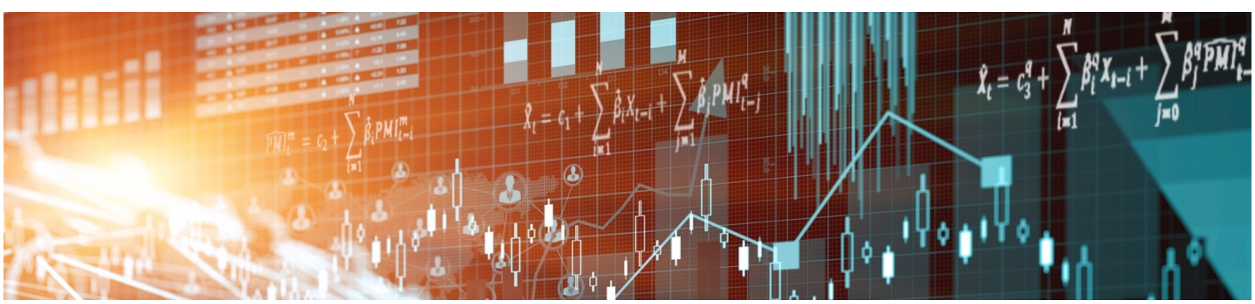
BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2017-38

A Counterfactual Valuation of the Stock Index as a Predictor of Crashes



by Tom Roberts

www.bank-banque-canada.ca

Bank of Canada Staff Working Paper 2017-38

September 2017

A Counterfactual Valuation of the Stock Index as a Predictor of Crashes

by

Tom Roberts

Financial Stability Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
troberts@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

I would like to thank Toni Ahnert, Jason Allen, Gabriel Bruneau, Thibaut Duprey, José Fique, Timothy Grieder, Rodrigo Sekkel, and Virginie Traclet for useful comments and suggestions, and William Leung for research assistance. All errors are my own.

Abstract

Stock market fundamentals would not seem to meaningfully predict returns over a shorter-term horizon—instead, I shift focus to severe downside risk (i.e., crashes). I use the cointegrating relationship between the log S&P Composite Index and log earnings over 1871 to 2015, combined with smoothed earnings, to first construct a counterfactual valuation benchmark. The price-versus-benchmark residual shows an improved, and economically meaningful, logit estimation of the likelihood of a crash over alternatives such as the dividend yield and price momentum. Rolling out-of-sample estimates highlight the challenges in this task. Nevertheless, the overall results support the common popular belief that a higher stock market valuation in relation to fundamentals entails a higher risk of a crash.

Bank topics: Asset pricing; Financial stability
JEL codes: G0, G01, G12, G17, G19

Résumé

Comme les facteurs fondamentaux du marché boursier ne semblent pas permettre d'effectuer des prédictions significatives des rendements à court terme, nous nous intéressons plutôt, dans cette étude, aux importants risques à la baisse (risques de krach). Nous analysons donc la relation de cointégration entre l'indice composite S&P (en logarithme) et les revenus (en logarithme) concernant la période 1871-2015, que nous combinons aux revenus lissés afin d'établir, pour commencer, un niveau de prix de référence contrefactuel. Le résidu, c'est-à-dire l'écart entre les prix réels et les prix de référence, procure une meilleure estimation (significative sur le plan économique) de la probabilité d'un krach dans un modèle logit, comparativement à d'autres indicateurs comme les dividendes et la dynamique des prix. Les estimations hors échantillon sur fenêtre glissante mettent en lumière la difficulté d'une telle tâche. Néanmoins, les résultats globaux obtenus corroborent l'idée répandue selon laquelle des cours boursiers élevés au regard des facteurs fondamentaux laissent supposer un risque accru de krach.

Sujets : Évaluation des actifs ; Stabilité financière
Codes JEL : G0, G01, G12, G17, G19

Non-technical summary

I propose an approach to estimating crash risk in an aggregate equity market. I use the cointegrating relationship between the log Standard & Poor's (S&P) Composite Index and log earnings over 1871 to 2015, combined with smoothed earnings, to first construct a counterfactual valuation benchmark. The price-versus-benchmark residual shows an improved, and economically meaningful, logit estimation of the likelihood of a crash over alternatives such as the dividend yield, volatility, and price momentum.

