MODEL-FREE DETECTION OF A SPECULATIVE ASSET BUBBLE: EVIDENCE FROM THE WORLD MARKET FOR SUPERSTAR WINES

TOR N. TOLHURST

Abstract. Bubbles occur when an asset price deviates from its fundamental value. Economists have shown asset bubbles are consistent with neoclassical theory and can arise in a variety of laboratory settings; yet cogent, direct evidence of a bubble in an actual market has remained elusive. The challenge for the empiricist is that fundamental values are necessarily unobservable. I propose a bubble test for one of two nearly identical assets assuming the difference in their fundamental values is a latent exchangeable random variable. Their close relationship implies the difference in fundamentals between the two assets is bounded, bounds which I can find using a concentration inequality. Using this test, I find strong evidence of a bubble in the price of the Bordeaux wine Lafite Rothschild, relative to other fine wines exchanged on global secondary markets. In at least three consecutive periods the upper bound is violated. This is the first direct evidence of a bubble which is both consistent with rational bubble theory and independent of a structural valuation model.

Keywords. Asset pricing, rational bubbles, concentration inequalities, wine.

1. Introduction

Asset bubbles damage the real economy: the distorted signals they create interfere with the coordination function of prices; the frenzy they spark promotes recklessness and facilitates fraud; their burst often precipitates wider financial crises; and they otherwise increase uncertainty and erode trust in markets. Economic agents also behave in intriguing ways during bubbles. This behavior—reflected in the eloquent titles of prominent works, from *Extraordinary Popular Delusions and the Madness of Crowds* (Mackay, 1852) to *Irrational Exuberance* (Shiller, 2015)—stimulated work in behavioral economics and the efficient market hypothesis. However, evidence of an asset bubble outside of a laboratory setting has been tarnished by joint testing: a nonzero bubble value has yet to be separated from possible misspecification of the chosen structural valuation model. Giglio, Maggiori, and Stroebel (2016) developed a model-free bubble test to avoid the endemic joint testing

E-mail address: `tntolhurst@ucdavis.edu`.
Date: July 19, 2019.

problem, but failed to find evidence of a bubble in two asset markets, the United Kingdom and Singapore housing markets, that are commonly thought to have experienced bubbles.

In this manuscript I use a test similar to, but more general than, Giglio, Maggiori, and Stroebel's to look for a bubble in a particular asset class, fine wines exchanged on global secondary markets. I find compelling evidence of a bubble. In the wake of the 2008 financial crisis, fine wine received attention as a promising alternative investment in both the popular press (e.g. Lister, 2011; Rabinovitch, 2010) and academia (e.g. Cardebat and Figuet, 2010; Faye, Le Fur, and Prat, 2015; Le Fur, Ameur, and Faye, 2016). Beginning in 2010, some industry insiders began to opine that there was a bubble in one particular Bordeaux wine, *Chateau Lafite-Rothschild* (e.g. Stimpfig, 2010; Temperton, 2011; Authers, 2012). It is easy to see why this conjecture emerged in Figure 1, which plots monthly prices of first growth Bordeaux wines, widely considered to be of equivalent quality. These five wines are the only ones classified as first growth in Bordeaux: Lafite, Margaux, Haut Brion, and Latour since 1855 and Mouton since 1973. Figure 1 clearly illustrates a boom in Lafite prices. Prior to 2006, their prices were roughly equivalent: in June 2005, they all cost roughly £75 per bottle (roughly $135 at the time) with Margaux £79, Lafite £75, Latour £74, Mouton £62, and Haut Brion £62.[1] Prices moved mostly together until June 2009. Then Lafite boomed: at the peak in February 2011, Lafite cost £810 per bottle, compared to £469 for Latour, £380 for Margaux, £360 for Mouton, and £314 for Haut Brion. Returns to Lafite over the nearly 70-month boom were astounding: 51.3 percent annualized for 5.75 years. As would be expected in a bubble, the relative price of Lafite then dropped precipitously. By January 2017, Lafite was down to £490 and again aligned with other first growth prices: Latour £495, Margaux £356, Mouton £350, and Haut Brion £337.[2] The ultimate purpose of this manuscript is to determine whether or not the price changes in Lafite provide direct evidence of a bubble. I will argue they do.

The possibility of bubbles has been well-established by economists.[3] Rational bubble theory shows explosive solutions for the market-clearing price are compatible with a neoclassical setup

[1]Prices are denominated in UKP because London is the historical hub of the secondary market. All prices are real in 2015 terms, deflated by the UK CPI reported by the Great Britain Office for National Statistics. Available at: `https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/d7bt/mm23`.

[2]The annualized rate of return from June 2005 to Jan. 2017 was a healthy but more reasonable 17.5 percent.

[3]The literature on bubbles is vast and rich. Note that while the literature demarcates several "bubble-like" behaviors—such as sunspots (Le Roy and Porter, 1981), fads (Camerer, 1989), information bubbles (Bikhchandani, Hirshleifer, and Welch, 1992), and irrational bubbles (Le Roy, 2004)—I limit the scope of the paper to focus on the leading conception, the rational bubble. For more complete reviews of the literature emphasizing theory see Camerer (1989) and Le Roy (2004) or for empirics West (1988), Flood and Hodrick (1990) and Gürkaynak (2008). For review-type discussions of bubbles in the context of the 2007–2008 financial crisis, see especially Bhattacharya and Yu (2008), O'Hara (2008), and Carvalho, Martin, and Ventura (2012).

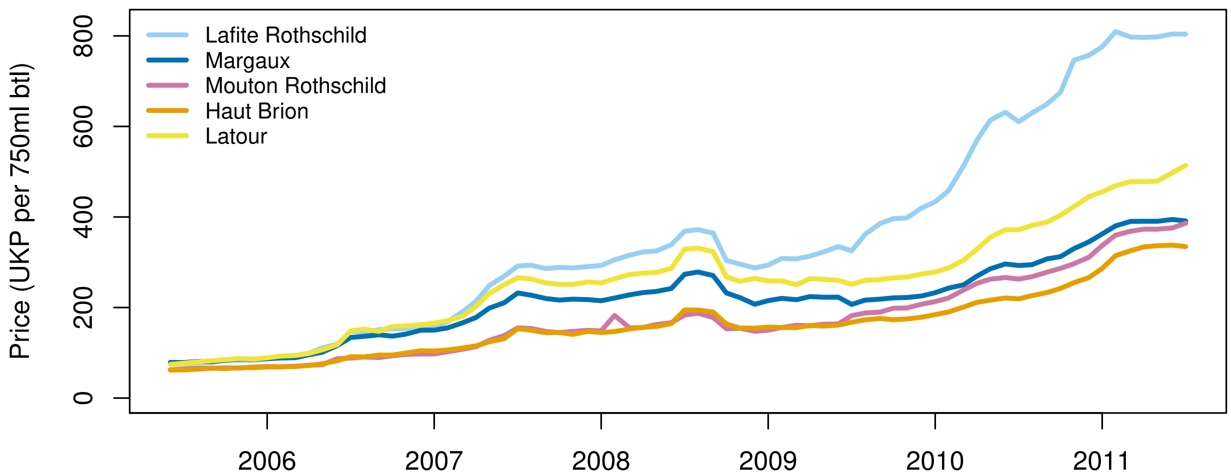

FIGURE 1. Monthly average price of a ten-vintage portfolio of the five first growth Bordeaux wines, June 2005 – July 2011.

(Samuelson, 1958; Diamond, 1965; Blanchard, 1979; Blanchard and Watson, 1982; Tirole, 1982, 1985; Froot and Obstfeld, 1991). The mechanism behind rational bubbles is a form of self-fulfilling prophecy: agents expect a bubble return, which leads to a realized bubble return, which leads to further expected bubble returns, continued participation in the bubble trade, and so on. Experimental evidence, reviewed by Palan (2013), has largely supported the theoretical possibility of rational bubbles (Smith, Suchanek, and Williams, 1988; Lei, Noussair, and Plott, 2001; Dufwenberg, Lindqvist, and Moore, 2005; Kirchler, Huber, and Stöckl, 2012; Palfrey and Wang, 2012; Moinas and Pouget, 2013; Eckel and Füllbrunn, 2015; Andrade, Odean, and Lin, 2015). Bubbles arise in laboratory settings even with knowledgeable traders who experienced bubbles in the past (Hussam, Porter, and Smith, 2008). Despite the common association of bubbles with economically irrational behavior, economists have shown bubbles are possible with rational, fully informed agents.

The fundamental value of an asset is the discounted present value of all future expected prices; a bubble occurs when the market price exceeds this fundamental value. Rational bubbles are the most important conceptualization of bubbles and the theoretical cornerstone of the empirical bubble literature. A rational bubble is a "particular type of explosive indeterminancy" (Flood and Hodrick, 1990, p.g. 86) where the bubble component of the asset price grows proportionate to the discount rate, trapping rational agents in a self-fulfilling cycle of rising prices. To fix ideas, consider the simple framework of Flood and Hodrick (1990).[4] Let P_t be the market-clearing asset price,

[4]Despite its simplicity, rational bubble theory is quite versatile. For example, agents can recognize the bubble will eventually burst (Camerer, 1989), face multiple sources of uncertainty (Doblas-Madrid, 2012), or endogeneous credit constraints (Miao and Wang, 2018).

and D_t its dividend, in period t. Assume homogeneous risk-neutral agents share a time-varying information set, $\mathcal{I}_t$, and a time-invariant, positively-valued discount rate r. The bubble component of the asset price, nonnegative given free disposal, is denoted B_t. There is an asset bubble when $B_t > 0$. The first-order equilibrium condition of the utility-maximization problem is:

$$P_t = \sum_{i=1}^{\infty} \mathbb{E}[D_{t+i}|\mathcal{I}_t](1+r)^{-i} + B_t, \tag{1}$$

which can be satisfied by a positive bubble component value, $B_t > 0$, provided:

$$B_t = \mathbb{E}[B_{t+1}|\mathcal{I}_t](1+r)^{-i}. \tag{2}$$

That is, an asset bubble can satisfy the equilibrium condition (1) provided the bubble component is expected to grow proportional to the discount rate. This is a "rational bubble." The term

$$F_t := \sum_{i=1}^{\infty} \mathbb{E}[D_{t+i}|\mathcal{I}_t](1+r)^{-i}$$

is referred to as the asset's fundamental value. The central empirical challenge in testing for a bubble (rational or otherwise) is that fundamental values are latent.

Treatment of the fundamental value distinguishes the two classes of empirical bubble tests: direct and indirect. Direct tests incorporate B_t into the hypothesis test using, for example, coefficient restrictions (Flood and Garber, 1980) or two-step, Hausman-style tests (West, 1987). While the ideal direct test would provide the most credible evidence for a bubble, direct tests are plagued by a joint hypothesis problem: because F_t is unobservable it must be modeled with a structural valuation model, but then the the no-bubble null is joint with this model. In other words, evidence of a bubble cannot be separated from a misspecified structural valuation model or otherwise unobservable changes in fundamentals.

Indirect tests circumvent the challenge of unobservable F_t by examining observed prices P_t for statistically anomalous behavior (e.g. explosive unit roots). Given the difficulties with unobservable fundamentals and advancements in statistical methods, indirect tests have received the most attention over the past 30-odd years.[5] Prominent examples of indirect tests include variance-bounds (Shiller, 1981; Le Roy and Porter, 1981), stationarity (Hamilton and Whiteman, 1985; Kleidon, 1986), cointegration (Diba and Grossman, 1987, 1988a,b), vector autoregression (Campbell and

[5] Empirical applications of heterogeneous agent models (Harrison and Kreps, 1978), reviewed in Xiong (2013), are increasingly prominent in the literature but subject to the same joint hypothesis testing pitfalls.

Shiller, 1987, 1988a,b), Markov switching (Hall, Psaradakis, and Sola, 1999; Al-Anaswah and Wilfling, 2011), and most recently, right-sided unit roots (Phillips and Yu, 2011; Phillips, Wu, and Yu, 2011; Phillips, Shi, and Yu, 2014, 2015a,b) based on the asymptotic theory of Phillips and Magdalinos (2007). A recent application of the Phillips *et al.* approach is to the cryptocurrency Bitcoin (Cheung, Roca, and Su, 2015). However, rejection of the null hypothesis in an indirect test is not equivalent to detection of a rational bubble: for example, Giglio, Maggiori, and Stroebel (2016) find statistical evidence of a bubble using the indirect Phillips *et al.* test, but find a precisely estimated no-bubble null result on the same data with their direct test (which overcomes the joint hypothesis testing problem).

To circumvent the joint hypothesis problem, Giglio, Maggiori, and Stroebel (2016) compare real estate prices within a jurisdiction where there are two different forms of ownership: freeholds (infinite-maturity) and leaseholds with a very long maturity. Provided the form of ownership does not generate unobservable differences in prices (i.e. the houses differ only in their maturity), the difference in latent fundamental values would be arbitrarily small. A statistically nonzero difference in observed prices could then be directly attributed to a rational bubble, without imposing a model on the fundamental values. A more general way to look at how Giglio, Maggiori, and Stroebel (2016) approach testing for a bubble is that, for the right kind of assets, we can replace valuation models with identifying assumptions. This is an attractive trade. Valuation models imposed to estimate fundamental values are necessarily joint to a direct bubble test, whereas identifying assumptions need not be joint. Further, identifying assumptions may be quantitatively or qualitatively verifiable, or at least not rejected, independently of a bubble test.

The contributions of this manuscript are twofold. First, I propose a direct test of a bubble for one of two nearly identical assets that is independent of a structural valuation model. Intuitively, the fact that the two assets are nearly identical implies that their fundamental values should move (mostly) together. Thus, I assume the relationship between the two products enforces bounds on the difference in their fundamental values. To construct the bounds, I assume only that the difference in their fundamental values is an exchangeable random variable.[6] This implies I can find probabilistic bounds (i.e. random deviations outside the bounds occur with a known probability) using a concentration inequality without any other assumptions about the latent data generating

[6] Saw, Yang, and Mo (1984) define a weakly exchangeable random variable as a $X_1, X_2, \ldots, X_n, X_{n+1}$ such that $P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n, X_{n+1} \leq x_{n+1})$ is invariant over all permutations of $x_1, x_2, \ldots, x_n, x_{n+1}$ for all values of $x_1, x_2, \ldots, x_n, x_{n+1}$. For example, an i.i.d. random variable is by definition weakly exchangeable.

process. The bounds are based on previous observed price differences (with appropriate finite sample adjustments) and can be time-varying. This means I assume that, in the absence of a bubble, changes in fundamental values cannot be simultaneously large *and* fast. As the test is direct and independent of a structural valuation model, it is model-free in the sense of Giglio, Maggiori, and Stroebel (2016). Additionally, my more general identifying assumptions are applicable to any nearly identical assets, a wider class of assets than leasehold and freehold real estate.

Second, I apply this test to Lafite and find strong robust evidence of a bubble. I show the close peers to Lafite are nearly identical in a variety of dimensions, including wine type, style, scarcity, price tier, historical classifications, and perhaps most notably, expert review scores. The prices I use to test for a bubble are a ten-vintage rolling portfolio of most recent releases, which mitigates the possibility of the change in the price of a single vintage (such as a production shock, stock exhaustion, or an exceptional expert review) creating a non-bubble surge in prices. I consider other possible explanations for a Lafite-specific shock common to all ten vintages, but find them to be largely without merit. For example, Lafite is one of the largest fine wine producers in Bordeaux and even old vintages are readily (and widely) available for purchase. With time-invariant bounds, the test detects a Lafite bubble beginning in late-2009 and ending more than two years later. With (much lower power) time-varying bounds, the test detects the beginning of a bubble in at least three consecutive periods in late-2009. Three consecutive periods of detection imply the probability of a random false detection is less than 0.1 percent. The findings are robust to reconducting the test using detrended or first-differenced data, bootstrapping estimates of the concentration inequality, alternative bandwidths for the concentration inequality, as well as alternative specifications for the group of close peers.

