Predicting the Unpredictable: Forecastable Bubbles and Business Cycles under Rational Expectations

Eduard Gracia

Abstract A popular interpretation of the Rational Expectations/Efficient Markets hypothesis states that, if it holds, market valuations must follow a random walk; hence, the hypothesis is frequently criticized on the basis of empirical evidence against such a prediction. Yet this reasoning incurs what we could call the ‘fallacy of probability diffusion symmetry’: although market efficiency does indeed imply that the mean (i.e. “expected”) path must rule out any cyclical or otherwise arbitrage-enabling pattern, if the probability diffusion process is asymmetric the observed path will most closely resemble not the mean but the median, which is not subject to this condition. In this context, this paper develops an efficient markets model where the median path of Tobin’s q ratio displays regular, periodic cycles of bubbles and crashes reflecting an agency problem between investors and producers. The model is tested against U.S. market data, and its results suggest that such a regular cycle does indeed exist and is statistically significant. The aggregate production model in Gracia (Uncertainty and Capacity Constraints: Reconsidering the Aggregate Production Function, 2011) is then put forward to show how financial fluctuations can drive the business cycle by periodically impacting aggregate productivity and, as a consequence, GDP growth.

JEL E22, E23, E32, G12, G14

Keywords Rational expectations; efficient markets; financial bubbles; stock markets; booms and crashes; Tobin’s q; business cycles; economic rents

Correspondence Eduard Gracia, 39 Palace Court, London W2 4LS, United Kingdom; e-mail: eduard.gracia@booz.com


© Author(s) 2012. Licensed under a Creative Commons License - Attribution-NonCommercial 2.0 Germany
“The next major bust, 18 years after the 1990 downturn, will be around 2008, if there is no major interruption such as a global war”

Fred Foldvary (1997)

1 Introduction: The Fallacy of Diffusion Symmetry

In the wake of the 2008 recession, as of every major recession for the last 150 years, the question of why the downturn happened and whether the dismal science should have predicted it has been posed again. In a way, this follows an old tradition: the Long Depression in the 1870s triggered the Marginalist Revolution, the Great Depression in the 1930s gave birth to Keynesianism, and the Oil Crisis in the 1970s marked the ascendancy of the Rational Expectations Hypothesis. In its 21st Century edition, a substantial part of the challenge seems to focus on the latter hypothesis, and whether a return to adaptive expectations / bounded rationality would improve the models’ explanatory power respective to empirical observations. Unfortunately the debate, at least at a popular level, is vitiated by a logical fallacy: what we could call the ‘fallacy of probability diffusion symmetry’.

The popular controversy goes as follows. Mainly under the neoclassical banner, defenders of the Rational Expectations / Efficient Markets hypothesis claim that, should markets not behave according to it, they would create arbitrage opportunities that would enrich anyone clever enough to spot them, and it would therefore suffice to add a few rational players in the mix for their irrational competitors to be driven out of business. This, they argue, means that observed market values must follow a random walk, and therefore stock market bubbles and crashes be utterly unpredictable, for any predictable patterns should already be discounted out. Those in the opposite camp take then this conclusion as the central contention point, against which they pose numerous examples of departures from the random walk hypothesis in the observed data, as well as names of economists who, against the mainstream opinion, were able to forecast the 2008 crisis well before it hit.

From a strictly logical perspective, however, both sides might well be wrong. Indeed, both camps are implicitly assuming that, if expectations of future market prices were rational, they would reflect the mean (or ‘expected’) path calculated on the basis of all the information available; therefore, if current prices reflect the fu-
ture expected path, their observed trajectory should also approximate the mean. Hence, if arbitrage ruled out any predictable patterns along the mean path, then they should also be absent from the observed path – and, as a consequence, proving their existence in the observed data series would also prove that expectations were not rational in the first place. Yet, sensible as it sounds, this implication is only true if the underlying probability diffusion process is symmetric.

An intuitive example may make this point clear. Imagine a game of triple-or-nothing: you make a bet of, say, $10, toss a coin and triple the investment if result is heads, or else lose it all if it is tails. Evidently, if the game is played only once, the distribution of results is symmetric, with a 50% probability of making triple or nothing, and with both mean (i.e. average) and median (i.e. the value that leaves 50% of the distribution on either side) being $15. Yet if we play the game, say, ten times in a row reinvesting the profits every time, the distribution changes: now there is a 0.098% probability of making $590,490 and a 99.9% of losing everything, so the mean value is $577 (a substantial profit respect to the $10 investment), but the median is obviously zero. Under Rational Expectations, $577 is of course the investment’s expected value; yet, for an external observer viewing the data series, there is a 99.9% probability that the observed value after ten tosses be nil i.e. closest to the median path. In other words: the prediction with the highest probability of success is not the mean, but the median.

This is not just the case for this straightforward example but for any asymmetric distribution, including those most frequently used in standard financial modeling. Appendix 1 illustrates this for perhaps the simplest random-walk asset return model: the Brownian motion. It follows that, if probability diffusion processes are asymmetric (as virtually all the standard asset pricing models are), then neither does the observation of predictable market patterns imply irrationality, nor does the Efficient Markets Hypothesis rule them out, for there is nothing preventing them from appearing on the median path.1 Hence, the classical papers by Fama (1965) and Samuelson (1965) postulating that valuations in a rational, efficient market must preclude any cyclical or otherwise arbitrage-enabling pattern remain completely valid – only, they apply solely to the mean path.

1 This, importantly, has nothing to do with the “fallacy of ergodicity” highlighted by authors such as Paul Davidson (e.g. Davidson 2009): the binomial random process in the example above, for instance, is totally ergodic and has a perfectly well understood probability distribution.
This, importantly, does not imply that we are in front of a money machine, for it does not mean investors can resort to those predictable patterns to ‘beat the market’ but, at most, to fine-tune their probability of loss. For example, in the game of triple-or-nothing, a rational investor may bet $10 and expect to end up with $577, but by not betting may avoid suffering a loss 99.9% of the times… and, by selling short (assuming this were allowed in the game), might achieve a gain 99.9% of the times even with a negative expected value of the position – more or less like the trapeze artist who makes a bit of money every night at the cost of risking life and limb. Similarly, investing in assets with, say, an abnormally high P/E ratio may not indicate any form of ‘irrational exuberance’ on the investors’ side but a rational valuation of an asset with, say, a high expected gain as well as a high probability of loss.2 Under these conditions, to be sure, there is nothing preventing as smart analyst from repeatedly issuing a successful contrarian prediction: in the game of triple-or-nothing, for instance, our analysts could consistently forecast a loss of $10, and be right 99.9% of the time.

There are unfortunately very few examples of papers in the literature where the association of the observed time series to the median instead of the mean trajectory plays a role in the core analysis. One of these few is Roll (1992), which proves that portfolio managers who are measured against their deviation from a market index are essentially being forced to track the market median path, which is suboptimal respective to the mean. A much more direct precedent, however, is Gracia (2005), which shows how an efficient market subject to a normal random-walk perturbation may display a persistent, periodic cycle of asset valuation bubbles and crashes along the median path, even though the mean remains cycle-free.3

The main purpose of this paper is to develop a model of financial valuations in an efficient, rational expectations market such that it would result in a persistent cycle along its median path – and hence, given a representative enough sample, in the observed path. The model is directly inspired in Gracia (2005), although it has been revised to make it both more general and more parsimonious. It portrays the

---

2 Note that this differs from the rational bubbles literature following Blanchard and Watson (1982), which requires asset values to diverge from their fundamentals – something the model in this paper does not require.

cycle as the result of rent-seeking behavior by those controlling the production process, which results in a predator-prey cycle; in this regard, it is also a descendant of Goodwin (1967) ‘class struggle’ model, albeit in a rational-expectations context. Its predictions are then tested on the basis of two closely-related indicators of U.S. stock market valuation: Robert Shiller’s Cyclically-Adjusted Price-Earnings ratio (CAPE) (see for instance Shiller 2005) and Stephen Wright’s long-run estimates of Tobin’s \( q \) ratio (Wright 2004). In this context, testing against Shiller’s CAPE metric is particularly relevant because it has proven to be a very good leading indicator of financial bubbles and therefore, if it happened to display a strong cyclical component, it would constitute \textit{prima facie} evidence that bubbles also do.

