Testing for Asset Price Bubbles using Options Data*

Nicola Fusari† Robert Jarrow ‡ Sujan Lamichhane§

First Version: August 2020
Updated: November 2021

Abstract

We present a new approach to identifying asset price bubbles based on options data. We estimate asset bubbles by exploiting the differential pricing between put and call options. We apply our methodology to two stock market indexes, the S&P 500 and the Nasdaq-100, and two technology stocks, Amazon and Facebook, over the 2014-2018 sample period. We find that, while indexes do not exhibit significant bubbles, Amazon and Facebook show frequent and significant bubbles. The estimated bubbles tend to be associated with high trading volume and earning announcement days. Since our approach can be implemented in real time, it is useful to both policy-makers and investors. As an illustration we consider two case studies: the Nasdaq dot-com bubble (between 1999 to 2002) and GameStop (between December 2020 and January 2021). In both cases we identify significant and persistent bubbles.

Keywords: Asset Price Bubbles, Option Pricing, Stochastic Volatility, Martingales, Local Martingales.
JEL classification: C51, C52, G12, G13.

*We thank Torben Andersen, Maria Teresa Gonzalez-Perez (MFA discussant), Kris Jacobs, Dimitris Papadimitriou (EFA discussant), Nagpurnanand Prabhala, Eric Renault, Olivier Scaillet, and Viktor Todorov for their comments and suggestions, as well as seminar participants at Dhuram University, 2021 MFA conference, 2021 Vienna Workshop on the Econometrics of Options Markets, and 2021 EFA conference.

†Johns Hopkins Carey Business School, Baltimore, M.D. 21202, email: nfusari1@jhu.edu.

‡S.C. Johnson Graduate School of Management, Cornell University, Ithaca, N.Y. 14853 and Kamakura Corporation, Honolulu, Hawaii 96815, email: raj15@cornell.edu.

§International Monetary Fund, email: sl2563@cornell.edu. The views expressed in this paper are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

1 Introduction

Price bubbles arise because buyers often purchase an asset not to hold it forever and capture its perpetual cash flows but to resell it at a higher price sometime in the future. The existence of a bubble is therefore closely related to investors' expectations about the future evolution of the asset price. Thus, options, with their forward-looking nature, are the natural instruments to estimate and test for the existence of bubbles in their underlying assets. In this paper we develop (i) an easy-to-implement *estimation* procedure that enables us to quantify the magnitude of a price bubble, and (ii) a *statistical test* that we use to assess the significance of the estimated price bubble. Importantly, option prices are the only input required to carry out the estimation. Increased liquidity on the option market over the last decade, along with an ever-increasing number of options traded and quoted over a large range of strikes,[1] facilitates the effective and reliable implementation of our methodology.

Our methodology builds on the fact that call and put options reflect an underlying asset's bubble in different ways (see Jarrow et al. (2010)). This insight is an implication of the local martingale theory of bubbles, as developed in Loewenstein and Willard (2000b), Loewenstein and Willard (2000a), Cox and Hobson (2005), Heston et al. (2007), and Jarrow et al. (2010). According to this theory, a price bubble is defined as the difference between the market price of an asset S_t and its fundamental value S_t^*, i.e. the asset's expected discounted cash flows under the risk neutral measure. A bubble arises when the (discounted) price process is a supermartingale. In this case, while the standard risk neutral (RN) valuation methodology applies to put options (i.e. the market prices of a put option P_t corresponds to its RN value P_t^*), the RN valuation does not apply to the price of a call option because its market price C_t deviates from its RN valuation C_t^*.[2]

Figure 1 shows an example of such a deviation for the Nasdaq index on January 2, 2001 for an option cross-section with 17 days-to-maturity. The figure plots the corresponding Black and Scholes implied volatilities for ease of presentation.[3] The left panel shows the observed market (green triangles) implied volatilities versus the model-implied risk-neutral

[1] As shown, for example, by Andersen et al. (2017) for the case of the SP&P 500 index.

[2] More precisely, the presence of asset price bubble implies that under the risk neutral measure $\mathbb{Q}$ the expected discounted value of the call option's payoff does not equal its market price, i.e. $C_t \neq C_t^*$. As such, the standard RN valuation approach cannot be used to match market call prices. However, if there is no bubble (i.e. when the discounted price process is a martingale) the market prices and RN values for calls coincide, i.e. $C_t = C_t^*$. Interestingly, put-call parity continues to hold regardless of a bubble in the underlying asset (see Section 2).

[3] We do not use the Black and Scholes (1973) model to price the options. The Black and Scholes model is simply used as a quoting device to transform prices into implied volatilities. Notice that there is a monotonic relationship between prices and implied volatilities.

ones for put options (i.e. the implied volatility computed using the put's fundamental value). As seen, the two lines, even if the market data is affected by some degree of noise, greatly overlap, pointing to the equality between the market and fundamental values of put options (i.e., $P_t = P_t^*$). In contrast, the right panel shows a different situation for the call options. We see that there is a consistent positive gap between the market price of call options (green triangles) and their fundamental values (i.e., $C_t > C_t^*$).[4]

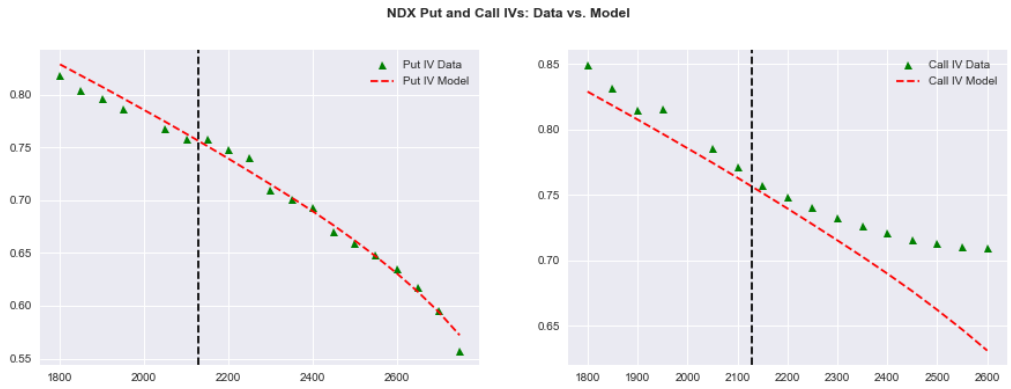


Figure 1: **Puts, Calls, and Bubble**: The figure reports the estimation results for Nasdaq on January 2, 2001 (17 days-to-maturity option cross-section). The figure shows in the left (right) panels, the observed implied volatility of put (call) options, green triangles, and the implied volatility computed using puts (calls) fundamental values, red dashed line.

Moreover, it can be shown that the resulting difference $(C_t - C_t^*) \geq 0$ is directly related to the magnitude of the bubble in the underlying asset $(S_t - S_t^*)$. As shown in Section 2.2, the relationship between the bubble in the option and the bubble in the underlying is a direct function of the option's time-to-maturity, with the difference $(C_t - C_t^*)$ vanishing when the maturity of the option approaches zero. As such, the option market can both be used to detect and to provide a *lower bound* for the size of the bubble in the underlying asset.

We develop a three-step procedure to estimate a bubble in the asset price. First, we propose a flexible parametric model for the evolution of the asset's price process. The model, for different values of its parameters, needs to admit both a martingale and a strict local martingale representation. This requirement stems from the fact that, as shown in Jarrow et al. (2010), a process admits a price bubble if and only if it is a strict local martingale. We use the family of processes posited in Andersen and Piterbarg (2007). To address model misspecification concerns, we augment this class of models with price jumps in order to better fit option prices. We call the resulting model the generalized stochastic volatility jump

[4]The details on how we compute the fundamental value of put and call options are described in Section 2.

diffusion (G-SVJD) model. The G-SVJD model subsumes many of the models commonly used in the option-pricing literature, such as the Merton (1976), the Heston (1993) stochastic volatility, and the Bates (1996) stochastic volatility with jump model. Second, we estimate the G-SVJD model using *only* put options[5] because, as noted above, put options do not exhibit bubbles and standard risk neutral valuation applies. Thus, in this step the G-SVJD model is both estimated and tested for accuracy on put options. Third, we use the estimated G-SVJD to price call options, which embed the underlying asset price bubble (if one is present). The bubble magnitude is then computed as the difference between the call's market price and its risk neutral value. Finally, we develop a statistical test to assess the significance of the estimated bubble which takes into account potential model misspecification and measurement errors.[6] Because our approach can be implemented in real time, it is useful for both policy-makers and investors who need a forward-looking approach for daily monitoring of the financial system for effective risk management.[7]

We apply our methodology to four assets over the 2014-2018 period: the S&P 500 (SPX) and the Nasdaq-100 (NDX) – two indexes that represent broad market movements – and Amazon (AMZN) and Facebook (FB) – two popular technology stocks. Our analysis reveals that SPX and NDX do not show episodes of price bubbles.[8] In sharp contrast, both AMZN and FB exhibit many more instances of large and persistent positive bubbles.

Our results challenge the widespread tendency to label periods marked by rapidly rising stock prices as representing bubbles. For example, from 2016 to mid-2018, FB's price increased from around \$100 to over \$200, exceeding a 100% price increase. Similarly, AMZN rose from \$500 in early 2016 to around \$1200 by late 2017, exceeding a 140% increase. However, we did not find any significant bubbles during these periods. From a theoretical point of view (see Section 2 for details) a bubble exists when volatility rises faster than the price itself. As higher asset volatility is usually associated with a higher level of buying/selling, this can be interpreted as representing an economic motive to capture the future resale value. We provide evidence of this fact by showing that bubbles tend to be associated with periods of increased volatility (as measured by the risk-neutral quadratic variation) and increased trading activity. Furthermore, we show that the probability of a price bubble increases in

[5]We use both in-the-money and out-of-the-money put options in this step in order to span the entire support of the risk neutral distribution.

[6]Figure 1 exemplifies the extent of measurement error in option implied volatilities and, in turn, in option prices.

[7]This is especially true because our approach can be applied to options data on equity indexes, individual stocks, and also ETFs.

[8]We measure the magnitude of a bubble as a percentage of the underlying asset price. See section 3.3 for more details.

the period preceding earnings announcements.

We then study the relationship between the parameter estimates for the underlying asset price process and the option-based bubble estimates. The theory imposes strict conditions on the parameter values that are compatible with the presence of a bubble. Empirically, even if we do not find a prefect one-to-one relationship, we find a significant link between the estimated model's parameters and the identified bubbles. In particular, we show that the estimated parameters help explain both the magnitude and the probability of the price bubbles, providing additional evidence supporting the existence of bubbles and the validity of the local martingale theory of bubbles.

Finally, to illustrate that our methodology can detect bubbles during periods widely perceived as exhibiting bubbles, we consider the Nasdaq (NDX) dot-com bubble in the early 2000s and the more recent meteoric rise of GameStop (GME). We show that our approach detects a large and persistent bubble during the dot-com boom of early 2000. For GME, we detect a bubble starting in December 2020 that remains large in magnitude and significant throughout January. These results are consistent with the short-run rise (and consequent fall) of GME's stock price between January and February 2021. Both the dot-com bubble and the GME case studies further provide empirical support for our proposed approach.

Discussion: Before we present our analysis, we consider the economics underlying the existence of bubbles. Given investors' heterogeneous beliefs, bubbles arise because buyers purchase an asset, not necessarily to hold it forever, but to resell it at a premium sometime in the future. This idea can be traced back to Harrison and Kreps (1978) and has been discussed recently in Scheinkman (2013), Xiong (2013), and Jarrow (2015). We can conceive of a risky asset's fundamental value as the price an investor would pay for holding the asset until its maturity, i.e. purely a buy-and-hold value. A price bubble reflects the idea that the resale value, the market price, exceeds the buy-and-hold value.[9] Thus, a price bubble is not an absolute concept, but rather a relative one with respect to some fundamental value (which may be unobservable). Intuitively, a bubble exists when the motive to buy an asset is not to hold it forever and consume the asset's cash flows (i.e. dividends) but rather to resell the asset at some (potentially random) future time for a (potentially) higher price.[10]

[9] A concrete example is the amount one is willing to pay to buy a work of art if after the purchase it must be held forever (and cannot be resold) – this is the fundamental value. The market price is the amount one is willing to pay if there is a possibility of selling the art at a higher price sometime in the future.

[10] In fact, we rarely see investors hold onto an asset forever and never liquidate it. Even the most notable long-term investor, Warren Buffett, periodically liquidates existing stock positions and/or adds new positions whose holding periods are longer than for most other investors, but nevertheless they are still finite. For a

As a consequence, one might expect that the stronger the reselling motive, the larger the bubble magnitude.

Related Literature: There is a related strand of the econometrics literature that develops estimators of price bubbles using historical time-series asset prices and alternative methods (e.g. see Phillips et al. (2015a), Phillips et al. (2015b), and Phillips and Shi (2018))[11]. These papers use a different model structure and a different definition of an asset price bubble. Our model structure features continuous trading over a finite horizon; theirs features discrete trading over an infinite horizon. The distinction between discrete and continous trading is important because in discrete time only two types of asset price bubbles are possible, while in continuous time three types of bubbles can exist.

Indeed, as discussed in Jarrow et al. (2010) and Jarrow (2015), in continuous time models there are three types of bubbles, called type 1, type 2, and type 3 bubbles. A type 1 bubble exists only in infinite horizon models, and it captures fiat money (a security with zero cash flows but strictly positive value). A type 2 bubble also exists only in infinite horizon models and it corresponds to an asset whose price process (under the risk neutral probability measure) is a martingale but not a uniformly integrable martingale. Intuitively, the risk-adjusted expected discounted cash flows and liquidation value at time "infinity," i.e. the fundamental value, do not equal the market price. These are the only two price bubbles possible in discrete time models. Lastly, a type 3 bubble exists only in continuous trading models and it corresponds to an asset whose price process is a local martingale but not a martingale. In economic terms, the risk-adjusted expected discounted cash flows and liquidation value at any finite time, i.e. the fundamental value, do not equal the market price. We study type 3 bubbles, while the above mentioned econometric literature focuses on type 2 bubbles. Importantly, type 2 and 3 bubbles capture distinct economic phenomena.

Type 2 bubbles are studied within infinite horizon models where the market price of an asset is compared with its fundamental value, which is estimated using a model for the asset's dividends and discount rate. This approach may encounter any of three issues. First, because the model is set on an infinite horizon, the model estimation requires a large sample, and that creates a trade-off between the model asymptotics and possible structural breaks in the market. Second, as there is no consensus on the model for estimating the fundamental value, the traditional approach faces a joint-hypothesis issue. Third, the existing literature

more detailed discussion of this view, see Jarrow (2016).

[11] Also see Giglio et al. (2016) for an empirical analysis of the existence of housing bubbles in the United Kingdom and Singapore.

on bubble detection mainly exploits information contained in historical prices. However, this backward-looking methodology implicitly assumes time series stationarity. However, given the forward-looking nature of bubbles, the stationarity assumption may not be satisfied (unless additional assumptions are made).

In contrast, a type 3 bubbles correspond to short-horizon trading strategies as investors trade to capture the resale value at some finite future time. This means that type 3 bubbles can be estimated using short time periods.[12] In particular, our approach allows us to detect the presence of a bubble and quantify its magnitude, simply by observing the price of options on any given day, without resorting to historical or backward-looking time windows. Finally, notice that, although our estimation procedure is still potentially affected by a joint-hypothesis issue, in contrast to the traditional approach our three-step procedure allows for an independent validation of the model before testing for bubbles. That is, the quality of the fit of the G-SVJD model and thus its accuracy is assessed in the first step.

Jarrow and Kwok (2020) is another recent paper that attempts to detect price bubbles using options data. Although both papers use insights from Jarrow et al. (2010) to identify price bubbles, their approaches are quite different. Jarrow and Kwok (2020) assume that the market satisfies no free lunch with vanishing risk (NFLVR) but not no dominance (ND)[13] and they estimate the asset's price bubble using a non-parametric approach. In doing so they introduce a truncation bias into their bubble estimates due to the limited range of call and put strikes, which they have to adjust for in a second step. In contrast, we assume both NFLVR and ND within a flexible parametric model for the underlying asset price evolution. While the joint assumption of NFLVR and ND assures that European put-call parity holds on observed market prices, the use of a parametric model has three main advantages: (i) it enables an independent validation of a bubble's existence (as explained in Section 3.2,) (ii) it avoids the truncation bias in the estimation of a bubble's magnitude, and (iii) it allows us, in conjunction with stock price data, to predict bubbles by looking at changes of the underlying process's parameters (see Section 5.4). The two papers provide evidence supporting each other's results.