Equity markets are not a main direct concern for financial stability, but the underpricing of risk, in general, has been a key contributing factor to past financial crises. With risk-taking behaviour that would most likely permeate across asset classes, equity markets could provide a valuable lens into the possibility of a generalized lack of caution and search for yield. As such, the approach in this paper may help to inform and complement a vulnerability assessment of financial markets.

Rolling out-of-sample estimates highlight the challenges in this task. Nevertheless, a wide range of robustness checks support the conclusion that a higher stock market valuation in relation to through-the-cycle fundamentals entails a higher risk of a crash. To illustrate the results, using a 1920 to 2015 sample period, the estimated probability of a 25% year-over-year crash, starting within a pre-crash horizon of one year, peaks at 71% in December 1999, during the dot-com bubble. This compares to an unconditional probability of about 8.0%, and estimates of 55% and 21% prior to the Great Crash of 1929, and the recent crisis, respectively.

This paper relates to a number of strands of the literature: (1) the proliferation of research on early-warning models of financial crises and stress; (2) valuation measures as predictors of stock market returns; and (3) empirical studies on equity market crashes. Its main contribution is to combine these areas and provide a parsimonious approach to assessing downside risk based on through-the-cycle fundamentals.

1 Introduction

Severe corrections in equity markets, for instance, following the Internet bubble in 2000, and in the upheaval of the 2008-09 financial crisis, capture the public imagination. One could contend that they provide a lucid example of the literature's observation that stock markets exhibit much greater volatility than underlying fundamentals would seem to warrant (Shiller, 1981; LeRoy and Porter, 1981). But while elevated valuation measures would appear to be associated with below-average long-term returns (Fama and French, 1988; Campbell and Shiller, 1998, 2001; Cochrane, 2008), such a relationship is not apparent over a shorter-term horizon. The view of markets being efficient should preclude the possibility of any meaningful forward-looking insight into such events, but less empirical attention has been paid specifically to negative tail outcomes — i.e., crashes — in aggregate stock markets.

In this paper, I first estimate the cointegrating relationship between the log Standard and Poor's (S&P) Composite Stock Index and log earnings over the 1871 to 2015 period. I then construct a counterfactual benchmark index level that is instead reflective of log smoothed, through-the-cycle earnings, based on the historically estimated cointegrating relationship. The residual term between the actual log price level and the counterfactual benchmark acts as the main regressor in a logit estimation of nominal market crashes ranging from -15% to -30%, over the subsequent year. Although there is a substantial degree of noise in this forward-looking endeavour, this approach gives an economically meaningful in-sample improvement in the estimation of the likelihood of a crash, relative to a range of alternative indicators, including the dividend yield and price momentum. The results are robust to variations in specification, such as in the measure of smoothed earnings, crash magnitude, pre-crash horizon, in-sample period, and the inclusion of covariates. For out-of-sample performance, the results are mixed but hold some promise.

The current paper relates to a number of strands of the literature: (1) the proliferation of research that has proposed, or that has attempted to evaluate the usefulness of, early-warning models of financial crises and stress; (2) the aforementioned valuation measures as predictors of stock market returns; and (3) empirical studies on equity market crashes. Its main contribution is to combine the areas on early-warning and valuation ratios to examine "stress" (i.e., crashes) in equity markets and to propose a parsimonious approach for evaluating the conditional probability of this risk in relation to fundamentals. Thus risk is implicitly tackled as a phenomenon that arises from mispricing relative to through-the-cycle fundamentals. In contrast, ex-post market volatility or related measures are often used as a proxy for gauging financial risk — with the shortcoming that such volatility might increase only after a risk event has materialized.

The early-warning literature on the likelihood of financial crises or stress has considered leading indicators (e.g., Kaminsky and Reinhart, 1999; Borio and Lowe, 2002; Borio and Drehmann, 2009; Frankel and Saravelos, 2012; Pasricha et al., 2013), logit or probit models (e.g., Bussiere and Fratzscher, 2006;

Davis and Karim, 2008; Schularick and Taylor, 2012), or hybrid approaches (International Monetary Fund, 2010). Typically, asset prices and credit growth are used as indicators of potential build-up in vulnerabilities, in the hopes of signalling the occurrence of a financial crisis over some forward-looking horizon usually ranging from one to three years.[1] While some of these papers consider financial crises, others expand the dependent variable to include stress (Misina and Tkacz, 2009; Pasricha et al., 2013), owing to the limited number of crises in recent history with which to estimate a model. One motivation of this literature has been to provide policy-makers with tools to assess vulnerabilities in the financial system.