We resort then to the aggregate production function model developed in Gracia (2011) to explain how such financial market behavior would lead to macro-economic business cycles consistent with key stylized facts without implying any sort of irrational behavior.\(^4\) This production function, which is built on neoclassical, rational-expectations micro foundations, shows how a higher share of rents over total output value leads to lower aggregate productivity; hence, rent-seeking leads to production inefficiency. This is actually quite intuitive: economic rents, by definition, reflect constraints and therefore inefficiencies (as departures from perfect competition), so rent-seekers must increase those inefficiencies if they want to expand their overall share of the pie. As Gracia (2011) proves empirically (against U.S. aggregate GDP data) that this production function is a better specification than the Cobb-Douglas function, it is entirely legitimate to resort here to it.

Note that, since these empirical tests rejected both the Cobb-Douglas function as a specification and aggregate capital as an explanatory variable, the Gracia (2011) production function should not be regarded as complementary but as a better-fitting alternative to any function posing aggregate capital as an input. Counterintuitive as this might seem, it is in fact quite logical: in a world where capital may be invested to support both productive and rent-seeking activities, there is no way to know to what extent any given, additional unit of capital actually contributes to increasing aggregate output instead of creating rent-enhancing

constraints (such as oligopolistic rents or other barriers of entry)\(^5\) resulting in inefficiencies.

This paper is structured as follows. Section 2 outlines recent literature on successful market predictions, and elaborates a bit more on the questions this paper tries to answer. Section 3 provides an intuitive explanation of the model; for the sake of readability, analytical developments have been committed to Appendix 2. Section 4 proceeds then to test the model’s key predictions. Subsequently, Section 5 discusses the macroeconomic implications of this model when combined with Gracia (2011). Finally, Section 6 provides a summary of the paper’s main findings and conclusions.

2 Predicting the Unpredictable

As so many economists kept asserting that no one could have seen the latest downturn coming,\(^6\) it became almost a popular pastime in heterodox circles to quote those who had most egregiously missed it as well as those who had got it most spectacularly right. There is certainly no shortage of the latter. A 2010 contest in the Real-World Economics Review Blog, the “Revere Award”, shortlisted twelve economists who made particularly accurate predictions of the crash (Dean Baker, Wynne Godley, Michael Hudson, Steve Keen, Paul Krugman, Jakob Brøchner Madsen, Ann Pettifor, Kurt Richebächer, Nouriel Roubini, Robert Shiller, George Soros and Joseph Stiglitz). Far from being complete, this list has been heavily criticized due to some glaring omissions such as Marc Farber, Fred Foldvary, Fred

---

\(^5\) For example, when a company invests, say, in a marketing campaign it is impossible to know to what extent it contributes to improving its market recognition as an alternative supplier (which would enhance efficiency by fostering competition), to what extent it transfers some demand, and profits, from its competitors to itself (which might represent a mere transfer, leaving aggregate rents the same) and to what extent it enhances its oligopolistic market power through product differentiation (which would increase their oligopoly rents and therefore pull the market further away from perfect competition). As all three effects would take place simultaneously, net aggregate impact is uncertain.

\(^6\) For a good survey of such assertions see for example Bezemer (2009).
Harrison, Michael Hudson, Eric Janszen, Raghuram Rajan, Peter Schiff or Nassim Nicholas Taleb.  

Many if not most of these authors reached their conclusions by considering various financial indicators in the context of non-neutral-money, bounded-rationality or otherwise inefficient-markets models. Moreover, by relying on cyclical market patterns some were able to make uncannily accurate predictions even before the financial indicators would raise any grounds for concern. This is the case of Foldvary (1997) and Harrison (1997) who, on the basis of the 18-year cycle Hoyt (1933) observed in the Chicago real estate market, forecasted the 2008 global recession more than a decade in advance. Not that this was a one-off success either: Harrison (1983), for example, resorted to the same real estate cycle pattern to predict the 1992 downturn as a follow up from the one in 1974.

Even among the supporters of bounded rationality, the existence of predictable long-range cycles spanning many years or even decades is frequently regarded with skepticism. It has not always been so, though: until the 1930s it was commonly accepted in academic circles that economic waves existed and had well-defined frequencies. More than 150 years ago, Juglar (1862) observed a trade cycle with a wavelength of 7–11 years associated to fixed asset investment. Then, in the early 20th Century, similar findings came in quick succession: Kitchin (1923) identified a shorter, inventory-driven cycle lasting around 3–5 years, Kondratiev (1926) a long wave lasting 45–60 years, and Kuznets (1930) an intermediate swing lasting 15–25 years, which he associated to building activity (thus linking it to the 18-year property cycle Hoyt 1933 would identify shortly afterwards). Yet, as Burns and Mitchell (1946) argued, under more strict empirical tests the evidence was far from conclusive, so the view that business cycles are, as Zarnowitz (1992) put it, “recurrent but non-periodic” gradually took hold.

Identifying cycles on the basis of aggregate GDP data is indeed fraught with technical difficulties, as the data series are either not long enough or not homogeneous enough. Yet stock market datasets, which are more granular, have been empirically proven to reflect mean-reversion patterns akin to business cycles. Thus, for example, Fama and French (1988) as well as Poterba and Summers (1988) identified a 3–5 year cycle in stock market returns i.e. the same frequency

---

7 References for each one of these authors’ predictions can be found, for example, in Bezemer (2009) and Gaffney (2011) as well as in the Real-World Economics Review Blog: http://rwer.wordpress.com/2010/03/31/shortlist-for-the-revere-award-for-economics-3/
Kitchin had reported 65 years before. Furthermore, those who successfully predicted the crash without relying on a cyclical pattern generally did this on the basis of their observation of anomalous values for certain publicly-available indicators such as the CAPE ratio (e.g. Shiller 2005) or the debt-to-GDP ratio (e.g. Keen 2006). Under market efficiency, the information contained in those variables should already have been discounted from the mean path.

At the same time, the fact that these predictions foresaw not only fluctuations in the financial markets but also in the aggregate economy poses a different question: how can a financial crash impact the real economy in a strictly rational world? How could, for instance, a liquidity shortage impact GDP in a world where money is neutral along the expected path? One of the difficulties here is that, as Kydland and Prescott (1991) pointed out, around 70% of the cycle is statistically explained by (i.e. correlated to) fluctuations of Total Factor Productivity (TFP). On one side, this means that (re)introducing bounded rationality, market frictions and violations of Say’s Law to justify some sort of financial illusion impacting demand does not suffice, for one does not just need to explain how financial shocks might trigger recessions but also how they could cause output to fall through a productivity drop. On the other hand, however, it is not enough to assume, as Real Business Cycle models proposed (starting with Kydland and Prescott 1982 and then Long and Plosser 1983), that such productivity fluctuations obey to exogenous technology shocks for, as Galí (1996) and Shea (1998) found, there is virtually no correlation between TFP fluctuations and actual technology shocks. These rules out models (e.g. Kiyotaki and Moore 1997) that portray credit crises as a result of technology shocks, as they reverse the causality in a way that is inconsistent with the actual experience of financial downturns.

In sum, the questions the following sections will try to answer are the following:

---

8 Of course, once one has abandoned the constraints of rational behavior it is always possible to find a set of assumptions to fit any given set of observations, but this goes against Occam’s Razor just as a regression with zero degrees of liberty achieves perfect fit but explains nothing at all.
Can an efficient stock market display a periodic cycle of bubbles and crashes?
How could such financial phenomena impact aggregate productivity (and, through it, aggregate output) in a rational expectations economy?

Section 3 outlines how (a) could happen; a rationale for (b) is offered in Section 5.

3 A Model of Periodic Stock Market Cycles

Let’s imagine a market with two kinds of players:

• Producers who manage and have control of a company’s productive process.
• Investors who contribute their resources in exchange for a return.

This of course defines an agency problem i.e. a situation where agents (producers in this case) act on behalf of principals (that is, investors) but, at the same time, can also use their privileged position and knowledge to exploit them. In the absence of any control mechanism by investors, to be sure, the information asymmetry in favor of producers would open an easy route for them to exploit investors simply by raising cash, transferring it to their own private accounts or otherwise spending it for their personal purposes and then declaring bankruptcy. Even with controls, this exploitation can take many forms. Managers, for example, may assign to themselves salaries and perks above market level, or may decide to make empire-building investments that increase their power but yield poor returns, or may use their privileged information to do some insider dealing; other employees, similarly, may also use their privileged information to shirk their duties and/or to generate efficiency rents for themselves.