We also contribute to the empirical option-pricing literature, which usually employs models with stochastic volatility and jumps, as in Andersen et al. (2015), which belong to the affine family, as described in Duffie et al. (2000). However, by construction, these models

[12]Further, the notion of a fundamental value and bubble have well-defined economic meanings for our type 3 bubbles: bubbles arise as investors trade to capture the resale value at some finite future time. However, type 2 bubbles reflect an abstract definition of fundamental value (the expected value of selling at infinity) and consequently the economic notion of a type 2 bubble is problematic/debatable.

[13]Both assumptions are discussed in Section 2.

do not allow for the existence of price bubbles. Here, we introduce the G-SVJD model, which subsumes many of the models used in the option-pricing literature as special cases. In particular, the G-SVJD is a superset of the affine family and it also allows for non-affine dynamics. The estimation results of the G-SVJD on option data reveal strong evidence against the affine model specification.

This paper also adds to a recent strand in the literature that uses option prices to infer characteristics of the underlying asset process. For example, Carr and Wu (2003) use time decay in option prices to detect continuous and discontinuous components, Medvedev and Scaillet (2007) relate the slope of the implied volatility of short-term at-the-money options to estimate the leverage effect. More recently, Todorov (2019) uses cross-sections of options with short maturities to nonparametrically estimate the asset's volatility, while Bandi et al. (2021) use short-term options to estimate asset's volatility, leverage effect, and volatility of volatility. Here, we exploit the differential pricing of put and call options to determine whether the underlying process displays a bubble.

The rest of the paper is organized as follows. Section 2 presents the local martingale theory of bubbles, while Section 3 describes our estimation procedure. Section 4 describes the data and Section 5 presents the results of the empirical analysis. In Section 6 we apply our estimation approach to GameStop. Section 7 concludes.

2 A Theory of Bubbles

In this section we review the local martingale theory of bubbles and present the arbitrage-free option valuation formulas used to test for the existence of bubbles in underlying asset prices. The local martingale theory of bubbles is developed in a sequence of papers by Loewenstein and Willard (2000b), Loewenstein and Willard (2000a), Cox and Hobson (2005), Heston et al. (2007), and Jarrow et al. (2010). This theory extends the characterization of bubbles from discrete time, infinite horizon models to continuous time, finite and infinite horizon models. It is shown that rational bubbles, which are consistent with no-arbitrage and equilibrium, can exist in such settings (see Jarrow (2015) for a review of this literature). This approach facilitates the use of commonly used continuous time stochastic process to help estimate and test for the existence of asset price bubbles. This paper exploits these insights to do just this.

2.1 The Market

We consider a continuous time, continuous trading model on a finite horizon $[0, T]$ given a complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. We choose a continuous time framework because discrete time can be considered a special case,[14] and because continuous time allows for more types of asset price bubbles to exist within the model. The markets are assumed to be frictionless and two assets trade: a risky asset (stock) and a money market account (mma) that is locally riskless. By frictionless we mean that there are no transaction costs nor trading constraints (e.g. short-sale restrictions). The market price of the risky asset is given by a non-negative semimartingale $\mathbb{S}_t$.[15] We denote the time t value of mma as $B_t = e^{\int_0^t r_s ds}$ where r_t is the default-free spot rate of interest. Finally, for tractability, in the remainder of the paper we consider the normalized (discounted) asset price $S_t = \frac{\mathbb{S}_t}{B_t}$, i.e. the asset price denominated in units of the numeraire (mma). This transformation makes S_t a local martingale under the risk-neutral measure and facilitates the application of the theory in Jarrow et al. (2010). For simplicity we assume that no dividends are paid on the underlying stock (we relax this assumption in Section 3).

There are three key assumptions. First, we assume that the economy satisfies no free lunch with vanishing risk (NFLVR). This is just a technical extension of the standard no-arbitrage assumption. Second, we assume that the economy satisfies no dominance (ND) – originally used by Merton (1973). ND is a stronger assumption than NFLVR but much weaker than assuming market equilibrium. ND means that there is no trading strategy that can be generated whose payoff and current cost of construction dominates any single asset's payoffs and current price.[16] In an economy that satisfies only NFLVR, with the existence of price bubbles, ND can be violated. A violation of ND implies that European put-call parity may not hold. It is well known, however, that European put-call parity is rarely violated empirically. Our assumption of ND is consistent with this observation because ND implies European put-call parity.[17] Third, note that the assumptions of NFLVR and ND imply that in a complete market bubbles cannot exist. We want to study bubbles in markets that satisfy NFLVR and ND where European put-call parity holds, so we assume that our market is incomplete when trading occurs in only the risky asset and the mma.

Under NFLVR, by the first fundamental theorem of asset pricing, there exists an equiva-

[14] This special case can be obtained by setting the price process to be a constant between successive time points.

[15] For simplicity of presentation, we omit the necessary measurability and integrability conditions on the assumed price processes; see Jarrow (2021) for these details.

[16] For a more technical discussion see Jarrow et al. (2010) and Jarrow (2021).

[17] A generalized version of put-call parity, involving inequalities, holds for American options as well.

lent local martingale measure (ELMM) that can be used to value options. A local martingale is a stochastic process such that, when stopped on a sequence of stopping times approaching infinity, the stopped process is a martingale.[18] However, in an incomplete market, there are infinitely many such ELMMs. We assume that the market chooses a unique ELMM to value the options..[19] For example, with options trading, the option prices often determine a unique ELMM consistent with their values. A representative trader's equilibrium determined ELMM is such a choice (see Jarrow (2015)).

Given the unique ELMM $\mathbb{Q}$ chosen by the market, the fundamental value (FV) of the asset at time t is defined as

$$FV_t = S_t^* = E_t[S_T], \tag{1}$$

where $E_t[.] = E^{\mathbb{Q}}[.|\mathcal{F}_t]$. This is the expected (risk-adjusted) discounted cash flow from liquidating the stock at the model's horizon, time T. Alternatively stated, it is the value agents would pay if, after purchase, they had to hold the stock until liquidation. This fundamental value is consistent with the standard definition of an asset's fundamental value used in the classic economics literature (see Jarrow (2015), and Jarrow (2021)). Indeed, in the investor's standard portfolio optimization problem, the first order condition yields an investor's ELMM, dependent upon her preferences, beliefs, information, and endowment. A representative trader's ELMM, in an economic equilibrium, is consistent with this definition.

Another definition of an asset's fundamental value appears in the literature (see Heston et al. (2007) , Herdegen and Schweizer (2016), Loewenstein and Willard (2013)). There, the fundamental value is defined as the super-replication cost of the stock. In a complete market the two definitions agree, however, in an incomplete market the super-replication price corresponds to the largest such ELMM (see Jarrow (2021)). Our market selected ELMM is therefore less than or equal to the super-replication cost of the stock implying a potentially larger bubble using our definition. This would be the case, for example, in an equilibrium determined by a representative trader's ELMM.

In what follows, quantities with a "*" superscript denote fundamental values, while those without denote market prices. Given the above result, the asset price bubble is given by:

$$\beta_t = S_t - S_t^*. \tag{2}$$

[18]For the formal definition of a local martingale see Jarrow (2021).

[19]A sufficient condition for this selection is that the market is in equilibrium or that the market studied is embedded in a larger market, which includes options, where this embedding uniquely determines the ELMM (see Jacod and Protter (2010)).

This is also the classic notion of a bubble, a deviation of the market price from its fundamental value. Further, for any $t \leq t+\tau \leq T$, we have that

$$S_t = E_t[S_{t+\tau}] + (\beta_t - E_t[\beta_{t+\tau}])\,. \tag{3}$$

Given these definitions, it follows directly that an asset price bubble exists (i.e. $\beta_t > 0$) if and only if the discounted stock price process S_t is a strict local martingale – a local martingale which is not a martingale. This implies that the bubble component β_t is also a strict local martingale. Since a local martingale bounded below is a supermartingale, β_t is also a non-negative supermartingale, i.e. declining in expectation over time with $\beta_T = 0$ at the terminal time T. In fact, if a bubbles exist, $(\beta_t - E_t[\beta_{t+\tau}]) > 0$ for any $t \leq t+\tau < T$ in equation (3). We use this fact below when designing a test for detecting bubbles using call options with maturity τ. Figure 2 provides a graphical depiction of Equation (3). The figure shows that using options with time-to-maturity $\tau < T$ one can detect the existence of a bubble, and that the magnitude of the estimated bubble is positively related to the option's maturity. Hypothetically, if one could observe options with maturity $\tau = T$, then the option market could, in principle, reveal the full extent of the bubble in the underlying asset, however this is impossible in practice.[20] This implies that for a given τ our estimates can be interpreted as a lower bound on the size of the bubble. The tightness of the bound will depend, among other things, on the difference $T - \tau$. Lastly, as formally shown in Section 2.2.1, while the bubble depends on the option's time-to-maturity it is independent of the option's strike price. This implies that the largest bubble that one can detect is bounded above by the lowest option price observed across a given set of strikes. This apparently tight upper bound, however, needs to be understood with respect to two distinct aspects: first, for large strikes, the market for call options is generally very illiquid with large bid/ask spreads. Quotes for such deep out of the money options are generally not deep and might not be fully representative of the underlying market. Second, and more importantly, as formally shown in Appendix C, even small bubbles detected over a short horizon (i.e. using short maturity options), can be the reflection of a much larger bubble over longer horizons.

Necessary and sufficient conditions on the stochastic process' local characteristics to be a strict local martingale are known only for a handful of stochastic processes (see Jarrow et al. (2020) for a collection of such processes). For example, if the price process is represented by $dS_t = \sigma(S_t)dW_t$ where W_t is a standard Brownian motion and $\sigma(S)$ is the stock's volatility, then S_t is a strict local martingale under the ELMM $\mathbb{Q}$ if and only if $\int_\epsilon^\infty \frac{s}{\sigma^2(s)}ds < \infty$ for

[20] Recall that T corresponds to the horizon of the model which is not know.

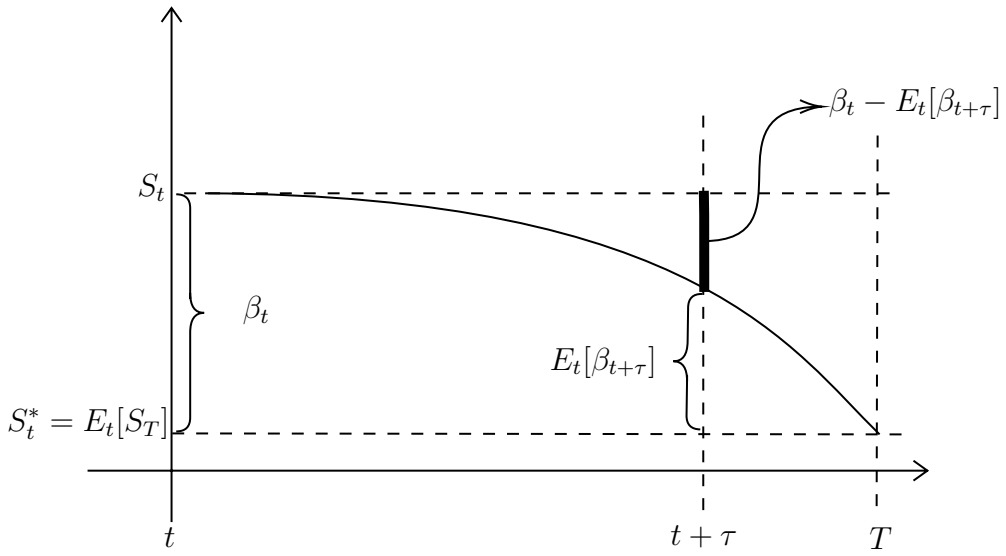


Figure 2: Graphical Representation of Equation (3)

some $\epsilon > 0$ (see Delbaen and Shirakawa (2002), also see Jarrow (2021), chapter 3). This example implies that a bubble exists if and only if the local volatility increases faster than the stock price level S_t, so that above integral is finite. This intuition, that bubbles are generated by increasing volatility as the stock price increases, is quite general and utilized below to help explain our empirical results.

2.2 Option Valuation

We now briefly present the relevant theorems, contained in Jarrow et al. (2010), that are used to carry out the empirical estimation. We abstract from dividend payments, but all the results can be extended to the case of dividend-paying stocks (see Section 3). Let $\mathbb{K}$ denote the strike price on an option (call or put) with expiration date $t+\tau \leq T$. Let $K = \frac{\mathbb{K}}{B_{t+\tau}}$ denote the strike price in units of the mma.

Let $C_t(K,\tau)$ and $P_t(K,\tau)$ denote the price, in units of the mma, of the call and put options, respectively, observed at time t with strike price K and maturity τ. At expiration, the call and put option payoffs are equal to $C_{t+\tau}(K,0) = max\{S_{t+\tau} - K, 0\}$, and $P_{t+\tau}(K,0) = max\{K - S_{t+\tau}, 0\}$, respectively. In what follows, in order to simplify the notation, unless explicitly noted, we remove the option dependency from strike price and maturity. The following sections describe how a bubble in the underlying asset affects the prices of American and European call and put options.

2.2.1 European Option

The fundamental values of European put and call options (all in units of mma) are:

$$P_t^{E^*} = E_t\left[\frac{P_{t+\tau}}{B_{t+\tau}}\right] \quad \text{and} \quad C_t^{E^*} = E_t\left[\frac{C_{t+\tau}}{B_{t+\tau}}\right], \tag{4}$$

respectively. The fundamental values correspond to the expected (after risk adjustment) discounted payoffs of the options at maturity. Theorems 6.4 and 6.5 in Jarrow et al. (2010) show that, because put payoffs are bounded above by the strike price, European puts have no bubbles, i.e.

$$P_t^E = P_t^{E^*}. \tag{5}$$

Intuitively, a bubble can exist only when an asset can be bought with the expectation that it can be sold in the future at a higher price. But, if the price of the asset (or option) is bounded above, then at some future time this can no longer happen. By backward induction, rational traders will realize this possibility at earlier times as well, and the motive for this purchase disappears at all earlier times.

However, the above argument does not apply to European calls, since their payoffs are unbounded above. It can be shown that call options exhibit the same bubble as the underlying asset (as seen on the right-hand side of Equation (3)), i.e.

$$C_t^E = C_t^{E^*} + \beta_t - E_t\left[\beta_{t+\tau}\right]. \tag{6}$$

Thus, the market price of a European call option C_t^E deviates from its risk neutral counterpart $C_t^{E^*}$. The difference $\beta_t - E_t[\beta_{t+\tau}]$ reflects the size of the bubble β_t at time t, relative to the expected size of the bubble $E_t[\beta_{t+\tau}]$ at time $t+\tau$, which is the maturity date of the option. Note that this difference is independent of the option's strike price, which means that different strike options with the same maturity should exhibit the same bubble. This independence of the strike price depends on three assumptions, frictionless markets, NFLVR (no arbitrage) and ND (no dominance). The no dominance assumption guarantees the satisfaction of put call parity, from which expression (6) is derived. We note that, for large strike options, the frictionless market assumption is less likely to be satisfied, suggesting that this relation is more likely to hold for strike prices near the money.

In expression (6), the difference is dependent on the option's maturity date. Although the existence of the bubble (if the difference is positive or zero) does not depend on the option's maturity date, the magnitude of the price differences does. This means that the magnitude

of the bubble, i.e. $\beta_t - E_t[\beta_{t+\tau}]$, is *positively* related to the option's maturity, with the bubble vanishing in the limit as the option's maturity τ reaches zero since $E_t[\beta_t] = \beta_t$. colorred In Appendix C we show in a simulation exercise how the estimated bubble's magnitude depends on the option's time-to-maturity.

Note that in our setting the put-call parity (for maturity τ) still holds. In fact, we have:

$$C_t^E - P_t^E = S_t - Ke^{-r_t\tau}, \tag{7}$$

which, in the presence of a bubble, reads as

$$\underbrace{C_t^{E^*} + \beta_t - E_t\left[\beta_{t+\tau}\right]}_{Call\ with\ bubble} - P^{E*} = \underbrace{S_t^* + \beta_t - E_t\left[\beta_{t+\tau}\right]}_{Underlying\ with\ bubble} - Ke^{-r_t\tau}. \tag{8}$$

From the above equation we can clearly see that the bubble simultaneously influences the market price of the call option and the underlying stock price.

2.2.2 American Option

Payoffs for American puts are bounded above, so their valuation follows the well-known results from standard option-pricing theory formulated without the inclusion of price bubbles. But this is not true for American calls. As shown by Theorem 6.7 in Jarrow et al. (2010), the market price and fundamental value of an American call are always equal, i.e.

$$C_t^A = C_t^{A^*} \tag{9}$$

where C_t^A and $C_t^{A^*}$ denote the market price and the fundamental value of the American call, respectively. This result follows because of the definition of the American call option's fundamental value as equal to the supremum of the expected discounted payoffs across all early stopping times. A price bubble makes the stock price process a supermartingle under the ELMM. Hence, the market price for the asset under the ELMM is expected to decline. With bubbles, early stopping increases the fundamental value of an American call option to match the market price of the option (see Jarrow et al. (2010) for a detailed exposition of this argument).