To the author's knowledge, fewer papers have looked specifically at the risk of aggregate equity market crashes. Baron and Xiong (2017) examine bank credit expansion and find it to be a predictor of increased bank equity crash risk, and Chen et al. (2001) show that trading volume and positive past returns can forecast negative skewness in individual stock returns, though with less statistical significance for the aggregate stock market.[2] Greenwood et al. (2016) examine sharp industry price increases, and net-of-market returns, and find that subsequent returns are not out of the ordinary but that the probability of a crash increases, while Goetzmann (2015) finds that the risk of a crash after a boom is only slightly higher. This paper differs in methodology, such as by borrowing from the early-warning literature and emphasizing the cointegration of prices with fundamentals, as well as in its interpretation that a large deviation from fundamentals has negative implications for both (1) shorter-term crash risk and (2) medium-term returns, and that (3) strong price increases per se, according to some particular threshold, are less important as an indication of crash risk or a bubble. In contrast, Goetzmann (2015) and Greenwood et al. (2016) do not estimate crash risk, and instead examine outcomes conditional on a price run-up.[3]

Stock market crashes by themselves are likely not a cause of financial crises (Mishkin and White, 2002), but they are certainly an element of financial system turbulence — and the underpricing of risk, in general, has been a key contributing factor to past financial crises. The underpricing of risks in any asset class has implications for market participants through the risk of portfolio losses, contagion, and confidence effects, but could also provide information about a more generalized lack of caution in financial markets.[4] Thus, while

[1]Indicators have often been based on growth rates or the gap from an HP trend.

[2]A recent literature on individual stock price crash risk, reviewed by Habib et al. (2017), has identified weak financial reporting and corporate disclosures, managerial incentives and characteristics, weak corporate governance, and other factors as potential sources of negative skewness in returns. Kim et al. (2011a; 2011b) are examples pertaining to corporate tax avoidance, and option sensitivity of chief financial officer incentives.

[3]The literature has also looked at price corrections in other assets, e.g., housing. See, for example, Bauer (2014). Borgy et al. (2009) investigate the possibility of detecting asset-price booms, and estimate the probability that such a boom turns out to be costly or low cost, according to subsequent sub-par real gross domestic product (GDP) growth. They find that above-trend growth in real GDP and house prices increases the probability of a costly stock-price boom.

[4]Negative equity returns are also commonly incorporated into indices of financial stress

equity markets are not the main direct concern for financial stability, they can provide a valuable lens into risk-taking behaviour that would most likely permeate across asset classes.[5]

Equity markets have real economy implications, as well. Equity prices could affect corporate investment decisions, including merger and acquisition activity (Edmans et al., 2012, 2015), and financing decisions, in terms of secondary or initial public offering (IPO) equity issuance (Baker and Wurgler, 2000, 2007). An ability to "take the temperature of the market" would be relevant in these decisions, and an effective leading indicator of corrections could, by extension of informing on the stage of the market cycle, provide insight into this. Furthermore, Mishkin and White (2002) remind us that wealth and the cost of capital are both standard channels in the monetary transmission mechanism.[6] And in terms of expectations, Galbraith (1979, 1994) and Chancellor (1999) highlight the profound effects of the 1929 crash on consumer and business confidence and expectations of the period.

Apart from issues of financial stability and the real economy effects of equity markets, investors would undoubtedly *wish* to have a better understanding of the elusive probability distribution of future returns. While long-term investors will be more interested in the expected equity premium over the long term, crashes can have confidence-shaking effects or affect those with shorter-term horizons. Crashes could also be further exacerbated by the presence of highly leveraged market participants with short-term liquidity needs (Shleifer and Vishny, 2011). Furthermore, I find that heightened crash risk has correlated with lower medium-term returns.