2. Empirical Test

2.1. **Setup.** Consider a well-functioning marketplace with homogeneous risk-neutral participants. Two nearly identical, dividend-earning, infinitely-lived assets share a time-varying information set, $\mathcal{I}_t$, and stochastic discount rate, r_t, expressed as $\delta_{t,t+i} = \prod_{j=0}^{i-1}(1+r_j)^{-1}$. Their market-clearing prices are denoted P_t^* and P_t. The price of an asset today is the discounted expected price of the asset tomorrow, which by iterated expectations and recursion gives the present values:

$$(3) \qquad P_t^* = \sum_{i=1}^{\infty} \mathbb{E}[\delta_{t,t+i} D_{t+i}^* | \mathcal{I}_t] + B_t,$$
$$P_t = \sum_{i=1}^{\infty} \mathbb{E}[\delta_{t,t+i} D_{t+i} | \mathcal{I}_t].$$

Let $F_t^* = \sum_{i=1}^{\infty} \mathbb{E}[\delta_{t,t+i} D_{t+i}^* | \mathcal{I}_t]$ with dividends D_t^*, and F_t defined analogously, be the fundamental values of the two assets. An asset bubble occurs when an asset's price exceeds its fundamental value. By free disposal, $B_t \geq 0$. I test for a bubble in P_t^* (with $P_t^* \geq P_t$), where a bubble occurs when $B_t > 0$.[7]

2.2. **Generalized Model-Free Test.** I seek a test of $B_t > 0$ based on observed prices (P_t^*, P_t). For the test to be model-free, it must be independent of a structural valuational model on the latent fundamental values (F_t^*, F_t). Let $\pi_t = F_t^* - F_t$ denote the difference in fundamental values. Assume $D_t^* = D_t + q_t$ with $q_t \geq 0$, such that $\pi_t \geq 0$. Suppose further there exists a finite upper bound on this difference, π_t^*, which arises because the two assets are related and is enforced by arbitrage. We then have a bubble test based on observable prices and π_t^*:

$$(4) \qquad \text{For } \pi_t \in [0, \pi_t^*], \text{ any } (P_t^* - P_t) - \pi_t > 0 \text{ implies } B_t > 0.$$

So long as π_t^* is independent of a valuation model on the fundamental values, this test is model-free.

The focus of the test then becomes finding the upper bound π_t^*. The upper bound could be the cost of transforming asset one into asset two: for example, transporting or augmenting asset two to generate the same cashflow as asset one.[8] This transformation need not be physical or tangible: for example, it could be the compensation required to make consumers indifferent between two goods. Hypothetically, for some assets π_t^* may be directly observable or estimated with reasonable

[7] Suppose the second asset price also had a bubble component value, $b_t \geq 0$. I can normalize $b_t = 0$ when testing for $B_t > 0$ if a bubble is detected: if $(P_t^* - P_t) - (F_t^* - F_t) = B_t - b_0 > 0$, obviously $B_t > 0$ for any $b_0 \geq 0$. If not and $(P_t^* - P_t) - (F_t^* - F_t) = B_t - b_0$ cannot reject zero, we would not be able to differentiate between $B_t = 0$ or $B_t \approx b_t > 0$. If both assets are subject to a common bubble such that $P_t^* = F_t^* + B_t + b_t$ and $P_t = F_t + b_t$, then the common bubble terms, b_t, cancel out, but the test for $B_t > 0$ remains.

[8] The approach to probabilistically bounding fundamental values is based on the Baulch (1997) parity bounds framework for spatial arbitrage—specifically, the movement of a homogeneous good across two geographically distant markets. Given the different context, many details of Baulch (1997) are not pertinent. Perhaps the most important difference between Baulch (1997) and the proposed approach is that the former uses exogenous information to estimate π_t^*. Specifically, Baulch uses data to estimate transport costs in one period, then extrapolates single-period transport costs into a time-series using a consumer price index, and finds probabilistic bounds using parametric estimates. Given difficulties with specifying the information required to estimate a threshold, I will use prior observations and non-restrictive, nonparametric assumptions on the data-generating process to define the upper bound.

confidence.[9] But generally, this will be difficult in practice. Certainly in my empirical application, it is not obvious how to do so without a hedonic valuation model—which would invalidate the requirement that π_t^* accord with all structural valuation models.I will pursue a different strategy.

Assume π_t is a random variable from some unknown underlying distribution. Other than that π_t is latent and bounded below by zero, very little is known about it. Specifically, there is nothing like the transversality condition in Giglio, Maggiori, and Stroebel (2016) that implies π_t is arbitrarily small. Nor can a central limit theorem be applied to reveal its underlying distribution. But for any specified level of false detection control, α, a concentration inequality can be used to guarantee that $P(\pi - \mu < \lambda\sigma) > 1 - \alpha$, for some constant λ, assuming only that the first two moments, μ and σ^2, are finite. The Cantelli (i.e. one-sided Cheybychev) inequality is appropriate because it is one-sided and imposes no more than two finite moments on the underlying distribution.[10] For $\lambda > 1$:

$$P\left(\frac{\pi_t - \mu_t}{\sigma_t} > \lambda\right) < \frac{1}{1+\lambda^2}, \tag{5}$$

which implies there is no more than a ten percent probability of an observation $\lambda = 3$ standard deviations above the mean, or a five percent probability of an observation $\lambda = 4.36$ standard deviations above the mean. Thus, if π_t is standardized based on its first two moments, then $\pi_t^* = 3$ and $\pi_t^* = 4.36$ would be nonparametric upper bounds for tests at $\alpha = 0.10$ and $\alpha = 0.05$, respectively. Note these are extremely conservative bounds: with a standard half-normal distribution, the probability of a realization 3 (4.36) standard deviations above the mean is only 0.00270 (1.3×10^{-5}). Clearly these bounds are independent of any structural valuation model.

Population moments are unobservable, so I derive a one-sided version of Saw, Yang, and Mo (1984) for the Cantelli inequality in finite samples (Appendix A). The finite-sample versions of the Cheybychev and Cantelli inequalities find finite-sample analogs that apply to observation X_{t+1} based on estimates of the first two moments from a weakly exchangeable sample $X_1, X_2, \ldots, X_t$. As in Saw, Yang, and Mo (1984) and its multivariate extension Stellato, Van Parys, and Goulart (2017), adjusting the inequality for a finite sample creates a minor complication: the sample inequality as a

[9]The following illustrative example comes to mind. Two identical firms harvest or mine resource X, for which demand is inelastic, stock is inexhaustible, sites are open access, and there are two hierarchical quality grades: A and B. Firm A harvests a site which produces grade A resource X, the price of which is possibly bubbly. Firm B harvests a different site which gives only grade B quality. Then π_t^* is the maximal cost for firm B to switch sites, which could be directly observed or reasonably well estimated.

[10]Tighter bounds can be achieved with more moments, see for example Bhattacharyya (1987).

function of λ is a step-function, and this step-function converges nonmonotonically to the population inequality. This has two consequences for reporting and generating results. First, controlling the false detection rate at exactly α is not always possible. For example, with 24 observations, $\lambda \in [2.61, 3.26)$ occurs with probability 0.1154 and $\lambda \in [3.26, 4.70)$ occurs with probability 0.0769, so I cannot control false detection at exactly ten percent. In this situation, I always choose the lowest value of λ that attains the desired false detection rate (i.e. $\lambda = 3.26$ in the proceeding example) and report the exact control probability where possible. Second, I do not conduct tests at the one percent level because to do so at the sample sizes considered would be misleading: for example, with 24 observations, $\lambda \in [4.71, \infty)$ occurs with probability 0.0385 and testing at one percent is not possible.[11]

The final consideration is the sample on which to calculate the analogs of the population moments. I calculate the sample analogs using a bubble-free pre-period, $\boldsymbol{P}_0$, which is a subset of all price pairs occurring before the possible bubble. Ideally we want $\boldsymbol{P}_0$ to exclude any observations where $B_t = 0$ might be violated—their inclusion biases $(\pi_t - \mu_t)/\sigma_t$ *downwards*, making detection more difficult—but because the bias is downward, misspecification will not increase false detection, and so is not critical if a bubble is detected in spite of the bias.[12] Specifically, define $\boldsymbol{P} = \{(P_t^*, P_t)\}_{t=1}^{T}$ as all price pairs we observe, indexed by $t = 1, \ldots, T$, and a bubble-free subset (of sequential observations), $\boldsymbol{P}_0 \subset \boldsymbol{P}$. I use two specifications of $\boldsymbol{P}_0$ to test for a bubble in P_t^*: time-invariant fundamental values based on a pre-determined bubble-free period, such that $\boldsymbol{P}_0^I = \{(P_j^*, P_j)\}_{j=T_0}^{T_1}$ where $1 \leq T_0 < T_1 < t \leq T$; and time-varying fundamental values via a (forward recursive) rolling window of length H, such that $\boldsymbol{P}_0^V = \{(P_j^*, P_j)\}_{j=t-H-1}^{t-1}$. Specifically, for time-invariant fundamental values, I assume all observations prior to 2009 form a bubble-free pre-period; that is, π_t^* is assumed to be fully described by pre-2009 data and otherwise time-invariant. For time-varying fundamental values, I use a base specification of $H = 24$ months. Note that rolling windows are a common approach to testing bubbles because they simulate testing for a

[11]A third complication affects some ancillary results with sample sizes 18 and 40. Nonmonotonic convergence implies that, at times, a *lower* λ threshold attains the desired rate of control in the finite sample than would be required in the population inequality. For example, with 18 observations, an observation 2.77 standard deviations above the mean occurs with probability 0.1000, whereas with the population inequality, this probability is attained 3 standard deviations above the mean. In this situation, I use the λ threshold given by the finite-sample inequality because it is exact for that sample size. See Appendix A for further detail.

[12]Consider a sample X_n representing the pre-period, which with no loss of generality has mean zero and unit variance. Add $B > 0$ to all observations: the mean increases and variance stays constant, implying a lower Z-score. Add $B_i > 0$ to one or more (but not all) observations: the mean and variance increases, again implying a lower Z-score. Thus, given $B_t \geq 0$, introducing bubbly observations into the pre-period results in a downward biased statistic and therefore more difficult detection.

bubble concurrent with new data appearing from one period to the next (e.g. Phillips and Yu, 2011; Baur and Glover, 2014). For both specifications, I use the sample mean and standard deviation of the respective $\boldsymbol{P}_0$. The robustness checks re-conduct the tests under alternative bandwidth values (i.e. different T_1 and H). Additionally, the robustness checks include a bootstrap to account for sampling variability in calculating the sample mean and variance. Again, these operations accord with any structural valuation model of the fundamental values, and thus the test is model-free.

2.3. **Discussion.** In summary, there are four identifying assumptions of the generalized model-free test: (i) the two assets must be sufficiently well related so that the difference in their fundamental values is bounded and that they share a common information set and discount rate; (ii) for $(P_t^* - P_t) - \pi_t^*$ to identify $B_t > 0$, it must be that $P_t^* \geq P_t$ and $F_t^* \geq F_t$ in period t; (iii) for an accurate π_t^*, the finite-sample Cantelli inequality must be calculated on a exchangeable sequence of random variables; and (iv) the bubble-free pre-period $\boldsymbol{P}_0$ must be well-defined.

For two nearly identical assets, (i) should hold without much complication. As in Giglio, Maggiori, and Stroebel (2016, pg. 1055, footnote 8), (i) also implies very fast, temporary price changes due to changes in the discount rate or information set are netted out and will not confound inference on a bubble. In fact, (i) does not even need to hold exactly: as in Baulch (1997), the relationship between the assets is allowed to vary independently within the bounds for whatever reason, including (possibly time-varying) differences in discount rates, information sets, and even preferences. When the test is conducted on a forward recursive basis, net changes in these relationships could be quite large, so long as they are relatively gradual (or quite fast, so long as they are relatively small). Similarly, assumptions (ii) and (iv) can be easily addressed. Assumption (ii) should not create much complication for the right assets: $P_t^* - P_t$ observed and π_t^* is derived independent of (F_t^*, F_t), so we can simply ignore the test results when either is negative. Assumption (iv) does make detection more difficult if it is violated by contaminating $\boldsymbol{P}_0$ with small bubbly observations. However, because a violation of (iv) does not make detection easier, it is inconsequential in the sense that it will not increase the probability of false detection. That is, if we find $B_t > 0$, the validity of assumption (iv) does not matter.

The main identifying assumption, which merits more attention, is (iii): the latent difference in fundamental values, π_t, is a weakly exchangeable sequence of random variables. This assumption ensures that the bound from the finite-sample Cantelli inequality is correct, where a non-exchangeable data generating process could confound bubble detection if it biases the upper bound π_t^* downwards. A sequence of random variables is exchangeable when its joint distribution is invariant to permutations of its sequencing order (i.e. permutating its index). The primary concern in a time-series context is that exchangeability does not admit most forms of serial dependence, making it a relatively strong assumption on the data.

Stepping back, in essence I am trading imposing a specific structural valuation model for independent (w.r.t. valuation models) identifying assumptions. In light of how little I impose on the theoretical interpretation of $P_t^* - P_t$ (and the latent distribution of π_t), I require only one relatively strong assumption on the data. Making this trade-off more attractive is that the veracity of the key assumption can be examined, perhaps not perfectly but at least to a reasonable extent (obviously not possible when imposing a structural valuation model), by comparing the two assets. Furthermore, the setup and context should alleviate some of the non-exchangability concerns arising from serial dependence. In general, because π_t is a (cross-sectional) difference, time-varying unobservable factors with common influence on both prices (e.g. parallel trends) are automatically differenced out. Definitionally, because the assets are nearly identical, latent factors which influence one price but not the other should be relatively minor. But nevertheless, any remaining influences might be satisfactorily accounted for with a deterministic trend or nth order-differencing. Autoregressive processes are perhaps the most common indirect bubble tests (e.g. Phillips, Wu, and Yu, 2011): their presence is more likely indicative of a bubble below the detection threshold than a false detection caused by an downward biased π_t^*. I will address these concerns with robustness checks.

Overall, the generalized model-free test presents a number of advantages. Most notably, the upper bound in the difference in fundamental values, π_t^*, is independent of a structural valuation model, and thus the test circumvents the joint hypothesis problem along the same lines as the test in Giglio, Maggiori, and Stroebel (2016). Additionally, it is plausibly applicable to a much wider class of assets than that of Giglio, Maggiori, and Stroebel. Other advantages include: the test is determined without distributional assumptions and requires only one observation above π_t^* to detect a bubble (in contrast to tests using serial dependence); the test is computationally simple in that it requires only the calculation of a sample mean and standard deviation; the test requires a relatively small series of training observations (at minimum, enough to reliably calculate a standard deviation); and under the population concentration inequality, the test can be inverted to give the implied probability of a bubble.[13] So long as a bubble is detected in the first asset, P_t^*, misspecification of the bubble-free pre-period is irrelevant. So too is an otherwise possibly confounding nonzero bubble component in the second asset.

The test is limited to a situation with two related assets, but even with two assets related in the required ways, we will not be able to detect a bubble in one of the two assets (or a bubble common to both assets). Of course, the test also necessarily depends on the frequency of observations, which is often determined by the availability of the data. But even with an ideal data set, the test has two major weaknesses.

The test's first weakness is low power. Concentration inequalities such as Cantelli and Cheyby-chev assume very little about the underlying distribution and are therefore extremely conservative. As a result, the upper bound is an extremely difficult threshold to reach; in other words, *a priori*, I know the test is likely to fail to reject the null even if $B_t > 0$. If the bubble evolves slowly over time, these low power problems will be exacerbated. If the test is conducted using an updating rolling window, they will be exacerbated even further: with the rolling window updating from one iteration to the next, a small (not sufficiently large to exceed π_t^*) nonzero bubble value will artificially make it more difficult to detect a bubble in the subsequent period. This is true even if $\boldsymbol{P}_0$ is correctly specified and exacerbated even further if not. Furthermore, when we fail to reject we cannot be certain there is no bubble: the result could be type II error, the consequence of a misspecified $\boldsymbol{P}_0$, or a countervailing bubble in the second asset.