Finally, given no dividends on the underlying asset, we have the standard result (which simply follows using ND) that

$$C_t^A = C_t^E, \tag{10}$$

because an American call option is never exercised early. In conjunction, the two previous equations imply that

$$C_t^A = C_t^{E^*} + \beta_t - E_t\left[\beta_{t+\tau}\right]. \tag{11}$$

This means that even in the case of an American call option we can use the price of the corresponding European-style option to *identify* and *quantify* the presence of a stock-price bubble.

2.3 Bubble Estimation

Sections 2.2.1 and 2.2.2 show how bubbles on the underlying asset are embedded in the prices of European and American call options. Specifically, Equations (6) and (11) show that the size of the bubble is given by the difference between the market price of the call option (either European or American, which we denote by $C_t^{E/A}$) and the call's fundamental value. For both option types, the fundamental value is given by the fundamental price of the European-type call option, i.e. $C_t^{E^*}$.

The above relationships allow us to define a simple procedure for identifying a bubble that entails three steps. (I) First, we specify a parametric model for the underlying asset S_t under the risk-neutral measure $\mathbb{Q}$. This amounts to specifying a parameter vector θ that governs the dynamics of the asset which, in conjunction with the observed underlying price S_t, allows us to generate model-implied put option prices, $P_t^E(\theta)$; (II) second, we estimate the parameter vector θ by minimizing the distance between observed, either European or American, put prices ($P_t^{E/A}$), and option-implied European Put option prices ($P_t^E(\theta)$). It is crucial for this step that observed put market prices do not exhibit bubbles, as discussed in Section 2.2; (III) third, we use the parameter vector estimated in the previous step, $\hat{\theta}$, to compute model-implied European call option prices ($C_t^{E^*}(\hat{\theta})$) that are generated under the risk neutral measure and thus represent the calls' fundamental value, that is $C_t^{A^*/E^*} \approx C_t^{E^*}(\hat{\theta}, K, \tau)$.

We estimate the price bubble as the difference between the observed call market price ($C_t^{A/E}(K,\tau)$) and the call's fundamental value (which is given by the call model price $C_t^{E^*}(\hat{\theta}, K, \tau)$):

$$\mathcal{B}_t(K,\tau) = \beta_t - E_t[\beta_{t+\tau}] = \left(C_t^{A/E}(K,\tau) - C_t^{E^*}(\hat{\theta}, K, \tau)\right) \times 100. \tag{12}$$

Notice that we multiply the bubble magnitude by 100, because one standardized option contract is equivalent to owning 100 shares of the underlying stock.[21] As seen above, a price

[21] Also, to obtain the magnitude of the bubble in dollars, we should multiply Equation (12) by the value

bubble exists in the underlying asset if and only if the market price of a call option exceeds the model price (which represents the call's fundamental value).

It should be noted that our identification strategy crucially depends on the parametric model specified in step (I). As such, the above testing procedure for identifying the presence of a bubble would seem to be affected by the so-called joint-hypothesis problem. Fortunately, step (II) of our procedure allows us to mitigate the potential concerns about model misspecification as the quality of the pricing fit on put options provides a direct validation of the model. This enables us to decouple the testing of the specific option-pricing model from the subsequent tests for the existence of a bubble in the underlying asset. In Section 3 we describe in detail how the above procedure can be implemented using either European or American options.

3 The Estimation Procedure

We now discuss the implementation of the three-step estimation procedure described in Section 2.3.

3.1 Step I: The Model

This section presents the parametric model we use for our empirical analysis. The crucial requirement is that the parametrized process for the dynamics of the underlying asset needs to admit both martingale and a strict local martingale possibilities. This requirement rules out many well-known processes such as those with independent increments (e.g. Levy processes), the Merton (1976) jump diffusion, the Heston (1993) stochastic volatility, the Bates (1996) model, and affine jump-diffusion models (see Duffie et al. (2000)).

There are a limited number of known candidate processes (see Jarrow et al. (2020)). We select the following stochastic volatility model under the ELMM $\mathbb{Q}$:[22]

$$dS_t = S_t\sqrt{V_t}dW_t \tag{13}$$

$$dV_t = \kappa(\bar{\nu} - V_t)dt + \sigma_v V_t^p dZ_t \tag{14}$$

where W_t and Z_t are correlated standard Brownian motions with $\rho \in [-1, 1]$ such that

of the mma, B_t. However, by setting the current time t as the initial time, where $B_t = 1$, this adjustment is not needed.

[22]Since the normalized stock price is $S_t = \frac{\mathbb{S}_t}{B_t}$, applying integration by parts, we get $\frac{dS_t}{S_t} = \frac{d\mathbb{S}_t}{\mathbb{S}_t} - r_t dt$. Hence, we obtain the discounted price evolution by removing the riskless drift term rdt.

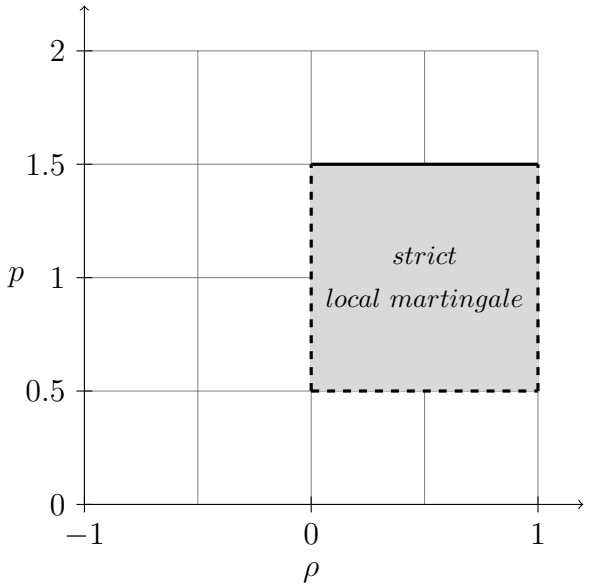


Figure 3: **G-SVJD model**: The gray area represents the parameter space that makes the model in equation (15) and (16) a strict local martingale.

$dW_t dZ_t = \rho dt$, $\sigma_v > 0$ is the volatility of the variance process V_t (i.e. vol of vol), $\kappa > 0$ is the mean-reversion speed of the variance process, and $\bar{\nu} > 0$ is the long-term mean of the variance.

Figure 3 shows the varying combinations of the parameters p (which enters Equation (14) by affecting the volatility of volatility) and the correlation coefficient between the Brownian increments, ρ, that jointly govern whether the stock price is a strict local martingale or just a martingale under $\mathbb{Q}$. Specifically, from the results in Andersen and Piterbarg (2007) and Lions and Musiela (2007), we know that:

(i) when $p \leq 1/2$ or $p > 3/2$, S_t is a martingale;

(ii) when $1/2 < p < 3/2$, S_t is a martingale for $\rho \leq 0$ and a strict supermartingale (and hence a strict local martingale) for $\rho > 0$.[23]

As discussed above, the intuition for the existence of bubbles is that the stock's volatility increases at a faster rate than its price, resulting in explosive behavior in the stock price process. In our case, this intuition holds true under the parameter restrictions (i)-(ii). This "non-linear" combination of parameter values for the existence of a bubble results from a

[23]Further, when $p = 3/2$, S_t is a martingale for $\rho \leq \frac{1}{2}\sigma_v$ and a strict supermartingale for $\rho > \frac{1}{2}\sigma_v$ (see Andersen and Piterbarg (2007) for more details). Because this condition involves an exact value of the parameter p, in what follows we focus on conditions (i) and (ii), which involve parameter ranges and thus are empirically more relevant.

tradeoff induced by the parameters ρ and p. Under condition (i) for parameter p, the volatility is not explosive enough relative to the stock price even if $\rho > 0$. In contrast, under condition (ii), $\rho > 0$ results in explosive behavior in the stock's volatility relative to the stock price increase.

While the processes in Equations (13) and (14) are technically sufficient for our purposes, a large body of asset-pricing literature highlights the importance of price jumps, especially when confronting a model with option prices and the corresponding high level of implied volatilities for out-of-the-money options (see Merton (1976), Bates (1996), and more recently Andersen et al. (2015)). To address concerns about model misspecification, we augment the model by adding an independent and discontinuous martingale to Equation (13) above. Adding a martingale to a strict local martingale retains the strict local martingale behavior of the original process, so the parameter values of the original diffusion process will again govern the existence, or lack thereof, of bubbles. Thus, we consider the following generalized stochastic volatility jump-diffusion (G-SVJD) model:

$$\frac{dS_t}{S_t} = \sqrt{V_t} dW_t + dJ_t - \lambda M dt \tag{15}$$

$$dV_t = \kappa(\bar{\nu} - V_t)dt + \sigma_v V_t^p dZ_t \tag{16}$$

where $(dJ_t - \lambda M dt)$ is the martingale jump term, $dJ_t = d\left(\sum_{j=1}^{N_t}(Y_j - 1)\right) = (Y_t - 1)dN_t$ with Y_i being i.i.d random variables representing the price jump size, with $Y_j \sim \mathcal{LN}(\mu_y, \sigma_y)$ such that $log(Y_j) \sim \mathcal{N}(\mu_y, \sigma_y)$. $N_t \sim Poisson(\lambda t)$ is the counting process with intensity given by λ counting the total number of jumps up to time t. Finally, $J_t - \lambda M t$ is a martingale with $M = E[Y_j - 1] = e^{(\mu_y + 0.5\sigma_y^2)} - 1$ being the so-called jump compensator. Note that jumps in volatility cannot be added to the G-SVJD model. Volatility jumps would affect the current stock price evolution and invalidate this model of bubble identification. As discussed above, only a handful of stochastic (volatility) processes are currently known with a complete characterization of their martingale/strict local martingale behavior that govern the existence of bubbles. The G-SVJD model is arguably quite general within this class of processes, and it is also the closest to processes employed successfully in previous option-pricing empirical studies.[24]

[24]In particular, note that the G-SVJD nests several well-known models as special cases: (i) the Merton (1976) model for $\sigma_v = 0$, (ii) the Heston (1993) model for $p = 0.5$ and $\lambda = 0$, and (iii) the Bates (1996) model for $p = 0.5$ and $\lambda > 0$.

3.2 Step II: Model Estimation

Given the above G-SVJD model, we now discuss the parameter estimation. As closed-form or analytical expressions for the price of European options (i.e. the call's fundamental values) are not available, we resort to Monte Carlo simulation. Specifically, we simulate MC trajectories from the discretized versions of the model in Equations (15) and (16) according to:

$$V_{i,t+\Delta t} = V_{i,t} + \kappa(\bar{\nu} - V_{i,t})\Delta t + \sigma_v V_{i,t}^{p}\sqrt{\Delta t}B_{i,1} \tag{17}$$

$$S_{i,t+\Delta t} = S_{i,t}\exp\left[-\left(\lambda M + \frac{1}{2}V_{i,t}\right)\Delta t + \sqrt{V_{i,t}\Delta t}\left(\rho B_{i,1} + \sqrt{1-\rho^2}B_{i,2}\right) + \mathcal{J}\right], \tag{18}$$

where $B_{i,1}$ and $B_{i,2}$ are independent standard normal variables, for $i = 1, \cdots, MC$ and $\Delta t = \frac{1}{365\times 5}$.[25] The model-implied put price $P_t(\theta, K, \tau)$ for strike K and maturity τ is obtained from the simulated prices $\{S_{i,t+\tau}\}_{i=1,\cdots,MC}$ as:

$$P_t(\theta, K, \tau) = \frac{1}{MC}\sum_{i=1}^{MC}(K - S_{i,t+\tau})^{+}. \tag{19}$$

If the underling asset pays a dividend yield we transform the price process S_t by augmenting it by the dividend yield earned over the maturity of the option, i.e. $\tilde{S}_t = S_t e^{\delta t}$. The new stock price $\tilde{S}_t$ will be S_t plus accumulated dividends, i.e.

$$\frac{d\tilde{S}_t}{\tilde{S}_t} = \frac{dS_t}{S_t} + \delta dt \tag{20}$$

Equation (20) transforms the asset with cash flows into an equivalent asset without cash flows by reinvesting all cash flows into the risky asset itself (see Jarrow (2021)). Price trajectories for $\tilde{S}_t$ can be generated following the same discretization scheme outlines in Equations (17) and (18). Finally, European put option prices are computed following standard risk neutral valuation as follows:

$$P_t(\theta, K, \tau) = \frac{1}{MC}\sum_{i=1}^{MC}(K - \tilde{S}_{i,t+\tau}e^{-\delta(t+\tau-t)})^{+}.$$

[25] This amounts to simulating the process in five time steps each day, or every 4.8 hours.

where $\tilde{S}_{t+\tau}e^{-\delta(t+\tau-t)}$ represents the value without accumulated dividends over $(t, t+\tau)$, because the option holder does not receive the dividends paid over the option's life.[26]

The G-SVJD model has a total of 9 parameters:

$$\theta = \{V_0, \bar{\nu}, \kappa, \sigma_v, \rho, p, \mu_y, \sigma_y, \lambda\}.$$

As in Andersen et al. (2017) we estimate the parameter vector on each day t using one option's maturity per-date. [27] As explained in Section 4, on each day we select the maturity date (between 7 and 50 calendar days) with the highest cumulative traded volume in put options.[28] Hence, for each underlying asset we obtain a sequence of parameter vectors $\{\theta_t\}_{t=1,\cdots,T}$, where T represents the total number of days in our sample. We estimate θ_t as in Andersen et al. (2015) and Andersen et al. (2019) by minimizing the root mean squared errors between market and model Black-Scholes implied volatilities ($RMSE_{IV}$), i.e.:

$$\hat{\theta} = \min_{\theta_t} \sqrt{\frac{1}{N_t^p}\sum_{n=1}^{N_t^p}\left[IV_t(K_{t,n},\tau_t) - IV_t(\theta_t, K_{t,n},\tau_t)\right]^2} \tag{21}$$

where N^p is the number of market put prices on date t with strike price $K_{t,n}$ and maturity τ_t. Note that N^p comprises both ITM and OTM put options which span the entire support of the risk neutral distribution. Finally, $IV_t(K_{t,n},\tau_n)$ and $IV_t(\theta_t, K_{t,n},\tau_t)$ are the implied volatilities of the market and model-implied put prices $P_t(K_{t,n},\tau_t)$ and $P_t(\theta_t, K_{t,n},\tau_t)$, respectively. We minimize implied volatilities instead of prices because doing so normalizes the objective function across strikes/maturities and the errors are interpretable as volatility percentages (see Christoffersen and Jacobs (2004) for more details on the choice of this type of an objective function).[29]

To reduce the dimension of the parameter space, and given that we are using only one option maturity per day in our estimation, we apply the following parameter restrictions to discipline the identification of $V_{0,t}$, κ_t, and $\bar{\nu}_t$. First, we fix the speed of mean reversion at

[26] For dividend-paying individual stocks, our approach applies unchanged to American put options. For American call options on stocks that pay dividends, our approach can be extended in piecewise fashion to options whose maturity dates are prior to the next ex-dividend date.

[27] Ideally, at the estimation stage, one could use a few day window over which the parameters are kept constant, allowing only the spot volatility to change day-by-day. Also, one could price options with multiple maturities on a given day. While these alternatives are worth exploring, the computational complexity that they entail would make the estimation infeasible given that option values can only be computed via simulation. Choosing the most liquid cross-section on any given day, and re-estimating the parameters every day, strikes a balance between robustness and computational feasibility.

[28] Call options also experience similarly high volumes on these very dates.

[29] We impose the Feller condition $\sigma_{v,t}^2 \leq 2\kappa_t\bar{\nu}_t$ to ensure the non-negativity of the variance process V_t.

