology. 2007;20(6):2075–2091.
- 403 [23] McGill BJ, Maurer BA, Weiser MD. Empirical evaluation of neutral theory.
404 Ecology. 2006;87(6):1411–1423.
- 405 [24] Chave J, Alonso D, Etienne RS. Theoretical biology: comparing models of
406 species abundance. Nature. 2006;441(7089):E1–E1.

- 407 [25] Chisholm RA, Condit R, Rahman KA, Baker PJ, Bunyavejchewin S, Chen
408 YY, et al. Temporal variability of forest communities: empirical estimates
409 of population change in 4000 tree species. *Ecology letters*. 2014;17(7):855–
410 865.
- 411 [26] van Kampen NG. *Stochastic processes in physics and chemistry*. Amster-
412 dam ; New York : New York: North-Holland; 1981.
- 413 [27] Pal S. Analysis of market weights under volatility-stabilized market models.
414 *The Annals of Applied Probability*. 2011;21(3):1180–1213.
- 415 [28] Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A,
416 Stombaugh J, et al. Moving pictures of the human microbiome. *Genome*
417 *Biol*. 2011;12(5):R50.
- 418 [29] Pardieck K, Ziolkowski Jr D, Hudson M. North american breeding bird
419 survey dataset 1966-2013, version 2013.0. US Geological Survey, Patuxent
420 Wildlife Research Center < www.pwrc.usgs.gov/BBS/RawData. 2014;.
- 421 [30] White H. A heteroskedasticity-consistent covariance matrix estimator and
422 a direct test for heteroskedasticity. *Econometrica: Journal of the Econo-*
423 *metric Society*. 1980;p. 817–838.
- 424 [31] Washburne A. *Competition and Coexistence in an Unpredictable World*;
425 2015.
- 426 [32] Hubbell SP. Neutral theory in community ecology and the hypothesis of
427 functional equivalence. *Functional ecology*. 2005;19(1):166–172.
- 428 [33] Pflughoeft KJ, Versalovic J. Human microbiome in health and disease.
429 *Annual Review of Pathology: Mechanisms of Disease*. 2012;7:99–122.

- 430 [34] David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe
431 BE, et al. Diet rapidly and reproducibly alters the human gut microbiome.
432 Nature. 2014;505(7484):559–563.
- 433 [35] Chicheportiche R, Bouchaud JP. Goodness-of-fit tests with dependent
434 observations. Journal of Statistical Mechanics: Theory and Experiment.
435 2011;2011(09):P09003.
- 436 [36] Marteau P, Vesa T. Pharmacokinetics of probiotics and biotherapeutic
437 agents in humans. Bioscience and microflora. 1998;17(1):1–6.
- 438 [37] Fisher RA. The genetical theory of natural selection. Clarendon; 1930.
- 439 [38] Wright S. Evolution in Mendelian populations. Genetics. 1931;16(2):97.
- 440 [39] Durrett R. Probability models for DNA sequence evolution. Springer Sci-
441 ence & Business Media; 2008.
- 442 [40] Everitt BS, Skrondal A. The Cambridge dictionary of statistics. Cambridge:
443 Cambridge. 2002;.