The test's second weakness is that the exact nature of the bubble may be ambiguous. Finding $B_t > 0$ is consistent with a rational bubble—a rational bubble is in the set of bubbles identified by the test—but also other concepts such as a sunspot (Camerer, 1989). This is also true in Giglio, Maggiori, and Stroebel (2016): the overarching objective is to find a test that is model-free and consistent with a rational bubble.[14]

[13]The probability of a bubble is (weakly) greater than $1 - 1/(Z_t^2 + 1)$ where $Z_t = (P_t^* - P_t - \hat{\mu}_t)/\hat{\sigma}_t$. Given the nonmonotonic convergence of the finite-sample inequality (Appendix A), and that variation in Z_t tends to be smaller than the steps in the step-function, using the finite-sample inequality for this probability is not particularly insightful.

[14]Given the asset in my empirical application is a consumable good, it is worth noting that the test alone cannot differentiate between a bubble and a fad. This is a concern of ongoing work but consider the following from Mustacich (2015). The broad consensus among industry experts is that a bubble was initiated by demand from mainland China where, during the bubbly period: Lafite frauds were widespread, deceptive trademark names were filed, many stores opened with completely uninformed merchants, and Chinese investors purchased more than 30 estates in Bordeaux, many of which included variations of Lafite in their name such as "Lafitte." These activities involved non-trivial investments and are thus more consistent with a bubble than a fad.

3. Conducting the Generalized Model-Free Test

I divide this section into four subsections. I begin by addressing whether or not wines are an asset in which a rational bubble could arise. Then I address how Lafite is nearly identical to the wines I chose to compose the second asset. In the final two subsections I describe the data used to conduct the test and consider the plausibility of alternative explanations for a price surge.

3.1. **Applicability of Rational Bubble Theory to Wine.** Wine is an alternative asset and a transferable and tangible property. Tirole (1982, 1985) and Le Roy (2004) set out a number of criteria for rational bubbles, which superstar wines meet. According to Tirole (1982, 1985), rational bubble theory applies to assets that have resale value and are inelastically supplied in the short-run—both true for fine Bordeaux wines. All the data I use comes from secondary markets, so clearly the assets have resale value. Wines are inelastically supplied in the short-run as output is constrained to yields on a particular estate in a particular vintage. Expansions to production are restricted by law. Bordeaux estates can expand their production by purchasing neighboring vineyards, though inventory from previous vintages acquired in the acquisition cannot be sold under the acquiring estate's label. Alternatively, a Bordeaux estate could plant new vineyards on a contiguous property, but properties must be within the fixed area defined by appellation laws and even then they must wait several years for the vineyard to achieve adequate production quantity and quality.

Le Roy (2004) requires the asset pay dividends. Dividends need not be pecuniary to have fundamental value: at least a portion of the dividend's value might reflect the increased utility from consuming the wine at a later period or the appreciation in quality (or price) of the wine due to aging. Similarly for discount rates, at least a portion of the discount rate could represent the forebearance cost of deferring consumption from one period to the next. Even if the fundamental value of wine was entirely nonpecuniary, nonpecuniary value is not explicitly ruled out in rational bubble theory and regardless, nonpecuniary value in a wine can be readily exchanged for money on secondary markets.

Wine is technically not infinitely-lived, but practically it may be considered so.[15] Ultra-premium wines such as first growth Bordeaux retain their value for decades. Using an aggregation database, a cursory search of wines for sale on secondary markets retrieved 17 vintages of Lafite prior to 1900

[15]Rational bubbles can arise in finite-period games—even when all agents know the asset will eventually be worthless—under asymmetric information (Conlon, 2004).

going back to 1812 and vintages going back to 1892 for Haut Brion, 1884 for Margaux, 1865 for Latour, and 1853 for Mouton. Examples beyond the first growths include vintages going back to 1825 for Yquem and 1849 for Ausone.[16] All of these very old wines are valued at thousands of dollars.[17] Furthermore, the wineries under study (especially the five first growth) are centuries old and, as I use a rolling ten-vintage portfolio, I am effectively testing for a bubble in the "brands" (e.g. Lafite, Latour, etc.) rather than a specific wine-vintage pair. In the case of Lafite and Latour, for example, their names have been prominent since the mid-17th and early-18th century, respectively. There is no reason for consumers to expect these names to disappear in the foreseeable future.

3.2. **Two Nearly Identical Assets.** Assumption (i) of the generalized model-free test asserts that the two assets considered in the test are nearly identical. Most importantly, their fundamental values must be related so that the difference in fundamental values is bounded. This assumption is inherently qualitative because fundamental values are latent. I will test for a bubble in Lafite using two separate comparisons: (i) an equally weighted composite asset of the four other first growth Bordeaux wines and (ii) a composite asset formed from up to 46 other fine Bordeaux wines using synthetic control methods (for details see data description). The two composite assets end up being formed from different combinations of five wines: the four first growths, Latour, Margaux, Mouton, and Haut Brion, plus Ausone. I will consider eight dimensions on which wines may be differentiated: wine color, region, varietal composition (grapes used to make the wines), style, scarcity, price tier, historical classification, and expert review scores. In all respects the five wines are similar to Lafite. Ausone is the least similar, but its price behavior was one of the two wines chosen from 46 in the data-driven synthetic control to match the pre-2009 price behavior of Lafite.

All six wines are red wines produced in the Bordeaux region of France. The vineyards of the other five are less than 40 miles from Lafite: Ausone is the furthest at 38 miles, Haut Brion the second furthest at under 30 miles, and Margaux the third furthest at less than 14 miles. All share a maritime climate and are close to the banks of the Gironde estuary or one of its two main offshoots. All six wines are produced almost exclusively from a blend of two or more of the three grapes which define Bordeaux (cabernet sauvignon, merlot, and, to a lesser extent, cabernet franc). All six wines embody the Bordeaux style: rich, full-bodied, tannic wines aged for extended periods

[16] `www.wine-searcher.com` on Aug. 16, 2018.

[17] Two other notable examples of very old (perhaps dating back to the 18th century) and valuable wines are the controversial Jefferson bottles (see: `https://www.newyorker.com/magazine/2007/09/03/the-jefferson-bottles`), and champagne found in a Baltic Sea shipwreck (see: `https://www.theguardian.com/world/2010/jul/18/champagne-found-sea-oldest-vintage`).

prior to release in new oak barrels. Production volumes are much higher in Bordeaux than other famous French regions such as Champagne and Burgundy. Vintages of each wine would be readily available in luxury wine shops or at auction. On grounds of color, region, varietal composition, and scarcity, all the wines are similar.

To examine prices, consider the ranking of ten-vintage average prices in 50 top Bordeaux wines at the first and last points of my sample (June 2005 and January 2017), before and after the possible bubble. Ausone is the fourth most expensive wine in June 2005 and the fifth most expensive in January 2017. Haut Brion is rank 11 and then rank 10; Lafite jumps from 8 to 4; Latour 9 to 3; Margaux drops 6 to 8; and Mouton from 10 to 9. In terms of tiers, two microproduction wines (whose prices are driven more by scarcity), Le Pin and Petrus, clearly form a top tier: their prices are at least three times more expensive than the second tier. The six wines I use are all in the second tier in both periods. The second tier is considerably higher than the third tier. For example, the 12th most expensive wine in June 2005 is Mission Haut Brion at £40, considerably lower than Haut Brion at £62. In January 2017, the next three highest prices after Haut Brion (£337) are also considerably lower: Angelus (£229), Pavie (£195), and Yquem (£176). Overall, these wines are far more expensive than most Bordeaux. More importantly with respect to assumption (i), the six wines are clearly within the same price tier before and after the possible Lafite bubble.

These wines have long been considered to be of equivalent quality. Although definitions of wine quality are nebulous, I will consider two enduring markers: historical classifications and expert review scores. Lafite, Latour, Margaux, Mouton, and Haut Brion are the five wines in Bordeaux with the highest possible classification, known as *premier cru* or first growth. The 1855 Classification of Bordeaux identified 61 top properties within what was, at the time, unequivocally France's top wine growing region. The classification divided the properties into five hierarchical rankings based on price and reputation, with rankings conducted by a committee of intermediary suppliers. The intermediaries were chosen because they dealt with all producers and wholesalers, and so were informed and considered sufficiently impartial. Wines within a ranking level were considered by the committee to be of equivalent quality. In 1855, four wines were ranked as first growths: Lafite, Latour, Margaux, and Haut Brion. The 1855 classification has been continuously held in high esteem and has been modified only twice since: one addition to the fifth growth category in 1855, and Mouton was upgraded from second to first growth in 1973. The connection between the five first growths is undoubtedly strong. Ausone was also considered a superstar Bordeaux

Table 1. Average Standardized Expert Review Scores by Producer.

	(1) Pre-2009 Reviews	(2) All Reviews
Lafite	94.879***	94.457***
	(0.456)	(0.373)
Ausone	−0.627***	−0.061
	(0.125)	(0.072)
Carruades Lafite	−6.836***	−6.218***
	(0.208)	(0.108)
Haut Brion	−0.691***	−0.248***
	(0.138)	(0.068)
Latour	0.002	0.342***
	(0.160)	(0.071)
Margaux	0.069	0.214***
	(0.121)	(0.062)
Mouton	−1.267***	−0.749***
	(0.126)	(0.059)
Vintage FE	✓	✓
Review Dates	1996–2008	1996–2018
N	1,148	3,105
R^2	0.815	0.855

Note: Table shows estimated average standardized expert review score from regressing score against producer and vintage fixed effects. The dependent variable is a standardized expert review score, provided by Global Wine Score (Cardebat and Paroissien, 2015). The critic's score for a wine is standardized across critics by evaluating the score on the critic's empirical c.d.f. and then aggregated by averaging standardized scores across critics. One observation is the standardized expert review score for a given wine-vintage pair at a particular point in time. Column (1) focuses on scores prior to 2009 using data from January 1996 to December 2008 (vintages 1992–2006). Column (2) is the full sample of scores using data from January 1996 to December 2018 (vintages 1992–2016). I use vintages going back to 1993 because this is the oldest vintage in the ten-vintage portfolio in June 2005 for the main data set. Robust standard errors reported in parentheses. Significance levels: * - $p < 0.10$, ** - $p < 0.05$, and *** - $p < 0.01$.

wine long before 2005. The 1855 Classification only considered Bordeaux wines from the "Left Bank" of the river which bisects Bordeaux's wine regions, whereas Ausone is located on the Right Bank.[18] Ausone was one of four wines ranked at the highest tier in 1955, 1969, 1986, 1996, and 2006 classifications in its Saint-Émilion sub-region of Bordeaux. Since at least 1973, the six wines have been considered amongst the highest quality wines made in Bordeaux.

Finally, I consider expert review scores, which vary from one vintage to the next, map production circumstances into a measure of quality (implicitly including changes in production techniques, management, etc.), and can be highly influential on prices (Cardebat and Figuet, 2004; Ali, Lecocq,

[18]Le Pin and Petrus are also located on the Right Bank wines and so were excluded from the 1855 Classification. Also, Le Pin was formed only in 1979 and Petrus was not held in high regard until the 1960s.

and Visser, 2008). To do so, I collected standardized scores from Global Wine Score (GWS) (Cardebat and Paroissien, 2015), specifically the GWS Evolution score for each wine and for all vintages in the main data set (1992 to 2016) from January 1996 to December 2008. GWS standardizes raw scores on a per critic basis (using the empirical c.d.f. of an individual critic's rankings) and then aggregates by averaging across critics. Table 1 reports the estimated average score of each wine from regression scores against producer and vintage fixed effects. In addition to the six wines, I also include scores for Lafite's second wine, Carruades Lafite. Overall, wine and vintage fixed effects explain over four-fifths of the variation in scores in both models. Model (1) report results from a subset of scores conducted prior to a possible bubble (January 1996 to December 2008), while Model (2) repeats the estimates on a full sample of reviews spanning from January 1996 to December 2018. Across models from (1) to (2), the estimates are quite stable, with no dramatic changes: Ausone, Haut Brion, and Mouton get a half-point closer to Lafite, while Latour and Margaux get a third of a point higher.

Consider the estimates in Table 1 in the context of whether or not the six wines can be fairly called nearly identical to Lafite. In an average vintage, Lafite scores over 94 points (i.e. the average critic gives 94 percent of their reviews lower scores). Immediately, we see Carruades Lafite is not in the same class as the other six wines: an average score of only 88 compared to all the other wines being within 1.3 points of Lafite. The 1.267 point deficit in pre-2009 reviews of Mouton drops to a less than 0.75 point difference over the full sample. With the exception of pre-2009 reviewed Mouton, all five wines in both samples are within one point of Lafite over many vintages. Wine reviews are presented in integer values, so while differences may be statistically significant, I would argue their magnitudes are quite small. Taken together, the results of Table 1 suggest the six wines I consider are nearly identical in quality, as measured by expert review scores. Overall, Lafite is nearly identical to Ausone, Haut Brion, Latour, Margaux, and Mouton on eight categories which differentiate wines. Thus, I am comfortable assuming the assets are sufficiently similar so as to bound the difference in their fundamental values.

3.3. **Data.** For the bubble test, I use data provided by the London International Vinters Exchange (Liv-Ex) from June 2005 to January 2017. Liv-Ex collects daily transaction records on the price of fine wines exchanged through global secondary markets. Price updates are collected through over 400 member distributors and auction houses around the world, accounting for approximately 35,000 transactions worth \$30 million daily. These indices, which aim to reflect current and past market

Table 2. Component wines of the Fine Wine 500 for Jan. 2017, Wines 1–250.

Liv-Ex Group	Components	Price (£/750mL) Jun. 2005	Mar. 2011	Jan. 2017
First Growth 50 (FW50)	Lafite Rothschild	74.73	797.54	489.65
	Margaux	78.72	390.28	356.30
	Mouton Rothschild	62.42	368.49	349.88
	Haut Brion	61.82	325.47	337.20
	Latour	74.34	478.27	494.96
Left Bank 200	Beychevelle	14.47	54.14	63.54
	Calon Segur	15.95	34.08	56.73
	Cos d'Estournel	21.66	74.36	95.09
	Ducru Beaucaillou	22.77	67.97	102.16
	Duhart Milon	10.98	79.63	49.42
	Grand Puy Lacoste	15.22	28.55	37.81
	Gruaud Larose	16.53	30.21	39.59
	Haut Bailly	12.65	33.24	59.01
	Leoville Barton	19.01	42.99	50.15
	Leoville Las Cases	35.24	104.14	111.87
	Leoville Poyferre	15.00	46.51	59.71
	Lynch Bages	21.61	69.67	80.00
	Mission Haut Brion	40.38	185.96	197.22
	Montrose	19.56	61.63	80.87
	Palmer	29.85	98.10	149.78
	Pape Clement	–	–	77.98
	Pichon Baron	17.32	57.35	76.14
	Pichon Lalande	26.76	69.40	75.50
	Pontet Canet	12.51	42.60	70.36
	Smith Haut Lafitte	–	–	67.45

Note: Data from the London International Vinters Exchange. Reported price is the average price of the ten most recently released vintages at that point in time (e.g. for June 2016–July 2017, the ten most recently released vintages are 2005 to 2014 for most wines). Prices are real in 2015 terms deflated with UK CPI. All wines are red except Sauternes, which are dessert wines.

conditions, are reported on professional data vendor services including Bloomberg and Thomson Reuters.

Liv-Ex provided monthly mid-point prices for the component wines and vintages used to construct their *Fine Wine 50* (FW50) and *Bordeaux Fine Wine 500* (FW500) indices. An exhaustive list of the component wines and vintages (used in the Jan. 2017 index) is provided in Tables 2 and 3. The FW50 is composed of the most recently released ten vintages of the five first growth Bordeaux wines. The FW500 is ten vintages of all of the wines in the FW50 as well as five 'super brands'

Table 3. Component wines of the Fine Wine 500 for Jan. 2017, Wines 251–500.