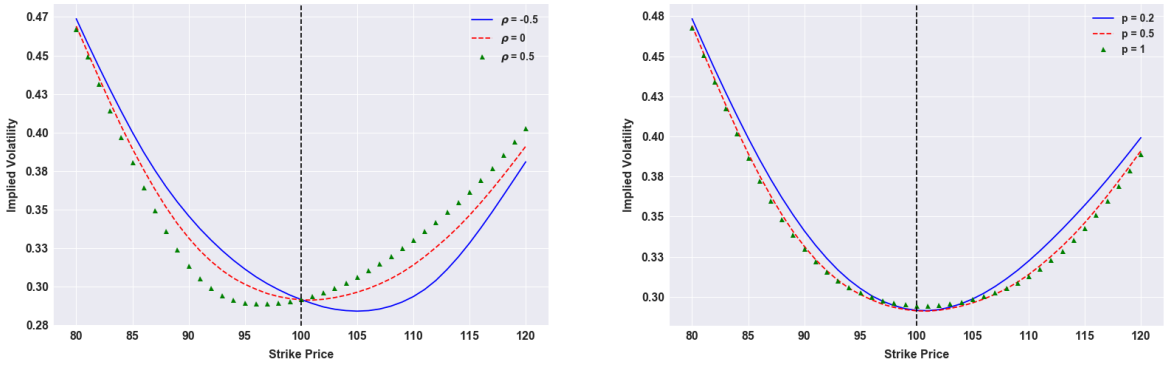


Figure 4: **Parameter Identification**: The left (right) Panel shows the implied volatility generated by the G-SVJD model in equations (16) and (15) for varying values of the parameters ρ and p. We set the remaining parameters to their estimated time-series averages for Amazon, as reported in Table 2.

$\kappa_t = 5$, which corresponds to 22 days half-life of the variance process, which is comparable to what has been estimated in the option-pricing literature.[30] Second, we set $\bar{\nu}_t = IVATM_t^2$, where $IVATM_t$ is the implied volatility of the option closest to at-the-money on day t.

As explained in Section 2, the identification of a bubble in the underlying asset hinges on the value of the parameters ρ and p. To understand how option prices and their corresponding implied volatilities help in identifying these parameters, Figure 4 shows the implied volatility generated by our model as a function of strike prices for varying values of ρ and p. From the left Panel of Figure 4 we can see that the correlation coefficient ρ is identified by the slope of the implied volatility of the options that are around at-the-money. A negative (positive) ρ is associated with a downward (upward) sloping IV curve (around at-the-money). From the right Panel of Figure 4 we see that the parameter p, which affects the volatility of volatility of the price process, affects the thickness of the tail of the return's risk-neutral distribution (i.e. its kurtosis), which is then reflected in the concavity of the IV curve.

3.3 Step III: Bubble Estimation

Next, given the estimated parameters $\hat{\theta}_t$, on each day we generate call option prices with maturity τ_t over the set of available strikes $\{K_{t,n}\}_{n=1,\cdots,N_t^c}$ (where N_t^c represents the number of available call options at time t with maturity τ_t). The bubble for a given strike $K_{t,n}$ is

[30] In unreported results, available upon request, we confirm that the size of the estimated bubble is only marginally affected by the speed of mean reversion of the variance process.

estimated as:

$$\hat{\mathcal{B}}_t(K_{t,n}, \tau_t) = \left[C_t^{A/E}(K_{t,n}, \tau_t) - C_t(\hat{\theta}_t, K_{t,n}, \tau_t)\right] \times 100. \tag{22}$$

As there are N_t^c call options (which can differ from the number of put options N_t^p), we take the average of the estimated bubble across the various option strikes $K_{t,n}$ within the same maturity τ_t, and we normalize them by the value of the underlying asset. Thus, we define the average bubble magnitude as a percentage of the asset value, i.e.

$$\hat{\mathbb{B}}_t = \frac{1}{\mathbb{S}_t} \sum_n^{N_t^c} \frac{\hat{\mathcal{B}}_t(K_{t,n}, \tau_t)}{N_t^c}. \tag{23}$$

Since options frequently trade within the bid-ask spread but not necessarily at the mid point of the bid-ask spread, it is not clear what the "true" market price is. We take this into account by employing the following classification:

- If the model call price lies between the bid-ask prices, i.e. $C_t^{Bid}(K_{t,n}, \tau_t) \leq C_t(\hat{\theta}_t, K_{t,n}, \tau t) \leq C_t^{Ask}(K_{t,n}, \tau_t)$, we set $\hat{\mathcal{B}}_t(K_{t,n}, \tau_t) = 0$
- If $C_t^{Bid}(K_{t,n}, \tau_t) > C_t(\hat{\theta}, K_{t,n}, \tau t)$ or $C_t^{Ask}(K_{t,n}, \tau_t) < C_t(\hat{\theta}_t, K_{t,n}, \tau t)$ we compute the bubble according to Equation (22)

This is a more conservative approach relative to taking differences using market mid-prices and makes our inference more robust to measurement errors and possible liquidity concerns. In the next section we describe how to build formal statistical tests to assess the null hypothesis that $\mathbb{B}_t$ (or functions of $\mathbb{B}_t$) is equal to zero.[31]

3.4 Testing For Bubbles

In this section we define two statistical tests that can be used to test for the presence of bubbles in the underlying assets. Both tests build on the observation that, in the local martingale theory of bubbles, negative bubbles (i.e. $\mathbb{B}_t < 0$) are precluded because the fundamental value always represents a lower bound on the market price.[32]

[31] Our framework assumes frictionless markets without arbitrage. In reality, there could be instances in which short-lived arbitrage opportunities arise across different option contracts, or instances in which put-call-parity is temporarily violated, creating a gap between the call's market and fundamental values. In our methodology the above instances would also be labeled as a bubble.

[32] Note that some equilibrium models could generate negative bubbles in the presence of trading constraints. However, the economic mechanisms are quite different from those considered here (for example, see Jarrow and Lamichhane (2021)).

However, negative bubbles might arise empirically because of three issues. First, we measure a bubble as the difference between the market and the model's call price. Even though the G-SVJD model is quite general, possible model misspecification could introduce measurement error. Second, market prices are observed with errors. Options, especially OTM options, show significant bid-ask spreads. Moreover, the option and the underlying markets could be affected by asynchronous trading and market participants' trading constraints. Third, when applying our methodology to American options we still compute the put prices according to equation (19), which does not account for the early exercise premium. Thus, we interpret negative bubbles as a reflection of noise caused by either model misspecification or measurement error. We use the estimated negative bubbles to measure the noise in our bubble estimates, which is then used to compute confidence intervals to test the null hypothesis of zero bubbles.

Given the earlier discussion in the introduction, empirically we expect to see bubbles fluctuating in the vicinity of zero most of the times, with occasional large spikes that we can interpret as periods of heightened reselling motives.[33] We build our tests on the following statistical model:

$$\hat{\mathbb{B}}_t = \mathbb{B}_t + \varepsilon_t,$$

where $\hat{\mathbb{B}}_t$ is the estimate of a bubble magnitude and $\varepsilon_t \sim N(0, \sigma_t)$ represents the model error plus noise. This test is constructed to test the null that the bubble is equal to zero on a given day, i.e.

$$(H_0) : \mathbb{B}_t = 0.$$

We use σ_t to construct a time-varying threshold. A bubble is considered significant if it is above this threshold on any given day t. Importantly, since the threshold is estimated using only the available information up to time t, this makes it possible to estimate a bubble in real time.

To implement the above test, σ_t is estimated over a small time window immediately before time t. The estimator relies on the following observation. If $\hat{\mathbb{B}}_t < 0$, then $\mathbb{B}_t + \varepsilon_t < 0$. Since bubbles must always be non-negative ($\mathbb{B}_t \geq 0$), this implies the error must be negative and large, i.e.

$$\varepsilon_t < -\mathbb{B}_t \leq 0.$$

Under the null, where $\mathbb{B}_t = 0$, the bubble estimate is just the error, i.e. $\hat{\mathbb{B}}_t = \varepsilon_t < 0$. This

[33] More formally, we observe one sample path $\omega \in \Omega$ over a given observation period $t \in [1, T]$, which implies that, given any $t \in [1, T]$, $\mathbb{B}_t(\omega)$ is a constant. Hence, in what follows, we can drop the ω argument.

insight allows us to create an estimator for the variance of the noise process. Specifically, within a small window of time (equal to k days) before time t, we partition the daily bubble estimates $\{\mathbb{B}_i\}_{i=t-k,\cdots,t}$ and collect observations at only those times $i = t-k, ..., t$ where $\hat{\mathbb{B}}_i < 0$. Call this set Υ_-. Let the number of elements in this set be N_-. For this set of observations, an unbiased estimator of σ_t^2 is

$$\hat{\sigma}_t^2 = \sum_{i \in \Upsilon_-} \frac{1}{N_-} \hat{\mathbb{B}}_i^2. \tag{24}$$

The proof of this statement is contained in Appendix B. Even though under the null hypothesis the observations $\hat{\mathbb{B}}_t > 0$ are also just error, since we observe $\hat{\mathbb{B}}_t > 0$, it must be the case that if bubbles exist, $\mathbb{B}_t > 0$ is larger than any negative errors included within this set (which occur with positive probability, roughly half the time). This implies that including these observations in our σ_t^2 estimator would generate a biased statistic if bubbles exist, because $\mathbb{B}_t$ itself varies over time.

4 Data

We focus our empirical analysis on two sets of underlings: (i) two broad market indexes, S&P 500 (ticker: SPX) and Nasdaq-100 (ticker: NDX), and (ii) two individual stocks, Amazon (ticker: AMZN) and Facebook (ticker: FB), which are routinely associated, in the financial press, with possible price bubbles. Our sample period runs from January 2, 2014 through December 31, 2018. The starting point of our sample period is dictated by the quality of the available option data. As shown in Andersen et al. (2017), over the last decade the equity index option market has exhibited tremendous growth in terms of both trading volume and number of available contracts. This growth owes, in part, to the introduction of short maturity options, the so-called weeklies. More recently, the popularity of weekly options has expanded from index options to options on individual stocks. Starting our sample period in 2014 allows us to have, every day, a sufficient number of options over a wide strike range, which in turn facilitates a robust and reliable implementation of our proposed methodology, as explained in Section 3.

We obtain daily end-of-day put and call option quotes from OptionMetrics. On each day, and for each asset, we keep only the cross-section with the highest cumulative volume.[34] The

[34]To remove highly illiquid contracts, for SPX and NDX we keep only options with standardized moneyness, defined as $m = \ln(K/S)/(IV_{ATM}\sqrt{\tau})$, between -10 and 5 and Black-Scholes implied volatility less than 1. Ror AMZN only, we keep options with Black-Scholes delta between 0.01 and 0.99 (in absolute value) given

choice of option maturity is dictated by two observations. First, as discussed in Section 2.2.1, because the magnitude of a potential bubble decreases with the option's tenor, we discard options with very short maturities (i.e. $\tau < 7$). Second, because most of the option liquidity is concentrated over relatively short horizons, we discard options with maturities longer that 50 calendar days. Table 1 reports summary statistics for the final sample of the option data. As can be seen from Table 1, which is relevant to to our empirical analysis, both puts and calls, on all four assets, span a large and similar standardized moneyness range that covers, at least, the range between -4 and 3. This means that both option types span a large portion of the risk-neutral distribution.

In terms of number of options, although SPX and NDX exhibit, on average, a larger number of available contracts with a daily average of around 120 unique strikes, both AMZN and FB still exhibit a large number of quoted options with at least 20 unique strikes per day that are always available. Interestingly, while for indexes we see more put options than call options, for single stocks the two option types are represented equally. A similar pattern can be observed in trading interest (i.e. trading volume and open interest). For indexes these are more concentrated in put options while for single stocks these are divided equally between put and call options. The average option maturity is slightly higher for the equity indexes (23 for SPX and 18 for NDX) than for single stocks (15 for AMZN and 17 for FB). This reflects the considerable popularity of standard monthly options for indexes, while weeklies absorb most of the trading volume in equity options. In terms of observation errors, both SPX and NDX have a relative bid-ask spread of around 33% while in AMZN and FB the relative bid-ask spread is around 14%. A possible explanation of the latter fact is that the average at-the-money implied volatility (and, as a consequence, the average at-the-money option price, which enters at the denominator of the relative bid-ask spread) is much lower for indexes (between 12% and 16%) than for single stocks (around 28%). In summary, since our empirical analysis is based on both in- and out-of-the-money put and call options, it is important to note that both option types are available over a large moneyness range with high trading volume and open interest.

Regarding the underlying assets, our data sources differ between individual stocks and indexes. Since equity options and their corresponding underlying assets (AMZN and FB) trade in a synchronized fashion with both the option and equity markets closing at 4:00 p.m. EST, we use the closing prices provided by OptionMetrics. For indexes (SPX and NDX), however, the option market closes at 4:15 p.m. while the cash markets close at 4:00 pm. Even

highly illiquid contracts beyond that range.

Table 1: **Summary Statistics for Put and Call Options**: The sample size is 1,258 for both puts and calls. First five columns report put statistics and second five columns report call statistics. Moneyness (normalized) are OTMmin = $log(K_{min}/S)/(\sigma_{ATM}\sqrt{\tau})$, ITMmax = $log(K_{max}/S)/(\sigma_{ATM}\sqrt{\tau})$. K_{min} is out-of-the-money (OTM) for puts and in-the-money (ITM) for calls. K_{max} is ITM for puts and OTM for calls. $N^p|N^c$ are the number of put/call options. Volume and Open Interest (reported as $\times 1000$) are sums of all the volume and open interests across all strikes between K_{min} and K_{max} of the maturity (in days) with highest volume. Bid-Ask-Spread shows the average bid-ask spread of option prices as a percentage of the option mid-point. Panels A, B, C, D correspond to SPX, NDX, AMZN, and FB, respectively.

	Puts					Calls				
	Mean	SD	P25	P50	P75	Mean	SD	P25	P50	P75
Panel A: SPX										
OTMmin	-9.66	0.35	-9.84	-9.75	-9.59	-6.65	2.83	-9.56	-7.08	-3.86
ITMmax	2.61	1.38	1.33	2.25	3.99	3.24	0.75	2.66	3.13	3.81
$N^p\|N^c$	132.86	45.84	99	130	164	110.77	45.82	77	105	138
Volume	127.7	63.7	83.97	113.21	154.72	74.37	40.6	45.02	66.59	94.13
Open Interest	1255.37	663.76	772.87	1116.27	1833.27	824.39	526.25	421.25	618.23	1289.96
Maturity	23.11	12.06	11	23	32	23.11	12.06	11	23	32
Bid-Ask Spread	0.33	0.19	0.17	0.33	0.48	0.20	0.09	0.14	0.18	0.24
ATM IV	0.12	0.04	0.09	0.11	0.14	0.12	0.04	0.09	0.11	0.14
Panel B: NDX										
OTMmin	-8.33	1.53	-9.43	-8.96	-7.63	-5.87	2.8	-8.92	-5.75	-3.14
ITMmax	2.25	1.26	1.24	1.67	3.23	2.73	0.87	2.03	2.59	3.26
$N^p\|N^c$	119.9	49.73	86	111	140	110.1	46.92	78	104	137
Volume	5.18	8.92	1.98	3.09	5.19	2.32	2.91	0.8	1.46	2.64
Open Interest	28.15	28.65	9.49	18.14	36.14	16.71	15.3	3.84	11.81	25.37
Maturity	18.12	9.88	9	16	25	18.12	9.88	9	16	25
Bid-Ask Spread	0.32	0.18	0.20	0.28	0.43	0.18	0.11	0.10	0.16	0.23
ATM IV	0.16	0.05	0.12	0.14	0.17	0.16	0.05	0.12	0.14	0.17
Panel C: AMZN										
OTMmin	-4.05	0.8	-4.55	-4.03	-3.52	-4.92	2.71	-6.8	-4.34	-2.64
ITMmax	4.31	2.85	1.82	3.49	6.29	2.96	0.66	2.51	2.85	3.3
$N^p\|N^c$	89.61	56.87	48	73	113	87.41	57.25	47	71	109
Volume	8.68	5.71	4.77	7.16	10.78	10.45	7.54	5.51	8.57	12.95
Open Interest	30.85	18.87	14.4	29.1	41.34	34.76	22.94	15.09	32.63	47.01
Maturity	15.27	8.93	8	11	23	15.27	8.93	8	11	23
Bid-Ask Spread	0.13	0.08	0.07	0.11	0.18	0.14	0.08	0.08	0.12	0.17
ATM IV	0.28	0.11	0.20	0.25	0.32	0.28	0.11	0.21	0.25	0.32
Panel D: FB										
OTMmin	-5.45	1.87	-6.43	-5.13	-4.11	-6.33	3.65	-8.62	-5.72	-3.17
ITMmax	5.26	3.72	1.83	4.34	7.82	3.32	1.06	2.56	3.12	3.9
$N^p\|N^c$	36.61	15.42	25.25	35	44	34.34	14.88	24	32	41
Volume	22.86	16.22	12.69	18.37	27.55	38.49	27.22	20.77	31.57	46.91
Open Interest	132.38	86.23	52.65	127.64	184.41	210.68	137.18	86.51	210.93	283.97
Maturity	17.1	10.18	8	15	24	17.1	10.18	8	15	24
Bid-Ask Spread	0.16	0.08	0.11	0.15	0.20	0.11	0.06	0.07	0.11	0.14
ATM IV	0.28	0.11	0.20	0.25	0.33	0.28	0.11	0.20	0.25	0.33

though OptionMetrics collects option quotes as closely as possible to 4:00 p.m. , in order to avoid potential issues of asynchronous prices[35] we follow Andersen et al. (2015) and use put-call parity to compute the *implied price* of the underlying index. More specifically, because options on SPX and NDX are European-style, assuming that the put-call parity holds in the data, we compute the implied forward price $F_{t,\tau}$ on a common set of strikes K across put and call options pairs with the same maturity τ via $F_{t,\tau} = K + e^{r\tau}(C_t(K,\tau) - P_t(K\tau))$. We then take the median $F_{t,\tau}$ using the five option contracts that are closest to at-the-money because they tend to be the most liquid. Finally, using the risk-free rate (obtained by linearly interpolating the zero-yield curve provided by OptionMetrics to match the option maturity) and the continuously compounded dividend yield (provided by OptionMetrics), we compute the implied spot price as $S_t = F_{t,\tau}e^{(r_t-\delta_t)\tau}$.