To counter these dangers, investors implement a system of punishments and rewards including, for example, regulation against fraud, audits and bureaucratic controls, the threat of takeover and dismissal, etc. To be sure, these controls are neither free nor foolproof, so their implementation will be subject to a cost-benefit analysis: investors will only impose them up to the point where their marginal benefits (i.e. the expected reduction of depredation costs) equal their marginal costs. Ultimately, the benchmark against which investors will gauge the firm’s performance is its liquidation value i.e. whether the assets the enterprise manages

9 See Appendix 2 for an analytical version of the reasoning in Section 3.
would be worth more or less if sold in the market instead of being managed within the company’s framework. A straightforward way to measure this cost of opportunity is the ratio of a company’s market price divided by its assets’ replacement value i.e. what is known as Tobin’s q: if the ratio falls below unity, the firm is actually destroying value and therefore liquidation becomes an attractive option. Since liquidation (or, in a less dramatic version, reorganization and dismissal of the worst-offending producers) puts a hard stop to the producers’ rent extraction, it represents the investors’ ultimate weapon; yet, to the extent it is neither instant nor cost-free, its deterrence power is also limited.

Under these conditions, the producers’ rent extraction takes place as follows. In a given production process structure, the producer has a certain degree of control that translates into a given percentage of the output being siphoned out as rents. At every given point in time, a certain number of new opportunities to modify this productive structure will randomly pop up. Other things being equal, the producers’ decisions are of course biased in favor of the options that generate the highest level of rents, but their power to choose is limited by the control mechanisms imposed by investors. Rational producers will therefore ‘pluck the goose’ just enough to maximize their future expected rents, which also means leaving just enough to keep investors happy. Markets will then price these companies accordingly, so that the expected growth of their value equals their discount rate less (plus) the weight of the net cash flow they distribute to (raise from) investors. The higher this price stands above the threshold value (i.e. the higher Tobin’s q ratio), to be sure, the lower the probability that investors liquidate, and therefore the more freedom of action producers will have to maximize their rents within the boundaries of the investors’ control mechanisms.

In an efficient market, the mean path is of course the one where both investors’ and producers’ expectations are fulfilled, so asset values grow precisely at their market discount rate less the dividends they cash out. Yet, as Appendix 1 proves for a geometric Wiener process, when the probability diffusion process is asymmetric the best approximation to the observed time series is the median, not the mean path… And along the median path (as along any path different from the mean) the players’ expectations consistently fail to be met, and thus constantly need to be realigned.

Needless to say, Tobin’s q is the standard metric first put forward by James Tobin (1969).
Hence, if one assumes now a geometric Wiener perturbation on market valuations (which is an asymmetric function such that the median always falls below the mean), the observed interaction of producers and investors will result in a cycle. Indeed, as new opportunities to pluck the goose appear, producers take them only to the extent they do not expect them to bring the company’s return to investors below the market discount rate – yet this leads them to overshoot more often than not, as the underlying business’ median return is lower than the mean. When this happens, as long as the companies’ market price remains above liquidation value, the surge of liquidations resulting from this market correction remains small. Sooner or later, though, the accumulated impact of excessive rent extraction pulls valuations below asset replacement values. As a result, the firms whose $q$-ratio is the lowest, which are in principle those whose producers have been the most aggressive rent-extractors, are gradually liquidated (as liquidation is not an instant process), thereby eventually reducing the weight of producer rents in the system until growth can be resumed.\textsuperscript{11}

Appendix 2 develops this reasoning analytically and shows how, under the stated assumptions, the median path of the company’s $q$-ratio displays a predator-prey cycle punctuated by bubbles and crashes. Specifically, it behaves according to the following expression:

\[
\begin{align*}
\frac{dq_t}{q_t} &= \left( \pi - \frac{\sigma^2}{2} \rho_t \right) dt \\
\frac{d\rho_t}{\rho_t} &= \lambda (q_t - 1) dt
\end{align*}
\]

where $\pi, \sigma^2, \lambda > 0$ are positive parameters

\textsuperscript{11} To be sure, when a company’s assets are liquidated they are bought by (and thus incorporated into) another company, so liquidation does not necessarily lead to a reduction in the volume of assets put to productive purposes; yet, to the extent the firms liquidated are also those that extracted the highest rents, the average rate of rent extraction must come down as a result. The same outcome is of course to be expected from “low key” liquidations i.e. reorganizations where the worst-offending producers are fired and the company processes are rearranged to further limit shirking opportunities.
Where ‘\( q_t \)’ represents Tobin’s \( q \) ratio, ‘\( \rho_t \)’ is the ratio of producers’ rents divided by investors’ income, and ‘\( \pi \)’, ‘\( \sigma^2 \)’ and ‘\( \lambda \)’ are parameters capturing, respectively, the median equity premium, the variance on equity returns and the speed of liquidation. Figure 1 illustrates how solvency and the rent ratio behave in such a model.\(^{12}\)

Five comments are worth making at this point:

• In this model, the parameters ‘\( \pi \)’, ‘\( \sigma^2 \)’ and ‘\( \lambda \)’ determine the frequency of the cycle, so that its wavelength is longer the smaller they are, and becomes infinite if any of them is zero.\(^{13}\)

• Regular as these waves appear, they are not deterministic. What makes them periodic is that, as the relative weight of rents climbs up, it progressively takes a smaller negative shock to trigger a solvency crisis, so that, at the point where its probability exceeds 50\%, the median path starts a downturn.

• The timing of bubbles and crashes would therefore be predictable to some extent (albeit never with absolute certainty) both on the basis of the timing since the last crash and of the behavior of key variables (e.g. Tobin’s \( q \) or any related time series, such as CAPE, shooting above its long-term central value).

• The fall of \( q \) as a result of the “crash” is always steeper than its increase during the preceding boom – which of course fits well the historical experience of financial bubbles (after all, they are called ‘bubbles’ because they ‘pop’).

• There is no need that the market price depart from the net present value of future cash flows, as in Blanchard and Watson’s (1982) rational bubbles: here, the mean path may be perfectly consistent with its fundamentals, yet the periodic bubbles and crashes depicted in Figure 1 would appear all the same.

Theoretical models are of not much use, though, without empirical confirmation. Hence, in the next section we test whether this model’s main predictions hold.

---

\(^{12}\) Note that, to develop this diagram, the specific values of the parameters ‘\( \pi \)’, ‘\( \sigma^2 \)’ and ‘\( \lambda \)’ have not been selected to be realistic but merely to make the shape more evident to the viewer.

\(^{13}\) Although, incidentally, if any of them is infinite then the cycle also disappears by collapsing to the trivial solution \( q_t = \rho_t = 0 \).
Figure 1: Predator-prey dynamics in the model in Appendix 2 (median path)

4 Testing the Model

Testing Strategy

Separating the value of producers’ rents from the marginal productivity of their services is inherently tricky, if nothing else because it is precisely their privileged knowledge that allows them to extract them from investors, so they must necessarily be difficult for an external observer to detect. Tobin’s $q$ is, on the other hand, not too hard to calculate through various, fairly standard calculation methods. Several authors (e.g. Wright 2004) argue convincingly that the value obtained through these methods may be downward-biased due to systematic overvaluation of asset values at replacement cost (more or less in the same way that a used car is not worth the same as a new one, so its liquidation value is below replacement cost). Yet, although this bias might quite possibly compromise a test of Tobin’s $q$ convergence to unity, it poses no particular challenge if all we want to test is whether it does indeed display a cyclical behavior.

The key challenge is finding a data time series that is granular enough yet covers a period long enough to actually test this. Indeed, however one may want to
estimate them, realistic values for the parameters in expression (1) are always small (equity premiums and average annual variances are typically single-digit, and the process of liquidating or reorganizing the management of an underperforming company can also be quite time-consuming), which means cycles may well stretch over many years. If we want to test whether such cycles do actually exist, therefore, we need a reliable, homogeneous sample that covers the span of the underlying cycle several times over so that it can provide a reasonable degree of confidence to accept or reject the hypothesis.

In addition, there is a question on what test to perform. To prevent any bias towards acceptance of the hypothesis, we will rule out any test that somehow forces a positive answer: if nothing else, we should remind ourselves that, per Fourier’s theorem, any continuous function can be decomposed into an aggregation of sinusoidal curves, so “fitting” a cyclical function to the sample, or even decomposing it as a spectrum, does not necessarily prove the point either. This is also why filters and calibrations are dangerous, as it is not always clear whether the cycle actually exists in the underlying data or is just a result of the interaction between the sample and the filter.

To avoid these pitfalls, this paper will resort to a simple autocorrelation test on the unfiltered data, and demand that its correlograms display the same pattern as the theoretical model, as illustrated in Figure 2.

Specifically, the autocorrelation function (ACF) should display:

- Positive and negative autocorrelation intervals alternating over time.
- Regular time lags between positive and negative autocorrelation intervals.
- Symmetric, sinusoidal-looking autocorrelation waves.\(^{14}\)
- Decreasing autocorrelation amplitude (as the function is stationary).

