I present a modified present value model that can explain a wide variety of asset pricing phenomena. The central ingredient is a noisy signal about the future economic fundamentals. Not distinguishing noise from economic fundamentals, agents face a signal extraction problem. As a result stock prices include a bubble component. However, market are efficient in the sense that prices optimally reflect all available information; in other words, prices are “efficiently wrong”. The model can generate interesting asset pricing phenomena: episodes of volatility clustering, boom-and-bust, underreaction, overreaction, etc. I show that excess volatility, forecast rationality and violation of cross-equation restrictions can be reconciled with market efficiency. Finally, I propose a new procedure to decompose stock prices into a fundamental and a bubble component. I find that market can stay under- or overvalued for an extended period of time. In particular, the US stock market was undervalued during the 1970s and overvalued during the 1990s. Interestingly, I find no evidence that the market was overvalued before the Wall Street Crash of 1929.

*The author is deeply indebted to Carlos Velasco for guidance and encouragement. I also benefited most from the comments of Fabio Canova, Luca Gambetti, Jesus Gonzalo, Matthias Kredler, Sophocles Mavroeidis, and Hernan Seoane.
1 Introduction

The announcement that both Fama and Shiller, along with Lars Peter Hansen, had been awarded 2013 Nobel Prize in Economics came as a surprise to many. Fama and Shiller are two giants of financial economics. The core of modern finance theory is based on Fama’s “Efficient Market Hypothesis” which asserts that prices rationally incorporate all available information. Robert Shiller, on the other hand, is the father of “Behavioral Finance” and a diehard critic of rationality, so much so that he described market efficiency as “one of the most remarkable errors in the history of economic thoughts”. This sparked a heated debate between these two scholars, and economics at large, over market efficiency (or equivalently rationality). Given the crucial role of financial sector in the modern economy, as is evident from the world-wide financial crisis of 2008-2009, the importance of this question could not be overemphasized. My contribution is then to present a modified Present Value (PV) model that can explain a wide variety of asset pricing phenomena, while retaining the simplicity of rational expectations, symmetric information and homogeneous beliefs framework.

Fama (1970) defines market efficiency as: “a market in which prices always “fully reflect” (his emphasis) available information”. Many economists and practitioners have misinterpreted Efficient Market Hypothesis (EMH) as stating that prices are “correct” This interpretation is wrong. EMH does not imply that prices are correct; it is a mechanism by which available information is reflected into prices. I argue that the central issue is the “available information”. Markets have a variety of sources for information. As with any other source of information, the information used by market participants is very often noisy. Market efficiency implies that any source of information should be exploited. Therefore, market participants face a problem: How to use the noisy information? I argue that the noisy information generates patterns that may look like market inefficiency to an outside observer.

To formalize this point, I introduce a noisy signal (news) into a simple present value

\footnote{See Malkiel (2003) for a comprehensive review of EMH and its critics.}
model. Assuming that agents only observe a noisy signal, requires solving a *signal-extraction* problem. The presence of noise in the signal produces expectational errors, generating temporary deviations of stock prices from their fundamental values. Therefore, consistent with Shiller, prices include a component totally unrelated to economic fundamentals. In contrast to Shiller, however, we can not interpret the deviation of actual prices from fundamental values as evidence against market rationality, as in my framework, market rationality holds by construction. In other words prices are “efficiently wrong”. So Shiller meets Fama.

What about other phenomena that EMH struggles to explain? Motivated by Fama’s joint hypothesis problem, which asserts that any test of efficiency is a joint test of efficiency and the asset pricing model, a long tradition in asset pricing has therefore opted to rationalize the observed phenomena by developing new preferences and discount factors. I believe that the literature interpretation of Fama’s joint hypothesis problem is too narrow. I would go further: I argue that market efficiency must be jointly tested with an asset pricing model and an information set. Given a correct asset pricing model, rational responses of market participants may seem irrational to outside econometricians, if the information sets of the agents and the econometricians are misaligned. To avoid confusion, I refer to this as the “triple hypothesis problem”.

The triple hypothesis problem opens the door for rational explanations to puzzling stock market phenomena. Specifically, I show that stock prices are not excessively volatile. I start with Shiller’s plots. I show that the lack of correspondence between prices and the fundamental values can be interpreted as a rational response to noisy information. Campbell and Shiller derive upper bounds on the variance of stock prices. They find that the bounds are grossly violated, and conclude that stock prices are excessively volatile. They do not provide a formal proof for their bounds. In this paper’s framework, I prove that Campbell-Shiller’s bound are reversed. Therefore, Campbell-Shiller findings support market efficiency, not the other way around.

---

2 See Cochrane (2011) for an accessible review of this literature.
I move on to the econometrics testing of market efficiency. I argue that these rejections may simply reflect the misalignment of the information sets of the agents and the econometricians. Conventional econometric tests of EMH involve comparing observed variables (e.g., prices) to some metrics of the market (e.g., fundamental prices) based on the agents’ true expectations. Since agents’ expectation is unobservable, in practice these metrics are compared to the econometrician’s estimate, usually obtained from linear projections onto an information set observed by econometricians. The equilibrium solution of our simple PV model takes the form of a linear forward-looking representation. I show that forward-looking equilibrium dynamics pose substantial challenges to econometric tests of market efficiency. An econometrician, who uses standard econometric procedures based on linear projections (such as VARs), will find that asset prices display rejections of cross-equation restrictions, and forecast irrationality.

The model can generate interesting asset market behavior internally. Generating asset pricing dynamics internally, that is not from exogenous variation in the probability distribution of dividend growth and returns, is a non-trivial contribution. In the data, some booms in asset prices are followed by busts. Others are not. My model is consistent with this observation. Further, underreaction and overreaction could be understood qualitatively as rational response to noisy information. Finally, the model naturally produces stock prices with volatility clustering.

Finally, I approximate fundamental values by projecting the observed price-dividend ratio onto the linear space spanned by future dividends and returns. Beside simplicity, my proposal has the advantage of using realized returns instead of proxies and therefore is not sensitive to the misspecification of unobserved discount rates. Applying my procedure to US stock market data, I decompose price-dividend ratio into a fundamental component and a bubble component. I find that stock prices can remain overvalued or undervalued for an extended period of time. The largest positive bubble was the dot-com bubble, starting in 1985. Negative bubbles were present in at least two periods: 1938-1949, and 1965-1980.
Interestingly, I find no evidence that the market was overvalued before the Wall Street Crash of 1929.

Relation to the Literature. This paper relates to the news literature as an important source of economic fluctuations. Beaudry and Portier (2006) provide empirical evidence supporting the view that news about the future is an important source of aggregate macroeconomic fluctuations. Jaimovich and Rebelo (2009), Schmitt-Grohé and Uribe (2012), Den Haan and Kaltenbrunner (2009), and others developed models that could generate business cycle dynamics in response to news. Lorenzoni (2009) emphasize the relevance of noise shocks as well as fundamental shocks. Blanchard et al. (2013) highlights the challenges that noisy news about the future economic fundamentals pose to standard econometrics techniques. Forni et al. (2016) extend this literature to asset pricing by introducing a noisy signal into a simple present value model, and show that the equilibrium price includes a bubble. Blanchard et al. (2013) and Forni et al. (2016) emphasize the difference between fundamental and noise shocks in explaining temporary and permanent fluctuations. My paper, instead, emphasizes the role of the noisy information in explaining asset market puzzles.

Within the rational expectations paradigm, a few papers propose discount rate models to address asset market puzzles. The external-habit formation of Campbell and Cochrane (1999) and the long-run risk model of Bansal and Yaron (2004) are two successful examples. These studies have related the puzzles to the joint hypothesis problem. Outside the rationality paradigm, Barberis et al. (1998) and Hong and Stein (1999) present models that accommodate overreaction and underreaction, but as Fama (1998) argues, perform poorly otherwise. A particular advantage of my framework is that within the class of rational, representative agent framework explains a wide range of asset market puzzles.

The remainder of the paper is organized as follows. Section II presents a simple present value model around which the discussion is best organized. Section III discusses the implication for the market efficiency tests. Section IV presents the results of the historical

\footnote{See also Leeper et al. (2013) and Kasa et al. (2014).}
decomposition of US stock prices. The last Section provides concluding comments.