5 Empirical Analysis

This section discusses the estimation results of the G-SVJD model and the estimated bubbles for the four assets analyzed in our main empirical exercise: SPX, NDX, AMZN, and FB.

5.1 G-SVJD Model Estimation Results

Table 2 shows the summary statistics for the daily estimated parameters of the G-SVJD model and the $RMSE_{IV}$. The average parameter estimates are quite similar across the assets, with a few exceptions. The equity index volatilities, on average, are between 10% and 13%, while AMZN and FB display much larger values of around 30%. This is consistent with the fact that the ATM IV (as reported in Table 1) is much higher for equity than for index options. All assets have similar average negative jump sizes, with the single stocks' jump intensities being twice as large as those of SPX and NDX. The correlation coefficient ρ is negative, on average, for all the assets, but less so for AMZN and FB; the P99 of the latter group is actually positive. Finally, the parameter p is approximately 0.2 for indexes and 0.35 for single stocks. Both values are significantly smaller than the value of 0.5 which is commonly used in the option-pricing literature, pointing to the fact that the affine specification for the variance process seem to be rejected when confronted with option data.

In terms of the option-pricing fit, the G-SVJD model does a remarkable job. The average $RMSE_{IV}$ is 0.57% and 0.49% for SPX and NDX, respectively. These are small percentage errors which are similar to the best performing model in Andersen et al. (2017), especially

[35] See IvyDB US Reference Manual v4.0, p. 19.

Table 2: **G-SVJD Model Estimation Results**: This table reports the means, standard deviations (SD), the 1st (P01), 25th (P25), 50th (P50), 75th (P75), 99th (P99) percentiles of the daily parameter estimates as well as the corresponding $RMSE_{IV}$ (%). Panels A, B, C, D correspond to SPX, NDX, AMZN, and FB, respectively.

	Mean	SD	P01	P25	P50	P75	P99
Panel A: SPX							
V_0	0.011	0.013	0	0.003	0.007	0.013	0.064
σ_V	0.322	0.148	0.07	0.229	0.296	0.396	0.773
ρ	-0.805	0.176	-0.99	-0.933	-0.839	-0.724	-0.305
p	0.205	0.138	0.1	0.132	0.189	0.243	0.47
μ_y	-0.219	0.18	-0.791	-0.296	-0.177	-0.1	0.102
σ_y	0.229	0.179	0.003	0.119	0.179	0.277	0.91
λ	0.135	0.208	0.007	0.036	0.079	0.156	0.888
$RMSE_{IV}(\%)$	0.57	0.28	0.18	0.38	0.52	0.70	1.47
Panel B: NDX							
V_0	0.019	0.02	0	0.007	0.013	0.022	0.103
σ_V	0.398	0.178	0.071	0.289	0.372	0.469	0.956
ρ	-0.717	0.194	-0.989	-0.836	-0.744	-0.634	-0.19
p	0.189	0.16	0.1	0.114	0.155	0.211	0.642
μ_y	-0.221	0.19	-0.787	-0.3	-0.177	-0.09	0.11
σ_y	0.238	0.22	0.001	0.101	0.16	0.273	0.982
λ	0.176	0.324	0.007	0.039	0.083	0.188	1.666
$RMSE_{IV}(\%)$	0.49	0.24	0.13	0.32	0.44	0.59	1.35
Panel C: AMZN							
V_0	0.079	0.074	0.004	0.035	0.055	0.091	0.381
σ_V	0.586	0.321	0.047	0.358	0.546	0.754	1.6
ρ	-0.578	0.374	-0.99	-0.889	-0.615	-0.381	0.864
p	0.4	0.396	0.1	0.13	0.251	0.509	1.939
μ_y	-0.213	0.284	-0.798	-0.379	-0.192	-0.051	0.7
σ_y	0.53	0.346	0.001	0.172	0.549	0.874	1
λ	0.241	0.52	0.003	0.044	0.082	0.177	2.886
$RMSE_{IV}(\%)$	0.95	0.62	0.21	0.50	0.78	1.21	2.99
Panel D: FB							
V_0	0.08	0.084	0.001	0.033	0.057	0.096	0.404
σ_V	0.646	0.313	0.114	0.438	0.591	0.805	1.599
ρ	-0.613	0.327	-0.99	-0.881	-0.637	-0.447	0.818
p	0.327	0.339	0.1	0.122	0.195	0.394	1.815
μ_y	-0.157	0.306	-0.794	-0.352	-0.135	0.024	0.69
σ_y	0.551	0.339	0.001	0.224	0.58	0.884	1
λ	0.249	0.61	0.004	0.032	0.065	0.165	3.647
$RMSE_{IV}(\%)$	1.17	0.86	0.20	0.63	0.94	1.40	4.53

when considering that the P99 is around 1.4%. The $RMSE_{IV}$ for AMZN and FB are larger, 0.95% and 1.17%, respectively. This result needs to be interpreted while bearing in mind that the implied volatilities of the equity options are much higher than those of the index options (see Table 1). Nevertheless, more than 50% of the days have a $RMSE_{IV}$ of less than 1% and the 75th percentile is also close to 1%.

5.2 Bubble Estimation Results

In this section we analyze the daily bubble estimates. We first describe the results for the equity indexes (SPX and NDX) and then for the individual stocks (AMZN and FB).[36] Because our estimation is done day-by-day using only one option cross-section per day, and with the number of available options changing over time, the estimation results are subject to noise. As a result, interesting periods are those characterized by a *persistent* sequence of estimated positive bubbles rather than days with *isolated* positive bubbles.

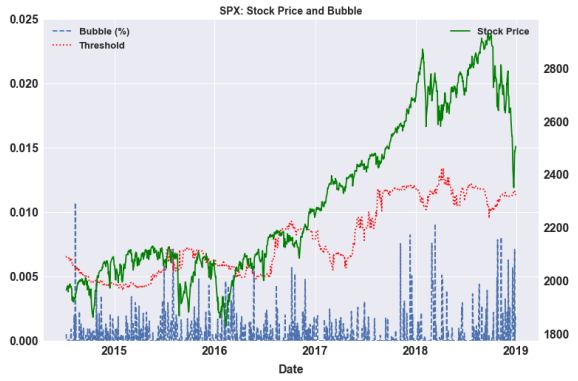
For each assets, Figure 5 shows on the right axes the daily asset price (solid lines, green color) and on the left axes the estimated bubble magnitudes as a percentage of stock prices, $\hat{\mathbb{B}}_t$, (dashed lines, blue color) and the time-varying critical values that corresponds to the conditional test described in section 3.4 for a one-side probability of 10%, i.e. we define a time varying threshold as $\mathbb{T}_t = 1.65\hat{\sigma}_t$.[37] We estimate the volatility of the error, σ_t, on each day, using the estimated negative bubbles, $\hat{\mathbb{B}}_t$, over the previous 180 days as described in Section 3.4. This implies that, at each time t, a bubble can be considered statistically significant whenever it exceeds the threshold $\mathbb{T}_t$.[38]

From Panels (a) and (b) of Figure 5 we notice that bubbles in both the SPX and NDX often lie in a tight range above zero, with occasional isolated spikes. We observe a different behavior in Panels (c) and (d) of Figure 5 for AMZN and FB. For both stocks the estimated bubbles are in the vicinity of zero most of the time, but they also show periods with *clusters* of large significant bubbles, in particular at the beginning of 2016 and at the beginning and at the end of 2018. The estimated bubbles for AMZN and FB are both more frequent and

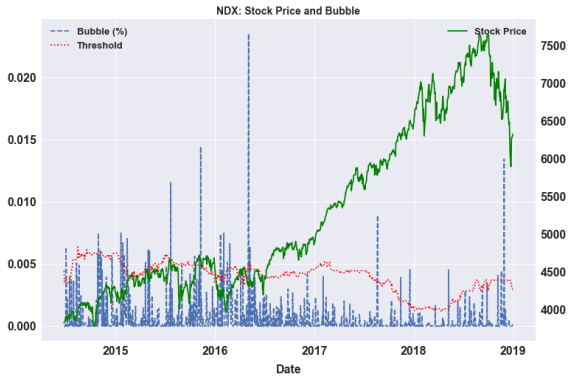
[36] When presenting the time-series plot, coherently with our framework, we report only the estimated positive bubble because, as described in Section 3, negative values reflect estimation error and market microstructure noise. The average size of the negative bubble is, of course, reflected in the time-varying threshold used in the conditional test described in Section. 3.4 and depicted with a red dotted line in Figure 5. Also, in order to reduce the impact of noisy data and to reduce concern about model mispecification, we exclude from our analysis the days with the 5% of highest $RMSE_{IV}$ for all the assets (these are the days in which the G-SVJ model has the poorest fit).

[37] In order to reduce the impact outliers we estimate σ_t via $\left(\sqrt{median(\hat{\sigma}^2_{N\_})}\frac{3}{2}\right)$ which is theoretically consistent with the assumption of a normal distribution for the error term ε_t.

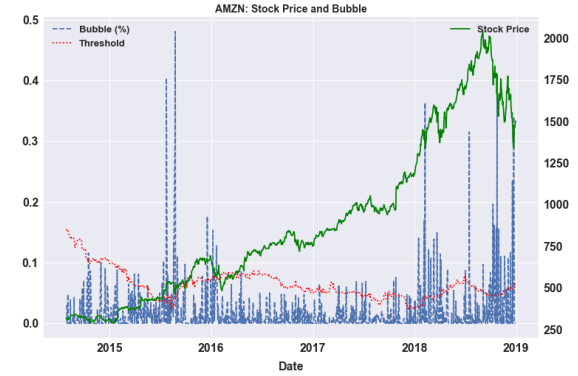
[38] Changing the probability level to 5% or 1% leads to quantitatively similar results.



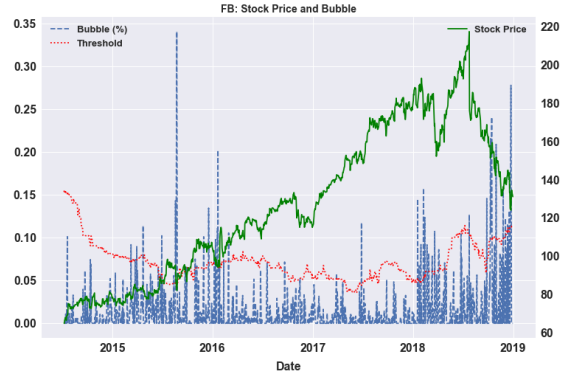
(a) SPX: Stock Price and Bubble



(b) NDX: Stock Price and Bubble



(c) AMZN: Stock Price and Bubble



(d) FB: Stock Price and Bubble

Figure 5: **SPX, NDX, AMZN, and FB: Stock Prices and Bubbles**. Panels (a), (b), (c) and (d) show on the right axes the stock prices (solid lines, green color) and the left axes the estimated bubble magnitudes as a percentage of stock prices (dashed lines, blue color). Each Panel also shows the time-varying threshold $\mathbb{T}_t = 1.65\hat{\sigma}_t$ (dotted lines, red color).

larger in magnitude than the bubbles estimated for both SPX and NDX. This finding can be explained by the fact that because indexes are weighted averages of their constituent stocks, the corresponding estimated bubble reflects the "average bubble" among its constituents, which will be much smaller and more stable over time relative to bubbles in single stocks.[39]

In order to illustrate how within our methodology the bubble estimates hinge on the differential pricing between put and call options, in Figure 6 we report the results on two representative days for AMZN and FB. Formally, our bubble estimates depend on the difference between option prices. However, in this figure we graph implied volatilities (which are

[39]For all four assets, the patterns in the bubble estimates reported in Figure 5 remain qualitatively and quantitatively similar when we compute bubbles as in Equation (23) using only ITM or OTM options, or when using only call options with positive trading volume.

non-linear functions of option prices) as they allow one to better visualize the gap between market and model call prices. On both days the estimated bubble is positive and significant. As can be seen in the left panels, on both days and for both assets, the fit of the G-SVJ model on the put options is extremely accurate, with a $RMSE_{IV}$ for AMZN equal to 1.5% and 0.8% for FB. For the same days, the right panels show the observed implied volatility of the call options along with the implied volatility generated by the G-SVJ model based on the parameter estimates obtained by fitting the model to put options. It is evident that there is a large and persistent (in the moneyness dimension) gap between the two lines. Intuitively, our bubble estimate is proportional to the area between the market observed call's implied volatility and the model implied volatility.[40]

The formal results of the conditional test are reported in the Table 3. The first row shows that while SPX and NDX have very rare bubble episodes, with a frequency of less than 1% over the entire sample, AMZN and FB display a larger number of significant bubble instances which are identified in around 5%-8% of the days. The bottom Panel of Table 3 reports the average $RMSE_{IV}$ option over the entire sample and over days with negative, positive, and significantly positive bubbles.

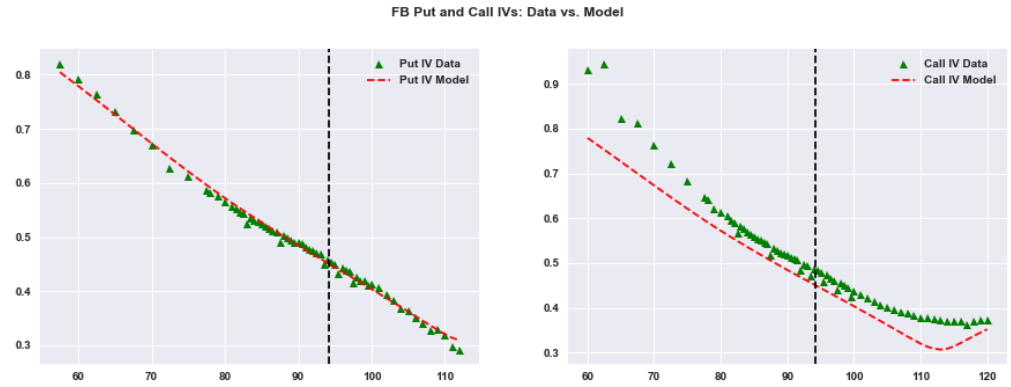
As seen from Figure 5, the only significant SPX bubble episodes is in 2014. NDX experienced slightly more instances of positive bubbles relative to SPX with relatively larger magnitudes. However, the few significant bubble in both indexes tend to be isolated. The latter fact, coupled with the fact that days with estimated positive bubble also tend to have larger pricing error than the average over the full sample (0.7% and 0.5% versus 0.51% and 0.44%, respectively), indicates that those few episodes should be considered with caution. In contrast, single stocks experience much more frequent and larger bubbles. Also, Table 3 shows that on days in which we detect a bubble (for both AMZN and FB), the $RMSE_{IV}$ is actually significantly smaller (0.68% and 0.71%) than the average $RMSE_{IV}$ over the entire sample (0.80% and 0.99%). Importantly, the $RMSE_{IV}$ for days with negative bubbles is higher than the average $RMSE_{IV}$. This provides supporting evidence for our assumption that days with negative bubbles are affected by estimation noise.

As for the bubbles magnitude, AMZN's bubble ranged between 10 and over 20 bps from mid-2015 to early 2016, followed by numerous instance of over 10 bps spread over the course of 2018. FB's bubble ranged above 10 bps from mid-2015 to early 2016, followed by numerous instances of over 15 bps in 2018.[41] As explained in Section 2.2, the magnitudes

[40] Recall from Equation (22) that when the call's market price is higher (lower) than the model call's price the estimated bubble is positive (negative).