Whereas the partial autocorrelation function (PACF) should show:

- A high, positive partial autocorrelation for the first lag.
- A smaller, negative autocorrelation for the second lag, potentially followed by a short series of gradually smaller autocorrelations for the next lags.\(^{15}\)

\(^{14}\) This is so even if the underlying cycle is not symmetric (as in a predator-prey model). The reason, intuitively, is that the diagram reflects the correlation with both low and high past values, so even if, say, crashes are steeper than booms, the lags compensate each other and thus the result is a symmetric wave on the correlogram.
Figure 2: Theoretical correlograms for the $q$-ratio in Appendix 2 (median path)

Note that although, faced with this correlogram, it might seem natural to resort to the Box-Jenkins methodology and fit a linear ARIMA function to the sample, we will not do so here, as it would not help to prove or disprove the paper’s hypothesis that the underlying process follows a non-linear expression such as (1).

**Test Data and Results**

The most complete dataset available for Tobin’s $q$ is perhaps Wright (2004), which spans from 1900 to 2002; Wright put forward several computations\(^\text{16}\) of which we will take here the one he calls “equity $q$”. Unfortunately an annual series such as this is not granular enough to test long-wave autocorrelations, so we resort

\(^{15}\) Although this may or may not be observable as it quickly falls below the significance threshold.

\(^{16}\) The data can be downloaded from [http://www.econ.bbk.ac.uk/faculty/wright/](http://www.econ.bbk.ac.uk/faculty/wright/).
to Shiller’s Cyclically Adjusted Price Earnings (CAPE) ratio, which has a
correlation coefficient of 85% with the equity $q$ but provides a long monthly
series,\textsuperscript{17} as a proxy. The CAPE (just a standard P/E except that the last 10 years’
earnings average is taken as the denominator to minimize short-term fluctuations)
has proven to be a very powerful tool to identify market bubbles, as was in fact the
basis of Shiller’s successful prediction of both the 2008 ‘subprime’ and the 2000
‘dot.com’ crashes. It is thus particularly intriguing to test whether the CAPE
contains a cyclical component both due to its link to stock market bubbles and
because, to the extent its behavior can be explained through a rational expectations
model, it would not be necessary to resort to bounded rationality or any other form
of market inefficiency as an explanatory hypothesis. In short, Shiller’s CAPE
yields the correlogram shown in Figure 3.

Allowing for the usual dampening impact of random noise, these results are
entirely consistent with the theoretical pattern in Figure 2. Moreover, the
autocorrelations are quite high (up to $\pm 0.3$ i.e. 30% autocorrelation coefficient)
and exceed comfortably the normal 95% significance threshold. The wave displays
a very low frequency: the average wavelength for the cycles completed within
the sample is close to 403 months (i.e. \textit{circa} 33.5 years) with a standard deviation
of just 17%. This is consistent with the observed values of the constants in (1)
typically being single-digit rates (particularly the equity premium $\pi$ and the return
variance $\sigma^2$, as the liquidation rate $\lambda$ may be more difficult to assess), which is
bound to lead to a slow-motion cycle.

Now, of course, to check whether this cyclical behavior is also present in the $q$
ratio or exclusive of CAPE we need to run the same test on the equity $q$ data series
(Figure 4).

Not only does the cyclical pattern reappear, but its average wavelength (just
over 31 years, with a standard deviation of only 13.5%) is remarkably close to the
CAPE sample,\textsuperscript{18} and so is also the maximum autocorrelation of \textit{circa} $\pm 30\%$ ...

\\textsuperscript{17} The data are regularly updated in Shiller’s website, \url{http://www.econ.yale.edu/~shiller/data.htm}.
The high correlation with Tobin’s $q$ has been repeatedly highlighted before even in non-academic
circles e.g. the Smithers & Co. website \url{http://www.smithers.co.uk/page.php?id=34}.

\\textsuperscript{18} This is particularly remarkable because the CAPE data tested are monthly instead of annual as in
the $q$ ratio, and cover a longer time span (1881 to 2011 for CAPE vs. 1900 to 2002 for the equity
$q$ data provided by Wright 2004). The coincidence of identifying a cycle of just over 30 years in
both samples therefore suggests significant long-term stability in the underlying parameters.
only, due of course to the paucity of data, in this case the correlation wave exceeds the threshold for 95% confidence a lot less comfortably. This may also explain why, unlike the CAPE partial autocorrelation function (which essentially matched the theoretical diagram in Figure 2), the $q$ ratio PACF only displays the first autocorrelation lag as statistically significant: arguably, this is due to the data scarcity in this series, which pulls most autocorrelations under the significance threshold.

Furthermore, this long-wave cycle has quite a high degree of explanatory power: in fact, even a simple sinusoidal curve with a 31-year wavelength (i.e. the cycle time suggested by the analysis in Figure 4) has a 72% correlation coefficient respective to the equity $q$ series – or, what is the same, can be said to explain more

*Figure 3: CAPE autocorrelation diagrams (monthly data, Jan. 1881 to Oct. 2011)*
Figure 4: Equity $q$ autocorrelation diagram (annual data, 1900 to 2002)

than half ($R^2 = 52\%$) of its observed variation. This, however, poses another question: although we know the correlogram yields symmetric waves even against an asymmetric wave, if a simple sinusoidal curve already has such a high explanatory power, how do we know the underlying waves are not sinusoidal anyway? After all, with a very small change in the basic assumptions of Appendix 2 one could have turned the model’s median path into a sinusoidal (i.e. symmetric) wave.\footnote{Specifically, it would suffice to replace the expression $d\rho_t = \lambda(q_t - 1)\rho_t \, dt$ in Assumption 7 with the expression $d\rho_t = \lambda(\ln q_t) \, dt$ to turn the median path into a pure sinusoidal wave.} The evidence, however, suggests that the actual path behaves just as in a predator-prey wave, where the fall is steeper (and therefore also shorter in time) than the climb-up. Indeed, according to Wright’s time series there was an increase of the equity $q$ in 57\% of the years in the sample and a fall in only 43\% of them. In
fact, even if we restrict this calculation to the period 1900 to 1993 (as the tracking sinusoidal wave appears to reach in 1993 the same point the cycle as in 1900), the average of climb and fall years still holds: 57% upwards and 43% downwards.

In sum, the evidence supports the hypothesis that the bubbles and crashes observed for the 100+ years on record could largely be explained through an efficient-markets, predator-prey cycle model. The underlying model itself may not necessarily be exactly the one put forward in this paper (to begin with, a model such as Gracia 2005 would behave in a similar way), but the hypothesis that a long-wave cycle with steeper slopes downwards that upwards (i.e. with a predator-prey shape) explains most of the observed equity q behavior cannot be rejected.

Two side comments are probably appropriate at this point:

- Although this long cycle is consistent with the model in this paper (for equity premium and return variances are usually single-digit over long periods of time), it is worth mentioning once again that its identification is not model-dependent. The finding of any cycle under these conditions is hence all the more robust, as it was not “imposed” through any filtering, regression or calibration that could make something visible when it actually does not exist.
- This wavelength of just over 30 years is, on the other hand, quite different from that of most classical studies: it definitely does not match the waves identified by Kitchin (3-5 years), Juglar (7-11 years), Kuznets (15-25 years) or Kondratiev (45-60 years). Whether the frequency we have identified here might result from the combination of others or represent something entirely different, however, is not a question we can answer on the basis of the evidence above, and will therefore remain out of scope for this paper.

5 Aggregate Productivity Impact

The next question is how these oscillatory financial phenomena could, under rational expectations, have an impact not only on GDP growth but also on TFP (which, per Kydland and Prescott 1991, typically explains 70% of the business cycle). The standard Cobb-Douglas production function does not lend itself very well to this (unless modified through ancillary assumptions, of course), as its TFP growth rate is exogenous ex hypothesi. The answer is, conversely, quite straightforward on the basis of the aggregate production function put forward in
Gracia (2011). As this function has been empirically proven to be a better specification of aggregate U.S. GDP behavior (and the Cobb-Douglas function was in fact rejected as a valid specification in a non-nested model comparison test), it is entirely legitimate to resort here to it.

Both the analytical development and an extensive discussion of the rationale for this aggregate function are available online, so here we will only outline its intuitive logic. The key difference respective to the Cobb-Douglas function is that Gracia (2011) does not assume perfect competition and therefore allows for the existence of economic rents. Economic rents may result from any form of resource constraint: physical availability (i.e. Ricardian rents), control over supply (monopoly rents), information asymmetry (agency rents), etc. By definition, these constitute fixed costs (or fixed assets) above marginal costs (or income), so they cannot be instantly flexible to unexpected changes in demand volumes: one cannot change the capacity of a plant from one day to the next. A positive demand shock, therefore, will face higher unit costs (i.e. lower productivity) as extra production is initially subject to capacity constraints, whereas a negative shock will not reduce costs (i.e. increase productivity) in the same proportion due to the rigidity of fixed costs.