2 The Econometric Model

Let me begin by considering a general one-period Euler equation asserts that:

\[ P_t = \mathbb{E}\left[ M_t (P_{t+1} + D_t) | I_t^s \right] \]  \hspace{1cm} (2.1)

where \( P_t \) is the asset price at the beginning of period \( t \), \( D_t \) is the dividend paid at the end of period \( t \), and \( M_t \) is the discount factor. Moreover, \( \mathbb{E} \) denotes expectations conditional on all available information to the market participants at the beginning of period \( t \), denoted by \( I_t^s \).

To complete the description of the model, we must specify a process for discount rates, dividends, and how agents make optimal forecast based on their information set.

Discount Rates. For ease of exposition assume that the discount factor is constant over time, that is \( \mathbb{E}[M_t | I_t^s] = \beta \), for all \( t \). The setting in this section could be generalized to allow for time-varying discount factors, at the cost of a more complicated notation and probably the analytical solution.

Dividends. I consider a typical unit root process

\[ D_t = D_{t-1} + \epsilon_t, \]

where the dividend change, \( \epsilon_t \), follows an independent and identically distributed (iid) process. Although there is an extensive empirical evidence supporting this assumption, my setting can accommodate both stationary and nonstationary ARMA process.

\[ ^4 \text{In what follows, I interchangeably use “dividends”, “cash flows”. Similarly, “market participants”, “agents”, “traders”, and “investors” are all the same thing here.} \]
Agents’ Information Set. I assume that agents (and econometricians) observe the complete history of prices and dividends. Moreover, at the beginning of each period and right before prices are determined, agents (but not econometricians) observe some noisy signals about the future dividend changes. Following the notation that \( s_{t|t+k} \) denotes the news at time \( t \) about time \( t+k \) dividends, we have

\[
s_{t|t+k} = \epsilon_{t+k} + \nu_{t|t+k} \quad (k \geq 0),
\]

where \( \nu_{t|t+k} \sim iid(0, \sigma^2_{\nu}) \) denotes the noise at time \( t \) about the dividend change at time \( t+k \), and \( \nu_t \) is orthogonal to \( \epsilon_t \) at all leads and lags. Thus, the agents’ information set at the begging of period \( t \) (say \( I^a_t \)), encompass the current and past signals, as well as the history of observed prices and dividends. The econometricians’ information set (say \( I^o_t \)), on the other hand, only includes the history of observed prices and dividends.

Equation (2.2) nests some interesting cases commonly used in the literature. The closest is Forni et al. (2016), which assume that agents observe a noisy signal

\[
s_t = \epsilon_t + \nu_t,
\]

instead of the structural shocks \( \epsilon_t \). In order to prevent the agents from perfect signal-extraction, FGLS assume that dividends are driven by a structural shock whose effects are delayed

\[
D_t = D_{t-1} + \epsilon_{t-q}, \quad (q \geq 1)
\]

This way of representing news has some drawbacks. Beside posing awkward questions about the source of delayed effect of the shock, (2.3) restrict the learning process: agents do not learn until dividend changes. In contrast, (2.2) allows for the size of the noise to change (usually decrease) when we get closer to the actual dividend change.

What is the source of news about future economic fundamentals? This includes public
release of information such as macroeconomic reports. It is well known that survey data, such as stock market and consumer confidence indexes, can predict key macroeconomic variables. Public statements by policy makers affects the market when setting the prices. For example, when Alan Greenspan, then the chairman of the Federal Reserve Board, used the term *irrational exuberance* on his speech on December 5, 1996, the markets dropped as the investors were fearful that the FED might increase the interest rate to bring down the bullish stock prices. Macroeconomic forecasts are also an important source of information highly valued by the markets.

In the following, I illustrates how investors may use economic forecasts as a noisy signal about the future economic fundamentals. The object of interest here is the China GDP growth \( (\epsilon_{t+j}, j \geq -1) \). The Wold bank (and other organizations, such as IMF, OECD, etc) announces the forecasts \( (s_{t|t+j}, j \geq -1) \) in January each year. Since GDP is a lagging economic variable, that is the accurate measure of it is released several months after the years end, the forecast also includes the last year GDP growth forecast. As Table (1) indicates, however, the forecasts are noisy. Moreover, as expected, the size of the noise is decreasing in the revisions of the forecast, released in the following years.

Do markets react to revisions? My answer is a solid yes. To see this, note that the January 2015 forecast is overestimating the GDP growth of China, and the January 2016 revises the forecast down. Therefore, somewhere in between markets should have declined. And that is exactly what happened in August 2015. Figure (1) plots the daily S&P500 for the year 2015. Consistent with our intuition, a wave of selling hit markets around the world in August and the S&P500 crashed. Many market analysts and specialists associated the drop to fears of China’s economic slowdown.

*Rational Expectations.* I have not yet specified how agents make optimal predictions. For convenience, I assume that agents’ optimal forecast is linear, that is \( \mathbb{E}[\epsilon_{t+j}|I_t^s] = \mathbb{P}[\epsilon_{t+j}|I_t^s] \).
Table 1: China Real GDP Growth ($\epsilon_t$) and its Forecast ($s_{t|t+j}$)

<table>
<thead>
<tr>
<th></th>
<th>2015</th>
<th></th>
<th>2016</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GDP Growth</td>
<td>Forecast</td>
<td>Noise</td>
<td>GDP Growth</td>
</tr>
<tr>
<td>$\epsilon_{2014} = 7.3$</td>
<td>$s_{2015</td>
<td>2014} = 7.4$</td>
<td>$\nu_{2015</td>
<td>2014} = 0.1$</td>
</tr>
<tr>
<td>$\epsilon_{2015} = 6.9$</td>
<td>$s_{2015</td>
<td>2015} = 7.1$</td>
<td>$\nu_{2015</td>
<td>2015} = 0.2$</td>
</tr>
<tr>
<td>$\epsilon_{2016} = ?$</td>
<td>$s_{2015</td>
<td>2016} = 7.0$</td>
<td>$\nu_{2015</td>
<td>2016} = ?$</td>
</tr>
<tr>
<td>$\epsilon_{2017} = ?$</td>
<td>$s_{2015</td>
<td>2017} = 6.9$</td>
<td>$\nu_{2015</td>
<td>2017} = ?$</td>
</tr>
</tbody>
</table>


Figure 1: Daily S&P500 Index, Year 2015

Notes: The S&P500 index over the period January 1, 2015 and December 31, 2015. These data are obtained from finance.yahoo.com.
Therefore, $\mathbb{E}[\epsilon_{t+j}|I_t^s]$ is simply the projection of $\epsilon_{t+j}$ on $s_{t|t+j}$, which implies that

$$
\mathbb{E}[D_{t+j}|I_t^s] = D_{t-1} + \sum_{j=0}^{\infty} \gamma s_{t|t+j}, \text{ for all } j \geq 0,
$$

$$
= D_{t-1} + \sum_{j=0}^{\infty} \gamma (\epsilon_{t+j} + \nu_{t|t+j}), \text{ for all } j \geq 0, \quad (2.4)
$$

where $\gamma = \frac{\sigma^2}{\sigma^2 + \nu^2}$.

We can now find equilibrium stock prices. Solving (2.1) forward, together with the transversality condition we obtain a simple Present Value (PV) model with constant discount factor

$$
P_t = \sum_{j=0}^{\infty} \beta^{j+1} \mathbb{E}(D_{t+j}|I_t^s),
$$

(2.5)

where $\beta = \frac{1}{1+R}$. Substituting (2.4) into (2.5), the generalized form of equilibrium price is given by

$$
P_t = \kappa D_{t-1} + \kappa \gamma \sum_{j=0}^{\infty} \beta^j s_{t|t+j}
$$

$$
= \kappa D_{t-1} + \kappa \gamma \sum_{j=0}^{\infty} \beta^j \epsilon_{t+j} + \kappa \gamma \sum_{j=0}^{\infty} \beta^j \nu_{t|t+j}. \quad (2.6)
$$

From (2.6) we see that the stock price can be decomposed into two components, the fundamental part, associated with the best possible predication of future dividends, and a bubble component, associated with the limited information of the agents. Here, I refer to any asset prices that exceed an asset’s fundamental value as bubble.