		Price (£/750mL)		
Liv-Ex Group	Components	Jun. 2005	Mar. 2011	Jan. 2017
Right Bank 50	Petrus	324.81	1336.21	1722.72
	Ausone	87.56	613.69	471.94
	Cheval Blanc	79.49	302.58	368.86
	Pin	348.86	1134.73	1759.87
	Lafleur	95.26	363.04	416.02
Right Bank 100	Angelus	31.87	121.90	228.65
	Pavie	30.36	116.59	195.86
	Clinet	–	–	67.58
	Fleur Petrus	26.57	102.91	126.34
	Evangile	35.18	82.47	103.39
	Conseillante	24.31	65.22	78.34
	Vieux Chateau Certan	26.80	64.59	109.50
	Clos Fourtet	13.65	34.74	69.00
	Troplong Mondot	15.15	46.36	65.58
	Eglise Clinet	41.55	107.51	137.46
Sauternes 50	Yquem	77.20	171.19	175.68
	Climens	26.82	59.82	37.63
	Coutet (Barsac)	12.57	16.86	20.32
	Suduiraut	15.28	22.84	28.02
	Rieussec	20.38	26.28	24.01
Second Wines 50	Carruades Lafite	14.10	272.14	165.67
	Petit Mouton	24.82	93.74	140.46
	Forts Latour	18.89	157.20	126.83
	Pavillon Rouge	17.47	107.71	106.95
	Bahans/Clarence Haut Brion	14.57	76.46	65.29

Note: Data from the London International Vinters Exchange. Price is the average price of the ten most recently released vintages at that point in time (e.g. for June 2016–July 2017, the ten most recently released vintages are 2005 to 2014). Prices are real in 2015 terms deflated with UK CPI. All wines are red except Sauternes, which are dessert wines.

from the Right Bank, five second growth brands, five Sauternes (dessert wine) brands, and 30 well-regarded brands from the greater left (20 brands) and right (10 brands) bank regions. All records in the FW50 and FW500 are complete except for three wines, which do not enter the FW500 until 2016 (Pape Clement, Smith Haut Lafitte, and Clinet) and are hence dropped. Hereafter I exclude the ten vintages of my superstar wine, Lafite, from the FW50 and thus refer to it as the FW40. Similarly, I exclude the FW50 from the FW500 and refer to it as the FW450. The synthetic control considers all wines in the FW500 except for Lafite, Pape Clement, Smith Haut Lafitte, and Clinet. Due to sample availability, the period of consideration is July 2005 to January 2017 ($T = 140$).

Figure 2 plots the prices (£/750mL) of the individual component wines over time in a number of different ways. The top panel plots Lafite against the four other first growth wines. Lafite is clearly the most expensive of the five wines, though after 2014 Latour begins to catch up. The middle panel plots the prices of the FW450 component wines, with the eight most expensive component wines highlighted with colors and all other wines in gray. Three wines in the FW450—Petrus and Le Pin, and at times Ausone—are priced at or above the level of the first growth wines. The bottom panel plots the individual components of the FW450 in index terms (standardized to a value of 100 for July 2005) where the potential "Lafite bubble" is evident even in Lafite's second-tier wine, Carruades Lafite. All values are in real terms (2015 base), deflated using the monthly UK consumer price index.

By most standards the wines are expensive: in January 2017 the average price for wines in the FW500 was £203 (roughly \$317 at 2015 exchange rates), while the median wine cost £93.78 (roughly \$146). At its maximum, the price of Lafite increases more than ninefold relative to July 2005, compared to nearly fivefold for FW40 and over threefold for FW450. In terms of downside, no prices drop below their value in July 2005. The coefficient of variation in prices for Lafite (0.425) is considerably higher than that of FW40 (0.309) and FW450 (0.267). Returns for these wines were also quite high: holding Lafite provided, on average, an annualized return of over 15 percent ($1.012^{12} \approx 1.154$). The average returns were 12.7 percent for FW40 and 11.4 percent for FW450. Holding a rolling portfolio of Lafite whilest shorting FW40 or FW450 would result in monthly returns of 2.0 or 5.3 percent (26.8 and 85.8 percent annualized), respectively—large in magnitude by any standard.

To implement the model-free test, I convert all prices to index terms with a value of 100 in July 2005. Lafite is asset one in (3), with price denoted P_t^*. I use two specifications for asset two. First, an equally-weighted index of FW40, the other four first growth Bordeaux wines: Margaux, Mouton, Haut Brion, and Latour. These four wines have long been considered to be of equivalent quality to Lafite and have a long history of roughly equivalent prices.

Second, I use synthetic control (Abadie and Gardeazabal, 2003; Abadie, Diamond, and Hainmueller, 2010, 2015) to construct a composite "synthetic Lafite." Consider the bubble-free period $\boldsymbol{P}_0^I$ as the analog to a preintervention period. Weights are assigned to K other Bordeaux wines

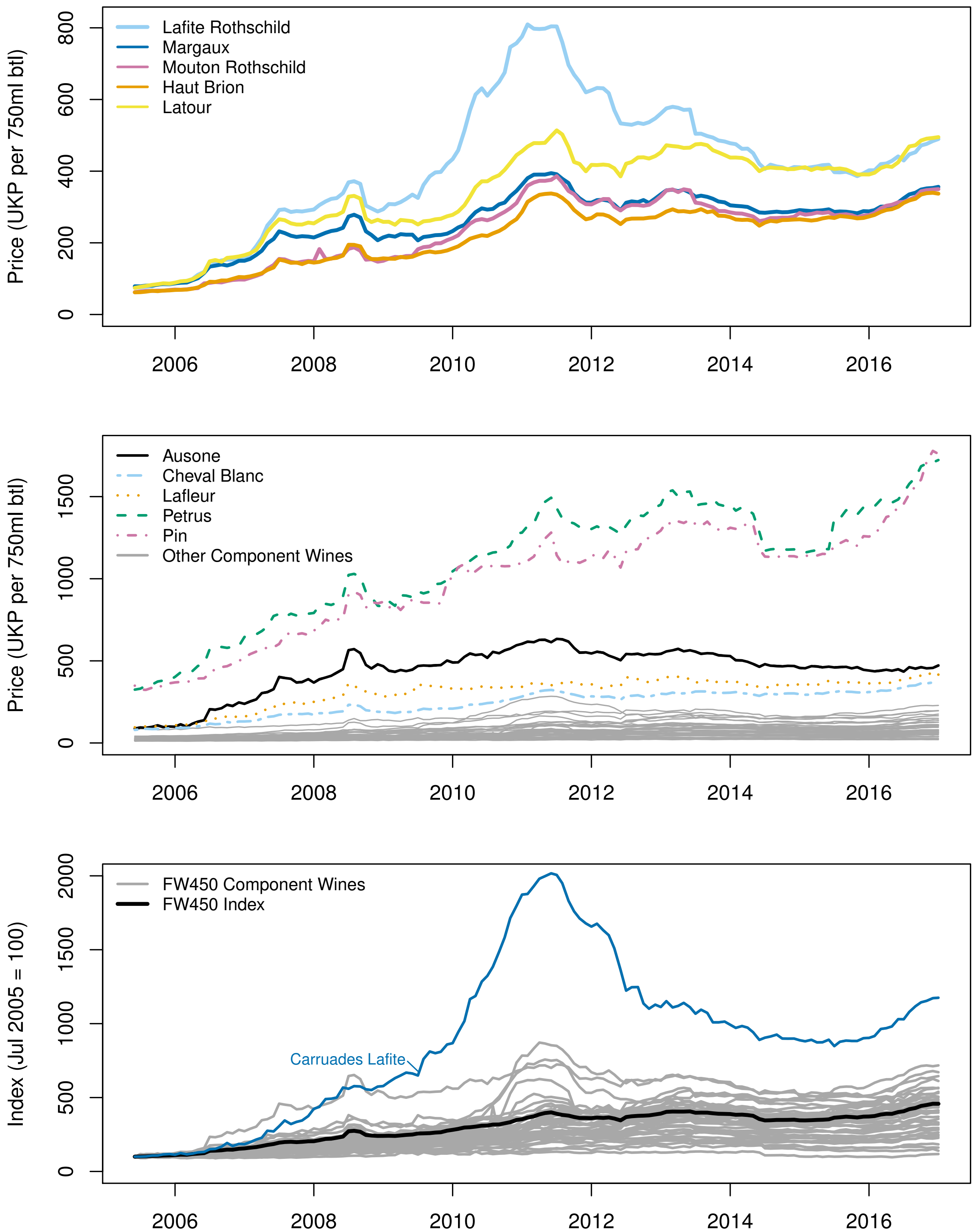

Figure 2. Top: prices of the FW50 components. Middle: prices of the FW450 (FW500 exc. FW50) components. Bottom: index values for the FW450 components (Base: July 2005 = 100).

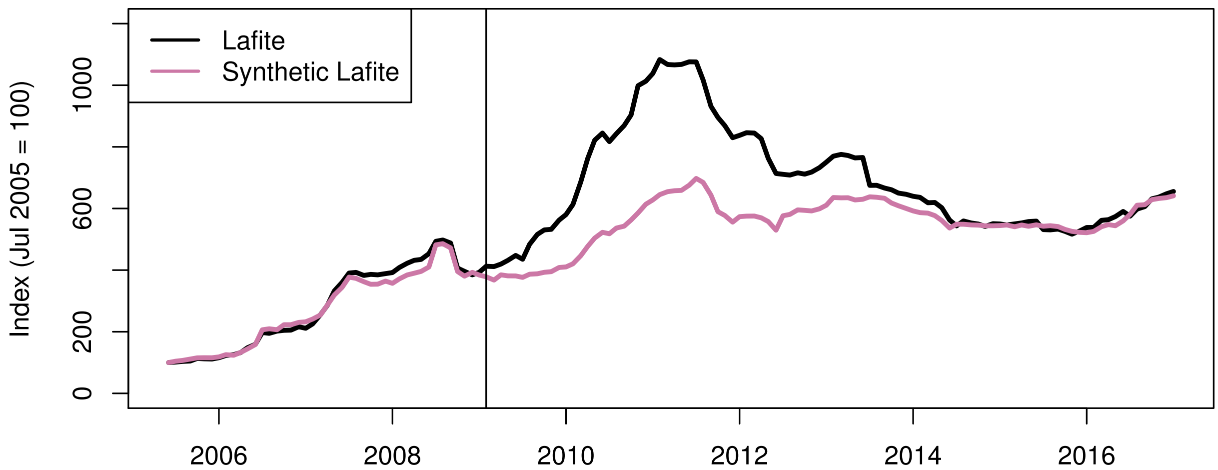

Figure 3. Synthetic Lafite constructed with synthetic control methods based on July 2005–Jan. 2009 pre-period.

to minimize the mean squared prediction error in the bubble-free period. That is, select weights $w_1, \ldots, w_K$ to minimize $T_1^{-1} \sum_{t=T_0}^{T_1} \left(P_t^* - \sum_{k=1}^{K} w_k P_{kt} \right)^2$ subject to $0 \leq w_k \leq 1$ for $k = 1, \ldots, K$ and $w_1 + \cdots + w_K = 1$. Intuitively, the method uses data-driven weights to create a series that mimics the behavior of Lafite prices during the bubble-free period. The $K = 46$ wines used to estimate the synthetic Lafite are the FW500 wines excluding Lafite and the three wines which do not report a complete price history. Figure 3 presents the synthetic control estimate of Lafite based on a July 2005–Jan. 2009 pre-period. The robustness checks include different specifications on the end of the bubble-free pre-period. For the main specification, the synthetic control procedure selects nonzero weights on two wines—19.5 percent Ausone and 80.5 percent Latour—which Figure 3 shows provides an excellent representation of Lafite in the pre-period (the average monthly prediction error is only 2.9 units relative to an average index value of 281.1).

3.4. **Plausibility of Alternative Explanations.** Before proceeding to the test, it is worth quickly addressing alternative explanations for the surge in Lafite prices early-2009 and mid-2011. The ten-vintage rolling portfolio gives any one vintage ten percent weight in the price, so it is relatively difficult for one vintage to dominate the price. One exception is when the portfolio transitions and the new vintage (three years old) replaces the oldest vintage (13 years old) between May and June: a much more expensive new vintage replacing the old vintage could create a discontinuous jump in the ten-vintage price. Examining the transition points, Figure 4 makes very clear there are no dramatic price changes at transition points and, in fact, the prices of the vintages leaving the portfolio were higher than the ones entering in 2009 and 2010. Furthermore, examining the

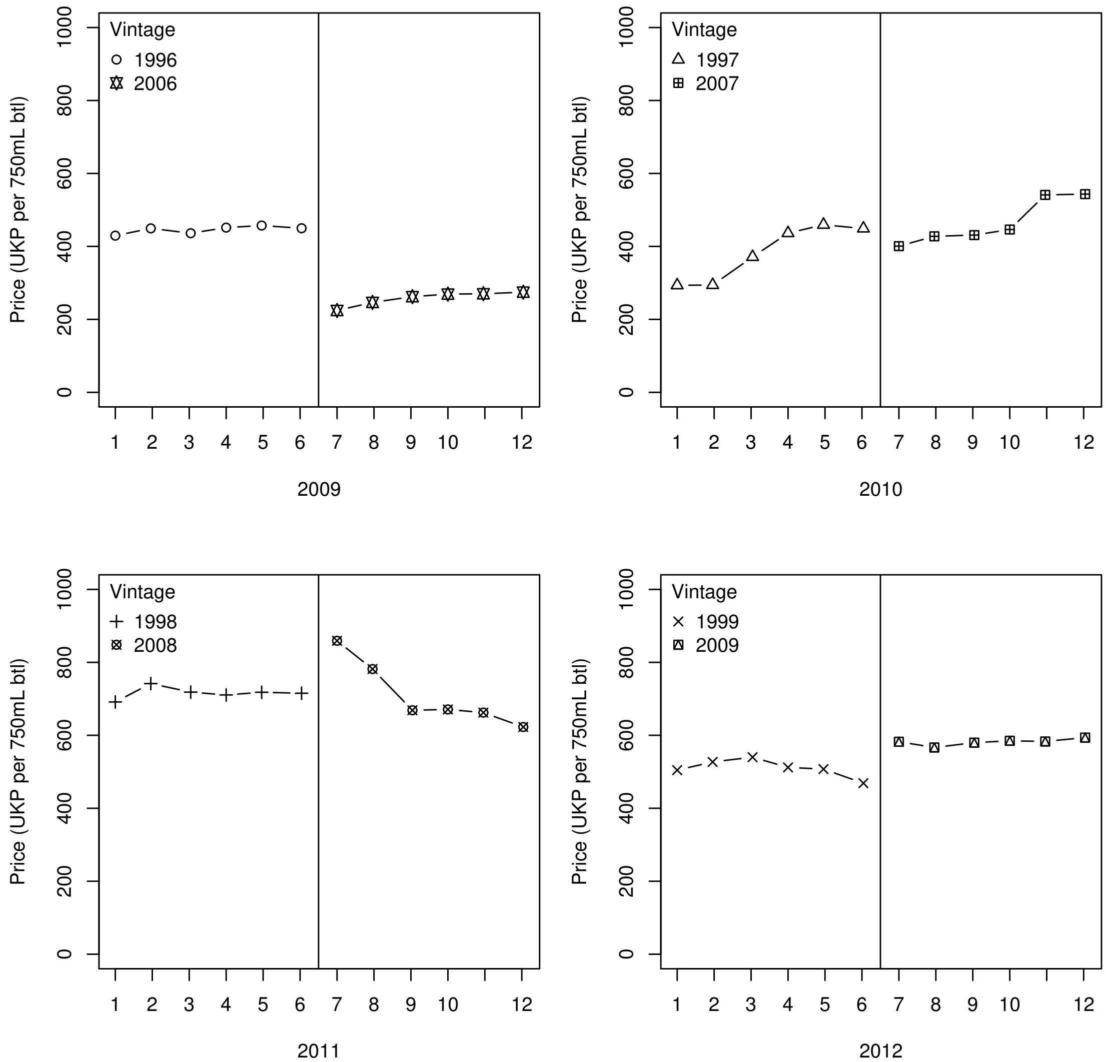

FIGURE 4. Prices of Lafite when ten-vintage portfolio drops oldest vintage, 2009–2012.