Hence, if we consider a closed Walrasian economy where Say’s Law applies so that an increase in one product’s demand must equal a decrease in another product’s, any unexpected change in the composition of demand, swapping demand from one product to another, will result in a fall of productivity. This fall will be more severe the higher the weight of rents over total income (for rents represent the economic translation of capacity constraints) and the higher the variability of demand composition. In other words: rents are a measure of the productive system’s rigidity so, for a given level of demand uncertainty, the higher this rigidity, the lower the system’s productivity will be. Conversely, in such a world capital aggregates do not tell us much about productive capacity, because investments can be made both to support strictly productive and rent-seeking

---


21 By introducing ancillary assumptions it is of course possible to develop (as the literature does extensively) a Cobb-Douglas-based model where economic rents exist; yet, per Gracia (2011), this raises methodological questions related to the aggregate production function’s micro foundations.
activities: as the former enhance overall productive capacity, but the latter diminish it, the aggregate impact of capital on output is ambiguous, and so its statistical significance must be (as it actually is) low.

This links up directly with the rent-driven cycle model we developed in Section 3 and Appendix 2. In good times, producers (meaning anyone with some control over the means of production) gradually find ways to increase their rent extraction, which of course imposes additional rigidity on the productive process. During periods of growth, the weight of these rents over the rest of output (i.e. the producers’ rent ratio \( \rho \)) does not grow very quickly, so its impact on productivity is also small. Yet, per Figure 1, when the bubble finally bursts and Tobin’s \( q \) starts its downward spiral, the rent ratio shoots up because their variable portion (i.e. the marginal cost) drops faster than the fixed costs (that is, the economic rents). Hence, as the rent ratio goes up, productivity must fall.

In short, the prediction is that financial recessions will impact output growth through the fall in aggregate productivity they cause. We can now resort to the production function in Gracia (2011) to show how. Analytically, the function is as follows:

\[
\bar{Y}_t = \bar{A}_t \bar{H}_t^{1+\bar{\rho}_t}
\]  

(2)

Where \( \bar{Y}_t \) represents GDP, \( \bar{H}_t \) work hours, \( \bar{\rho}_t \) the average rent ratio\(^{22} \) and \( \bar{A}_t \) the productivity factor, which follows the expression:

\[
\frac{d\bar{A}_t}{\bar{A}_t} = \left( \Gamma_t + \frac{\sigma^2 - \delta_t}{2} \bar{\rho}_t \right) dt + \bar{s}_t d\bar{W}_t
\]  

(3)

Where \( \Gamma_t dt + \bar{s}_t d\bar{W}_t \) is the technology-driven component of productivity growth (which, as in the standard literature, is broken down into a deterministic growth

\(^{22}\) This ratio includes not only producers’ rents but also any other type of economic rents in the overall economy: in fact, Gracia (2011) tested the model against another type of rent (namely, the risk-free interest rate net of monetary dilution) which was easier to measure. For convenience, in the remainder of this section we will assume all the rents other than producers’ rents to be constant, but of course in a more sophisticated analysis such an assumption logically ought to be relaxed.
rate $\Gamma_t$ plus a serially-uncorrelated Gaussian perturbation component $d\bar{W}_t$ whose standard deviation is $\bar{s}_t$), $\sigma_t^2$ the non-technology-driven portion of overall demand growth variance and $\bar{\delta}_t$ the weighted sum of individual product demand variances.\(^\text{23}\)

Thus, for example, if this were a closed Walrasian economy with a known resource endowment, overall demand would only vary with productivity shocks, which means that $\sigma_t^2 = 0$, whereas the weighted sum $\bar{\delta}_t$ would be positive, for it can only be nil if each and every product’s demand is deterministic. It follows that, at least in a closed economy (as would be the whole world, or an economy whose foreign sector is proportionally small, such as the USA or the EU) the factor $\frac{\sigma_t^2 - \bar{\delta}_t}{2}$ should have a negative sign and therefore, the higher the rent ratio $\bar{\rho}_t$, the lower the rate of productivity growth would be. Hence, if we assume technology growth as well as the difference $(\sigma_t^2 - \bar{\delta}_t)$ to be constant, the median rent we depicted in Figure 1 translates into the aggregate productivity path shown in Figure 5.

\[\text{Figure 5: Productivity Growth Rate over the Median Path (illustrative)}\]

\[\text{time} \rightarrow\]

\[\text{23} \text{ Technically, this is defined as an aggregate} \quad \bar{\delta}_t \equiv \sum_{j=1}^{m} \alpha_{j,t} \delta_{j,t} \text{ of all the products / production units } j \in \{1..m\}, \text{ where } \alpha_{j,t} \text{ is the relative share of output value of every individual production unit and } \delta_{j,t} \text{ is defined as } \delta_{j,t} = \sigma_{j,t}^2 \cdot \beta_{j,t}, \text{ defining, in turn, } \sigma_{j,t}^2 \text{ as the portion of the demand variance of each product / production unit } j \text{ that is not driven by a technology shock and } \beta_{j,t} \text{ as the relative share of labor input corresponding to each one of those products / production units.}\]
In other words, productivity must behave over time just like the rent ratio curve in Figure 1 (only upside down) i.e. displaying a cycle whose dips are steeper than the previous or subsequent climb up periods, and deeper respective to the trend than the booms stand above it. This behavior is consistent with the findings of Neftçi (1984), Sichel (1993), Ramsey and Rothman (1996), Verbrugge (1997) or Razzak (2001), who conclude that business cycles present “deepness” (i.e. recessions tend to fall deeper than expansions are tall respective to the trend) and “steepness” (i.e. the fall into recession is steeper than the climb up back to expansion).

One is reminded at this point of Simon Johnson’s views on crises in emerging economies, which he largely based on his experience as chief economist of the IMF (e.g. Johnson 2009 or Johnson and Kwak 2010). Johnson explains that, despite the wide diversity of their triggering events, economic crises always look depressingly similar, because they all result from powerful, privileged elites overreaching in good times to maximize their rents, but resisting the pressure to cut back on them when their excessive risk taking results in a credit crisis. Johnson also makes a very strong case that the U.S. 2008 credit crisis presents exactly the same profile, with the U.S. financial sector playing the role of the privileged elite. This is, of course, precisely the sort of mechanism portrayed in Appendix 2 as well as in Gracia (2005).

Furthermore, these crises are usually associated to low productivity growth rates during as well as in the years before the recession – just as Figure 1 would suggest. This correlation was first noticed when applying growth accounting to the Soviet economy (e.g. Powell 1968 or Ofer 1987): post-war growth rates looked impressive for a long while but, despite the progress of Soviet technology, productivity barely grew at all from around 1950, and actually fell from 1970 onwards, leading to the system collapse in 1991. Except perhaps for a few die-hard Marxists, the causal link between the decline and fall of the Soviet economy and its massive, rent-seeking state apparatus is just beyond question. This famously led Krugman (1994) to state, on the basis of the observation of comparably low productivity growth patterns among the so-called East Asian “Tiger” economies in the ‘80s and early ‘90s (e.g. Young 1992, 1994 and 1995, or Kim and Lau 1994), that the East Asian rapid economic expansion trend was unsustainable. When, as predicted, the 1997-98 Asian financial crisis hit and brought the growth period to an abrupt halt, its root cause was, unsurprisingly, also found largely in the rent-
seeking “crony capitalism” that politically well-connected elites practiced in those countries.

More recently, the same failure to increase productivity during a boom, leading to loss of competitiveness and hence excessive debt accumulation to finance a large current account deficit, has been highlighted in a recent World Bank special report (Gill and Raiser 2012) as the driver of the financial crisis’ severity in Greece, Italy, Portugal and Spain. This report finds, for instance, that the low quality of government services (low rule of law, low government effectiveness, low control of corruption) correlates strongly with these countries’ sluggish productivity growth. Such institutional failures, of course, translate into rent extraction opportunities for various power groups: wasteful public investments that benefit well-connected construction moguls, abnormally high salaries and perks for public servants, rigid labor laws, guild privileges, onerous business regulation, etc.