Shiller (2014) defines bubbles differently: “I define a speculative bubble as a situation in which news of price increases spurs investor enthusiasm, which spreads by psychological contagion from person to person, and, in the process, amplifies stories that might justify the price increase and brings in a larger and larger class of investors, who, despite doubts about the real value of the investment,

---

5 For an accessible survey of the bubble literature see Brunnermeier (2009).
are drawn to it partly through envy of others’ successes and partly through a gambler’s excitement.” This definition does not accommodate negative bubbles.

Noise bubbles have some interesting properties: First, unlike speculative bubbles driven by irrational anticipations, noise bubbles are rational: they are driven by noisy information about the future economic fundamentals. Second, noise bubbles are defined within the class of representative agent framework and benefit from the simplicity of symmetric information and common beliefs assumption. In contrast asymmetric information and heterogeneous beliefs bubbles have a reputation for being difficult to handle. Third, unlike standard rational bubbles, noise bubbles satisfy the transversality condition and therefore the model lends itself to a unique equilibrium. Finally, noise bubbles can explain when and why bubbles start, and why they deflate over several months.\(^6\)

The size of the bubble is controlled by three factors: \(\kappa, \nu_t, \gamma = \frac{\sigma^2}{\sigma^2 + \nu^2}\). If \(\kappa\) is big (\(\beta\) close to 1), even an small \(\nu_t\) can have an arbitrary large effect on the price. The variance of the noise is a particularly important parameter in quantifying the noise component. This is because the variance of the noise affects not only the weight of the noise in the bubble component (\(\gamma\)), but also the size of the noise (\(\nu_t\)). If the variance of the noise is very small, i.e. the signal is precise, then \(\gamma \approx 1\). However, \(\nu_t\) can not be too large if \(\sigma^2\) is small. As a result, the noise component will be negligible. On the other hand, if the variance of the noise is very large, then \(\gamma \approx 0\). However, in this case big realizations of \(\nu_t\) is more probable. As a result, if the signal is imprecise, there will always be a bubble component in the price.

\textit{Irrational Expectations.} Although the modeling approach in this paper is based on the rational expectations, the analysis can be extended to accommodate irrational theories. For example, with delusional agents believing that the noisy signal is precise, the irrational

\(^6\)See also Forni et al. (2016) and Brunnermeier (2009).
version of (2.4) is given by

\[ \mathbb{E}[D_{t+j} | I_t^s] = D_{t-1} + \sum_{j=0}^{\infty} s_{t|t+j}, \quad \text{for all} \ j \geq 0, \]

\[ = D_{t-1} + \sum_{j=0}^{\infty} \epsilon_{t+j} + \nu_{t|t+j}, \quad \text{for all} \ j \geq 0, \]

where I denote irrational expectations by \( \mathbb{E}[\cdot] \). Analogously, the irrational price is given by

\[ \hat{P}_t = \sum_{j=0}^{\infty} \beta^{j+1} \mathbb{E}(D_{t+j} | I_t^s), \]

\[ = \kappa D_{t-1} + \kappa \sum_{j=0}^{\infty} \beta^j \epsilon_{t+j} + \kappa \sum_{j=0}^{\infty} \beta^j \nu_{t|t+j}. \quad (2.7) \]

One important take away from (2.6) and (2.7) is that both include a bubble component. This simple example sheds a serious doubt over claims that stock markets are inefficient because prices deviate from fundamental values.

Linear Dynamic Representation. Linear models, such as Vector Autoregressions (VAR), are widely used in macroeconomic and financial econometrics literature. Since some of the evidence against the market efficiency (for instance cross-equations restriction) have been obtained using linear models, it is useful to drive the linear representation of (2.6). In order to reconcile the modeling approach of this paper with the VAR literature, we need a simplifying assumption, namely

\[ \nu_{t|t+k} = \nu_t, \quad k \geq 0. \]

This assumption implies that the errors in the signal are fixed in each period. Under this
[2.6] reduces to

\[ P_t = \kappa D_{t-1} + \kappa \gamma \sum_{j=0}^{\infty} \beta^j \epsilon_{t+j} + \kappa \gamma \sum_{j=0}^{\infty} \beta^j \nu_{t+j}, \]

which has a linear forward-looking (also known as non-causal) representation

\[ (1 - \beta F)S_t = \kappa \gamma (\epsilon_t + \nu_t), \tag{2.8} \]

where \( F \) is the forward operator (i.e., \( FX_t = X_{t+1} \)). In what follows, it is often more convenient to consider the spread between prices and a multiple of dividends, \( S_t \equiv P_t - \kappa D_{t-1} \). The advantage of considering the spread instead of first differences of prices is that these variables summarize the joint history of prices and dividends, and at the same time not losing information on the levels of these variables \cite{CampbellShiller1987}. One can think of \( S_t \) as the log of the price-dividend ratio, i.e., \( \log(S_t) \approx \log(P_t/D_t) \) under the assumption that \( \kappa \approx 1 \), which has received the most attention in the literature.

Linear forward-looking representations are observationally equivalent (up to second order moments) to a backward-looking representation which emerges from flipping the root of the AR polynomial from inside to outside the unit circle via the Blaschke factor\footnote{See Hansen et al. (1981).}

\[ (1 - \beta L)S_t = \frac{(1 - \beta L)}{(1 - \beta F)} \kappa \gamma (\epsilon_t + \nu_t), \tag{2.9} \]

where the econometricians innovations, \( \xi_t \), are the statistical shocks obtained from linear projections of \( S_t \) onto its past.

By considering a backward-looking process \(2.9\), the econometrician unknowingly conditions on less information than the agents. The econometrician’s information set is the linear space spanned by the current and past values of the observables, \( S_t \), which corresponds to
the linear space spanned by the current and past values of the signals, \((\epsilon_t + \nu_t)\). Since the Blaschke factor, \(\frac{1-\beta_L}{1-\beta_F}\), is a two-sided filter, each element \(\tilde{\xi}_t\) is a function of past, current, and future values of \((\epsilon_t + \nu_t)\), the closed linear space generated by current and past values of \(\tilde{\xi}_t\) is no larger than the space generated by the observables.

Augmenting the econometrician’s information set by dividends does not solve the information misalignment. The equilibrium solution for the dividends and the spread is given by

\[
\begin{bmatrix}
1 & 0 \\
0 & 1 - \beta F \\
N(F) & X_t
\end{bmatrix}
\begin{bmatrix}
\Delta D_t \\
S_t
\end{bmatrix}
= 
\begin{bmatrix}
\epsilon_t \\
\kappa \gamma (\epsilon_t + \nu_t)
\end{bmatrix}
\]  

(2.10)

where \(N(F)\) is a forward-looking –also known as non-causal– AR polynomial.

By employing VAR analysis, the econometrician considers a backward looking process

\[
\begin{bmatrix}
1 & 0 \\
0 & 1 - \beta L \\
\tilde{N}(L) & X_t
\end{bmatrix}
\begin{bmatrix}
\Delta D_t \\
S_t
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 \\
0 & 1 - \beta L \\
0 & 0 & \frac{1}{1-\beta F}
\end{bmatrix}
\begin{bmatrix}
1 & 0 \\
0 & 1 - \beta L \\
\tilde{N}(L) & X_t
\end{bmatrix}
\begin{bmatrix}
\epsilon_t \\
\kappa \gamma (\epsilon_t + \nu_t)
\end{bmatrix}
\]  

(2.11)

where \(\tilde{N}(L)\) is a backward-looking AR polynomial of the same order as \(N(F)\). The econometrician’s information set is the linear space spanned by the current and past values of the observables, \(X_t \equiv (\Delta D_t, S_t)\), which corresponds to the linear space spanned by the current and past values of \(\tilde{\xi}_t\). Since each element of \(\tilde{\xi}_t\) is a linear function of past, current, and future values of \(\xi_t\), the closed linear space generated by current and past values of \(\tilde{\xi}_t\) is no larger than the space generated by the \(\xi_t\), which spans the agents’ information set.
3 Data and Model Implications

In this section, I show that our simple modification of the present value model can generate interesting asset pricing phenomena internally. Specifically, boom-and-bust episodes and volatility clustering can be understood qualitatively. To quantitatively evaluate the fit of the model, I calibrate the model \((2.11)\) to the US data, and show that the model can match the observed data. My general conclusion is that, the model can qualitatively and quantitatively account for many asset pricing phenomena.