individual vintage prices from 2009 to 2012 in Figure 5, they appear to move together very closely, implying no one vintage dominates the ten-vintage portfolio. As a result, any alternative explanation for a price surge based on a single vintage shock—such as production shock, stock exhaustion, or release of exceptional expert review score for one vintage—is clearly inconsistent with the data.[19]

[19]Over and above the construction and use of the ten-vintage average, there are a number of counterarguments to a single vintage production or information shock causing the price changes. For example, if a Lafite production shock caused the price increase, then we would expect the detection to be stronger for less similar wines. On the contrary, in the robustness check where the second asset is the FW450 index, the case for a bubble is stronger comparing Lafite to the FW40 than the FW450. The latter includes more diverse production conditions (white wines and red wines from the Right Bank) and wineries that tend to be further away from Lafite (for example, Petrus is 30 miles, Angelus 31 miles, and Yquem 60 miles). Also, in general information on vintage quality (including expert reviews) is marketed long before a vintage is released and internalized into prices on the *en premier* futures market prior to

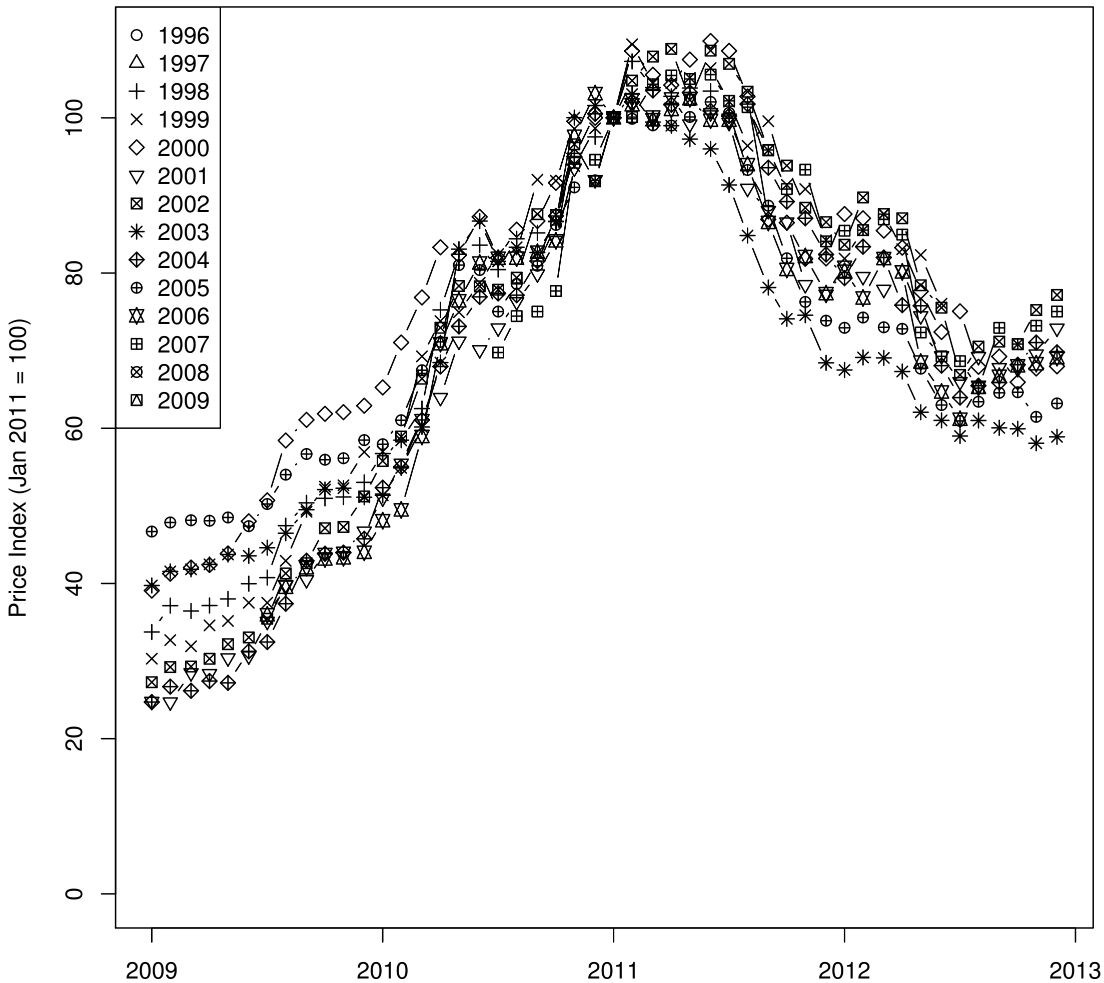

FIGURE 5. Lafite price indices by vintage during 2009–2012, showing no one vintage dominated (January 2011 = 100).

The set of feasible alternative explanations for the observed price surge is then limited to a Lafite-specific shock common to all ten vintages in the portfolio. A Lafite-specific, ten vintage wide surge in the demand shock could be a plausible alternative explanation if not for the extremely fast rate at which prices increased (i.e. more than 3.26 standard deviations). The obvious causes of a sudden, large demand shock clearly do not apply to Lafite: for example, its substitutes (i.e. the five other wines) were widely available on the market and there were no instantaneous changes to expected

the three year waiting period before it enters the portfolio. Thus, wine- and vintage-specific information updates are likely to be incremental when the vintage is released after three years.

future supply that might dramatically alter convenience yields. It would take a considerable month-to-month jump in the number of market participants or their wealth to induce the price increase. Perhaps such a shock would be plausible if the market were tiny, but exchange was conducted through a well-established and widely-accessible network including global entities such as Sotheby's and Christie's auction houses. Further, it is implausible that a shock of that magnitude and nature would be isolated to demand for Lafite and not other superstar wines. Even if one asserts that typically stable, well-behaved preferences can experience large *or* sudden shifts—an argument which conflicts with an axiomatic foundation of economic analysis since at least Stigler and Becker (1977)—the preference shift would need to be both large *and* fast (and exclusive to Lafite). When the upper bound in the test π_t^* is time invariant, these arguments have merit, but not so when π_t^* is time varying: the rolling window can account for large and gradual shifts, as well as small and fast ones, without confounding a preference shift and bubble.

A ten-vintage shortage is even less plausible. While its annual production is a closely guarded secret, Lafite is one of the largest winery estates in Bordeaux (possibly the largest) and produces roughly 210,000 standard 750mL bottles of its first wine each year and more of Carruades de Lafite (Johnson and Robinson, 2007). The oldest vintages in the ten-vintage portfolio are 13 years old and in practice, readily available for purchase (and distributed amongst many secondary market sellers).[20] Further, if stock exhaustion were to occur in all ten vintages, it would have taken longer than 18–24 months for prices to begin dropping: eight of the ten vintages that had approached exhaustion remain in the portfolio. Finally, any immediate, adverse shock to the availability of all ten vintages (i.e. policy shock) would not be Lafite-specific (especially given the large and active secondary market). In sum, I would argue that alternative explanations for the price surge based on the mechanics of the data or contemporaneous market conditions are not persuasive.

4. Results

I present results comparing Lafite to an index of the four other first growths, Lafite−FW40, and a data-driven Synthetic Lafite. The results are divided into three subsections as follows. In the first subsection, the concentration inequality is calculated based on a fixed bubble-free pre-period of July 2005–December 2008 (inclusive). In the second subsection, the concentration inequality is

[20]For example, June 5, 2019 search for the current 13 year old vintage, 2006, found 420 results on `https://www.wine-searcher.com/find/lafite+2006/1/united+states?Xsort_order=e`. Older vintages are also easy to find: a similar search for the 1996 vintage found 338 results.

calculated using a 24-month rolling window, allowing the upper bound π_t^* to change over time. I find bubbly periods in both, which the third subsection shows is robust to a variety of checks.

4.1. **Model-Free Test with Time-Invariant Bounds.** The first set of results uses the pre-2009 bubble-free pre-period, $\boldsymbol{P}_0^I$, to calculate the π_t^* bound and assumes the bound is constant throughout the sample. This allows us to identify, conditional on the validity of applying π_t^* from $\boldsymbol{P}_0^I$ to observations from January 2009 onwards, the beginning and end of a bubbly period. The panels in Figure 6 summarize the bubble-test at the five and ten percent level for Lafite$-$FW40 (top) and Lafite$-$Synthetic Lafite (bottom). Recall control of the false detection rate with the finite sample Cantelli-Cheybychev inequality is not exact: in this case with 43 pre-2009 observations, control is actually at 0.0444 for the five percent level and 0.0890 for ten percent. For Lafite$-$FW40, the test detects an uninterrupted bubble beginning in December 2009 at the ten percent level and March 2010 at the five percent level. The bubble is detected for 42 months at ten percent and 26 months at five percent. Using synthetic Lafite, the results are even stronger: the anomaly in Lafite prices is first detected in August 2009 and lasts until June 2013, a total of 47 months, all at the five percent level.

4.2. **Model-Free Test with Time-Varying Bounds.** Now I allow the fundamental value relationship between the two products to be time-varying through the use of a rolling window. Recall a rolling window approach is common to the literature, especially for predictive purposes (e.g. Phillips, Wu, and Yu, 2011) but necessarily involves a trade-off: by allowing the parameters used in the concentration inequality to update, B_t may be greater than zero but, if it is not sufficiently large to exceed the π_t^* threshold, it will be difficult to detect a bubble in the subsequent period. Thus, the probability of failing to detect a bubble—even when a bubble is present—is higher with the rolling window approach. For similar reasons, the rolling window can only detect the start of the bubble (not its end or duration). As a baseline, I use a 24-month rolling window. The results are again best summarized in a figure, where areas shaded in gray indicate detection of the beginning of the bubble at the five (dark gray) and ten (light gray) percent levels. With the finite-sample concentration inequality and 24 observations, control is actually at 0.0385 for the five percent level and 0.0769 for the ten percent level.

Figure 7 illustrates the bubble detection results for both price comparisons. In the Lafite$-$FW40 comparison illustrated in the top panel of Figure 7, the method detects a bubble in six months at

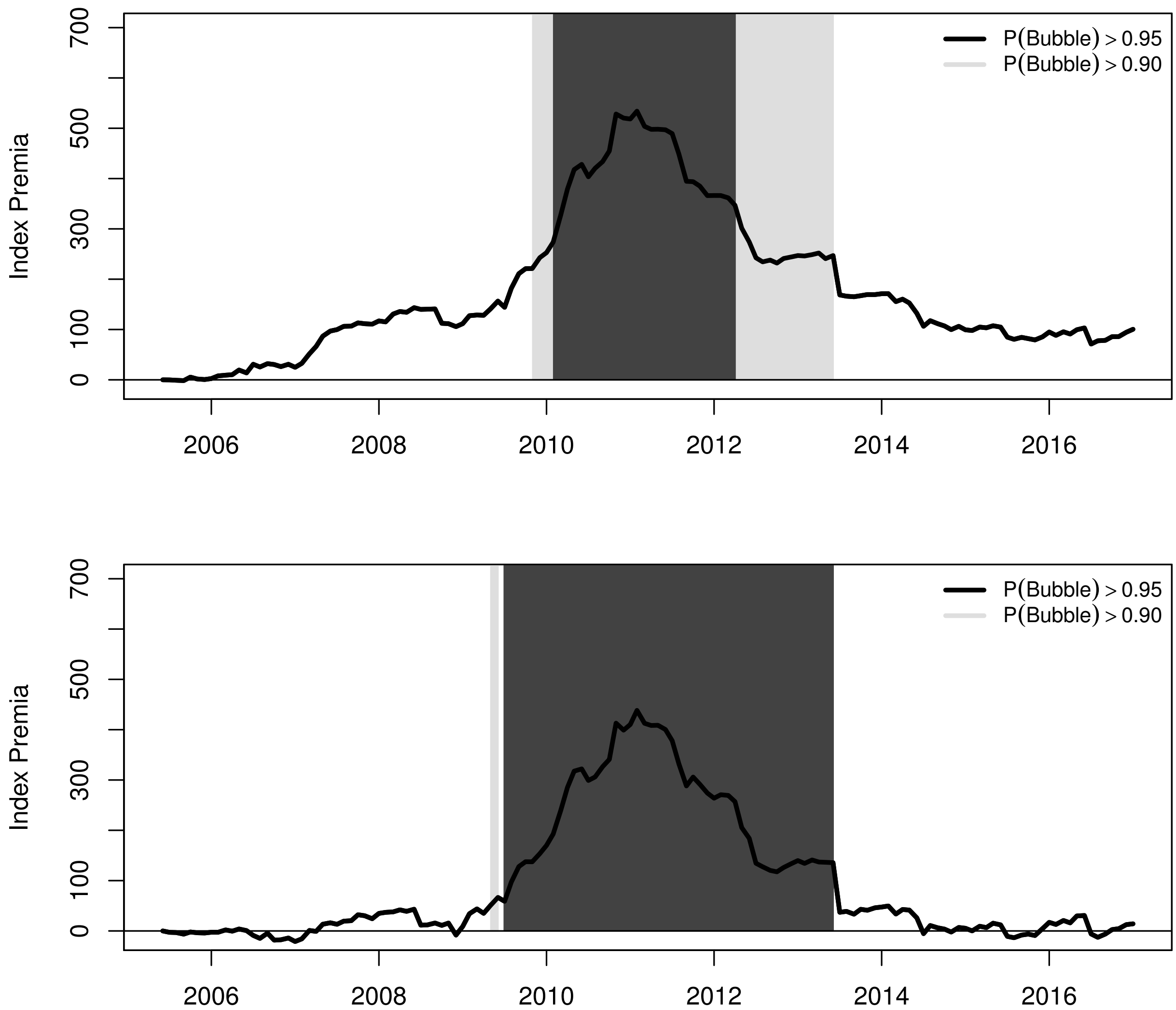

FIGURE 6. Bubble detection for Lafite with time invariant concentration inequality. Top panel: Lafite versus the four other first growth wines (FW40). Bottom panel: Lafite versus synthetic Lafite.
Note: Control attained at critical values with probability: 0.0444 and 0.0890 (c.f. Appendix A).

the ten percent in two groups: August 2009–October 2009 and March 2010–May 2010. Both sets seem to accurately predict the start of the bubble, which is all we can ask of the rolling window because in subsequent periods large anomolies are incorporated into its parameters, so subsequent detection becomes less likely (even if a bubble is present). Given the possibility of false detection (analogous to type I error), it is important to note both detection sets are three sequential periods long. Each sequential period of detection implies the possibility of a random false detection is less likely. Roughly speaking, the probability of randomly finding three sequential periods of detection at the ten percent level is less than $(0.1)^3 = 0.001$. Thus, we again have compelling evidence of a bubble in Lafite−FW40 during 2009–2010, even with time-varying parameters which we know