There is indeed abundant evidence of the impact of institutions controlled by rent-seeking groups on macroeconomic growth. Interpretations of historical events and macroeconomic fluctuations built on this premise and supported by both quantitative and qualitative evidence go at least as far back as North (1981), Olson (1982) or McNeill (1982), not to talk about more recent, extensive pieces of work such as North (1990), Olson (2000), Boix (2003), Ostrom (2005) or, most recently, Acemoglu and Robinson (2012). The reality and relevance of this phenomenon is therefore well established.

In this context, what the production function in Gracia (2011) implicitly highlights is that those rent-seeking groups are not irrationally imposing perverse rent-extraction institutions when they could choose more efficient redistribution mechanisms. Since any redistribution flow constitutes an economic rent, it can only exist on the basis of a rent-generating constraint, a form of rigidity imposed on the income-generation process – and, in an uncertain world, rigidity translates into inefficiency. In other words: productivity and redistribution are inextricably linked, and poor quality institutions do not survive despite their inefficiencies, but precisely because of them, as the constraints they impose on economic activity enable those who control them to capture additional rents from the system.

Sure enough, rent-generating rigidities are not only induced by government institutions but may also result from productivity-improving investments, such as process automation technology or avoidance of service redundancy through
centralization, for example. As discussed in Section 3 of this paper, one could even assume that rational agents will introduce such changes only up to the point where they neither add nor subtract productive efficiency and no further. Nevertheless, as long as these changes are rent-generating, they add rigidity to the process and, in a stochastic world, the higher the rigidity of the productive process (i.e. the heavier the weight of rents over total output), the greater its exposure to unexpected shocks, and therefore the higher the probability that a small stochastic shock lead to a major crash.

6 Conclusions and Final Remarks

The first objective of this paper was to highlight a serious fallacy (the Fallacy of Diffusion Symmetry) that underpins much of the debate around the Rational Expectations Hypothesis by fostering a belief that, if markets are efficient, predictable bubbles and crashes cannot exist. This, it has been shown, is only valid for symmetric probability diffusion processes, which are rarely, if ever, posed in modern financial models – although the fallacy itself all-too-frequently goes unacknowledged.

The second objective was to develop a rational-expectations model of periodic financial bubbles driven by an agency conflict between producers and investors along the lines of Gracia (2005), test its likelihood against U.S. stock market data, and then extend it to macroeconomic business cycles by combining it with the aggregate production function put forward in Gracia (2011).

Both objectives have substantial modeling implications. If the rational expectations debate is vitiated by the fallacy of diffusion symmetry, it may be unnecessary to resort to bounded rationality or price stickiness to explain many of the phenomena put forward as evidence of market inefficiency. On the contrary, it may make sense to review some classical rational expectations propositions (e.g. money neutrality, Ricardian equivalence, even perfect market clearing) to assess under what conditions they would still hold along the median path. Ultimately, the mean path be as it may, if the median trajectory is the one with the highest likelihood to match observations, then it arguably makes sense to target it instead of the mean as the primary objective of economic policy – which might, in turn,
vindicate a number of macroeconomic policies that a simplistic interpretation of efficient markets seemed to invalidate.

A potential research program building on the basis of the model put forward in this paper might therefore consider, among others, the following lines of work:

- Analyze the role of credit along the observed path (where it is not necessarily neutral) and its implications for monetary policy.
- Review the models that supported the successful predictions discussed in Section 2 to identify their key insights, and analyze whether they might be compatible with a rational expectations framework where the fallacy of diffusion symmetry has been cast away.
- Consider the impact of the production function in Gracia (2011) on the assessment of fiscal policy, redistribution and long-term growth.
- Conduct more empirical work to develop a working model for cycle statistical inference (subject of course to the limitations of forecasting the median path).

Rational expectations is a hypothesis with remarkable explanatory power but, by ignoring the impact of diffusion symmetry, its implications may have been oversold and stretched beyond what the theory granted or the observations supported. Neither does market efficiency necessarily entail random walk observations nor does their absence imply that the players’ rationality is somehow ‘bounded’. Recognizing this may lead to rejecting some popular views on the implications of market efficiency – yet, paradoxically, it may also vindicate rational expectations models against any bounded-rationality alternative.
The purpose of this appendix is to show how an empirical analysis of a time series generated by a geometric Gauss-Wiener diffusion process (i.e. a geometric Brownian motion) should be expected to yield an observed growth rate closer to the median, not the mean, path of the distribution.

Consider an asset whose market value $P_t$ follows a geometric Brownian motion i.e.:

$$\frac{dP_t}{P_t} = \mu dt + \sigma dZ_t,$$  \hspace{1cm} (1.1)

Where $\mu, \sigma$ are constants, $Z_t$ is a linear Brownian motion $dZ_t = \varepsilon_t \sqrt{dt}$ such that $Z_0 \equiv 0$, and the white noise variable $\varepsilon_t$ is normally-distributed, i.e., $\varepsilon_t \sim \mathcal{N}[0,1]$. Then, of course, the expected (or “mean”) growth rate of the asset’s market value is:

$$E_0 \left[ \frac{dP_t}{P_t dt} \right] = \mu + \sigma E_0 \left[ \frac{dZ_t}{dt} \right] = \mu$$ \hspace{1cm} (1.2)

We can now resort to Itô’s lemma to obtain the general expression of $P_t$:

$$d(\ln P_t) = \frac{dP_t}{P_t} - \frac{1}{2} \left( \frac{dP_t}{P_t} \right)^2 = \frac{dP_t}{P_t} - \frac{(\mu dt + \sigma dZ_t)^2}{2} = \frac{dP_t}{P_t} - \frac{\sigma^2}{2} dt$$

$$\frac{dP_t}{P_t} = d(\ln P_t) + \frac{\sigma^2}{2} dt = \mu dt + \sigma dZ_t$$

$$d(\ln P_t) = \mu dt - \frac{\sigma^2}{2} dt + \sigma dZ_t \iff \ln P_t = \left( \mu - \frac{\sigma^2}{2} \right) t + \sigma Z_t + \text{constant}$$

$$P_t = P_0 e^{\left( \mu - \frac{\sigma^2}{2} \right) t + \sigma Z_t}$$ \hspace{1cm} (1.3)
Since $Z_t$ is normally-distributed and therefore symmetrical, the diffusion process of $P_t$ must be asymmetrical, and the growth rate of its median path (i.e., the path that would cut across the distribution leaving 50% of the probability on each side) must be $\mu - \frac{\sigma^2}{2}$ i.e. different from that of the mean path.

If we now had a sufficiently long time series of empirical observations of $P_t$, we could calculate its average growth rate through a logarithmic regression under the following specification:

$$\ln P_t = a + bt + u_t$$

(1.4)

Where $a$ and $b$ represent the regression parameters and $u_t$ the series of residuals. It is therefore immediate that, assuming the data series is long and representative enough, the results of this regression should be expected to approximate the following:

$$
\begin{align*}
  a &= \ln P_0 \\
  b &= \mu - \frac{\sigma^2}{2} \\
  u_t &= \sigma Z_t
\end{align*}
$$

(1.5)

In other words, the result of the empirical analysis should be expected to be a growth rate equal to that of the median, not the mean.

Furthermore, the result would be the same if, instead of estimating the parameters of a logarithmic regression, we calculated the average continuous growth rate along the time series. Indeed: if, given a sample covering the time interval $t \in [0,T]$, the average continuous growth rate $g$ is defined as a magnitude such that:

$$
e^{gT} \equiv \frac{P_T}{P_0} \iff g \equiv \frac{\ln P_T - \ln P_0}{T}$$

(1.6)

This, in our case, yields the following expected result:
\[ g = \ln P_T - \ln P_0 = \left( \mu - \frac{\sigma^2}{2} \right) \frac{T + \sigma Z_T}{T} \]

\[ E_0[g] = \left( \mu - \frac{\sigma^2}{2} \right) + \frac{\sigma}{T} E_0[Z_T] = \left( \mu - \frac{\sigma^2}{2} \right) \]

(1.7)

Which is, once again, the median growth rate of the distribution, not the mean.

In sum, the observed path may, on average, be very different from this expected equilibrium without posing any challenge to the efficient markets hypothesis.\(^{24}\)

Q.E.D.

APPENDIX 2

The purpose of this appendix is to develop in analytical form the reasoning that in Section 3 of the main text was presented in an intuitive, discursive way.