3.1 Boom-and-Bust periods

High prices, relative to dividends, have preceded low returns, and vice versa, which has often been interpreted as evidence of return predictability. In the housing market, for example, \cite{Cochrane2011} concludes that high price-rent ratios before 2008 signal low future returns, not rising rents or prices that rise forever. We can also read these dynamics backwards: return is low when the actual price-dividend ratios were higher than fundamental values in the past. In other words, before 2008 agents were receiving false good news about the future. See footnote 3 for some evidence. As a result, prices were higher than the fundamental values.

**Example:** Suppose that agents receive a signal which gives good news, that is

\[ s_{t|t} = \epsilon_t + \nu_{t|t} > 0. \]

Agents not knowing which part of the news is noise, have to react to the information. As a result, the price is given by

\[ P_t = \kappa D_{t-1} + \kappa \gamma s_{t|t}, \]

where \( \kappa = \frac{\beta}{1-\beta} \) and \( \gamma = \frac{\sigma^2}{\sigma^2 + \sigma^2}. \) Let's consider two limit cases:
Case I. The news is false, i.e., \( s_{\text{t}|t} = \nu_{\text{t}|t} \). When agents observe \( \Delta D_t = 0 \), they know the news was false and will adjust prices to \( P_{t+1} = \kappa D_{t-1} \). It follows that

\[
1 + R_t = \frac{P_{t+1} + D_{t-1}}{P_t} = \frac{1}{\beta} \frac{D_{t-1}}{D_{t-1} + \gamma \nu_{\text{t}|t}} < \frac{1}{\beta}.
\]

Case II. The news is true, i.e., \( s_{\text{t}|t} = \epsilon_{\text{t}} \). When agents observe \( \Delta D_t = \epsilon_{\text{t}} \), price adjust to \( P_{t+1} = \kappa D_t \). It follows that

\[
1 + R_t = \frac{P_{t+1} + D_t}{P_t} = \frac{1}{\beta} \frac{D_t}{D_t - (1 - \gamma) \epsilon_{\text{t}}} > \frac{1}{\beta}.
\]

Both limit cases, and for that matter, any linear combination of them, give the impression that high price-dividend ratio predict low return, and vice a versa.

### 3.2 Volatility Clustering

Time series of financial asset prices often exhibit the volatility clustering property. That is, small price changes tend to follow small price changes, and large price changes follow large price changes, resulting in persistence of the amplitudes of price changes. Typically, nonlinear models such as the autoregressive conditionally heteroskedastic (ARCH) or the stochastic volatility models are suggested for such dynamics. See Bollerslev et al. (1992) for a comprehensive review of this literature.

From (2.8) we see that the spread is a forward-looking (also known as noncausal) process:

\[
S_t = \beta S_{t+1} + \xi_t,
\]

where, for convenience I set \( \kappa \gamma = 1 \), and \( \xi_t \equiv (\epsilon_t + \nu_t) \). Iterating forward, and imposing the
transversality condition

\[
\lim_{j \to \infty} S_{t+j} = 0
\]

we obtain

\[
S_t = \sum_{j=0}^{\infty} \beta^j \xi_{t+j}.
\]

There is a rich literature in statistical time series, that suggest forward-looking processes can display non-linear behavior. See, for instance, [Breidt et al. (2001)]. The following example illustrates this point.

Following [Gouriéroux and Zakoian (2016)], suppose in equation (3.1) \( \xi_t \) is an iid process following a standard Cauchy distribution. Then one can prove that

\[
\Delta S_t = \tilde{\xi}_t
\]

where \( \tilde{\xi}_t = \eta_t \left( \frac{1-\beta}{\beta} S_{t-1}^2 + \frac{1}{\beta(1-\beta)} \right)^{1/2} \), and

\[
\mathbb{E}[\eta_t | S_{t-1}] = 0, \quad \mathbb{E}[\eta_t^2 | S_{t-1}] = 1.
\]

Interestingly, the statistical innovations, \( \tilde{\xi}_t \), are not iid; they display GARCH-type effects which increase when \( \beta \) approaches 0 and vanish when \( \beta \) increases to 1.

4 Toward The Triple Hypothesis Problem

I now argue that many asset pricing phenomena can be understood within the triple hypothesis problem. In another word, many puzzling asset pricing regularities could reflect the misalignment of the information sets of egents and the econometricians. I focus on the following empirical regularities: (i) excess volatility; (ii) correspondence between prices
and perfect foresight prices; (iii) bubbles; (iv) under- and overreaction; (v) tests of forecast rationality; (vi) cross-equation restrictions.

It is useful to define the following information sets, where each set progressively includes more information:

\[ I^o_t = \text{span}(D_{t-j-1}, P_{t-j-1}; j \geq 0), \]
\[ I^s_t = \text{span}(D_{t-j-1}, P_{t-j-1}, s_{t-j|t-j+k}; j \geq 0, k \geq 1), \]
\[ I^*_t = \text{span}(D_{t+j}, P_{t+j}; j \geq 0), \]

such that \( I^o_t \subseteq I^s_t \subseteq I^*_t \), corresponding to the econometrician’s, agent’s, and perfect foresight information set, respectively. Accordingly, we can define

\[ P'_t = \sum_{j=0}^{\infty} \beta^{j+1} \mathbb{E}(D_{t+j}|I^o_t), \]
\[ P_t = \sum_{j=0}^{\infty} \beta^{j+1} \mathbb{E}(D_{t+j}|I^s_t), \]
\[ P^*_t = \sum_{j=0}^{\infty} \beta^{j+1} \mathbb{E}(D_{t+j}|I^*_t). \]

### 4.1 Excess Volatility

Volatility tests of market efficiency examines if the news about future economic fundamentals can explain the stock price movements. The basic idea of variance bound test can be characterized by the PV relation

\[ P_t = \mathbb{E}\left[ \sum_{j=0}^{\infty} \beta^{j+1} D_{t+j}|I^*_t \right] \quad (4.1) \]
which implies that
\[ \text{var}(P_t) \leq \text{var}(P^*_t). \quad (4.2) \]

By comparison of sample variance of stock price, \( P_t \), with its ex-post rational price, \( P^*_t \), Shiller (1981) argues that the variation in stock prices are too large to be justified by the variation in subsequent dividend payments. This ignited a flurry of responses, mainly criticizing the underlying assumptions of Shiller’s test. Kleidon (1986) have criticized the stationarity assumption. If the dividends process is nonstationary, the population moments can not be estimated by their sample counterparts. Flavin (1981) demonstrated that both the sample variance of \( P_t \) and that of \( P^*_t \) are estimated with downward bias in small samples, with the bias in the former estimate exceeding that in the latter. Finally, Marsh and Merton (1986) show that if managers smooth dividends, as there is extensive evidence of it, the variance bounds (4.2) will always be violated.

In response, Campbell and Shiller (1987, 1988a,b) propose an alternative testing procedure, by replacing \( \text{var}(P^*_t) \) with its counterpart obtained from a VAR, \( \text{var}(P'_t) \), where \( P'_t \equiv \mathbb{E}[P^*_t | I^*_t] \). Campbell and Shiller (1987) find that \( \text{var}(P'_t) \) is considerably more volatile than \( \text{var}(P_t) \), and conclude that: “our evaluation of the present value model for stocks indicates that the spread between stock prices and dividends moves too much”. Similar conclusions have been derived in Campbell and Shiller (1988a,b). They do not provide a formal proof. I prove it and interestingly I find that the Campbell-Shiller inequality is reversed.