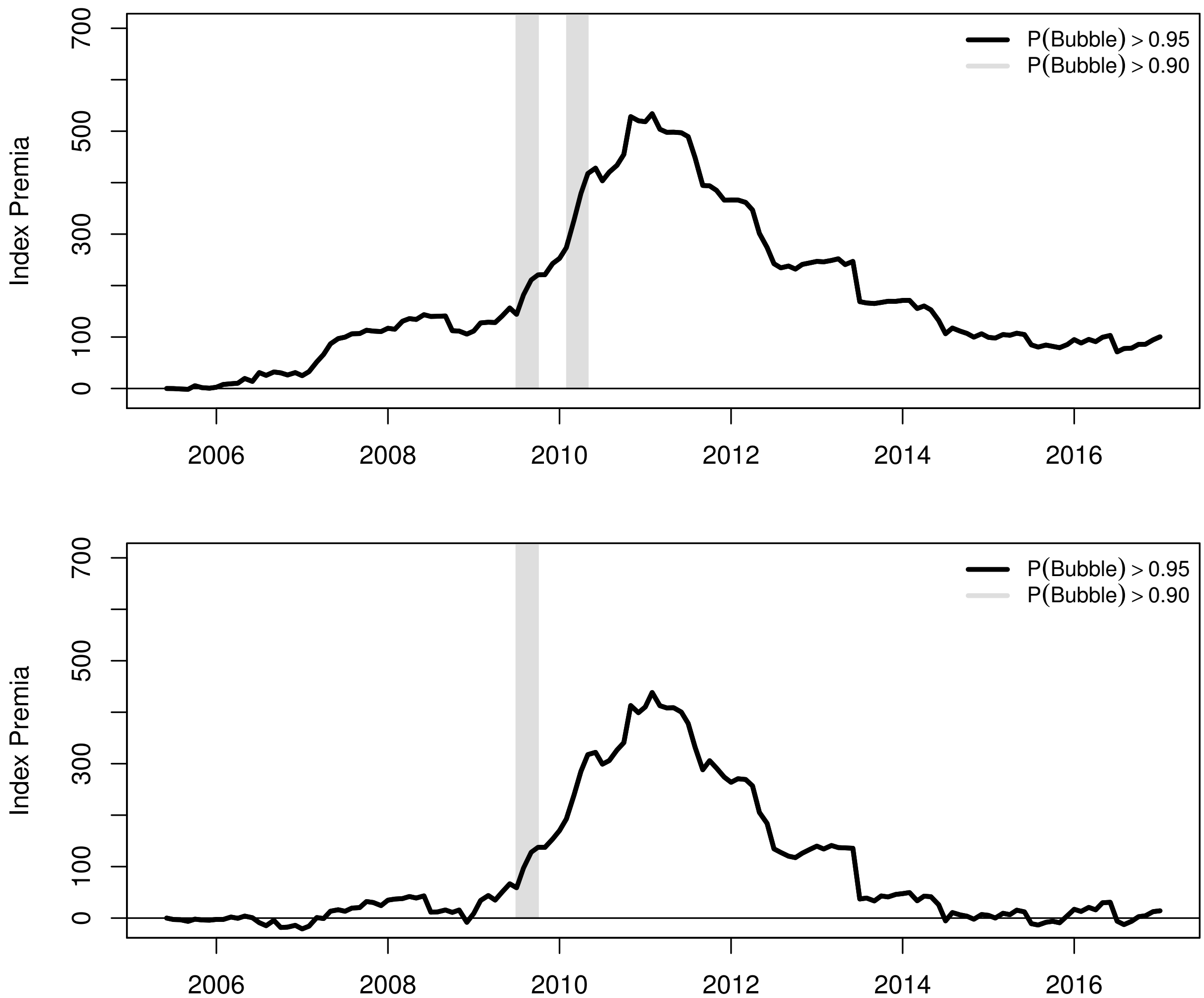

Figure 7. Bubble detection for Lafite with time varying concentration inequality. Top panel: Lafite versus the four other first growth wines (FW40). Bottom panel: Lafite versus synthetic Lafite.
Note: Control attained at critical values with probability: 0.0385 and 0.0769 (c.f. Appendix A).

reduce detection power.[21] The results are nearly identical for synthetic Lafite, illustrated in the bottom panel of Figure 7. Again, a bubble is detected in three consecutive months from August 2009 to October 2009 all at the ten percent level. The bubbly periods detected with the synthetic Lafite correspond exactly to the first detection set in Lafite−FW40.[22]

Before proceeding to the robustness checks, it is worth noting the time-invariant and time-varying results are important for at least three reasons. First, the early periods when an anomaly is detected

[21]Figure 8 in Appendix B illustrates the implied probability of a bubble under the population concentration inequality rather than using a discrete detection threshold. The finite sample inequality is not useful for this illustration because its steps are too large.

[22]Bubbly and non-bubbly observations receive the same weight in the rolling window calculations, which makes it even more difficult for the test to detect a bubble in sequential periods or when prices rise quickly, but just below the threshold. Future work could address this by down-weighting (or trimming) observations in proportion to their probability of a bubble or using a sup test.

correctly anticipate a large run-up in the relative price of Lafite, followed by a subsequent crash—exactly what would be expected in a bubble. Second, Lafite is compared to two peer groups: four other first growth wines (the FW40), which have long been considered of equivalent prestige and quality to Lafite, and a synthetic Lafite based on data-driven synthetic control methods using other fine Bordeaux wines as possible controls. Third, I am finding strongly significant, model-free direct evidence of a bubble. I find this evidence despite using the higher than conventional statistical thresholds given by the concentration inequality. Even with a time varying concentration inequality which reduces detection power, I find evidence of a bubble in Lafite using a model-free direct bubble test. The detection of three consecutive bubbly periods implies the probability of randomly detecting the anomalies observed in the data is much less than the nominal level of ten percent.

4.3. **Robustness Checks.** I conduct four robustness checks. First, I address the possibility of nonexchangeable $P_t^* - P_t$ by reconducting the tests after detrending and first-differencing the data. Second, I examine the use of estimates for the mean and variance in the concentration inequality using a bootstrap. Third, I examine different lengths of the bubble-free period. Finally, I conduct the bubble test on different comparison groups.

4.3.1. *Exchangeability of the Price Difference Series.* The accuracy of π_t^* depends on the exchangeability of the $\boldsymbol{P}_0$ sequence used to calculate it, where a trend or autoregressive process specific to one of the two assets could bias π_t^* downwards. To account for a possible deterministic trend, I estimate a pre-2009 linear trend line and reconduct the test on the detrended residuals (see Figure 9 in Appendix B). To account for a possible autoregressive process, I reconduct the test after first-differencing $\boldsymbol{P}_0$ (see Figure 10 in Appendix B).

The detection results of the model-free test are robust to being conducted using detrending residuals. With a time-invariant upper bound and detrending residuals, a continuous bubble is detected in FW40 from February 2010 to March 2012 at the 10 percent level (March 2010 to November 2011 at five percent) and in synthetic Lafite from August 2009 to June 2013 (September 2009 to July 2012). The first bubble detection is three months later for FW40 at ten percent, the same time for FW40 at five percent, the same time for synthetic Lafite at ten percent, and one month later for synthetic Lafite at five percent. Not surprisingly given the upward trend during pre-2009, the end of the bubble period is identified earlier in the robustness check: for example, at the ten percent level, 26 versus 42 months for FW40 and 21 versus 26 months at five percent. The trend coefficient on Lafite−Synthetic Lafite is practically zero and there is no change in the end of the bubble identified by the test. With the rolling window and detrended residuals, the test detects a March to May 2010 at the ten percent level in both FW40 and the synthetic control—the same periods detected in the main results.

The detection results are also robust to being conducted after taking a first difference i.e. $(P_t^* - P_t) - (P_{t-1}^* - P_{t-1})$. Note, however, that we will not detect a continuous bubble block because for a given time period, the cumulative level of the deviation from its prior observations is lost. For time-invariant upper bound and FW40, five observations detect a bubble at the ten percent level: August 2009, March to May 2010, and November 2010. March, April, and November 2010 remain detected at the five percent level. For time-invariant upper bound and synthetic Lafite, six observations are detected at the ten percent level: the same five months as FW40 plus September 2009. Again, March, April, and November 2010 remain detected at the five percent level. With time-varying upper bound and FW40, a bubble is detected at the ten percent level in August 2009, March 2010, and November 2010. With time-varying upper bound and synthetic control, a bubble is detected at the ten percent level in November 2010. Recall that these two robustness checks are over and above the control for time-varying unobservable factors with common influence on both prices (e.g. parallel trends) provided by differencing P_t^* and P_t, as well as the fact that for two nearly identical assets, by definition factors which influence only one of the prices should be quite small. Taken together, these results suggest robust detection of a bubble beginning somewhere between late-2009 and the end of 2010.

4.3.2. *Estimates in the Concentration Inequality.* The finite sample Cantelli inequality (Appendix A) adapts the inequality for use in finite samples, but assumes the estimates of the mean and variance are correct. To account for sampling variability, I conduct the following bootstrap procedure. For bootstrap iteration $m = 1, \ldots, M$, generate a bootstrapped sample $\tilde{\boldsymbol{P}}_0^{(m)}$ by sampling with replacement from $\boldsymbol{P}_0$ ($\boldsymbol{P}_0^I$ for time-invariant bounds and $\boldsymbol{P}_0^V$ for time-varying bounds). Then calculate $\tilde{Z}_t^{(m)} = (P_t^* - P_t - \bar{X}^{(m)})/S^{(m)}$ where $\bar{X}^{(m)}$ denotes the sample mean and $S^{(m)}$ the sample standard deviation of $\tilde{\boldsymbol{P}}_0^{(m)}$. For each price difference observation to be tested, repeat this M =4,999 times and then reconduct the bubble-test using the qth quantile of $\left(\tilde{Z}_t^{(1)}, \ldots, \tilde{Z}_t^{(M)}\right)$.

The results of the model-free test with time-invariant bounds are completely robust to the bootstrap. In the Lafite−FW40 comparison, there are no changes in the timing of the bubble detection at the five (i.e. using $q = 0.05$) or ten percent levels. In the Lafite−Synthetic Lafite comparison, the timing of the bubble detection changes only in that it is detected one period later (September 2009 instead of August 2009) at the five percent level. Certainly, these robustness checks do not bring into question the main qualitative conclusion of a bubble in Lafite.

The results of the model-free test with time-varying bounds are also robust to the bootstrap, though not surprisingly, less overwhelmingly. In the Lafite−FW40 comparison, six periods were detected at the ten percent level. The magnitude of deviations in the first two of these periods, August 2009 and September 2009, exceed 3.26 standard deviations and thus continue to be detected with the bootstrap at the ten percent level. In the Lafite−Synthetic Lafite, the first two of three periods, also August 2009 and September 2009, remain detected at the ten percent level. Given that the shorter sequences of $\boldsymbol{P}_0^V$ used in the rolling window likely lead to higher sampling variation, that observations remain anomalous is notable. Later observations face the further challenge that their sequence of preceding observations, $\boldsymbol{P}_0^V$, definitely contain bubbly observations.[23] In turn, the design of the bootstrap will necessarily exacerbate the lower detection power problems of the rolling window, because random samples with more bubbly observations will generate lower $\tilde{Z}_t^{(m)}$ (increase in variance more than offsets increase in mean) and thus more heavily represented at the lower tail of $\left(\tilde{Z}_t^{(1)}, \ldots, \tilde{Z}_t^{(M)}\right)$. Despite these non-trivial challenges, the finding of a bubble in Lafite remains robust.

4.3.3. *Length of Bubble-Free Sequences.* I rerun the model-free test using time-invariant bounds based on 36, 40, and 48 months, whereas the main results use all observations prior to 2009 (43 months). For these sample sizes, five (ten) percent control is attained at critical values with probability 0.0263 (0.0789), 0.0476 (0.0952), and 0.0400 (0.1000), respectively. A bubble is detected at the five percent level from April 2010 to March 2012 with 36 months, March 2010 to May 2012 with 40 months, and April 2010 to April 2012 with 48 months. At the ten percent level, the test first detects a bubble in December 2009 with 36, 40, or 48 months. The 36 and 40 month pre-periods detects continuously detect a bubble until June 2013. The 48 month pre-period also detects a bubble up to June 2013, except for one intermediate observation in October 2012. Overall, the time-invariant results with different bubble-free periods are broadly consistent with those in Figure 6.

[23] Observations which no longer meet the detection criteria with the bootstrap are: October 2009 (10 percent of bootstrap iterations have $\tilde{Z}_t^{(m)} \leq 2.73$), March 2010 (2.80), April 2010 (2.85), and May 2010 (2.59) for Lafite−FW40 and October 2009 (2.62) for Lafite−Synthetic Lafite.

For the model-free test using time-varying bounds, I use rolling windows of 18, 36, and 48 months, whereas the main results use 24 months. For these sample sizes, five (ten) percent control is attained at critical values with probability 0.0500 (0.1000), 0.0263 (0.0789), and 0.0400 (0.1000), respectively. Bubbles are detected after 2008 using rolling windows of all three window lengths: August 2009 to October 2009 and March 2010 to May 2010 with 18 months, March 2010 to June 2010 with 36 months, and March 2010 to June 2010 plus November 2010 with 48 months. Again, these robustness checks broadly support the main results. For brevity, results with Lafite−Synthetic Lafite are not presented, but they reach the same qualitative conclusion: changing the length of the bubble-free period does not markedly change the detection of a bubble in Lafite.

4.3.4. *Alternative Benchmark Indices.* For the final robustness check, I construct the comparison indices three different ways. Consider only the 24-month rolling window approach. First, I use the entire set of FW450 Bordeaux wines (excluding Lafite) instead of the FW40. A bubble is detected at the ten percent level for three consecutive months from March 2010 to May 2010. At first glance, more periods with a detected bubble in Lafite−FW45 than Lafite−FW450 is counterintuitive; after all, intuition suggests it should be easier to detect a Lafite bubble in *less* similar products. However, this finding is likely the result of the fact that being less related leads to Lafite−FW450 having a higher variance early in the sample, which in turn results in a higher detection threshold and the method finding the subsequent price changes "less anomalous." That being said, a bubble is detected for three consecutive periods at the ten percent level whether FW40 is the benchmark wine or the broader set of wines in the FW450.

Second and third, I reestimate the synthetic Lafite moving the bubble-free pre-period forward and backward by four months. With four fewer months the synthetic Lafite is formed with 27.3 percent Ausone, 4.1 percent Forts Latour, and 68.6 percent Latour, while with four more months, the synthetic Lafite is formed with 24.4 percent Ausone, 3.3 percent Forts Latour, and 72.3 percent Latour. In both cases, a bubble is detected from August 2009 to October 2009. The main qualitative result—direct, model-free evidence of a bubble Lafite—is robust to changes in any of these assumptions.

5. Conclusion

The latent nature of an asset's fundamental value has led to a longstanding challenge for empirical tests of rational bubble theory. Direct tests impose a structural valuation model on the asset's fundamental value, and are thus joint tests with model specification, while rejection of the null hypothesis in an indirect test is not equivalent to a test of rational bubble theory. I generalize the model-free test introduced by Giglio, Maggiori, and Stroebel (2016) to apply to any two assets which are nearly identical. I assume the nearly identical nature of the assets implies that the difference in their fundamental values is bounded. Assuming one of the assets is more valuable than the other, the lower bound is zero. Assuming the difference in their fundamental values is an exchangeable random variable, I find a probabilistic upper bound—that is, random deviations above this upper bound occur less than a known and fixed probability—using a concentration inequality. The main limitations of the test are that it has relatively low power due to its nonparametric and conservative assumptions, and that the exact nature of the bubble may be ambiguous. Applying this generalized model-free test to the case of superstar Bordeaux wines, I find strong and robust evidence of a bubble in one asset, a ten-vintage portfolio of the wine Chateau Lafite Rothschild, under a variety of approaches. The construction of the generalized model-free test implies the detected bubble is consistent with rational bubble theory. As Giglio, Maggiori, and Stroebel (2016) fail to detect a bubble in their empirical application, this is the first evidence of a bubble outside a laboratory setting that is consistent with rational bubble theory and circumvents the joint hypothesis problem of prior direct tests. Even allowing the relationship between the fundamental values of the two assets to change over time, this bubble is detected in three consecutive periods, implying a false detection probability of less than 0.1 percent.

References

Abadie, A., A. Diamond, and J. Hainmueller. 2015. "Comparative politics and the synthetic control method." *American Journal of Political Science* 59:495–510.

—. 2010. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American Statistical Association* 105:493–505.

Abadie, A., and J. Gardeazabal. 2003. "The economic costs of conflict: A case study of the Basque Country." *American Economic Review* 93:113–132.