Definitions and Assumptions:

Consider a firm whose total market value at instant \( t \) is \( K_t \), and that holds productive assets whose replacement cost at market prices would be \( \tilde{K}_t \). The

\(^{24}\) As a side note, it is worth noting that the mean and the median paths could be very different indeed: if, for example, we were in a situation where \( 0 < \mu < \frac{\sigma^2}{2} \), then the mean path would grow at a rate \( \mu > 0 \), whereas the median (and therefore the most likely observed path) would fall at a rate \( \mu - \frac{\sigma^2}{2} < 0 \). In other words, it might be perfectly rational for investors to buy and hold the asset (as they expect its value to grow at a rate \( \mu > 0 \) ) whilst an external observer wonders why they keep throwing their money at an investment that consistently loses value.
enterprise has two types of stakeholders: investors (i.e. principals) who hold ownership of the assets, and producers (i.e. agents) who control and manage those assets on their behalf. Hence, the net value added $Y_t$ the entity generates (i.e. the sum of the return it generates for labor and capital or, what is the same, the difference between its sales revenue and the market price of its non-labor inputs) may be broken down into four portions:

- The marginal cost of labor $m_t L_t$,
- The agency rents $R_t$ extracted by producers,
- The cash flow $C_t$ paid to investors (negative if the firm is raising capital) and
- The remaining value added, which is reinvested as retained earnings.

Net value added is therefore defined as follows:

$$Y_t = m_t L_t + R_t + C_t + \frac{dK_t}{dt}$$  \hspace{1cm} (2.1)

Of course this means the investors’ profit (‘$\Pi_t$’) is equal to $\Pi_t \equiv C_t + \frac{dK_t}{dt}$.

We now designate as $K_t^*$ the capitalized value of the producers’ control of the productive process i.e. of the “asset” represented by their ability to extract rents for themselves:25 Therefore, the firm’s overall rate of return $r_t$ is:

$$r_t \equiv \frac{Y_t - m_t L_t}{K_t + K_t^*} \equiv \frac{\Pi_t + R_t}{K_t + K_t^*}$$  \hspace{1cm} (2.2)

We also represent as $\tilde{r}_t$ the cost of opportunity of the assets $\tilde{K}_t$ i.e. the return they would yield if invested outside the structure of the company in a risk-free form – e.g. lending them for a fixed rent. In this context, $\tilde{C}_t$ represents the cash investors would extract from those assets while invested at a risk-free rate i.e.:

$$\tilde{C}_t = \tilde{r}_t \tilde{K}_t - \frac{d\tilde{K}_t}{dt}$$  \hspace{1cm} (2.3)

Finally, we define Tobin’s $q$ ratio as the market value of the company weighted by the replacement market value of its assets:

---

25 The asset that would be represented by the marginal cost of labour $m_t L_t$ has no financial value in this framework, as its cost would always equal its income.
And the producer rent ratio $\rho_t$ as the value of the producers’ rents divided by the value remaining for the investors in the company:

$$\rho_t \equiv \frac{K_t^*}{K_t}$$  \hspace{1cm} (2.5)

Now we introduce the following assumptions:

**Assumption 1 – Efficient Market Valuation:**

The company’s rate of return $r_t$ equals the market rate for assets of equivalent risk exposure, so its equity market valuation is such that, $\forall t \geq T$:

$$E_T[r_t, K_t] = E_T \left[ C_t + \frac{dK_t}{dt} \right]$$  \hspace{1cm} (2.6)

Where $E_T[c]$ indicates the mean value per the information available at instant $T$.

**Assumption 2 – Wiener Perturbation:**

The company’s rate of return follows a Wiener diffusion process with drift, i.e.:

$$r_t dt = \mu dt + \sigma dZ_t$$ \hspace{1cm} (2.7)

Where $\mu, \sigma > 0$ are positive parameters, $Z_t$ is a Brownian motion $dZ_t = \varepsilon_t \sqrt{dt}$ such that $Z_0 = 0$ and the white noise $\varepsilon_t$ is normally-distributed, i.e., $\varepsilon_t \sim N[0,1]$.

**Assumption 3 – Constant Risk-Free Rate:**

The risk-free rate at which assets could be lent/borrowed is constant i.e.:

$$\tilde{r}_t = \tilde{r} \quad (\text{where } \tilde{r} \rightarrow \text{constant} )$$  \hspace{1cm} (2.8)
Assumption 4 – Observable Market Valuations:

The market values $K_t$, $\tilde{K}_t$, and $K_t^*$ are observable at the instant $t$ where they take place, i.e., $K_t, \tilde{K}_t, K_t^* \in I_t$ (where $I_t$ is the set of information available) i.e. $\forall t \geq T\colon$

$$\lim_{T \to t} E_T[K_t] = K_t \quad \text{and} \quad \lim_{T \to t} E_T[\tilde{K}_t] = \tilde{K}_t \quad \text{and}$$

$$\lim_{T \to t} E_T[K_t^*] = K_t^*$$

(2.9)

Comment: This is quite intuitive, as these are all stock variables i.e. they are defined as values at a point in time (instead of flows between a point in time and the next).

Assumption 5 – Observable Cash Flows:

The cash flows paid to investors ($C_t$ and $\tilde{C}_t$) and to producers ($R_t$) are known at the instant $t$ in which they take place i.e. $C_t, \tilde{C}_t, R_t \in I_t$ (where $I_t$ represents the set of information available at time $t$) or, what is the same, $\forall t \geq T\colon$

$$\lim_{T \to t} E_T[C_t] = C_t \quad \text{and} \quad \lim_{T \to t} E_T[\tilde{C}_t] = \tilde{C}_t \quad \text{and}$$

$$\lim_{T \to t} E_T[R_t] = R_t$$

(2.10)

Comment: In other words, the risk of the return being different from expected at any given point in time is borne by the retained earnings $\frac{dK_t}{dt}$. 
Assumption 6 – Consistent Preference for Cash:

The investors’ willingness to sweat cash from their investments is the same for the company $K_t$ and for the assets $\tilde{K}_t$ i.e.:

$$\frac{C_t}{K_t} = \frac{\tilde{C}_t}{\tilde{K}_t}$$  \hfill (2.11)

Comment: There is quite a wide range of utility functions that would produce this result. As an example, Appendix 3 shows its derivation from a standard functional form (a time-additive discounted expected utility function with unity time elasticity).

Assumption 7 – Linear Investors’ Controls Function:

The degree of investors’ control on the producers’ ability to increase their rent extraction over time follows a linear function dependent on the investors’ relative gain or loss of value respective to investing outside the company i.e.:

$$d\rho_t = \lambda(q_t - 1)\rho_t dt$$  \hfill (2.12)

Where $\lambda$ represents a positive constant.

Comment: This postulate combines three intuitive ideas:

- Liquidation is not an instant process (hence $\lambda$ is assumed finite) but, the more investors find they are losing by not liquidating (i.e. the smaller Tobin’s $q$), the more companies will be reorganized or liquidated to cut down their rents.
- On the flip side, of course, the larger Tobin’s $q$ the least likely are investors to liquidate or to impose heavy controls, so more opportunities will pop up over time for producers to increase the rents they extract.
- On balance, liquidations will dominate when $q_t < 1$ (i.e. when it is more profitable for investors to liquidate), whereas producers will have more room to expand their rents when $q_t > 1$.

Although the functional form in (2.12) has been chosen primarily because of its simplicity, it can be justified intuitively if we assume that both the probability
of liquidation and that of increased rent extraction opportunities are distributed according to an exponential function.

**Analytical Development:**

Combining Assumption 1 (i.e. expression 2.6) with Definition (2.1) we find, \( \forall t \geq T \):

\[
E_T [r_i K_i] = E_T [\Pi_i] \iff E_T [r_i K_i^*] = E_T [R_i]
\] (2.13)

Which, per Assumptions 4 and 5 (i.e. expressions 2.9 and 2.10), becomes, for the special case \( T = t \):

\[
E_t [r_i K_i^*] = R_i
\] (2.14)

If we now combine Definitions (2.1), (2.2) and (2.5):

\[
R_i + C_i + \frac{dK_i}{dt} = r_i (K_i + K_i^*) = r_i (1 + \rho_i) K_i
\] (2.15)

This, combined with expression (2.14), becomes:

\[
E_t [r_i\rho_i K_i] + C_i + \frac{dK_i}{dt} = r_i (1 + \rho_i) K_i
\]

\[
\frac{dK_i}{dt} = r_i K_i - C_i + (r_i - E_t [r_i]) \rho_i K_i
\] (2.16)

By simple inspection, we can see that, along the expected path, the impact of \( q_i \) will be fully discounted out, for, if we write the expected value of (2.16) at point \( t \) and then apply Assumption 2 (i.e. expression 2.7), we obtain:

\[
E_t \left[ \frac{dK_i}{dt} \right] = E_t [r_i K_i - C_i] + \left( E_t [r_i] - E_t [r_i] \right) \rho_i K_i
\]

\[
E_t \left[ \frac{dK_i}{dt} \right] = E_t [r_i] K_i - C_i = \mu K_i - C_i
\] (2.17)
Which, after integration, yields the familiar Net Present Value formula i.e., $\forall t \geq 0$:

$$E_0[K_t] = A e^{\mu t} + \int_0^t E_0[C_t] e^{-\mu t} dt$$

(2.18)

Where $A$ represents an integration constant.26

This, to be sure, does not imply that the size of agency rents has no impact on the asset value, but simply that, in an efficient market, their expected impact has already been discounted from the asset market value at instant $t = 0$ and therefore, as long as the observed path matches the initial expectations, no further adjustment is necessary.