Proposition 1: Let \( P'_t \equiv \mathbb{E}[P^*_t | I^*_t] \) and \( I^*_t \subseteq I^*_t \). Then the present-value relation implies that
\[ \text{var}(P'_t) \leq \text{var}(P_t) \quad (4.3) \]

\(^8\)The reason for the downward bias in estimating the variances of \( P^*_t \) and \( P_t \) is that, the former is more highly autocorrelated than the latter, and therefore the reduction of degrees of freedom by one is inadequate correction for the induced downward bias in the sample variance.
Proof: From the law of iterated projections we have

\[ P'_t \equiv \mathbb{E}[P^*_t | I_t^o] = \mathbb{E}[\mathbb{E}[P^*_t | I_t^o] | I_t^o] = \mathbb{E}[P^*_t | I_t^o]. \]

The proof is complete upon noticing that \( \text{var}(\mathbb{E}(x | I_t)) \leq \text{var}(x) \) for any random variable \( x \). □

The intuition behind Proposition 1 is simple. \( P'_t \) and \( P_t \) are two different optimal forecasts of \( P^*_t \), based on two different information sets. Different forecasts decompose the variance of \( P_t^* \) into two components, the forecast component and the forecast error component. Specifically

\[
P_t^* = \mathbb{E}[P_t^* | I_t^o] + \xi_t \quad \Rightarrow \quad \text{var}(P_t^*) = \text{var}(P_t) + \text{var}(\xi_t)
\]

\[
P_t^* = \mathbb{E}[P_t^* | I_t^o] + \xi_t' \quad \Rightarrow \quad \text{var}(P_t^*) = \text{var}(P'_t) + \text{var}(\xi'_t)
\]

Since the variance of the forecast error with less information is at least as large as the variance of the forecast error with more information (i.e., \( \text{var}(\xi_t) \leq \text{var}(\xi'_t) \)), the inequality (4.3) must hold.

### 4.2 Correspondence between \( P_t \) and \( P_t^* \)

In his seminal paper, Shiller (1981) contrasts the plots of prices, \( P_t \), with its ex-post rational counterpart, \( P_t^* \), and argues that stock prices are moving too much to be justified by future dividends. However, there are economic reasons as well as empirical evidence against the constant discount factor assumption. For example, Grossman and Shiller (1981) consider a constant relative risk aversion utility of consumption function,

\[
U(C) = \frac{1}{1-A} C^{1-A}, \quad 0 < A < \infty,
\]
where $A$ is the coefficient of relative risk aversion, which is a measure of the concavity of the utility function. Thus, price equals the expected present value of dividends discounted by the marginal rates of substitution:

$$P_t = \mathbb{E} \left[ \sum_{j=1}^{\infty} \beta^j \frac{U'(C_{t+j})}{U'(C_t)} D_{t+j} | I_t \right].$$

Figure (2) plots the real stock prices, along with three different perfect foresight measures that differ from each other only in the assumed discount rates. All data have been obtained from Shiller’s website. These plots give the impression that stock prices deviate from the fundamental values, so much so that in his Noble lecture, Shiller (2014) concludes: “the figure reveals that there is little correspondence between any of these measures of ex-post rational price and actual stock price. People did not behave, in setting stock prices.”

I have an alternative explanation: $P_t$ does not correspond to $P_t^*$ if investors do have noisy information about the future dividends and discount factors, which is certainly the case. To see this point, note that

$$P_t = \kappa D_{t-1} + \kappa \gamma \sum_{j=0}^{\infty} \beta^j (\epsilon_{t+j} + \nu_{t+j} | t), \quad (4.4)$$

and

$$P_t^* = \kappa D_{t-1} + \kappa \sum_{j=0}^{\infty} \beta^j \epsilon_{t+j}, \quad (4.5)$$

From (4.4) and (4.5) we see that

$$P_t - P_t^* = \kappa (1 - \gamma) \sum_{j=0}^{\infty} \beta^j \epsilon_{t+j} + \kappa \gamma \sum_{j=0}^{\infty} \beta^j \nu_{t+j}. \quad (4.6)$$

Using simulations, Kleidon (1986) found similar patterns for nonstationary dividends and rationally determined stock prices, and concludes that plots of $P_t$ and $P_t^*$ should not be used as evidence of excess volatility or against constant discount factor.
How much $P_t$ deviates from $P_t^*$? Assuming that the discount rate is the correct one (i.e., ruling out the joint hypothesis problem), if the signal is accurate, i.e., $\sigma^2_{r_t} \rightarrow 0$, then the right-hand hide of (4.6) is negligible, and $P_t$ will be close to $P_t^*$. Alternatively, if the signal is very noisy, i.e., $\sigma^2_{r_t} \rightarrow \infty$, then $P_t$ will deviate from its ex-post rational counterpart, $P_t^*$. This is because none of the terms on the right-hand side of (4.6) will be small.

4.3 Detecting Bubbles

There is widespread belief among economists and investors that there are episodes of bubbles, during which security prices deviate considerably from their fundamental values. In case of dot-com bubble in the late 1990s, the price of Internet-related stocks soared rapidly before plummeting in 2000, destroying more than six trillion dollar wealth. Similarly, the rapid drop in the housing price in 2008 is believed to be a reason for the global financial crisis of 2008-2009. Shiller considered these to be obvious cases of “bubbles” and market irrationality.

There are several procedures for detecting bubbles, mainly focusing on the violation of the transversality condition and the restrictions that it imposes on the time series of the data. Yet, most of these studies fail to detect convincing evidence of its existence. It is easiest to illustrate the econometrics procedures testing for bubbles in the context of a simple PV model with constant discount factor. Without imposing the transversality condition, (2.5) is only one of many possible prices that solves the Euler equation (2.1). Once we drop the transversality assumption, any solution can be written as

$$P_t = \sum_{j=0}^{\infty} \beta^{j+1} \mathbb{E}(D_{t+j}|I_t^*) + B_t,$$

(4.7)

where $B_t$ satisfies

$$B_{t+1} = \frac{1}{\beta} B_t + e_t, \quad \mathbb{E}[e_t|I_t^*] = 0$$

(4.8)

which I refer to as classic rational bubble, to emphasize the difference with the noise bubble.
If $B_t \neq 0$, (4.8) implies that $P_t$ is explosive irrespective of dividends process. Applying unit root test to the level and first differences of stock prices, Diba and Grossman (1988) find that price is nonstationarity in levels but stationarity in differences. Since differences of an explosive process still manifest explosive characteristics, they reject the presence of an explosive rational bubbles. Further tests by Diba and Grossman (1988) provided confirmation of cointegration between stock prices and dividends, supporting the conclusion that prices did not diverge from long-run fundamentals and thereby giving additional evidence against explosive bubbles.

In contrast to classic rational bubbles, noise bubbles are stationary since the price and dividends are cointegrated. Therefore, common econometrics procedures that check for explosive behavior have no power in detecting noise bubbles.

Exploiting a unique feature of housing market in the UK and Singapore, where property ownership takes the form of a leasehold (finite-maturity) or freehold (infinite-maturity), Giglio et al. (2016) propose a model-free test to directly check the transversality condition. The price difference between the freehold and leasehold for otherwise identical properties thus approximate the classic rational bubble $B_t$. They find no such evidence. However, unlike classic rational bubbles, noise bubbles satisfy the transversality condition. This could be the reason why direct tests of bubbles usually fail to provide compelling evidence of bubbles. To conclude, noise bubbles are unpredictable in real time, which satisfies Fama’s (2014) critic of bubbles.

4.4 Under- and Overreaction

Another widely documented anomaly is the existence of under and overreactions. Bondt and Thaler (1985) is one of the first papers providing evidence for stock markets overreaction. They find that when stocks are ranked on the basis of past returns, the losers portfolio consistently overperformed, while the winners underperformed. Further studies, however,

10See also Phillips et al. (2011).
find that underreaction is about as frequent as overreaction, so much so that Michaely et al. (1995) assert that: “We hope future research will help us understand why the market appears to overreact in some circumstances and to underreact in others.”

Critics interpret underreaction and overreaction as clear evidence of market inefficiency. However, in an efficient market the speed of stock price adjustment to new information depends on the nature of information. If the information is 100% accurate, then prices must adjust immediately. On the other hand, markets treat the noisy information with caution, and therefore the optimal adjustment will be gradual. It is easiest to illustrate this point with a simple example.

**Example:** Suppose at the beginning of period $t$ agents receive a good news about a permanent increase in dividends

$$s_{t|t} = \epsilon_t + \nu_{t|t}$$

For simplicity, I only consider the good news, i.e., $s_{t|t} > 0$. If the news is accurate ($s_{t|t} = \epsilon_t$), prices adjust immediately to the new level

$$P_t = \kappa D_{t-1} + \kappa s_t = \kappa D_t$$  \hspace{1cm} (4.9)

where $\kappa = \frac{\beta}{1-\beta}$. However, both underreaction and overreaction are possible if information is noisy.