Al-Anaswah, N., and B. Wilfling. 2011. "Identification of speculative bubbles using state-space models with Markov-switching." *Journal of Banking & Finance* 35:1073–1086.

Ali, H.H., S. Lecocq, and M. Visser. 2008. "The impact of gurus: Parker grades and en primeur wine prices." *The Economic Journal* 118:F158–F173.

Andrade, E.B., T. Odean, and S. Lin. 2015. "Bubbling with excitement: An experiment." *Review of Finance* 20:447–466.

Authers, J. 2012. "Vintage strategy tempts palata of jaded investors." In *Financial Times*. The Long View, June 9, p. 18.

Baulch, B. 1997. "Transfer costs, spatial arbitrage, and testing for food market integration." *American Journal of Agricultural Economics* 79:477–487.

Baur, D.G., and K.J. Glover. 2014. "Heterogeneous expectations in the gold market: Specification and estimation." *Journal of Economic Dynamics and Control* 40:116–133.

Bhattacharya, U., and X. Yu. 2008. "The causes and consequences of recent financial market bubbles: An introduction." *The Review of Financial Studies* 21:3–10.

Bhattacharyya, B. 1987. "One sided Chebyshev inequality when the first four moments are known." *Communications in Statistics-Theory and Methods* 16:2789–2791.

Bikhchandani, S., D. Hirshleifer, and I. Welch. 1992. "A theory of fads, fashion, custom, and cultural change as informational cascades." *Journal of Political Economy* 100:992–1026.

Blanchard, O.J. 1979. "Speculative bubbles, crashes and rational expectations." *Economics Letters* 3:387–389.

Blanchard, O.J., and M.W. Watson. 1982. "Bubbles, rational expectations and financial markets." In P. Wachtel, ed. *Crisis in the Economic and Financial System*. Lexington, MA: Lexington Books, pp. 293–315.

Camerer, C. 1989. "Bubbles and fads in asset prices." *Journal of Economic Surveys* 3:3–41.

Campbell, J.Y., and R.J. Shiller. 1987. "Cointegration and tests of present value models." *Journal of Political Economy* 95:1062–1088.

—. 1988a. "The dividend-price ratio and expectations of future dividends and discount factors." *The Review of Financial Studies* 1:195–228.

—. 1988b. "Stock prices, earnings, and expected dividends." *Journal of Finance* 43:661–676.

Cardebat, J.M., and J.M. Figuet. 2010. "Is Bordeaux wine an alternative financial asset?" In *The 5th International Academy of Wine Business Research Conference*. pp. 8–10.

—. 2004. "What explains Bordeaux wine prices?" *Applied Economics Letters* 11:293–296.

Cardebat, J.M., and E. Paroissien. 2015. "Standardizing expert wine scores: An application for Bordeaux en primeur." *Journal of Wine Economics* 10:329–348.

Carvalho, V.M., A. Martin, and J. Ventura. 2012. "Understanding bubbly episodes." *American Economic Review* 102:95–100.

Cheung, A., E. Roca, and J.J. Su. 2015. "Crypto-currency bubbles: an application of the Phillips–Shi–Yu (2013) methodology on Mt. Gox bitcoin prices." *Applied Economics* 47:2348–2358.

Conlon, J.R. 2004. "Simple finite horizon bubbles robust to higher order knowledge." *Econometrica* 72:927–936.

Diamond, P.A. 1965. "National debt in a neoclassical growth model." *American Economic Review* 55:1126–1150.

Diba, B.T., and H.I. Grossman. 1988a. "Explosive rational bubbles in stock prices?" *American Economic Review* 78:520–530.

—. 1987. "On the inception of rational bubbles." *Quarterly Journal of Economics* 102:697–700.

—. 1988b. "The theory of rational bubbles in stock prices." *The Economic Journal* 98:746–754.

Doblas-Madrid, A. 2012. "A robust model of bubbles with multidimensional uncertainty." *Econometrica* 80:1845–1893.

Dufwenberg, M., T. Lindqvist, and E. Moore. 2005. "Bubbles and experience: An experiment." *American Economic Review* 95:1731–1737.

Eckel, C.C., and S.C. Füllbrunn. 2015. "Thar she blows? Gender, competition, and bubbles in experimental asset markets." *American Economic Review* 105:906–20.

Faye, B., E. Le Fur, and S. Prat. 2015. "Dynamics of fine wine and asset prices: evidence from short-and long-run co-movements." *Applied Economics* 47:3059–3077.

Flood, R.P., and P.M. Garber. 1980. "Market fundamentals versus price-level bubbles: The first tests." *Journal of Political Economy* 88:745–770.

Flood, R.P., and R.J. Hodrick. 1990. "On testing for speculative bubbles." *Journal of Economic Perspectives* 4(2):85–101.

Froot, K.A., and M. Obstfeld. 1991. "Intrinsic bubbles: The case of stock prices." *American Economic Review* 81:1189–1214.

Giglio, S., M. Maggiori, and J. Stroebel. 2016. "No-bubble condition: Model-free tests in housing markets." *Econometrica* 84:1047–1091.

Gürkaynak, R.S. 2008. "Econometric tests of asset price bubbles: Taking stock." *Journal of Economic Surveys* 22:166–186.

Hall, S.G., Z. Psaradakis, and M. Sola. 1999. "Detecting periodically collapsing bubbles: A Markov-switching unit root test." *Journal of Applied Econometrics* 14:143–154.

Hamilton, J.D., and C.H. Whiteman. 1985. "The observable implications of self-fulfilling expectations." *Journal of Monetary Economics* 16:353–373.

Harrison, J.M., and D.M. Kreps. 1978. "Speculative investor behavior in a stock market with heterogeneous expectations." *Quarterly Journal of Economics* 92:323–336.

Hussam, R.N., D. Porter, and V.L. Smith. 2008. "Thar she blows: Can bubbles be rekindled with experienced subjects?" *American Economic Review* 98:924–37.

Johnson, H., and J. Robinson. 2007. *The World Atlas of Wine*, sixth ed., G. Pitts, ed. Mitchell Beazley.

Kirchler, M., J. Huber, and T. Stöckl. 2012. "Thar she bursts: Reducing confusion reduces bubbles." *American Economic Review* 102:865–83.

Kleidon, A.W. 1986. "Variance bounds tests and stock price valuation models." *Journal of Political Economy* 94:953–1001.

Le Fur, E., H.B. Ameur, and B. Faye. 2016. "Time-varying risk premiums in the framework of wine investment." *Journal of Wine Economics* 11:355–378.

Le Roy, S.F. 2004. "Rational exuberance." *Journal of Economic Literature* 42:783–804.

Le Roy, S.F., and R.D. Porter. 1981. "The present-value relation: Tests based on implied variance bounds." *Econometrica* 49:555–574.

Lei, V., C.N. Noussair, and C.R. Plott. 2001. "Nonspeculative bubbles in experimental asset markets: Lack of common knowledge of rationality vs. actual irrationality." *Econometrica* 69:831–859.

Lister, E. 2011. "New ways to tap into fast-growing asset class." In *Financial Times*. FT Report - Buying and Investing in Wine, June 18, p. 3.

Mackay, C. 1852. *Extraordinary Popular Delusions and the Madness of Crowds*. Bentley (reprint ed., New York: Harmony Books 1980).

Miao, J., and P. Wang. 2018. "Asset bubbles and credit constraints." *American Economic Review* 108:2590–2628.

Moinas, S., and S. Pouget. 2013. "The bubble game: An experimental study of speculation." *Econometrica* 81:1507–1539.

Mustacich, S. 2015. *Thirsty Dragon: China's Lust for Bordeaux and the World's Best Wines*, 1st ed. Henry Holt and Company.

O'Hara, M. 2008. "Bubbles: Some perspectives (and loose talk) from history." *The Review of Financial Studies* 21:11–17.

Palan, S. 2013. "A review of bubbles and crashes in experimental asset markets." *Journal of Economic Surveys* 27:570–588.

Palfrey, T.R., and S.W. Wang. 2012. "Speculative overpricing in asset markets with information flows." *Econometrica* 80:1937–1976.

Phillips, P.C., and T. Magdalinos. 2007. "Limit theory for moderate deviations from a unit root." *Journal of Econometrics* 136:115–130.

Phillips, P.C., S. Shi, and J. Yu. 2014. "Specification sensitivity in right-tailed unit root testing for explosive behaviour." *Oxford Bulletin of Economics and Statistics* 76:315–333.

—. 2015a. "Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500." *International Economic Review* 56:1043–1078.

—. 2015b. "Testing for multiple bubbles: Limit theory of real-time detectors." *International Economic Review* 56:1079–1134.

Phillips, P.C., Y. Wu, and J. Yu. 2011. "Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?" *International Economic Review* 52:201–226.

Phillips, P.C., and J. Yu. 2011. "Dating the timeline of financial bubbles during the subprime crisis." *Quantitative Economics* 2:455–491.

Rabinovitch, S. 2010. "Chinese investors raise glasses to first wine fund." In *Financial Times*. World News, August 24, p. 6.

Samuelson, P.A. 1958. "An exact consumption-loan model of interest with or without the social contrivance of money." *Journal of Political Economy* 66:467–482.

Saw, J.G., M.C. Yang, and T.C. Mo. 1984. "Chebyshev inequality with estimated mean and variance." *The American Statistician* 38:130–132.

Shiller, R.J. 1981. "Do stock prices move too much to be justified by subsequent changes in dividends?" *American Economic Review* 71:421–436.

—. 2015. *Irrational Exuberance*. Princeton, NJ: Princeton University Press.

Smith, V.L., G.L. Suchanek, and A.W. Williams. 1988. "Bubbles, crashes, and endogenous expectations in experimental spot asset markets." *Econometrica* 56:1119–1151.

Stellato, B., B.P. Van Parys, and P.J. Goulart. 2017. "Multivariate Chebyshev inequality with estimated mean and variance." *The American Statistician* 71:123–127.

Stigler, G.J., and G.S. Becker. 1977. "De gustibus non est disputandum." *American Economic Review* 67:76–90.

Stimpfig, J. 2010. "Notes of recovery, hints of bubble." In *Financial Times*. FT Report - Buying and Investing in Wine, June 19, p. 1.

Temperton, P. 2011. "Try a useful asset bubble signal - the buy-to-not." In *Financial Times*. Fund Management, August 22, p. 8.

Tirole, J. 1985. "Asset bubbles and overlapping generations." *Econometrica* 53:1499–1528.

—. 1982. "On the possibility of speculation under rational expectations." *Econometrica* 50:1163–1181.

West, K.D. 1988. "Bubbles, fads and stock price volatility tests: A partial evaluation." *Journal of Finance* 43:639–656.

—. 1987. "A specification test for speculative bubbles." *Quarterly Journal of Economics* 102:553–580.

Xiong, W. 2013. "Bubbles, crises, and heterogeneous beliefs." Working paper No. 18905, National Bureau of Economic Research, Cambridge, MA.

Appendix A. Finite-Sample Cantelli-Cheybychev Inequality

Saw, Yang, and Mo (1984) derive a distribution-free probability bound Cheybychev applicable to finite samples. Their result was recently extended to the multivariate case (Stellato, Van Parys, and Goulart, 2017); however, a one-sided version of Saw, Yang, and Mo (1984) has yet to be considered. The extension to a one-sided inequality is non-trivial because relaxing symmetrical deviations introduces an additional dimension to the problem. We will find a Cantelli-Cheybychev-type inequality that applies to the variable X_{n+1} based on a sample $X_1, \ldots, X_n$ from some unknown population.

Assumption 1. *The finite sample, $X_1, X_2, \ldots, X_n, X_{n+1}$ with $n > 2$, is a sequence of weakly exchangeable random variables.*

Only assumption 1 is imposed in finding the finite sample Cantelli-Cheybychev-type bound. Of note: (a) the bound is distribution-free; (b) technically, the bound is valid even if the first two population moments do not exist; and (c) mutual exchangeablilty imposes less than, and subsumes, the important cases of an i.i.d. random variable or a random sample from a finite population.

As in Saw, Yang, and Mo (1984) and Stellato, Van Parys, and Goulart (2017), let

$$k^2 = \frac{n\lambda^2}{n - 1 + \lambda^2}$$

for a $k > 1$. Note $\lim_{n\to\infty} k = \lambda$. The parameter k is a finite-sample analog to the threshold parameter in the population inequality, λ.

The problem of finding a finite-sample Cantelli-Cheybychev-type inequality is equivalent to finding the maximum number of large positive deviations (LPD) in a mean zero, unit variance real vector $\boldsymbol{u}$ where large implies $u_j > k$. The maximum number of LPD depends on the number of large absolute deviations (LAD). The following theorem shows that, for any number of LAD, we can find the maximum number of LPD. The proof follows.

Theorem 1. *Let assumption 1 hold. Consider a $w = \#\{j : |u_j| > k\}$ where $w \leq v^*$ and $v^* = \max \#\{j : |u_j| > k\}$ as determined by Saw, Yang, and Mo (1984). Then for any m-length real vector $\boldsymbol{u}$ with $\sum_{j=1}^m u_j = 0$ and $\sum_{j=1}^m u_j^2 = m$,*

$$p = \max \#\{j : u_j > k\} = \frac{w + \delta^*}{2} \tag{6}$$

where

$$\delta^* = \underset{\delta \in \boldsymbol{\Delta}}{\operatorname{argmin}} \quad \ell^2(\delta) \quad s.t. \quad \ell^2(\delta) \geq k^2,$$

can be found with a grid search over $\delta \in \boldsymbol{\Delta}$ *given*

$$\ell^2(\delta) = \frac{m(m-w)}{w(m-w)+\delta^2}$$

$$\text{and} \qquad \boldsymbol{\Delta} = \begin{cases} \{0, 2, \ldots, w\} & \text{if } w \text{ is even,} \\ \{1, 3, \ldots, w\} & \text{if } w \text{ is odd.} \end{cases}$$

Proof. For each $\delta \in \Delta$, construct a $\boldsymbol{u}_\delta$ vector of length m by assigning $+\ell$ to $(w+\delta)/2$ elements, $-\ell$ to $(w-\delta)/2$ elements, and $-\varepsilon$ to the remaining $m-v$ elements, where $\ell^2 = m(m-w)/[w(m-w)+\delta^2]$ and $\varepsilon^2 = \ell^2\delta^2/[(m-v)^2]$. We now have a collection of valid $\boldsymbol{u}_\delta$ vectors, $\{\boldsymbol{u}_0, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_w\}$ (by construction each vector has $\sum_{j=1}^m u_j = 0$ and $\sum_{j=1}^m u_j^2 = m$). We want to find the vector with $\max \#\{j : u_j > k\}$. As ℓ^2 is decreasing in δ, indices with higher δ index (i.e. more asymmetric) have more positive deviations; however, when δ is too high, $\ell < k$, and so these positive deviations are no longer sufficiently large. Thus, eliminate from consideration all $\boldsymbol{u}_\delta$ with $\ell^2(\delta) < k^2$, and the vector with the largest δ index remaining, δ^*, satisfies $\max \#\{j : u_j > k\}$ with $(w+\delta^*)/2$ large positive deviations. Thus, $p = \max \#\{j : u_j > k\} = (w+\delta^*)/2$ for the given w. □

Theorem 1 gives the maximum number of LPD given some number of LAD in a mean zero, unit variance vector of length m. Intuitively, we can think of this value as the local maximum number of LPD. We can find the global maximum number of LPD, p^*, using the algorithm below. The strategy behind the algorithm is a simple grid search in three steps: (1) determine bounds on LAD for the grid search; (2) find the local maximum LPD for each LAD value in the grid; and (3) find the global maximum LPD across LAD values in the grid.

Algorithm 1.