[41] The overall bubble patterns as discussed above and as shown in Figure 5 are robust to bubbles computed

(a) AMZN: August 25, 2015



(b) FB: January 21, 2016

Figure 6: **AMZN and FB: Puts, Calls, and Bubbles**. The figure shows in the left (right) panels, the observed implied volatility of put (call) options, green triangles, and the implied volatility generated by the G-SVJ, red dashed line. Panel (a) corresponds to Amazon on August 25, 2015 (24 days-to-maturity). The $RMSE_{IV}$ in pricing the put options is equal to 1.5%. Panel (b) corresponds to Facebook on January 21, 2016 (29 days-to-maturity. and $RMSE_{IV}$ in pricing the put options is equal to 0.7%.

reported above reflect the amount of the bubble in the underlying asset that is embedded into the option contracts for a given maturity. Because the model horizon T is larger than the maturity of the option, the full size of the bubble can be orders of magnitude larger (see for example, the simulation study in Appendix C). For example, the average bubbles above 20 bps for AMZN in 2018 implies a bubble of ≈\$5.5 in dollar terms. Then, given that each option contract is written on 100 shares this in fact imply a ≈\$550 total dollar value. Similarly, the average of bubbles above 15 bps for FB in 2018 implies average bubble

over different delta ranges (e.g. deltas from 1 to 0.4 that mostly capture ITM options or deltas from 0 to 0.6 that mostly capture OTM options), pointing to the fact that, as theory suggests, the estimated bubbles do not depend from the option's strike price.

Table 3: **Conditional Tests**. The first row reports the percentages of days in which the null hypothesis of the conditional test in Section 3.4 (i.e. $\mathbb{B}_t = 0$) is rejected with a one-side probability of 10%. The second, third, fourth, and fifth rows report the average $RMSE_{IV}$ (i) over the entire sample period, (ii) on days when $\hat{\mathbb{B}}_t < 0$, (iii) on days when $\hat{\mathbb{B}}_t > 0$, and (iv) on days when the null is rejected, respectively. The sample size is equal to 1076 (there is a burn-in period of 180 days needed to compute the first estimate of the variance of ε_t dropping 5% of highest $RMSE_{IV}$ observations).

	SPX	NDX	AMZN	FB
% of days above threshold	0.37	3.43	8.00	5.48
	$RMSE_{IV}$(%)			
Entire Sample	0.51	0.44	**0.80**	**0.99**
Negative Bubbles	0.52	0.46	0.98	1.24
Positive Bubbles	0.50	0.43	0.66	0.76
Bubble Above Threshold	0.74	0.58	**0.68**	**0.71**

magnitude of ≈\$0.35 for a total of ≈\$35 value.

To summarize, our evidence points to the presence of short-lived but persistent bubble episodes in both AMZN and FB, over the 2014–2018. On the contrary, stock indexes, SPX and NDX do not show periods of persistent bubbles. In the next sections we focus on AMZN and FB and we investigate potential economic channels that can explain the presence of bubbles in those assets.

5.3 Implications of Price Bubbles

Given the above evidence about the presence of short-lived and persistent bubbles in AMZN and FB, in this section we discuss some economic implications of the estimated bubbles. We do so by analyzing four issues that are often associated with price bubbles: volatility, trading volume, earning announcements, and rapid price increases (or price bursts).

The volatility of the underlying asset plays a prominent role in the detection of bubbles because the explosive behavior of volatility (relative to the drift) determines whether a bubble exists. Our estimation results can be used to highlight the time-series relationship between bubbles and volatility. Indeed, the underlying *priced* volatility can be measured by the expectation of its risk-neutral quadratic variation:

$$\begin{aligned} QV_{[t,t+\tau]} &= \tfrac{1}{\tau} E_t \left[\int_t^{t+\tau} V_s ds + \textstyle\sum_{t<s\leq t+\tau} \Delta J_s^2 \right] \\ &\approx V_{0,t} + 2\lambda_t (e^{\mu_{y,t}+0.5\sigma_{y,t}^2} - \mu_{y,t} - 1), \end{aligned} \tag{25}$$

where the approximation relies on the fact that volatility tends to be highly persistent so that the integral and sum can be approximated by the spot variances. Figure 7 documents a large positive correlation between the estimated bubbles and the underlying quadratic variation. The correlation coefficient is equal to 0.58 for AMZN and 0.55 for FB.

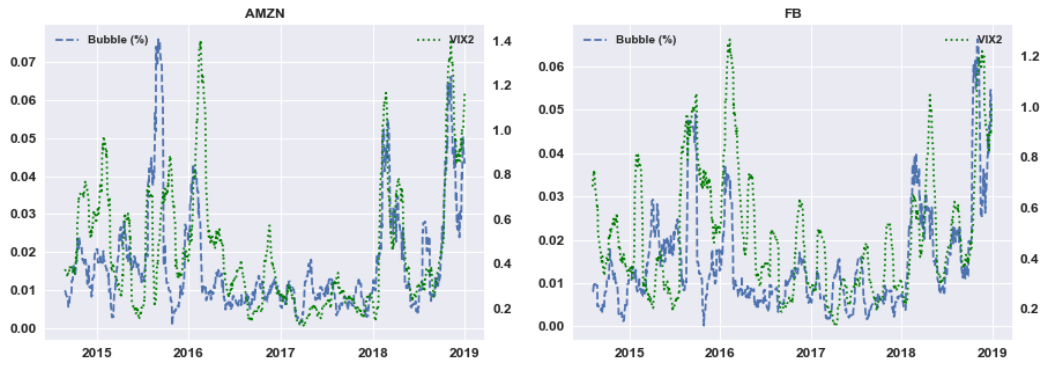


Figure 7: **Bubbles and Quadratic Variation**. The figure shows the time series of the smoothed (21-day backward looking moving average) positive bubble estimates (blue-line) and estimated quadratic variation (green-line) for AMZN (left panel) and FB (right panel).

Second, it is reasonable to believe that periods marked by pronounced price bubbles are also characterized by increased trading activity. The intuition is that higher trading volume is generally associated with increased volatility and that, as volatility increases, the asset price process is more likely to exhibit a large bubble.[42] To investigate the relationship between price bubbles and volume we estimated the following regression:

$$\hat{\mathbb{B}}_t^+ = \nu_0 + \nu_1 \times \ln\left(\text{Volume}_t\right) + \epsilon_t, \tag{26}$$

where $\hat{\mathbb{B}}_t^+$ refers to instances of the positive bubbles and $\ln\left(\text{Volume}_t\right)$ is the natural logarithm of the daily traded volume of the underlying asset (so that we can interpret the coefficient ν_1 as the effect on the bubble magnitude of a one percentage increase in the volume).

Table 4 shows the regression results for Equation (26). In particular, the coefficient ν_1 is positive and statistically significant for all four assets. A one percent increase in trading volume leads to an increase of about 0.0230%, and 0.0102% in the bubble magnitude for AMZN and FB, respectively. Not surprisingly the larger coefficient for AMZN is consistent with previous results showing that AMZN has the highest frequency of large and significant bubbles.

[42]This intuition is based, for example, on the fact that in a simple diffusive model for the stock price a bubble exists if and only if $\int_{\varepsilon}^{\infty} \frac{s}{\sigma^2(s)} ds < 0$, for some $\varepsilon > 0$ (see Delbaen and Shirakawa (2002)).

Table 4: **Bubble Magnitude and Trading Volume:** This table reports the results for the regression of the estimated positive bubbles on the trading volume of the underling asset. t-values are reported in parentheses. Coefficient estimates are reported as percents (%).

	AMZN	FB
Intercept	-0.3301	-0.1589
	(-8.28)	(-4.69)
Ln(Volume)	0.0230	0.0102
	(8.72)	(5.12)
R^2	0.06	0.02

Third, even though there could be numerous (and potentially unpredictable) factors that prompt the occurrence of a bubble, earnings announcements are of particular interest. These events are scheduled in advance and trading in both options and the underlying assets before earnings announcement days could be driven by expectations about price reactions to an announcement. In essence, earnings announcements allow us to directly test for an important instance of the buy-to-resell motive behind the bubble. We do so by running the following regression for AMZN and FB:

$$\hat{\mathbb{B}}_t^+ = \eta_0 + \eta_1 \times Dum_{1-14} + \eta_2 \times Dum_{15-28} + \eta_3 \times Dum_{29-42} + \epsilon_t, \tag{27}$$

where we define a dummy variable $Dum_{t_1-t_2} = 1$ if an earnings announcement is scheduled between time $t+t_1$ and $t+t_2$ and zero otherwise. The results of the regression are reported in the first and third columns of Table 5 (denoted by AMZN (1) and FB (1)). As seen, for both AMZN and FB the coefficients of the first and the second dummy variables $Dum_{1,14}$ and $Dum_{15,28}$, respectively, are positive and the first variable is particularly significant. Interestingly, the dummy variable coefficients are decreasing in general for both stocks, suggesting that, all else equal, the closer the earnings announcement date the larger the estimated bubble. Finally, the second and fourth columns of Table 5 (denoted by AMZN (2) and FB (2)), show that the effect of earnings announcements is robust to the inclusion of volume as a control.

Fourth, there is a widespread tendency to identify a period marked by a rapid asset price increase (or price burst) with the manifestation of a bubble. However, our framework and results show that rapid price increases do not necessarily imply the existence of a bubble. For example, from early 2016 to late 2017, FB's market price increased from around \$100 to nearly \$200, almost a 100% price increase. Similarly, AMZN's price increased from \$500

Table 5: **Bubbles and Earning Announcements**: This table reports the results for the regression of the estimated positive bubbles on a sequence of dummy variables that identify the number of days before an earning announcement. In particular, $Dum_{t_1-t_2} = 1$ if an earnings announcement is scheduled between time $t + t_1$ and $t + t_2$. t-statistics are reported in parentheses. Coefficient estimates are reported as percents (%).

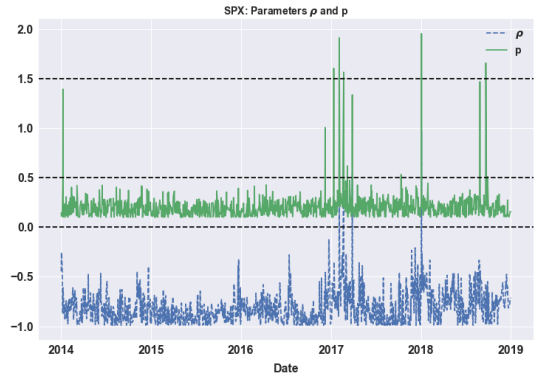
	AMZN	AMZN	FB	FB
	(1)	(2)	(1)	(2)
Intercept	0.0151	-0.3258	0.0116	-0.1362
	(7.51)	(-7.92)	(8.17)	(-4.27)
Dum_{1-14}	0.0056	0.0063	0.0078	0.0072
	(1.74)	(2.00)	(3.27)	(3.06)
Dum_{15-28}	0.0020	0.0045	0.0046	0.0049
	(0.61)	(1.43)	(1.89)	(2.06)
Dum_{29-42}	-0.0016	0.0025	-0.0022	-0.0010
	(-0.50)	(0.77)	(-0.91)	(-0.43)
Ln(Volume)		0.0224		0.0087
		(8.30)		(4.64)
R^2	0.01	0.07	0.02	0.04

in early 2016 to around \$1200 by late 2017, exceeding a 140% increase. However, we did not find many significant bubbles during this period. Similarly, NDX increased from around \$4000 in early 2016 to over \$7500 by mid-2018 (almost a 100% increase). SPX increased from around \$1800 to over \$2800 by mid 2018 (an increase in excess of 50%). Again, our estimates show no significant bubbles during this period. What matters for the occurrence of a bubble is the market's forward-looking expectations regarding company-specific and/or market-wide fundamentals and the future resale value of the asset. Our analysis shows that option prices, which are driven by investors' forward-looking expectations, can be effectively used to estimate price bubbles in underlying assets.

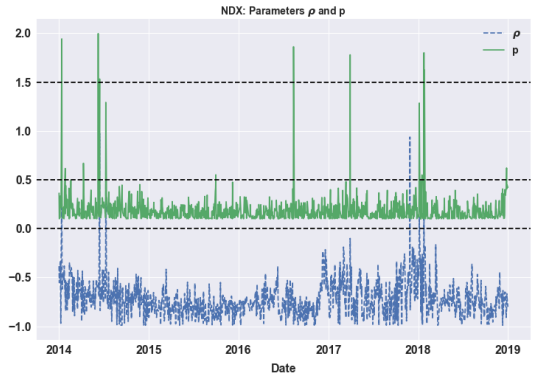
5.4 Robustness Check: Parameters and Bubbles

In this section we analyze the relationship between the estimated bubbles and the parameters of the underlying G-SVJD process. Given the limited sample and the small number of significant bubbles across all four assets, in order identify robust relationships in the data

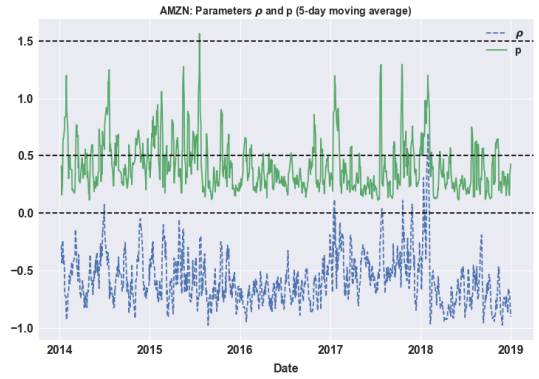
we focus on time-series averages of the estimated bubbles and the corresponding parameters ρ and p. Specifically, we first compute the 21-day moving averages of the estimated bubbles $\mathbb{B}_t$ (positive and negative) and parameters ρ_t and p_t for each asset and pool them together into one cross-section.[43]



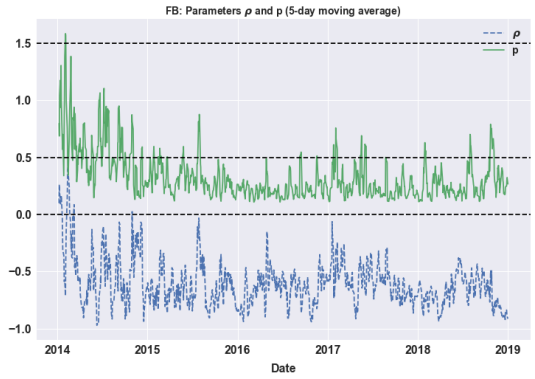
(a) SPX: Parameters ρ and p



(b) NDX: Parameters ρ and p



(c) AMZN: Parameters ρ and p



(d) FB: Parameters ρ and p

Figure 8: **SPX, NDX, AMZN, and FB: Parameter Estimates**. Panels (a), (b), (c), and (d) show the the time-series of the estimated parameters p (solid-lines, blue color) and ρ (dashed lines, green color). For AMZN and FB, five day moving averages are shown in Panels (c) and (d) to facilitate the observation of trends in parameter estimates.

Figures 8 shows the time-series of the estimated parameters p (solid-lines, blue color) and ρ (dashed lines, green color). For AMZN and FB, five day moving averages are shown in Panels (c) and (d) to facilitate the observation of trends in parameter estimates. Also, Table 2 reports summary statistics of the parameters' distribution. A few observations are

[43]Using a time-series average strikes a balance between controlling for negative bubble instances which are averaged out and producing a less noisy time-series bubble estimate. which allows to uncovering more stable relationships in the cross section of bubble and parameter estimates.

Table 6: **Correlation**: The table reports the correlation between ρ and p for all four assets (SPX, NDX, AMZN, and FB) and the pooled assets (All).

	SPX	NDX	AMZN	FB	All
$Corr(\rho_t, p_t)$	0.62	0.55	0.52	0.50	0.64

in order. First, while a large majority of option pricing models assumes both the correlation coefficient ρ and the variance's exponent in the volatility of volatility p to be constant, our empirical evidence points to a significant variation over time of these parameters. Second, as reported in Table 6, ρ and p are highly positively correlated, with the correlation coefficient ranging from 0.5 for FB to 0.62 for SPX. When we pool all the assets together as described in previous paragraph, the overall correlation is equal to 0.64 (last column of Table 6). It is important to notice that our estimation procedure does not force any ex-ante relationship between ρ and p. Hence, the evidence presented here suggests that the parameters governing the existence (or lack of) of a bubble have a tight empirical relationship.