**Overreaction.** If $s_{t|t} > 0$ and $\epsilon_t < \frac{\sigma^2}{\sigma^2_\epsilon} \nu_{t|t}$, then prices are given by

$$P_t = \kappa D_{t-1} + \kappa \frac{\sigma^2}{\sigma^2_\epsilon} (\epsilon_t + \nu_{t|t}) > \kappa D_t.$$  

However, at time $t + 1$ agents observe the true dividend change $\epsilon_t$, and prices adjust to the correct price (4.9). This gradual adjustment from time $t$ to time $t + 1$ gives the impression that prices overreact to good news.
Underreaction. If \( s_{t|t} > 0 \) and \( \epsilon_t > \frac{\sigma^2}{\sigma^2} \nu_{t|t} \), then prices are given by

\[
P_t = \kappa D_{t-1} + \kappa \frac{\sigma^2}{\sigma^2} (\epsilon_t + \nu_{t|t}) < \kappa D_t.
\]

When at time \( t + 1 \) agents observe \( \epsilon_t \) prices adjust to the correct price (4.9), which gives the impression that prices underreact to good news. Figure [3] presents these plots.

4.5 Tests of Forecast Rationality

EMH implies that the forecast errors are unpredictable given the information set available at the time when the forecast is made, that is

\[
E[S_t - E[S_t|I^s_t]|I^o_t] = E[\xi_t|I^o_t] = 0
\]

One straightforward way to test this hypothesis is by regressing \( \xi_t \) on variables included in \( I^o_t \) and testing that the coefficients are jointly zero. Empirical studies very often reject the null of market efficiency; see Rossi and Sekhposyan (2016) and the references therein.

A major difficulty in testing (4.10) is that \( E[S_t|I^s_t] \) is unobserved. In practice, this unobserved quantity is routinely replaced by an optimal forecast based on linear projections onto the linear space spanned by observable variables by the econometricians. However, econometricians’ information set, in general, is smaller than the agents’ information set, i.e. \( I^o_t \subseteq I^s_t \).

One might think, based on the conditioning down arguments of Hansen and Sargent (1991), that this would not create problems. Interestingly, this is not the case, and an econometrician who incorrectly uses the linear projection finds that (4.10) is rejected. Proposition 3 states that the innovations obtained from VAR are not martingale difference sequence which implies that asset returns are predictable by past returns or any other variable that contains some information on it.

**Proposition 3:** Let assumption A1 holds. An econometrician who uses linear projection to
Figure 3: Price Adjustment to New Information

(a) Precise Signal: Jump

(b) Imprecise Signal: Overreaction

(c) Imprecise Signal: Underreaction
construct excess returns, that is \( \tilde{\xi}_t \equiv S_t - \mathbb{E}[S_t|I_t^0] \), will find that \( \mathbb{E}[\tilde{\xi}_t|I_t^0] \neq 0 \).

### 4.6 Cross-Equation Restrictions

A distinguishing characteristic of rational expectations hypothesis is that the parameters describing the stochastic environment that the agents confront appears in the equilibrium solution. Campbell and Shiller (1987) propose a convenient method for characterizing the cross-equation restrictions that the PV relation imposes on the data. These restrictions are very often rejected by the data, which have interpreted as evidence against constant discount factors or sometimes rational expectation hypothesis.

In the following, I argue that very often the underlying assumptions used to drive these cross-equation restrictions are violated. PV model implies that

\[
S_t = \mathbb{E}[\kappa \sum_{j=0}^{\infty} \rho^j \Delta D_{t+j}|I_t^0].
\]

(4.11)

Assuming that the model has a backward-looking VAR representation

\[
\begin{bmatrix}
\Delta D_t \\
S_t \\
Z_t
\end{bmatrix} =
\begin{bmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{bmatrix}
\begin{bmatrix}
\Delta D_{t-1} \\
S_{t-1} \\
Z_{t-1}
\end{bmatrix} +
\begin{bmatrix}
\xi_{1,t} \\
\xi_{2,t}
\end{bmatrix},
\]

Campbell and Shiller (1987) drive these restrictions by projecting (4.11) onto \( I_t^0 \) and “assuming” that \( \mathbb{E}[Z_t|I_t^0] = A_i Z_t \)

\[
e_1' Z_t = \kappa \sum_{i=1}^{\infty} \beta^i e_2' A_i Z_t
\]

(4.12)

where \( e_1' = [0 \ 1] \) and \( e_2' = [1 \ 0] \). Since equation (4.12) holds for all realizations of \( Z_t \), we
have

\[ e_1' = \kappa \sum_{i=1}^{\infty} \beta^i e_2' A^i = \kappa e_2' \beta A (I - \beta A)^{-1} \]  

(4.13)

where the second equality follows by evaluating the infinite sum, which must converge because the elements of \( Z_t \) are stationary. Postmultiplying both sides of (4.13) by \((I - \beta A)\), we have

\[ e_1' (I - \beta A) = \kappa e_2' \beta A. \]  

(4.14)

When these restrictions are applied to PV asset pricing models, the finding is almost always a resounding rejection. However, these restrictions are derived by imposing the arbitrary assumption that \( \mathbb{E}[Z_t | I_t^p] = A^i Z_t \). If, as is certainly more plausible, the PV model is forward-looking, then these rejections are nothing but rejections of the underlying hypothesis used to derive (4.14). This proves the following proposition.

**Proposition 3:** Standard cross-equation restriction that exclude forward-looking representations, produce spurious rejections.

### 5 Historical Decomposition of Stock Prices

In this section, I propose an informal method for evaluating the fit of a present value model based on linear projection of the price-dividend ratio into the linear space spanned by future economic fundamentals. As far as I am aware of, Shiller (1981) is the first to propose a measure of fundamental value by constructing ex-post rational price using future dividends and a constant discount rate. Later, he proposed other measures of fundamental values using different time-varying discount factors; see Figure (2). Still, one might ask: why do we need another measure of fundamental value? Existing measures of fundamental values are using proxies for discount rates. From the joint hypothesis problem we know that the deviation of price from its fundamental value could be due to a bad discount rate model. I use realized returns, and therefore my proposal is immune to this critique.
The basic idea of the fundamental value is most easily explained in the context of the Euler equation:

\[
P_t = \mathbb{E}\left[ \frac{P_{t+1} + D_{t+1}}{1 + R_{t+1}} | I_t^s \right].
\]

Iterating forward and imposing the transversality condition, we obtain a general PV model

\[
P_t = \mathbb{E}\left[ \sum_{j=1}^{\infty} \left( \prod_{i=1}^{j} \frac{1}{1 + R_{t+i}} \right) D_{t+j} I_t^s \right].
\] (5.1)

Defining \( D_t = (1 + G_t) D_{t-1} \), we can write (5.1) as

\[
\frac{P_t}{D_t} = \mathbb{E}\left[ \sum_{j=1}^{\infty} \left( \prod_{i=1}^{j} Y_{t+i} | I_t^s \right) \right],
\] (5.2)

where \( Y_t = \frac{1 + G_t}{1 + R_t} \) is the discounted real dividend growth rate, which can be obtained directly from the data. Equation (5.2) is a modified Gordon growth model with time-varying dividend growth and discount rates. It has the usual interpretation that a high price-dividend ratio forecast dividends growth or lower future returns or some combination of the two.

I approximate the fundamental value by the fitted value of the linear projection of the price-dividend ratio into the linear space spanned by future \( Y_t \)'s:

\[
\frac{P_t}{D_t} = \gamma_0 + \gamma_1 \sum_{j=1}^{K} \prod_{i=1}^{j} Y_{t+i} + e_t.
\] (5.3)

Linear projection is appealing since it always exists and is unique, as long as the second moments exist and are bounded. Moreover, no particular assumption on the error term is required.

What should we expect from the correspondence between the observed price-dividend ratio and the fundamental value? Apart from the approximation error, the PV model implies that they must comove together if agents have some information about the future economic fundamentals. The price-dividend ratio will not coincide with the fundamentals
value, unless agents know the know future dividends and returns completely. Under the more realistic assumption that the market participants have some noisy information about the future fundamentals, price-dividend ratio and the fundamental value must comove together, with possible temporary deviations.