The median path, conversely, can be derived from expression (2.16) by applying the rule we developed in Appendix 1 i.e. since the median return is $\left(\mu - \frac{\sigma^2}{2}\right)$ then the median path of expression (2.16) is as follows:

$$\text{Median}_t \left[ \frac{dK_t}{dt} \right] = \left(\mu - \frac{\sigma^2}{2}\right) K_t - C_t - \frac{\sigma^2}{2} \rho_t K_t$$

(2.19)

At the same time, by combining Definition (2.3) with Assumptions 3, 4 and 5 (i.e. with expressions 2.8, 2.9 and 2.10) we obtain the (deterministic) path of $\tilde{K}_t$ i.e.:

$$\frac{d\tilde{K}_t}{dt} = \tilde{r} \tilde{K}_t - \tilde{C}_t$$

(2.20)

Hence, if we differentiate Definition (2.4) according to Itô’s lemma we obtain that:

$$\frac{dq_t}{q_t} = \frac{dK_t}{K_t} - \frac{d\tilde{K}_t}{\tilde{K}_t} + \frac{1}{2} \left( \frac{d\tilde{K}_t}{\tilde{K}_t} \right)^2$$

(2.21)

26 Note that the model would work just as well if one imposed, as is common in efficient markets asset valuation models, a transversality condition such that the expected value always equal the net present value of future cash flows (so that $A = 0$). This represents a key difference respective the ‘rational bubbles’ model Blanchard and Watson (1982) put forward, which can only work if no such transversality condition exists.
Whose median path is, applying Assumption 6 (i.e. expression 2.11):

\[
\text{Median}_t \left[ \frac{dq_t}{q_t} \right] = \left( \mu - \frac{\sigma^2}{2} - \tilde{\gamma} \right) dt - \frac{\sigma^2}{2} \rho_t dt - \left( \frac{C_t}{K_t} - \tilde{C}_t \right) dt
\]

\[
\text{Median}_t \left[ \frac{dq_t}{q_t} \right] = \left( \mu - \frac{\sigma^2}{2} - \tilde{\gamma} \right) dt - \frac{\sigma^2}{2} \rho_t dt
\]

(2.22)

For simplicity, we will designate by \( \pi \), the equity premium, which, for the median path in expression (2.22), will be the constant value \( \pi = \mu - \frac{\sigma^2}{2} - \tilde{\gamma} \).

On this basis it is now possible to close the dynamic system representing the median path by combining expression (2.22) with (2.12) in the following final expression:

\[
\begin{align*}
\frac{dq_t}{q_t} &= \left( \pi - \frac{\sigma^2}{2} \rho_t \right) dt \\
\frac{d\rho_t}{\rho_t} &= \lambda(q_t - 1) dt
\end{align*}
\]

(2.23)

Which is equal to expression (1) in Section 3.

Q.E.D.

Comment: Note that, if all three parameters \( \pi, \sigma^2, \lambda > 0 \) are all finite and positive, then expression (2.23) belongs to the family of Lotka-Volterra predator-prey dynamic systems.\(^{28}\) We know of course that the variance \( \sigma^2 \) is positive by

\(^{27}\) Of course this is just the non-trivial solution; there is also a trivial solution such that \( \rho_t = q_t = 0 \).

\(^{28}\) Lotka-Volterra predator-prey dynamic systems were originally developed in the context of biological studies (Lotka 1925, Volterra 1926) analysing the evolution of predator and prey populations in an ecosystem (hence their name). In economics, their best known instance of usage of a predator-prey process in economic modelling is of course Goodwin (1967).
definition (as it is the square of a real number) and, per Assumption 7, $\lambda$ is positive *ex hypothesi* (a negative value would mean that investors are more likely to liquidate the higher the $q$-ratio, which makes no sense). The median risk premium $\pi$, conversely, could theoretically also be zero or even negative (if investors are risk averse, risky assets will offer a positive risk premium along the mean path, but not necessarily along the median).\(^{29}\) Nevertheless, at least in countries that have been both politically stable and financially sophisticated for a very long time, such as Britain or the USA, historical equity returns have been above low-risk interest rates more than 50% of the time. Hence, as long as the starting values of $q_t$ and $\rho_t$ are positive, we can say that under Assumptions 1 to 7, *if the median equity premium is positive, then the median path of Tobin’s $q$ ratio will follow a Lotka-Volterra predator-prey cycle* such as the one plotted in the simulation in Figure 1 in the main text.

**APPENDIX 3**

The purpose of this appendix is to show how Assumption 6 in Appendix 2 can be derived from a standard representative consumer utility function within the parameters most usually applied in mainstream literature.

In the following example, we will assume that the representative consumer intends to maximize a von Neumann-Morgenstern time-additive discounted utility function with unity inter-temporal elasticity of substitution, i.e.:

$$\max E_0 \left[ \int_0^\infty (\ln C_t) e^{-\beta t} dt \right]$$

(3.1)

Where $\beta > 0$ is constant.

This maximization is then subject to the budget constraint:

\(^{29}\) In Gracia (2005) the assumption was that, instead of $\tilde{R}_t$ being constant, it had the same variability as $R_t$ so that the difference $\pi = R_t - \tilde{R}_t$ represented a mere “agency premium” which had the same value for the mean and the median and therefore, if investors were risk averse, would be guaranteed to be positive also in the median path. Here, conversely, we have adopted Assumption 2 instead, which is a bit simpler without making much of a difference.
Where $M_t$ represents a martingale such that:

$$\frac{dM_t}{M_t} \equiv -\left(\bar{r}dt + \sigma dZ_t\right) \iff M_t \equiv e^{-\left(\frac{\sigma^2}{2}\right)t - \alpha Z_t}$$  \hspace{1cm} (3.3)

Thus, the first-order condition for the resolution of this problem is, $\forall t \geq 0$:

$$\frac{\partial U(C_t)}{\partial C_t} e^{-\beta t} = e^{-\beta t} \left\{ e^{-\left(\frac{\sigma^2}{2}\right)t - \alpha Z_t} \right\}$$

$$\Lambda = C_t^{-1} e^{\left(\frac{\sigma^2}{2} - \beta\right)t + \alpha Z_t}$$  \hspace{1cm} (3.4)

Where $\Lambda$ represents the Lagrange multiplier. As this applies $\forall t \geq 0$, then:

$$\frac{C_t}{C_0} = e^{\left(\frac{\sigma^2}{2} - \beta\right)t + \alpha Z_t} \iff \frac{dC_t}{C_t} = (\bar{r} - \beta) dt + \sigma dZ_t$$  \hspace{1cm} (3.5)

If we now use this to replace into the budget constraint (3.2) we obtain:

$$K_0 = E_0 \left[ \int_0^\infty C_t \frac{M_t}{M_0} dt \right] = E_0 \left[ \int_0^\infty C_0 e^{-\beta t} dt \right] = \frac{C_0}{\beta}$$

$$\frac{C_0}{K_0} = \beta$$  \hspace{1cm} (3.6)

Hence, for any point in time $t$ taken as a reference, under this utility function the ratio $\frac{C_t}{K_t}$ equals the constant $\beta$ irrespective of the rate of return of the underlying asset, and therefore the ratio will apply all the same if the asset is $\tilde{K}_t$, so that:
\[ \frac{C_t}{K_t} = \frac{\tilde{C}_t}{\tilde{K}_t} = \beta \]  

(3.7)

Q.E.D.
References


Yale University Press, New Haven.

Oxford University Press.

Princeton University Press.


http://ideas.repec.org/a/mcb/jmonecb/v28y1996i1p1-21.html

http://ideas.repec.org/a/red/issued/v4y2001i1p230-243.html


http://ideas.repec.org/p/nbr/nberwo/6632.html


http://ideas.repec.org/a/oup/ecinqu/v31y1993i2p224-36.html


Please note:

You are most sincerely encouraged to participate in the open assessment of this article. You can do so by either recommending the article or by posting your comments.

Please go to:

http://dx.doi.org/10.5018/economics-ejournal.ja.2012-41

The Editor