Of course, a stock index already reflects the market's expectations for growth, and its determination of an appropriate discount rate. If the market's wisdom about future return prospects was unassailable at all times, then there would be nothing further to do — but there is no consensus on this point. There is a voluminous literature concerning how financial markets could, at times, exhibit irrational behaviour or limits to efficiency (e.g., Barberis and Thaler, 2003; Scheinkman and Xiong, 2003; Baker and Wurgler, 2007; Gromb and Vayanos, 2010). The theoretical nature of how this could come about is not the focus of this paper. Certainly, in retrospect, it seems that investors have severely misjudged return prospects at various junctures — for instance, the Internet bubble of 1999-2000 or, conversely, the market depths of the late

(e.g., Cardarelli et al., 2009; Duprey et al., forthcoming; Illing and Liu, 2006).

[5] The International Monetary Fund (2015) describes the increase in post-crisis correlations across equities, bonds, and commodity prices. Hennessy et al. (2011) emphasize that the repricing of risk by investors, a key contributor to the crisis, benefited risky assets in general.

[6] At their nadirs, the Internet crash and the recent crisis erased \$6.5 trillion and \$17.7 trillion (about one-third from declining house prices, and much of the rest from declining financial asset values) in U.S. household net worth, respectively, compared with GDP of \$14.4 trillion in 2008 (FCIC, 2011).

1970s.[7,8]

I find that, using a 1920 to 2015 sample period, the estimated probability of a -25% year-over-year correction, starting in a given month, peaks at 8.2% in December 1999, prior to the 2000 start of the unwinding of the Internet bubble. This compares with estimates of less than 0.1% in the latter half of the 1970s, prior to the bull market over the following two decades. The unconditional probability is about 0.7%.[9] Using a rolling out-of-sample prediction with data from 1871 onwards, the out-of-sample estimated crash probability leading up to the 2000 Internet bubble peaked sooner at 12.7% in April 1999, reflecting that valuation levels had reached unprecedented levels by the end of the 1990s. This highlights the main challenge that noise and false-positive signals or, at least, risk detection that is "too early," are a prevalent challenge in the early-warning literature. Of course, these figures are not sufficient to provide an assessment of the models, to which we will turn in the rest of the paper.

The model abstracts from the myriad of events that could abruptly contribute to a shift in sentiment and trigger a correction, which will almost always be inherently unpredictable. The goal is to retain a parsimonious specification that evaluates only whether there is a meaningful change in downside risk, as prices deviate further from what might be supported by through-the-cycle fundamentals. Implicitly, the assumption is that the susceptibility to shifts in sentiment increases as lofty valuations become burdened under their own weight. Some form of eventual mean reversion seems reasonable to expect — what has been less clear is whether a model could provide useful estimates of time-varying crash risk. That is where this paper aims to contribute.

Section 2 describes the data, section 3 the methodology, section 4 the results, a discussion is provided in section 5, and the conclusion follows.

2 Data

The data set describing the S&P Composite Index is available from Global Financial Data (GFD, 2017) or through Robert Shiller's homepage.[10] It includes monthly observations for the index level, dividends, earnings, the CPI, and the 10-year interest rate, with the stock price data being the monthly average of daily closing prices.[11] Sector indices from GFD are used for some further

[7] Admittedly, history is only one realization of what might have occurred; but while there may have been reasonable arguments to support either the optimism or pessimism of those periods, at the time, reconciling the range of market extremes stretches the imagination.

[8] In other ways, the market's insight deserves acclamation: for example, Paul Samuelson's quip about predicting 9 out of the last 5 recessions implies a better track record than economic forecasters (Summers, 2016).

[9] For this measure, I first-difference the crash dummy so as to take only the initial start of the crash, based on year-over-year returns. Using a one-year pre-crash horizon gives crash estimates peaking at 71% in December 1999, from lows of 0.5% in the late 1970s, with an unconditional probability of about 8.0%.

[10] `http://www.econ.yale.edu/~shiller/data.htm`.

[11] Essentially the only difference between these two data sources is in the CPI prior to 1913, and particularly prior to about 1890. I use Shiller's CPI, as it is more accessible.

analysis in Appendix B.

The original index of stock values by the Standard Statistics Corporation, weighted by capitalization, was introduced in 1918. This later became the S&P Composite Index in 1926 (Siegel, 2014), and was a 90-stock average consisting of 50 industrials, 20 rails and 20 utilities. The Cowles Commission back-calculated the series to 1871 using the Commercial and Financial Chronicle (GFD, 2017), with this extension being reported in Shiller (1989). In 1957, the index was expanded to 500 stocks, to become the S&P 500 Index, and contained 425 industrial, 25 railroad, and 50 utility firms — Siegel (2014) reports that it soon became the standard for measuring performance for investments in large U.S. stocks. It has reflected the evolution of the U.S. economy, which since that time has shown a diminished importance of the materials sector, and an expansion of the health care, information technology, and financial sectors (Siegel, 2014). This structural economic shift could be one of the many issues to keep in mind in analyzing these longer historical time series.