1. *(Saw, Yang, and Mo, 1984). Find* v^*. *Let* $v = \lfloor m/k^2 \rfloor$, *where* $\lfloor \cdot \rfloor$ *is the floor function, and* $a^2 = \frac{m(m-v)}{1+v(m-v)}$. *Then* $v^* = v$ *if* v *is even,* $v^* = v$ *if* v *is odd and* $v \leq a^2$, *or* $v^* = v - 1$ *otherwise.*
2. *Define the grid search support as* $\boldsymbol{W} = \{\frac{v^*}{2}, \frac{v^*}{2}+1, \ldots, v^*\}$ *if* v^* *is even or* $\boldsymbol{W} = \{\frac{v^*+1}{2}, \frac{v^*+1}{2}+1, \ldots, v^*\}$ *if* v^* *is odd. For each candidate value* $w \in \boldsymbol{W}$, *find* $p_w = \frac{w+\delta^*}{2}$ *as set out in Theorem 1.*
3. *Then* $p^* = \max\{w : p_w\}$.

Table 1 reports results of running the algorithm for a variety of n and λ. Note that, as $n \to \infty$, the algorithm converges to the population inequality. As in the case of Saw, Yang, and Mo (1984), convergence is not monotonic.

TABLE 4. Values of Algorithm 1 for $1 - P\left(Z_{n+1} > \lambda\right)$.

	n								
λ	5	10	20	50	100	1000	10,000	100,000	∞
1.1	0.6667	0.5455	0.5714	0.5490	0.5545	0.5485	0.5475	0.5475	0.5475
1.5	0.6667	0.7273	0.7143	0.7059	0.6931	0.6923	0.6923	0.6923	0.6923
2.0	0.8333	0.8182	0.8095	0.8039	0.8020	0.8002	0.8000	0.8000	0.8000
2.5	0.8333	0.8182	0.8571	0.8627	0.8614	0.8621	0.8620	0.8621	0.8621
3.0	0.8333	0.9091	0.9048	0.9020	0.9010	0.9001	0.9000	0.9000	0.9000
5.0	0.8333	0.9091	0.9524	0.9608	0.9604	0.9610	0.9615	0.9615	0.9615
10.0	0.8333	0.9091	0.9524	0.9804	0.9901	0.9900	0.9901	0.9901	0.9901

Note: Z_{n+1} is standardized based on the sample mean and variance of $X_1, \ldots X_n$. Values in the last column are the Cantelli-Cheybychev bound $1 - 1/(\lambda^2 + 1)$.

Table 2 reports the critical values and probabilities for select n relevant to the main manuscript. Note in many cases the probability is considerably tighter than the significance level. For example, for $\alpha = 0.10$ and $n = 24$, $P(Z_{n+1} > \lambda) \leq \alpha$ is given by $\lambda = 3.26$, for which $P(Z_{n+1} > \lambda) = 0.0769$. Also note that, due to nonmonotonic convergence, the critical value across sample sizes, but for the same α, may be higher or lower than the critical value at $n = \infty$. The table also shows why testing at the 1% level would be unreliable.

Table 5. Critical values for different sample sizes and significance levels.

	n						
λ_α	18	24	36	40	43	48	∞
A. Significance level $\alpha = 0.10$.							
Critical value	2.77	3.26	3.28	2.97	3.10	2.91	3.00
Probability	0.1000	0.0769	0.0789	0.0952	0.0890	0.1000	0.1000
B. Significance level $\alpha = 0.05$.							
Critical value	4.03	4.71	5.84	4.31	4.48	4.75	4.36
Probability	0.0500	0.0385	0.0263	0.0476	0.0444	0.0400	0.0500
C. Significance level $\alpha = 0.01$.*							
Critical value	4.03	4.71	5.84	6.18	6.42	6.79	9.95
Probability	0.0500	0.0385	0.0263	0.0238	0.0222	0.0200	0.0100

Note: Critical value is the λ which gives the probability $P(Z_{n+1} > \lambda) \leq \alpha$. Values in the last column are the Cantelli-Cheybychev bound $1 - 1/(\lambda^2 + 1)$.

* Sample sizes are too small to conduct tests at $\alpha = 0.01$, so table reports the critical value and probability at lowest probability possible for the given sample size.

Appendix B. Supplementary Figures

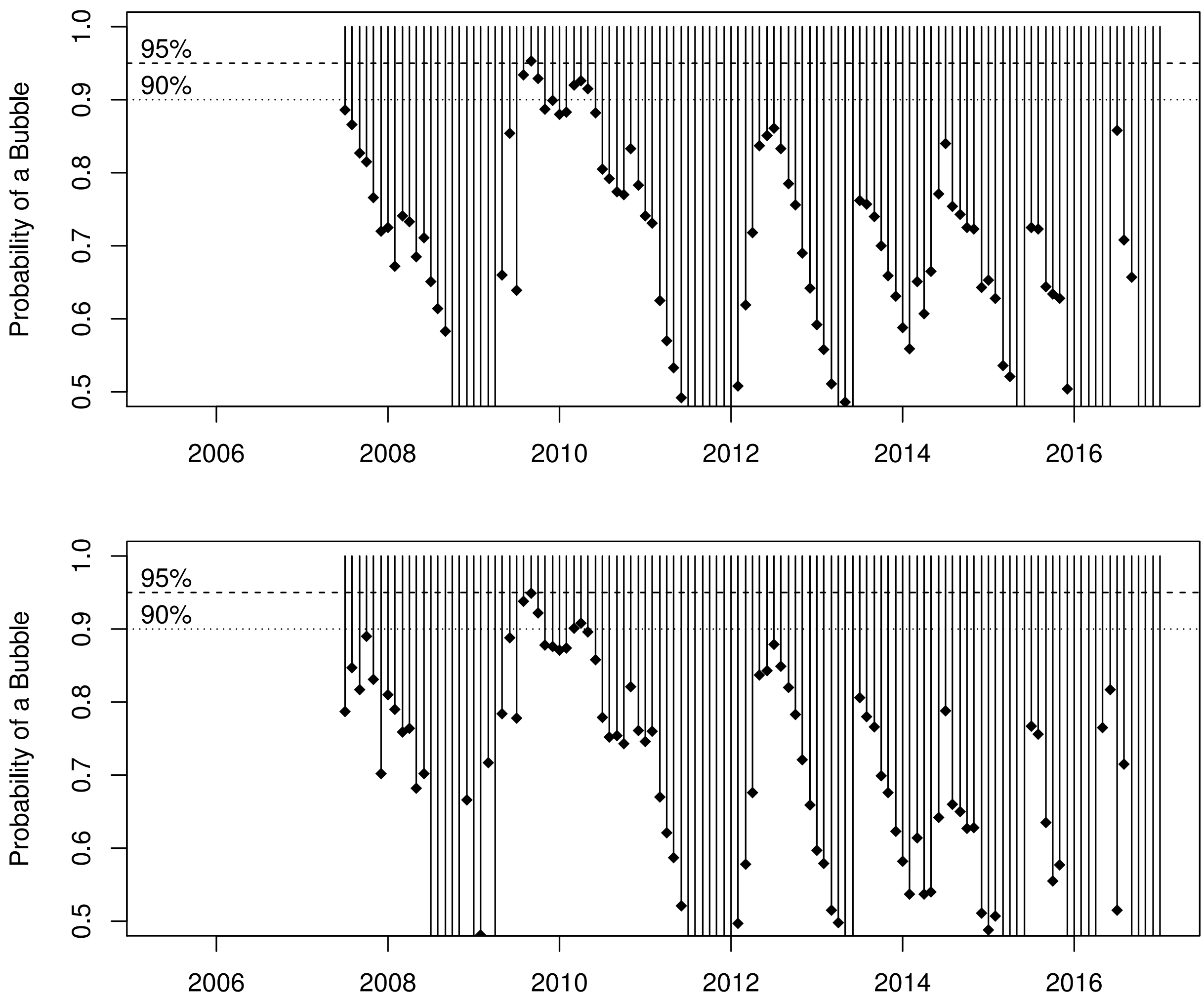

Figure 8. Implied probability of a bubble with time-varying population concentration inequality. Top: Lafite−FW40. Bottom: Lafite−Synthetic Lafite.
Note: Control is exact but based on large sample and population moments.

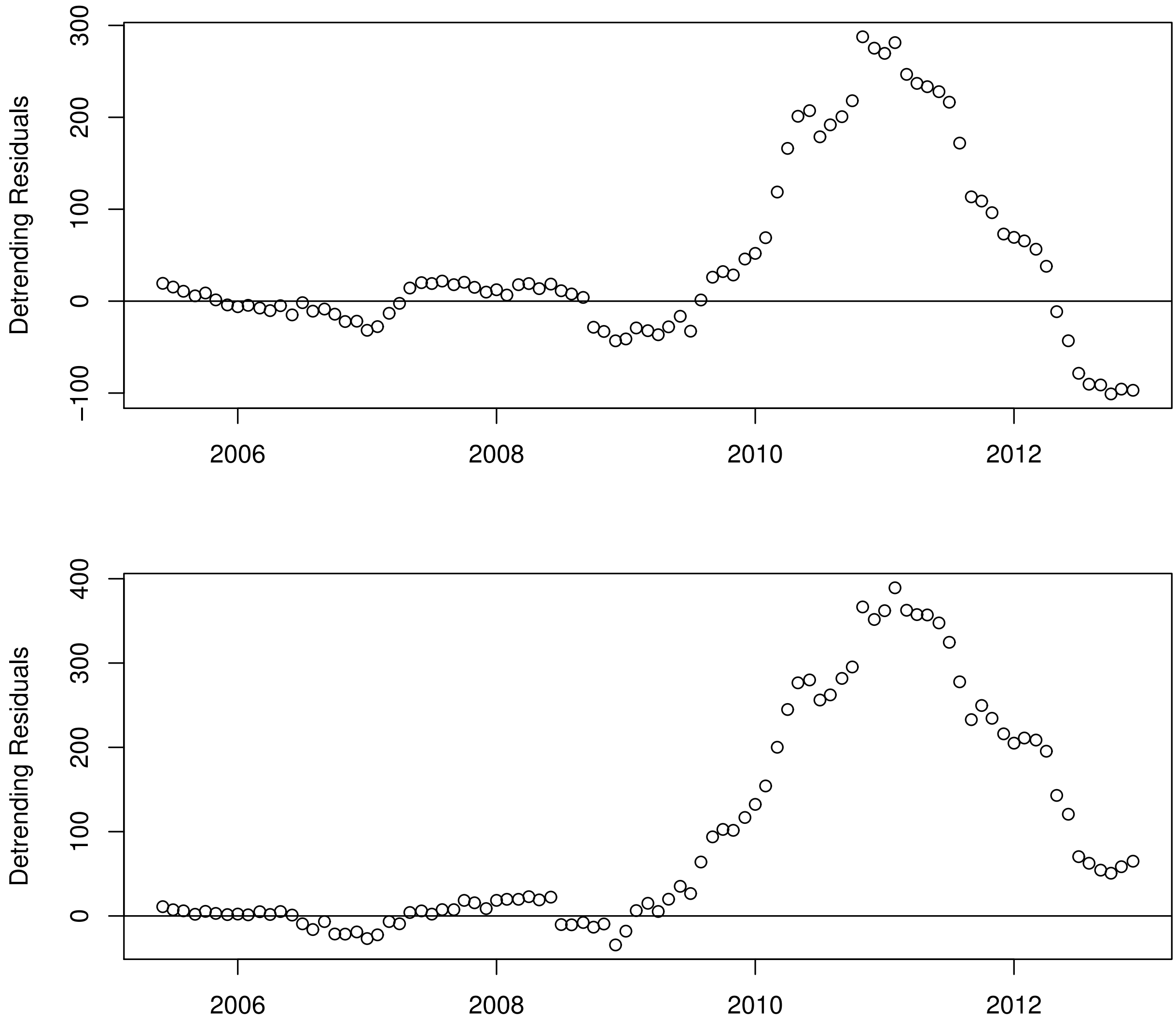

FIGURE 9. Detrending residuals of $P_t^* - P_t$ for exchangeability robustness check. Top panel: Lafite versus the four other first growth wines (FW40). Bottom panel: Lafite versus synthetic Lafite.

Note: Estimated linear trend using pre-2009 observations.

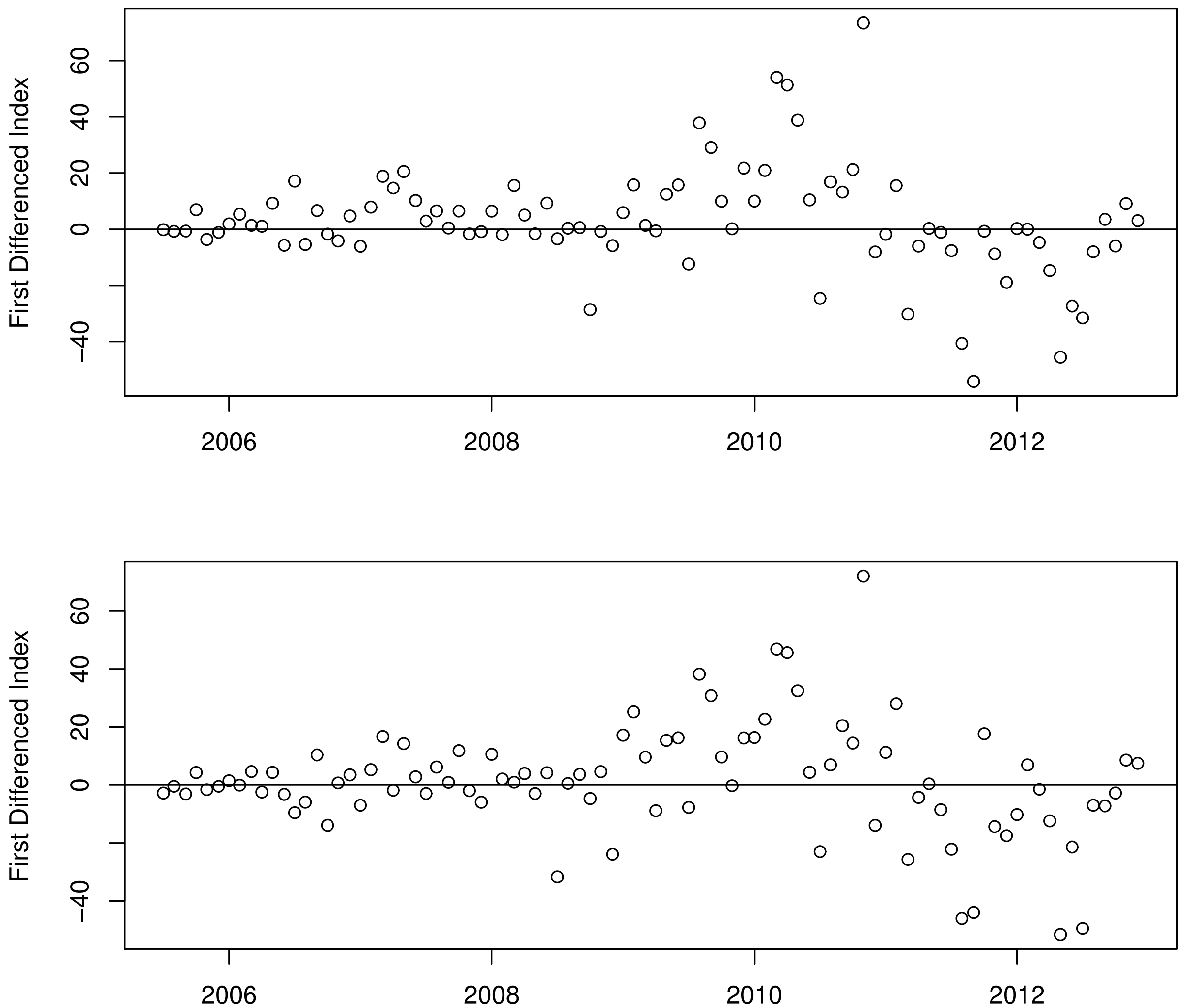

Figure 10. First difference of $P_t^* - P_t$ for exchangeability robustness check. Top panel: Lafite versus the four other first growth wines (FW40). Bottom panel: Lafite versus synthetic Lafite.