The next natural step is to investigate whether days in which we estimate positive bubbles also see ρ_t and p_t lying in the appropriate region of the parameter space. Recall that, according to our theoretical framework, a bubble exists in the underlying asset if and only if $\rho > 0$ and $0.5 < p < 1.5$. However, given the noise in our estimates generated by market microstructure noise (bid-ask spread, stale quotes, etc.) and possible model mispecification, as discussed in Section 3.4, the correspondence between these parameter estimates and the bubble existence/magnitude will not be one-to-one. As such, our goal is to focus on the overall patterns between parameter estimates and the estimated bubble in order to determine whether they are reasonably consistent with the theory. We do this by pooling together, across assets, the 21-day moving averages of the estimated bubbles and parameters as described earlier. We then run the following OLS and logit regressions, which are meant to capture the relationship between the parameters ρ and p and both the bubble *magnitude* and bubble *probability*:

$$\hat{\mathbb{B}}^{+}_{t,i} = \beta_0 + \beta_1 \times \rho_{t,i} + \beta_2 \times Dum_{p,t,i} + \beta_3 \times (Dum_{p,t,i} \times \rho_{t,i}) + \epsilon_{t,i} \tag{28}$$

$$\text{Logit}(1_{\{\hat{\mathbb{B}}_{t,i} \geq 0\}}) = \beta_0 + \beta_1 \times \rho_{t,i} + \beta_2 \times Dum_{p,t,i} + \beta_3 \times (Dum_{p,t,i} \times \rho_{t,i}) + \epsilon_{t,i} \tag{29}$$

where $i \in \{SPX, NDX, AMZN, FB\}$, $Dum_{p,t,i} = 1$ whenever $0.5 < p_{t,i} < 1.5$ and zero otherwise. $Dum_{p,t,i} \times \rho_{t,i}$ is an interaction term that captures the non-linear relation between parameters and bubble.

Table 7: **OLS and Logit Regressions of Bubble on Parameters**: This table reports the coefficients of the OLS regression (28) and logistic regression (29) with t-statistics in parentheses. Coefficient estimates of OLS regression are reported as percents (%).

	OLS	Logit
Intercept	0.0029	-0.3445
	(5.33)	(-1.83)
ρ	0.0015	1.3737
	(2.00)	(5.11)
Dum_p	0.0057	-0.1605
	(5.43)	(-0.48)
$Dum_p \times \rho$	0.0061	-0.8753
	(3.05)	(-1.38)
R^2	0.0219	0.0096

Table 7 reports the pooled OLS and logit regression estimates for the pooled data, along with the corresponding t-statistic in parenthesis. As seen, the signs of the coefficients are consistent with our hypothesis: the total impact of ρ given $Dum_p = 1$, i.e. the sum of β_1 and the interaction coefficient β_3, is positive and significant. To understand these results, consider the case with dummy variable $Dum_p = 1$ (i.e. when $0.5 < p < 1.5$). Here, a 0.1 increase in correlation ρ (e.g., from 0 to 0.1) increases the bubble magnitude by $\approx 0.08\%$. Also, a 0.1 unit increase in ρ increases the probability of seeing a bubble by $\approx 2.3\%$.[44]

In summary, the parameters ρ and p of the G-SVJD processes, the parameters that determine whether or not a bubble exists in the underlying assets, are highly correlated among themselves and help to explain the occurrence and the magnitude of the bubble estimated from option prices. These results are particularly important because our estimation procedure does not enforce any relationship between the values of the parameters p and ρ nor on the *sign* and the *magnitude* of the estimated bubble. In conjunction, these results provide strong evidence that is consistent with both the existence of positive price bubbles and the validity of the local-martingale theory of bubbles.

[44]We compute the effect of a change in ρ on the predicted bubble probability in two steps. First, we calculate the average of the predicted probabilities in the original pooled data. Second, we add 0.1 to the estimated ρ and again compute the average of the predicted probabilities. The difference between the two average probabilities gives the marginal change in the probability for a small change in ρ, which is what we report in the text.

6 Case Study: GameStop and Nasdaq Dot-com Bubble

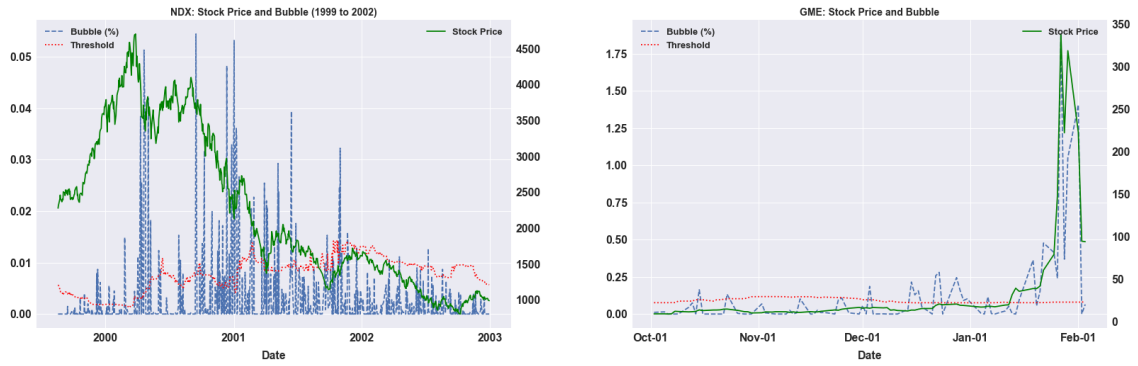
In this section we apply our methodology to NDX during the dot-com bubble in the early 2000s and GameStop (GME) between 2020 and 2021. The massive run up in the prices and subsequent crash of the technology sector, i.e. the alleged *dot-com* bubble of late 1990s and early 2000s is a perfect laboratory to check whether our methodology detects a bubble. We focus on the period from early 1999 to the end of 2002, which covers the full run up and crash cycle. Similarly, GME captured the public's and policy-makers' attention in January 2021 when its price exhibited unusual fluctuations and price appreciation for reasons not related to the company's fundamentals. GME's closing price increased from $35.50 on January 15th to $347.51 on January 27th and back to $92.41 by February 3rd. As such, GME represents another opportunity to see if our methodology can identify a price bubble.[45]

For NDX, we obtain data from Optionmetrics and apply the same filters as described in Section 4. We consider the period between January 1st, 1999 and December 31st, 2002. For GME, we obtain daily options quotes from the Chicago Board Options Exchange (CBOE) DataShop. Our sample period runs from April 1st, 2020 to February 3rd, 2021. Starting with April 2020 allows us to avoid the market turmoil due to the Covid-19 crisis. This choice provides sufficient sample size for our methodology, even after accounting for the burn-in period of 180 days needed to compute the time-varying threshold used in our conditional test for detecting bubbles. The data provided by the CBOE are similar to that provided by OptionMetrics. However, the CBOE also provides option quotes at 3:45PM Eastern Time, which are generally more reliable than closing prices because they have narrower bid-ask spreads. This is particularly important for stocks like GME which are generally illiquid. Thus, we use the 3:45PM quotes and we apply an analogous data filter as described in second paragraph of section 4.[46]

Panel (a) of Figure 9 shows the estimated bubble for NDX along with the time-varying threshold (left axis) and NDX price (right axis). Our method detects a bubble in NDX starting in March 2000. The bubble persists throughout the entire year and it slowly bursts

[45] News reports suggest that groups of retail traders fueled the trading activity in both GME stocks and options. Options are highly levered and if agents expect the prices to rise (fall) in the future, purchasing call (put) options is a well known strategy to benefit from such expectations. In a recent staff report (https://www.sec.gov/files/staff-report-equity-options-market-struction-conditions-early-2021.pdf) the SEC finds that both call and put options were actively traded: call options by investors who expected GME's price to keep increasing, and put option by investors who used them as an alternative say to short the stock.

[46] Also, in order to remove highly illiquid contracts we only retain put options with standardized moneyness ($= \ln(K/S)/(\sigma_{ATM}\sqrt{\tau})$) between - 4 and 5. This additional filter only removes 2% of observations which tend to have zero volume and open interest, and unusually high implied volatilities that exceeds 500%.



(a) NDX: Stock Price and Bubble

(b) GME: Stock Price and Bubble

Figure 9: **NDX & GME: Stock Prices and Bubbles**. In both panels, the right axis shows stock prices (solid lines, green color). The left axis show estimated bubble magnitudes as a percentage of stock prices (dashed lines, blue color) and time-varying threshold $\mathbb{T}_t = 1.65\hat{\sigma}_t$ (dotted lines, red color).

at the end of 2001. During this period the NDX price declined by more than 4500 points to around 1000 points. Beginning in 2002, the estimated bubbles reduce in magnitude and become insignificant.[47]

Analogous to NDX, Panel (b) of Figure 9 shows the estimation results for GME. As seen, while there are no significant bubbles in the months of October and November, there are a few instances of significant bubbles starting in mid to late December 2020. After that, as GME's price surges in the second half of January 2021, the estimated bubble increases almost in the same proportion. In particular, our estimates show that more than 1.5% of GME's price in late January is due to a bubble.[48] This is consistent with the extraordinary short term rise of the GME stock, daily rises in multiples of one hundred percent, from around $39 on January 20th, to $347 on January 27th, with an intra-day high in the upper $400s on January 28th. By February 3rd, GME's price dropped to $92, which is captured by our estimated bubble which drops significantly by the end of the period.[49]

Finally, the formal results of the conditional test as described in Section 3.4 are shown in Table 8. The first column refers to NDX while the second column is for GME. The first

[47] Even though the bubble's peak is generally identified after the bubble burst period, our approach is able to detect the ongoing/persistent bubble in NDX well before its ex-post identified peak, in real-time.

[48] In Appendix C we show that the parameters estimated from GME are compatible with a bubble ranging from 20% to 100% of the underlying asset over a one year horizon.

[49] Note that since NDX is an index where the bubble can also be seen as the weighted bubbles of individual stocks, its bubble magnitude will be generally smaller than those of the individual stocks. Despite this, it is interesting to observe high and persistent bubble magnitudes in NDX during the dot-com bubble era relative to occasional and smaller bubbles in a more recent period.

Table 8: **Conditional Tests**. The first row reports the percentages of days in which the null hypothesis of the conditional test (i.e. $\mathbb{B}_t = 0$) is rejected with a one-side probability of 10%. The second, third, fourth, and fifth rows report the average $RMSE_{IV}$ (i) over the entire sample period, (ii) on days when $\hat{\mathbb{B}}_t < 0$, (iii) on days when $\hat{\mathbb{B}}_t > 0$, and (iv) on days when the null is rejected, respectively.

	GME	NDX
% of days above threshold	9.98	27.71
$RMSE_{IV}(\%)$		
Entire Sample	**0.70**	**7.11**
Negative Bubbles	0.69	8.85
Positive Bubbles	0.75	6.21
Bubble Above Threshold	**0.71**	**5.28**

row of the table reports the percentage of days with estimated positive bubbles, while the lower part of the table reports the $RMSE_{IV}$ over the entire sample and sub-samples based on the signs and significance of the estimated bubbles. In the case of NDX, 10% of the days show a positive and significant bubble. Also, we notice that the average $RMSE_{IV}$ is equal to 0.70%. This is very similar to the pricing error reported in Table 3 in the case of NDX and SPX over the 2014–2018 period. Importantly, on days when we detect a bubble, the $RMSE_{IV}$ is similar to (0.71%), the average $RMSE_{IV}$ over the entire sample. In the case of GME, the average $RMSE_{IV}$ is equal to 7.11%. Although this is large when compared to the $RMSE_{IV}$ for AMZN and FB (which are around 1%), the average implied volatility of GME in our sample is 2.30, which is almost 6 times larger than those for AMZN and FB. Importantly, and consistent with the results in Table 3, on days when we detect a bubble, the $RMSE_{IV}$ is significantly lower (5.28%) than the average $RMSE_{IV}$ over the entire sample. This observation mitigates possible concerns about model misspecification. Also, the $RMSE_{IV}$ for days with negative bubbles is higher (8.25%) than the average $RMSE_{IV}$, thus providing support for our assumption that negative bubbles are the result of noise in the option data.[50]

Summarizing, our method identifies episodes of persistent bubbles during both the Nas-

[50] In the case of GME, one could be concerned that our results are influenced by specific and unusual market conditions. However, the SEC staff report (https://www.sec.gov/files/staff-report-equity-options-market-struction-conditions-early-2021.pdf) concludes that popular explanations for GME's price increase, such as a *short-squeeze* and/or a *gamma-squeeze* are not supported by the data. The SEC report concludes that the option market operated smoothly over the January–February 2021 period.

daq dot-com bubble and, more recently, during the rise and successive fall of GameStop price. Our analysis shows that our bubble detecting methodology can be easily implemented in real-time. Furthermore, we demonstrate that in the case of GME, options data are able to detect the onset of a bubble in mid December, when the stock price was still less than $15.

7 Conclusion

This paper presents a new approach to identifying asset price bubbles in real time using options data. We illustrate the method by applying it to the S&P 500 and Nasdaq-100 indexes and two highly liquid stocks: Amazon and Facebook between the 2014-2018 period. We also apply our methodology to the Nasaq dot-com bubble in the early 2000s and to the recent rise and fall in the price of GameStop stock. In both instances we detect large and significant bubble episodes. As such, our approach can be useful to both policy-makers and investors who need a forward-looking approach to daily, real-time monitoring of the financial system for risk-management.

A Appendix: Computational Details

The calibration discussed in section 3.2 is equivalent to solving a non-linear constrained and non-convex optimization problem (where parameters are restricted to certain bounds). Given the highly non-linear and non-convex nature of the objective function in Equation (21), we tested many optimization routines that are available in Python. We found Differential Evolution (DE), a global optimization algorithm, to be the best algorithm for our purpose. It is a derivative-free, stochastic population-based, optimization method that can search large areas of the parameter space.[51] A global optimization method is preferred to commonly used local optimizer (e.g. the widely used Nelder-Mead), as they are much less sensitive to the initial guess of the parameter values.

The accuracy and robustness of the optimization algorithm comes with a large computational cost. Despite the data filters discussed in Section 4, the estimation of the parameter vector on any single day for any of the underlying assets takes anywhere between two and 16 hours (on a system with an Intel-i7 processor at 3.60 GHz speed). We used the MARCC (Maryland Advanced Research Computing Center) system to carry out the entire estimation.[52]

B Appendix: Proofs

Theorem. Under the null hypothesis H_0 where $\mathbb{B}_t = 0$,

$$\sigma_t^2 = E\left(\sum_{t\in\Upsilon_-} \frac{1}{N_-}\hat{\mathbb{B}}_t^2\right) = \sum_{t\in\Upsilon_-} \frac{1}{N_-}\sigma_t^2.$$

Proof. The proof is based on the following lemma.

Lemma. Let x_t have mean zero, finite variance, with a symmetric and continuous distribution. Then, $Var(x_t) = E(x_t^2\,|x_t < 0)$.

Proof of Lemma.

$Var(x_t) = E(x_t^2) = E(x_t^2\,|x_t < 0)P(x_t < 0) + E(x_t^2\,|x_t > 0)P(x_t > 0)$

But, by symmetry, we have that $P(x_t < 0) = P(x_t > 0) = 1/2$, and $E(x_t^2\,|x_t < 0) = E(x_t^2\,|x_t > 0)$. Hence, $Var(x_t) = E(x_t^2\,|x_t < 0)$. This completes the proof of the lemma.

[51] See https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.differential_evolution.html for details on Python's implementation of DE. Note that, by default, DE refines the best parameters at the end with the L-BFGS-B algorithm.

[52] See https://www.marcc.jhu.edu/about-marcc/about-us/ for details about the MARCC system.

Given the lemma, we now prove the theorem.

Recall that $\hat{\mathbb{B}}_t = \mathbb{B}_t + \varepsilon_t$.

Under H_0 where $\mathbb{B}_t = 0$, conditioned on $t \in \Upsilon_-$.

$\sigma^2_{N_-} = E\left[\sum_{t\in\Upsilon_-} \frac{1}{N_-}\hat{\mathbb{B}}^2_t \mid t \in \Upsilon_-\right]$

$= E\left[\sum_{t\in\Upsilon_-} \frac{1}{N_-}\varepsilon^2_t \mid t \in \Upsilon_-\right]$

$= \sum_{t\in\Upsilon_-} \frac{1}{N_-} E\left[\varepsilon^2_t \,|\varepsilon_t < 0\right]$

This last inequality follows because the errors are independent.

By the lemma, letting $x_t = \varepsilon_t$, we have

$E\left[\varepsilon^2_t \,|\varepsilon_t < 0\right] = \sigma^2_t$.

Substitution yields

$= \sum_{t\in\Upsilon_-} \frac{1}{N_-}\sigma^2_t$, which completes the proof.

C Appendix: Simulation Exercise

In this section we study how the size of the estimated bubble depends on (1) the time-to-maturity of the options and (2) the value of the parameters p and ρ in the G-SVJD model.