3 Methodology

The value of a financial asset is straightforward, conceptually — it's derived from the stream of future cash flows, discounted to the present (Burr, 1938; Campbell et al., 1997). This discounting could include a time-varying one-period rate of interest and a premium for risk:

$$P_t = E_t\left[\Sigma_{k=0}^{\infty} \frac{D_{t+k}}{\prod_{j=0}^{k}(1+R_{t+j}+\varphi)}\right],$$

where D_t represents the cash flows, or specifically dividends, that are returned to the asset holder at the end of the period, R_t is the rate of interest and φ is a constant risk premium (Shiller, 2015).

For the unrealistic but nevertheless illustrative case where future growth in dividends and the discount rate are constant, or at least if we consider what would be the equivalent constant rates, the expression simplifies to the Gordon growth model (Gordon, 1962):

$$P_t = D_t E_t\left[\frac{1+G}{R-G}\right].$$

The expectations operator is often omitted, but here it expresses that future growth and the discount rate are, of course, unknown.[12] Taking logs, and adopting the convention of lowercase letters for logged variables:

[12]Campbell et al. (1997) provide a dynamic generalization of the Gordon growth model, allowing dividend growth and discount rates to vary across periods. This, along with a demonstration of the cointegration of prices and dividends and, by extension, prices and earnings, is provided in Appendix A.

$$\begin{aligned} p_t &= d_t + E_t\left[\ln(1+G) - \ln(R-G)\right] \\ &\approx d_t + E_t\left[G - \ln(R-G)\right] \\ &= d_t + E_t\left[G - r - \ln(1 - exp(g-r))\right]. \end{aligned} \tag{1}$$

From a practical standpoint, the $ln(R-G)$ term (or its derivations) presents a challenge, since it will approach negative infinity as $(R-G) \Rightarrow 0$. If the goal were to estimate an "intrinsic" value for the market index — as a counterfactual experiment to the market's own assessment — the derived price will be highly sensitive to the assumptions for R and G.[13]

As it turns out, estimations with smoothed historical values on the right-hand side (to proxy for through-the-cycle normalized values) do not reliably show statistical significance or the expected signs for growth and interest rate terms. Given that growth and interest rates exhibit broad and pronounced patterns over the sample period, in ways that do not consistently relate to stock prices, it is not surprising that estimating their role in valuation levels could be problematic.[14] Furthermore, the focus of investors is ostensibly on expectations of future growth and real interest rates, rather than naive extrapolations of historical trends.

Rather than risk a distraction, I focus instead on the cointegrating relationship between the log stock price index and log earnings. This leaves the coefficient on log earnings to absorb effects from trends in interest rates and other factors. Clearly this is not a perfect approach — for example, given that the decline in interest rates in recent years might not be suitably reflected — but it provides a starting point. Campbell and Shiller (1987) and Bauer (2014) provide examples from the literature where cointegrating relationships are similarly estimated between stock prices and dividends, and house prices and income, respectively.

Although OLS gives super-consistent estimates of the cointegrating relationship (Engle and Granger, 1987), the OLS estimator is not efficient and it may be biased for finite samples (Stock, 1987; Phillips, 1991). This paper uses a vector error correction model (VECM) for the main estimates of the cointegrating equation; this is expected to give more well-behaved estimates by taking into account the error structure of the underlying process (Johansen, 1988). Note also that when the series are cointegrated, the levels will contain predictive information beyond what can be achieved in any finite-order vector autoregressive representation (Hamilton, 1994).[15]

[13] One can see in Equation 1, for example, that a decline in the discount rate from 6% to 5%, other things being equal and holding the ratio of growth expectations to the discount rate constant, would add about 20% to the valuation level.

[14] Nominal rates show a large hump in the 1980s and smaller humps in the 1870s and 1930s, while real rates have at times gone negative — it is also debatable whether the concept of real rates was widely understood until the 1960s (Shiller, 2015). Real growth rates have trended downwards since peaking post-World War II.