Note that regression (5.3) is different from the long-run predictability regressions

\[
\sum_{j=1}^{K} X_{t+j} = \gamma_0 + \gamma_1 \frac{P_t}{D_t} + e_t,
\]

where \( X_t = R_t \), or \( \Delta D_t \). The literature almost exclusively focused on the relative importance of dividend growth and discount factors. More important, the Least Squares estimates of the parameters of the regression (5.4) has the usual interpretation. In contrast, parameters of the (5.3) are of no particular importance. Here, I am only interested in the comovement of the price-dividend ratio with future economic fundamentals. In other words, this is not a “forecasting” regression. It is a “backcasting” regression. Intuitively, it captures that part of the price-dividend ratio that can be explained by future fundamentals.

Table (2) presents summary statistics and the results, where I set \( K = 10 \). The results are not particularly sensitive to the choice of \( K \). I present empirical results both for the full sample 1871-2014 and for two subsamples 1871-1945 and 1946-2014, to avoid the effect of the outliers in the later sample period. Moreover, there is extensive empirical evidence of dividends smoothing in the postwar period than before, which might affect the performance of the PV model.

The historical decompositions of prices are presented in Figure (4). Consistent with the model laid out in Section 2, the price-dividend ratio comoves with the fundamental value. The \( R^2 = 0.50 \). Moreover, stocks can remain overvalued or undervalued for an extended period of time. Depending on the researcher’s taste, it can be interpreted as noise component or irrational behavior of the markets. In retrospect of Section 3.1, I prefer the

\[11\] Here, forecast refer to the prediction of future, nowcast refers to the prediction of current status and backcast refer to the prediction of the preceding time.
Table 2: Backcasting Regressions

<table>
<thead>
<tr>
<th>Sample</th>
<th>$\gamma_0$</th>
<th>$\gamma_1$</th>
<th>$R^2$</th>
<th>corr($\frac{P_t}{D_t}$, $\hat{P}_t$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1971-2014</td>
<td>$-10.4$</td>
<td>$4.4$</td>
<td>0.44</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>(3.5)</td>
<td>(0.43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1871-1945</td>
<td>$0.01$</td>
<td>$1.6$</td>
<td>0.67</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>(2.5)</td>
<td>(0.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1946-2014</td>
<td>$-6.5$</td>
<td>$4.7$</td>
<td>0.51</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>(5.1)</td>
<td>(0.6)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: The regression equation is $\sum_{j=1}^{10} X_{t+j} = \gamma_0 + \gamma_1 \frac{P_t}{D_t} + \epsilon_t$. $\hat{P}_t$ is the fitted value of the regression. Data are annual. The numbers in the parenthesis are the standard errors.

former interpretation.

The late 1990s and early 2000s market was overvalued. Shiller refers to this period as the millennium bubble or the dot-com bubble, where the price of Internet related stocks soared rapidly before plummeting in 2000, destroying more than six trillion dollar wealth. He uses Cyclically Adjusted Price-Earnings (CAPE) ratio, originally defined by [Campbell and Shiller (1998)](https://doi.org/10.1111/1540-6261.00048), as an informal measure to characterize bubbles. CAPE ratio equals the S&P Composite Index, divided by the ten-year moving average of real earnings on the index to smooth out the effects of economic cycles:

$$CAPE_t = \frac{P_t}{[(EARN_t + EARN_{t-1} + \cdots + EARN_{t-10})/10]}.$$  

Shiller date stamps the dot-com bubble in the 1982-2000 period. My proposed method dates the beginning at 1987. The difference could be due to two reasons. First, CAPE is a backward measure which does not include future information. Second, date stamping according to CAPE is arbitrary, since the fundamental value is not defined explicitly. Similar to Shiller, I also find that there was a positive bubble from 1949 to 1968.

Consistent with Irving Fischer’s observation, I find no evidence that the market was overvalued before the Wall Street Crash of 1929. Fisher was perhaps the greatest economist.
Figure 4: Historical Decomposition of Prices

(a) Sample 1971-2014

(b) Sample 1871-1945

(c) Sample 1946-2014

Notes: Price (solid, black), fundamental value (dashed, red), and noise component (dotted, blue). Annual data, 1871-2014.
of his time, but the crash cost him most of his academic reputation and personal wealth. Just before the crash, Fischer claimed that the stock market had reached “a permanently high plateau”, and that “security values in most instances were not inflated.” Of course, I am not the first one to point out this. Using data on productive capital on stocks and tax rates to estimate the fundamental value, McGrattan and Prescott (2001) also find that stock prices were undervalued, even at their pick. See also Donaldson and Kamstra (1996).

Figure (4) shows that stock market can remain undervalued for extended period of time. Negative bubbles were present in at least two periods: 1938-1949, and 1965-1980. Modigliani and Cohn (1979) hypothesize that the stock market suffers from money illusion, discounting real cash flows using nominal discount rates. An implication of such irrational behavior is that when inflation is high (as it was the case during 1970s), the stock market is undervalued. One could also give a rational interpretation of this phenomenon: high inflation serves as a signal predicting a decline in future economic activities, and the stock market rationally reflected the decline into prices. See also Fama (1981).

6 Conclusions

Within a perfectly rational representative agent model, I present a simple modification of standard present value model that explains a wide variety of asset pricing phenomena. The key insight is the role of noisy information about the future economic fundamentals. I argue that the information structure is key in understanding many asset pricing puzzles. In particular, the presence of noise in the signal leads to the triple hypothesis problem. I argue that many asset pricing puzzles can be understood through the lens of the triple hypothesis problem, which opens the door for rational explanations to puzzling stock market phenomena. Specifically, using the triple hypothesis argument, I show that stock prices are not excessively volatile. In the empirical section I propose a new method for approximating the fundamental value. Applying my procedure to the US data, I find extensive evidence of
Figure 5: Comparison of CAPE and Fundamental Value

Notes: Percentage deviation of fundamental values from prices (solid, black), Cyclical adjusted price-earnings ratio (dotted, blue). Annual data, 1871-2014.
stock market overvaluation and undervaluation. It remains to be seen if these proposals are burned out in empirical studies.

We could learn from engineers. Long ago, they realized that improving fossil fuel engines is not the answer to pollution; instead they shift their attention to design electric cars. So far, economics has failed to explain some economic phenomena. Analogous to engineers, economist may need to revise economic modeling, with more emphasis on information instead of more complicated models.
References


Hamidi Sahneh, M., 2015. Are the shocks obtained from svar fundamental?


Appendix A

I first prove Lemma 1, which is an extension of Theorem 5.4.1 [Rosenblatt (2000)], by dropping the identically distribution assumption. In Lemma 2, I use Lemma 1 to prove the univariate case of Proposition 2.1, and then show that under Assumption 1 the multivariate case can be reduced to the univariate case.

Lemma 1: Consider a univariate causal and non-invertible VARMA($p,q$) model, that is, $r_{\Phi} = r_p$ and $r_{\Theta} < r_q$. Let $\phi^t(\tau)$ denote the characteristic function of $\xi_t$ and $\phi^t_{\tau_0}(\cdot) = \frac{\partial \phi^t(\cdot)}{\partial \tau_0}$. Then linearity of the best predictor in mean square implies that

$$\sum_{k=-\infty}^{\infty} (\gamma_k - \sum_{l=1}^{\infty} \beta_l \gamma_{k-l}) h^{t-k} (\sum_{l=1}^{\infty} \tau_l \gamma_{k-l}) = 0 \quad (A.1)$$

where $h^t(\vartheta) = \frac{\phi^t_{\tau_0}(\vartheta)}{\phi^t(\vartheta)}$ and $\beta_l$'s are the coefficients of the best linear predictor of $x_t$ in mean square

$$x_t^* = \sum_{l=1}^{\infty} \beta_l x_{t-l}$$

Proof of Lemma 1: Writing (A.1) in the MA form we have:

$$x_t = \sum_{k=-\infty}^{\infty} \gamma_k \xi_{t-k}, \quad \gamma_k = 0 \quad \forall k < 0 \quad (A.2)$$