Regarding the first point, Figure 2 depicts the intuitive idea that the size of the detected bubble increase as the maturity of the options increase and, hypothetically, if we could observe options with time-to-maturity equal to T (the terminal time) the option market would be able to fully reveal the size of the bubble in the underlying asset. However, (i) T is conceptually quite large (as it corresponds to the end of the economy) and (ii) most of the liquidity in the option market is concentrated on short-medium maturity options (in our empirical exercise we focus on options with maturity between 7 and 50 calendar days). The higher liquidity of short maturity options makes our inference more robust, but, at the same time, it also implies that our estimates represents a lower bound for the magnitude of the bubble in the underlying asset.

Regarding the second point, within the lenses of the G-SVJD model, two parameters (ρ and p) jointly determine whether the underlying asset is a martingale or strict supermartingale (see Section 3.1). The sections aims at determining how the above two parameters impact the size of the estimated bubble. We perform a simulation study using the G-SVJD model. We take the parameters from the estimation results on the GME stock. In particular, we take the median of the estimated parameters on the days when the bubble magnitude is higher than the threshold level $\mathbb{T}_t$ as seen in Figure 9.[53] We then consider a set of parameters

[53]The parameters are: $V_0 = 8.65$, $\bar{\nu} = 13.86$, $\kappa = 5$, $\sigma_v = 4.57$, $\mu_y = -0.79$, $\sigma_y = 0.87$, $\lambda = 0.93$.

ρ and p that make the G-SVJD process a martingale (which implies the absence of a bubble) or a supermartingale (which implies the presence of a bubble), as discussed in Section 2.1. For the no-bubble case we fix $\rho = -0.90$ and $p = 0.3$, while for the bubble case we consider all the combinations between $\rho \in \{0.1, 0.5, 0.9\}$ and $p \in \{0.6, 1, 1.4\}$, for a total of nine parameter combinations. Finally, we normalize the initial stock price to be $S_0 = 100$, and we set the risk-free rate and dividend yield to zero.

For each parameter vector we simulate $N = 10,000,000$ trajectories over a 1-year horizon (i.e. 360 days) using a daily frequency. For each maturity τ ($\tau = 1, \cdots, 360$), we:

- compute the price of a set (N_K) of put options with strike prices $K \in \{90, 95, 100, 105, 110\}$ as $P_{\tau,K} = \frac{1}{N}\sum_{n=1}^{N} \max(K - S_{\tau,n}, 0)$. Recall that, as described in Section 2, put options are not affected by a bubble in the underlying asset.

- Using the price of the underlying asset (S_0) and the put prices from above, we compute the market price of the corresponding call options ($C_{\tau,K}$) via put-call parity. In this case, if the underlying asset is affected by a bubble, the resulting call price would be affected by it as well.

- Then, we compute the fundamental price of the call options as $C^*_{\tau,K} = \frac{1}{N}\sum_{n=1}^{N} \max(S_{\tau,n} - K, 0)$.

- Finally, the bubble magnitude is given by $B_\tau = \frac{1}{N_K}\sum_{j=1}^{N_K}\left(C_{\tau,K_j} - C^*_{\tau,K_j}\right)$.[54]

The blue solid line in Figure 10 reports the results of the numerical exercise for the nine different parameter combinations. Regardless of the parameter combination, the bubble magnitude (B_τ) increases with the maturity of the options. While both parameters ρ and p affect the size and the trajectory of the estimated bubble, the figure seems to indicate that ρ has the largest effect on the bubble size. Importantly, Figure 10 shows that even a modest bubble over a short horizons, can be the reflection of a much larger bubble over longer horizons. Moreover, the simulation study confirms that the size of the bubble is independent of the strike price K. As a benchmark, the green line in Figure 10 shows the results for the no-bubble parameters.

As a further graphical illustration, Figure 11 highlights the effect of a bubble on puts' and calls' implied volatilities. The figure shows that (i) the implied volatility computed using put

[54] Alternatively, one can compute the size of the bubble, for a given horizon τ, as the difference between the price of the underlying asset and its fundamental value which we can computed as $FV_\tau = \frac{1}{N}\sum_{n=1}^{N} S_{\tau,n}$. Although such a computation is only feasible in a simulation setting we confirm that both approaches leads to identical results.

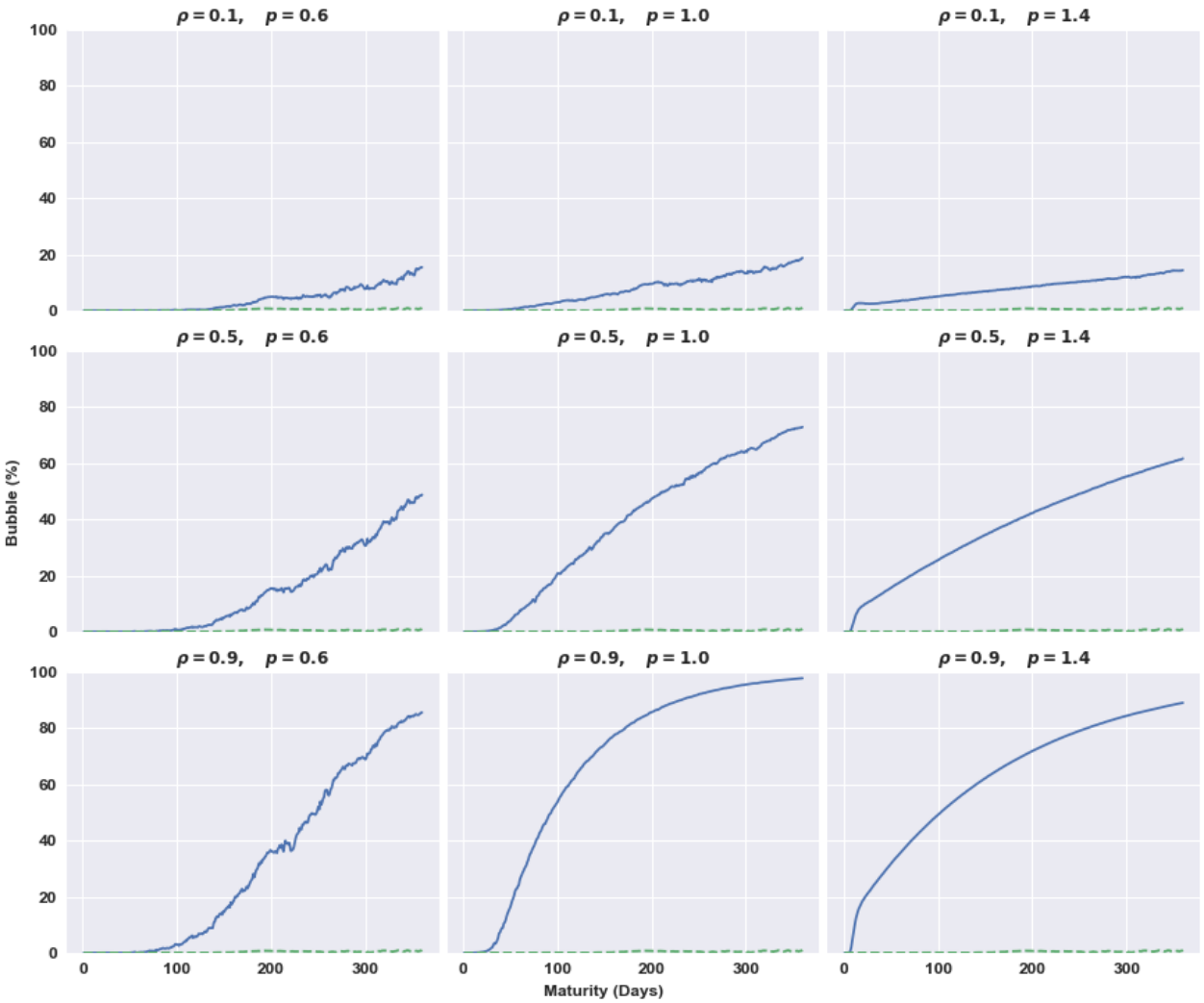


Figure 10: **Bubble's magnitude as a function of the option's time-to-maturity**: The x-axis shows the option maturity (in days) while y-axis shows the bubble magnitude (%). The blue solid-line corresponds to the case in which the parameters ρ and p are in the bubble range. (see shaded area in Figure 3).The green dotted-line shows the bubble magnitude (expected to be zero) when the above parameter are in the no-bubble region.

options (blue solid line) is always equal to the implied volatility computed using call prices obtained via put-call parity (red dots); (ii) in the presence of a bubble the implied volatility computed on call prices obtained via risk-neutral pricing (dotted black line) is consistently lower (thus, the price of those call options will be lower as well); and (iii) the gap between the IV curves (and, as a consequence, the magnitude of the bubble,) grows with the time to maturity of the options.

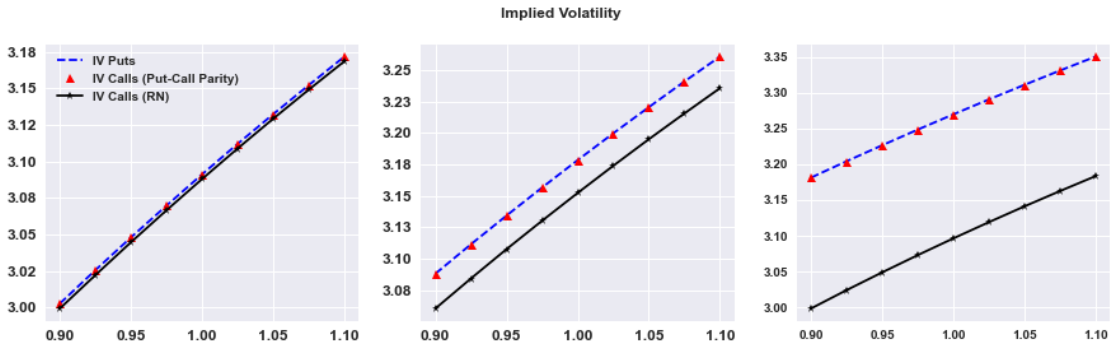


Figure 11: **Bubble and options implied volatility**: The three panels show the implied volatility according to the G-SVJ model with parameters $V_0 = 8.65$, $\bar{\nu} = 13.86$, $\kappa = 5$, $\sigma_v = 4.57$, $\mu_y = -0.79$, $\sigma_y = 0.87$, $\lambda = 0.93$. $\rho = 0.9$ and $p = 1$ are chosen to make the G-SVJ model a supermartingale (i.e. in the bubble region). The left, middle, and right panels correspond to options with time-to-maturity of 10, 20 and 30 days, respectively. In each panel, the solid (blue) lines show the implied volatility computed from put prices, the (red) dots show the implied volatility computed from call prices obtained via put-call parity, and the dotted (black) lines shows the implied volatility computed from call prices obtained via risk-neutral pricing.

References

Andersen, L. and V. Piterbarg (2007). Moment Explosions in Stochastic Volatility Models. *Finance and Stochastics 11*, 29–50. 1, 3.1, 23

Andersen, T. G., O. Bondarenko, and M. T. Gonzalez-Perez (2015). Exploring Return Dynamics via Corridor Implied Volatility. *The Review of Financial Studies 28*(10), 2902–2945. 4

Andersen, T. G., N. Fusari, and V. Todorov (2015). Parametric Inference and Dynamic State Recovery from Option Panels. *Econometrica 83*, 1081–1145. 1, 3.1, 3.2

Andersen, T. G., N. Fusari, and V. Todorov (2017). Short-term market risks implied by weekly options. *The Journal of Finance 72*(3), 1335–1386. 1, 3.2, 4, 5.1

Andersen, T. G., N. Fusari, V. Todorov, and R. T. Varneskov (2019). Unified Inference for Nonlinear Factor Models from Panels with Fixed and Large Time Span. *Journal of Econometrics, forthcoming 212*, 4–25. 3.2

Bandi, F., N. Fusari, and R. Reno' (2021). Structural Stochastic Volatility. working paper. 1

Bates, D. S. (1996). Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options. *Review of Financial Studies 9*, 69–107. 1, 3.1, 3.1, 24

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of political economy 81*(3), 637–654. 3

Carr, P. and L. Wu (2003). What Type of Process Underlies Options? A Simple Robust Test. *Journal of Finance 58*, 2581–2610. 1

Christoffersen, P. and K. Jacobs (2004). The Importance of the Loss Function in Option Valuation. *Journal of Financial Economics 72*, 291–318. 3.2

Cox, A. and D. Hobson (2005). Local Martingales, Bubbles and Option Prices. *Finance and Stochastics 9*, 477–492. 1, 2

Delbaen, F. and H. Shirakawa (2002). No Arbitrage Condition for Positive Diffusion Price Processes. *Asia-Pacific Financial Markets 9*, 159–68. 2.1, 42

Duffie, D., J. Pan, and K. Singleton (2000). Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica 68*, 1343–1376. 1, 3.1

Giglio, S., M. Maggiori, and J. Stroebel (2016). No-bubble condition: Model-free tests in housing markets. *Econometrica 84*, 1047–1091. 11

Harrison, M. and D. Kreps (1978). Speculative Investor Behavior in a Stock-Market with Heterogeneous Expectations. *Quarterly Journal of Economics 92*, 323–336. 1

Herdegen, M. and M. Schweizer (2016). Strong bubbles and strict local martingales. *International Journal of Theoretical and Applied Finance 19*. 2.1

Heston, S., M. Loewenstein, and G. Willard (2007). Options and Bubbles. *Review of Financial Studies 20*, 359–390. 1, 2, 2.1

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies 6*(2), 327–343. 1, 3.1, 24

Jacod, J. and P. Protter (2010). Risk Neutral Compatibility with Option Prices. *Finance and Stochastics 14*, 285–315. 19

Jarrow, R. (2015). Asset price bubbles. *Review of Financial Economics 7*, 201–218. 1, 1, 2, 2.1, 2.1

Jarrow, R. (2016). Asset price bubbles and the land of oz. *The Journal of Portfolio Theory Winter*, 1–6. 10

Jarrow, R. (2021). *Continuous-Time Asset Pricing Theory: A Martingale-Based Approach. 2nd ed.* Springer. 15, 16, 2.1, 2.1, 3.2

Jarrow, R. and S. Kwok (2020). Inferring financial bubbles from option data. Working paper, Cornell University. 1

Jarrow, R. and S. Lamichhane (2021). Asset price bubbles, market liquidity, and systemic risk. *Mathematics and Financial Economics 15*, 5–40. 32

Jarrow, R., P. Protter, and J. San Martin (2020). Asset price bubbles: An invariance theorem. Working paper, Cornell University. 2.1, 3.1

Jarrow, R., P. Protter, and K. Shimbo (2010). Asset Price Bubbles in Incomplete Markets. *Mathematical Finance 20*, 145–185. 1, 1, 1, 2, 2.1, 16, 2.2, 2.2.1, 2.2.2, 2.2.2

Lions, P. and M. Musiela (2007). Correlations and bounds for stochastic volatility models. *Annales de l'Institut Henri Poincare (C) Non Linear Analysis 24*, 1–16. 3.1

Loewenstein, M. and G. Willard (2000a). Local martingales, arbitrage and viability: free snacks and cheap thrills. *Economic Theory 16*, 135–161. 1, 2

Loewenstein, M. and G. Willard (2000b). Rational equilibrium asset pricing bubbles in continuous trading models. *Journal of Economic Theory 91*, 17–58. 1, 2

Loewenstein, M. and G. Willard (2013). Consumption and bubbles. *Journal of Economic Theory 148*, 563–600. 2.1

Medvedev, A. and O. Scaillet (2007). Approximation and calibration of short-term implied volatilities under jump-diffusion stochastic volatility. *Review of Financial Studies 20*, 427–459. 1

Merton, R. (1973). Theory of rational option pricing. *Bell Journal of Economics 4*, 141–183. 2.1

Merton, R. (1976). Option Pricing when Underlying Asset Returns are Discontinuous. *Journal of Financial Economics 3*, 125–144. 1, 3.1, 3.1, 24

Phillips, P. C. B. and S. Shi (2018). Real time monitoring of asset markets: Bubbles and crises. Cowles Foundation Discussion Paper 2152, Yale University. 1

Phillips, P. C. B., S. Shi, and J. Yu (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the s&p 500. *International Economic Review 56*, 1043–1078. 1

Phillips, P. C. B., S. Shi, and J. Yu (2015b). Testing for multiple bubbles: Limit theory of real-time detectors. *International Economic Review 56*, 1079–1134. 1

Scheinkman, J. (2013). Speculation, trading and bubbles. Working Paper 050, Princeton University. 1

Todorov, V. (2019). Nonparametric spot volatility from options. *Annals of Applied Probability 29*, 3590–3636. 1

Xiong, W. (2013). Bubbles, crises, and heterogeneousbeliefs. NBER Working Paper 18905, NBER. 1