[15] Stock and Watson (1993) propose dynamic OLS as another approach that is also asymptotically efficient and performs well in finite samples. The estimate of the cointegrating re-

The VECM is as follows, with p and e denoting log price and log earnings:

$$\begin{bmatrix}\Delta p_t \\ \Delta e_t\end{bmatrix} = \begin{bmatrix}\delta_p \\ \delta_e\end{bmatrix} + \begin{bmatrix}\Gamma_{pp} & \Gamma_{pe} \\ \Gamma_{ep} & \Gamma_{ee}\end{bmatrix} \begin{bmatrix}\Delta p_{t-1} \\ \Delta e_{t-1}\end{bmatrix} + \begin{bmatrix}\alpha_p \\ \alpha_e\end{bmatrix} (p_{t-1} - \beta e_{t-1}) + \begin{bmatrix}\epsilon_{p,t} \\ \epsilon_{e,t}\end{bmatrix}$$

Or, more succinctly,

$$\Delta X_t = \delta + \Gamma \Delta X_{t-1} - \alpha \beta' X_{t-1} + \epsilon_t$$

As in Campbell and Shiller (1998, 2001), I use earnings rather than dividends. While either variable works empirically for the logit estimations, dividend payout policy has noticeably trended downwards over the sample period. Lower dividends in the short term could be a harbinger of higher dividends in the future, as reinvested earnings lead to growth in earnings capacity, or they could be replaced with stock repurchases; either way, the relationship with dividends might be less apparent in a finite sample. The price-earnings ratio, meanwhile, has drifted upwards to a more modest degree. Furthermore, the model relates to the estimation of a valuation ratio rather than being a literal discounting of cash flows.

The estimated coefficient on log earnings is of particular interest since we might expect that valuation levels have generally risen over time in a sustained way. Siegel (2014) mentions a few reasons that could justify this: the historical decline in transaction costs; the low real returns on fixed-income assets, making them less attractive as an alternative; and broader participation in the stock market, through greater public awareness of the historical equity risk premium. Siegel (2014) also suggests that more conservative accounting since Sarbanes-Oxley may have served to understate earnings in recent years. These factors could be difficult to reliably measure, but the cointegration estimation could help to capture their effects. A coefficient even slightly greater than (less than) 1 would not be sustainable over an indefinite horizon, since the implied valuation ratio would then continue to rise (or fall) over time — but for a finite sample, it is arguably justifiable. Going forward, we should expect this coefficient to converge closer to 1, at some point.

3.1 Counterfactual valuation benchmark

Unfortunately, the cointegration residual, used directly, is not terribly useful for forecasting anything on its own. Instead, we would need a measure that looks through the economic cycle, by smoothing out the transitory hiccups in the economy, and corporate earnings. To make headway in that regard, I construct a "counterfactual valuation benchmark" — essentially replacing the

lationship, along with the implications from the logit models for corrections, is reasonably similar across these methods, and I use the dynamic OLS approach in Appendix B, while examining sectors.

contemporaneous log earnings observation with its value smoothed over some trailing period. This is the same type of approach as in Campbell and Shiller (1998, 2001), and follows the recommendation in Graham and Dodd (2009) to use a moving average as a way to normalize earnings.[16]

This paper uses mainly a one-sided 10-year moving average — this balances the goal of looking through the cycle, while not being overly long, and it has some precedent in Campbell and Shiller (1998, 2001). But the results are not particularly sensitive to this choice — varying the moving average window over a range of 7 to 15 years, or using a Hodrick-Prescott (HP) filtered gap, does not materially change the results, as shown in Appendix B. I use real earnings and prices for the benchmark and its residual, since this improves comparability for the moving average.

To minimize look-ahead bias (though this is still present in any in-sample estimation), I lag the earnings measure by three months.[17] Furthermore, I centre the benchmark such that the residual has a mean of 0 for the in-sample period — this implicitly assumes that the market, on average, has been fairly valued over that time.[18] As a result, the counterfactual valuation benchmark, and its residual, looks like this:

$$\begin{aligned} CVB_t &= \alpha + \beta_{p-e} e_{10,t-3} \\ \varepsilon_t &= p_t - CVB_t \end{aligned} \tag{2}$$

where $\alpha = \frac{1}{N}\sum p_t - \beta_{p-e} e_{10,t-3}$, such that $\sum \varepsilon_t = 0$, $e_{10,t-3}$ is the lagged 10-year moving average of log earnings, and β_{p-e} is the coefficient estimate for the cointegration of prices and earnings. Intuitively, α is the average price-earnings multiple that prevailed over the period, not counting the exponential effect of β_{p-e}.