The joint characteristic function of $\{x_{t-j}, j \geq 0\}$ is given by

$$\eta^t(\tau_0, \tau_1, \cdots, \tau_p, \cdots) = E\left\{exp\left(i \sum_{l=0}^{\infty} \tau_l x_{t-l}\right)\right\}$$

$$= \prod_{k=-\infty}^{\infty} \phi^{t-k} \left(\sum_{l=0}^{\infty} \tau_l \gamma_{t-l}\right) \quad (A.3)$$
while the joint characteristic function of \( \{x_{t-j}, j \geq 1\} \) is

\[
\bar{\eta}^t(\tau_1, \cdots, \tau_p, \cdots) = \prod_{k=-\infty}^{\infty} \phi^{t-k}(\sum_{l=1}^{\infty} \tau_l \gamma_{t-l})
\]  

(A.4)

Differentiating \( \eta^t(\tau_0, \tau_1, \cdots, \tau_p, \cdots) \) w.r.t. \( \tau_0 \) we have

\[
\frac{\partial}{\partial \tau_0} \eta^t(\tau_0, \tau_1, \cdots, \tau_p, \cdots) \bigg|_{\tau_0=0} = \eta^t_0(0, \tau_1, \cdots, \tau_p, \cdots) = \int ix_t \exp(i \sum_{l=1}^{\infty} \tau_l x_{t-l}) \, dF^t(x_t, x_{t-1}, \cdots, x_{t-p}, \cdots) 
\]  

(A.5)

\[
= i \int E[x_t|x_{t-s}, s > 0] \exp(i \sum_{l=1}^{\infty} \tau_l x_{t-l}) \, dF^t(x_{t-1}, \cdots, x_{t-p}, \cdots)
\]

where \( F^t(x_t, x_{t-1}, \cdots, x_{t-p}, \cdots) \) is the joint cumulative distribution function of \( x_{t-j}, j \geq 0 \).

Also by differentiating the logarithm of (A.3) w.r.t. \( \tau_0 \) we get:

\[
\frac{\eta^t_0(0, \tau_1, \cdots, \tau_p, \cdots)}{\eta^t(0, \tau_1, \cdots, \tau_p, \cdots)} = \sum_{k=-\infty}^{\infty} \gamma_k h^{t-k}(\sum_{l=1}^{\infty} \tau_l \gamma_{k-l}).
\]  

(A.6)

Similarly, differentiating the logarithm of \( \tilde{\eta}^t(\tau_1, \cdots, \tau_p, \cdots) \) w.r.t. \( \tau_j, j = 1, 2, \cdots \), we have

\[
\frac{\partial}{\partial \tau_j} \log \tilde{\eta}^t(\tau_1, \cdots, \tau_p, \cdots) = \sum_{k=-\infty}^{\infty} \gamma_{k-j} h^{t-k}(\sum_{l=1}^{\infty} \tau_l \gamma_{k-l}), \ j = 1, 2, \cdots
\]  

(A.7)

If the best predictor in mean square is linear we must have

\[
\eta^t_{\tau_0}(0, \tau_1, \cdots) = \sum_{k=1}^{\infty} \beta_k \tilde{\eta}^t_{r_k}(\tau_1, \tau_2, \cdots)
\]  

(A.8)

which implies

\[
\sum_{k=-\infty}^{\infty} (\gamma_k - \sum_{l=1}^{\infty} \beta_l \gamma_{k-l}) h^{t-k}(\sum_{l=1}^{\infty} \tau_l \gamma_{k-l}) = 0.
\]  

(A.9)

\( \square \)

**Lemma 2:** Let Assumption 1 hold. The univariate non-Gaussian AR model (??) is causal
if and only if the Wold innovations \( \{ \epsilon_t \} \) are MDS.

**Proof of Lemma 2:** A standard result for AR processes is that any AR\((p)\) process \( \{x_t\} \) which is non-causal with respect to the noise sequence \( \{ \xi_t \} \) can also be modeled as a causal AR\((p)\) with respect to a new noise sequence \( \{ \epsilon_t \} \) defined by\(^{12}\)

\[
\epsilon_t = \frac{\prod_{r<i \leq q} (1 - b_i L)}{\prod_{r<i \leq q} (1 - b_i^{-1} L)} \xi_t, \quad |b_i| < 1. \tag{A.10}
\]

which can be written as:

\[
\sum_{i=0}^{q-r} \alpha_i \epsilon_{t-i} = \epsilon_t \tag{A.11}
\]

where \( \epsilon_t = \sum_{i=0}^{q-r} \beta_i \xi_{t-i} \). Then (A.11) Lemma 1 and Corollary 5.4.2 of \cite{Rosenblatt} implies that the best one-step predictor of \( \epsilon_t \) is non-linear, i.e., \( E[\epsilon_t | \epsilon_{t-s}, s \geq 1] \) is non-linear.

If \( \epsilon_t \) were a MD, i.e. \( E[\epsilon_t | \epsilon_{t-s}, s \geq 1] = 0 \), Lemma 1 implies that:

\[
\sum_{k=\infty}^{\infty} \gamma_k h^{t-k} (\sum_{l=1}^{\infty} \tau_l \gamma_{k-l}) = 0 \tag{A.12}
\]

Since \( \mu_{a+1} \neq 0 \), we have

\[
\sum_{k=\infty}^{\infty} \gamma_k \gamma_{k-l_1} \cdots \gamma_{k-l_a} = 0, \quad l_1, \cdots, l_a = 1, 2, \cdots \tag{A.13}
\]

For the \( a \)th order partial derivative of the expression \( \text{(A.12)} \) w.r.t \( \tau_{l_1}, \cdots, \tau_{l_a} \) at \( \tau_{l_1} = \cdots = \tau_{l_a} = 0 \), \( i^{a+1} \mu_{a+1} \alpha! \) is multiplied by the expression \( \text{(A.13)} \) on the left. Since

\[
(1 - bz)(1 - b^{-1} z)^{-1} = b^2 + (b^2 - 1) \sum_{j=1}^{\infty} b^j z^{-j}
\]

\(^{12}\)See \cite{?}, page 103.
we have $\gamma_k = 0$ for $k > 0$. Therefore (A.13) is equal to

$$
\sum_{k=0}^{\infty} \gamma_{-k} \gamma_{-k-l_1} \cdots \gamma_{-k-l_a} = 0, \quad l_1, \cdots, l_a = 1, 2, \cdots.
$$

(A.14)

Also

$$\gamma_{-k} = \sum_{j=r+1}^{p} \alpha_j b_j^k, \quad k > 0$$

for some coefficients $\alpha_j \neq 0$, $j = r + 1, \cdots, p$. Therefore, equations (A.14) can be written as

$$\sum_{j_1, \cdots, j_a = r+1}^{p} \alpha_{j_1} \cdots \alpha_{j_a} b_{j_1}^{l_1} \cdots b_{j_a}^{l_a} \sum_{k=0}^{\infty} \gamma_{-k}(b_{j_1} \cdots b_{j_a})^k = 0$$

$l_1, \cdots, l_a = 1, \cdots, p$. Consider the set of equations obtained by letting $l_1, \cdots, l_a = 1, \cdots, s$. The matrix of this set of equations is

$$M = (M_{j,l}) = \{\alpha_{j_1} \cdots \alpha_{j_a} b_{j_1}^{l_1} \cdots b_{j_a}^{l_a}\}$$

where $j = (j_1, \cdots, j_a)$, $l = (l_1, \cdots, l_a)$, $j_1, \cdots, j_a = r + 1, \cdots, p$, $l_1, \cdots, l_a = 1, \cdots, s$. The determinant of this matrix is $(\prod_{u=r+1}^{p} \alpha_u)^{2a}$ multiplied by the $2a$-th power of the Vandermonde determinant

$$|b_j^l; j = r + 1, \cdots, q, l = 1, \cdots, s|$$

Since the determinant is nonzero, we must have

$$\gamma(b_{j_1}, \cdots, b_{j_a}) = \sum_{k=0}^{\infty} \gamma_k(b_{j_1}, \cdots, b_{j_a})^k$$

This implies $(b_{j_1} \cdots b_{j_a})$, for $j_1, \cdots, j_a = r + 1, \cdots, p$ are also zeros of $\gamma(z)$, a clear contradiction. Therefore the assumption that $E[\epsilon_t|\epsilon_{t-s}, s > 0] = 0$ cannot hold. ■

**Proof of Proposition 2.1:** The proof is similar to the Corollary 2.1 in [Hamidi Sahneh (2015)].