The residual then functions as the main regressor of interest in the logit models for an equity market crash. In this approach, market expectations for growth and the risk premium, insofar as they are different from what is implied by the counterfactual benchmark, become embedded in the residual term between actual log prices and the benchmark:

$$\varepsilon_{e,t} = \psi_t + e_t + E_t\left[G - ln(R - G)\right] - CVB_t$$

[16] Siegel (2014) describes a range of valuation yardsticks that have been used over time, including the cyclically adjusted price-earnings (CAPE) ratio (Campbell and Shiller, 1998, 2001), the Fed model (Lander et al. (1997), which compares the S&P earnings yield to 30-year government bond rates), and Tobin's Q, which compares market value to book value adjusted for inflation.

[17] Until 2002, the U.S. Securities and Exchange Commission (SEC) required that annual reports be filed within 90 days of the fiscal year-end, with quarterly reports to be filed within 45 days of the fiscal quarter-end. Since then, the SEC has phased in an acceleration of filing deadlines, down to 60 days and 35 days for annual and quarterly reports, for companies with a public float of $75 million or more (SEC, 2002).

[18] A trailing moving average of earnings will be a downward-biased proxy of through-the-cycle earnings if there is positive real economic growth in the average year.

where log dividends d_t equals the log of the payout ratio, ψ_t, plus log earnings, e_t. In a simpler case where $e_t = \beta_{p-e} e_{10,t-3}$,

$$\varepsilon_{e,t} = \psi_t - \alpha + E_t \left[G - ln(R - G) \right]$$

So if growth expectations and the discount rate are more optimistic than what is reflected in $\psi_t - \alpha$, the residual will be positive and the stock index will be higher than the counterfactual benchmark.

3.2 Identifying crashes

The focus on extreme negative tail outcomes requires a definition for what this is. The main definition that I adopt in this paper is a forward-looking year-over-year drop of 25% or more for the nominal S&P Composite Index; I also try thresholds ranging from 15% to 30%.[19,20] I first-difference this dummy to identify only the start dates for a group of such observations, thus avoiding serial correlation in the dependent variable.

These thresholds yield a limited but, one can hope, adequate number of events (the tally of distinct events, which I define as starting at least six months after the last previous observation with a forward-looking drop that exceeds the given threshold, are, at the 20% threshold, 17 for 1881-2015 and 11 for 1920-2015; for 25%, 11 and 7; and for 30%, 7 and 5). These thresholds generally identify the events that we would expect to see (e.g., the crashes of 1929, 1973-74, 2000, and 2008), with the results listed in Table 1.[21] Appendix C provides more detail on the timing and subsequent returns of these crashes, as well as a brief summary of the historical context for several of the major episodes.

Thresholds above 25% for the year-over-year price change start to miss some episodes that are commonly considered to be major corrections, e.g., following the 2000 Internet bubble. While a more nuanced crash definition,

[19] In most cases, a nominal measure of crashes should be more relevant — i.e., cashing-out or switching to bonds will not reduce an investor's exposure to inflation; on the other hand, periods of deflation occurred over the 1920s and 1930s. Results are similar using a real-price crash definition, though this works less well for the full sample period of 1881-2015, possibly owing to the greater volatility of inflation from 1881 to 1920.

[20] The maximum drawdown from time t looking over a given horizon is another example of how crashes could be defined. The identified periods end up being fairly similar to the year-over-year definition, with a correlation of 0.85 for crash start dates, using a one-year horizon.

[21] Mishkin and White (2002) similarly use a 20% nominal threshold to identify crashes, though they make use of the Dow Jones Industrial Average and the Nasdaq in addition to the S&P Composite. They identify 15 episodes, in the years 1903, 1907, 1917, 1920, 1929 and 1930-33, 1937, 1940, 1946, 1962, 1969-1970, 1973-74, 1987, 1990, and 2000-01. This lines up very closely with the dates I identify at the 20% threshold, accommodating the fact that they use backwards-looking returns, hence some discrepancy in timing. Exceptions are 1934, where I identify February 1934 as having a forward-looking year-over-year return of about -21%; 1990, where I do not identify a correction (Mishkin and White (2002) identify 1990 based mainly on the Nasdaq); and 1961/1962 and 1987, which fall slightly below the 20% threshold for year-over-year returns.

Table 1: Start Dates of Correction Events

Year	20%yoy	25%yoy	25% max drawdown
1876	Jan	Feb	Feb
1883	**Jun**, Nov		
1892	**Jul**	**Aug**	**Aug**
1893	Jan		
1895	**Aug**		
1902	**Jul**	**Sep**	**Sep**
1903	Jan		Jan
1906	**Aug**	**Sep**	**Sep**
1916	**Oct**	**Nov**	**Nov**
1917	Mar		Mar
1919	**Dec**		
1920	Mar		
1929	**Jul**, Dec	**Jul**, Dec	**Jul**, Dec
1932	Feb		
1934	**Feb**		
1936	**Oct**	**Oct**	**Oct**
1939			**Oct**
1940	**Apr**		
1946	**Apr**		
1969	**May**	**May**	**May**
1973	**Jul**	**Jul**	**Jul**
1987			**Aug**
2000	**Aug**	**Sep**	**Sep**
2001	**Jul**, Oct, Dec		**Dec**
2002		**Mar**	Mar
2007	**Oct**	**Oct**	**Oct**

Months are listed for the start dates of a crash event, defined by the first difference of those observations where the forward-looking drop exceeds the given threshold. Bolded months represent the start of a distinct crash event, defined by a crash that starts at least six months after the last observation with a forward-looking drop that exceeds the given threshold, for the 1881-2015 sample period.

e.g., identifying crash severity over a more flexible horizon, could avoid this, in practice the simpler definition works well and reduces the possibility of picking up a crash that is mostly transitory and therefore of less interest.

One notable episode that does not qualify at the 25% threshold is Black Monday, on October 19, 1987, when the Dow Jones Industrial Average fell 22.6%, with a partial recovery of 16.6% over the next two days. In the monthly data for the S&P Composite, this crash manifests itself as a forward-looking year-over-year change of -19.9% and -15.9% starting in August and September of 1987, respectively.[22] However, of greater interest to this paper are corrections that are somewhat more sustained. In contrast, while there was a significant rebound starting in April 2009 during the financial crisis, the price index did not recover to the peak 2007 level until about March 2013.

The five-month period following an observation with a forward-looking drop that exceeds the given threshold is excluded from the logit estimations, similar in spirit to Bussiere and Fratzscher (2006). This reflects that we might be somewhat agnostic as to how to judge an estimate while a correction is still in progress. Because observations with such a year-over-year drop tend to be serially correlated, the total period that is excluded can be longer than five months. Variations to this post-crash exclusion do not change the general conclusions, although given the first-differencing of the crash dummy, some measure like this is important to include.

4 Results

This section includes results for the estimation of the cointegrating relationship of prices and earnings, the likelihood of a crash and a rolling out-of-sample estimation, as well as the likelihood of a crash using a pre-crash horizon of one year. Additional robustness checks can be found in the Appendix B.

4.1 Estimation of the cointegrating relationship of prices and earnings

Estimates of the cointegrating relationship between the log stock price index and log earnings are shown in Table 2, for OLS, Dynamic OLS, and VECMs. The main underlying vector autoregression (VAR) lag order of 14, covering just over a year with the monthly data, is chosen for the VECM, as a mid-range lag order identified by some information criteria, subject to a maximum limit of 60 lags.[23]

Most estimates for the log earnings coefficient lie in a range between 1.05 and 1.20, which I find to be plausible, at least for the finite sample period.[24]

[22] This episode can be added manually to the 20% and 25% threshold definitions without much effect on the estimations.

[23] Schwarz's Bayesian information criterion chooses 3 lags, the Hannan and Quinn information criterion chooses 14 lags, and the final prediction error and Akaike information criterion identify 21 lags, for nominal prices and earnings.

[24] A coefficient of 1.11, for example, after applying it to smoothed earnings and centring to

The exception is log real earnings in the VECMs, where the coefficient is around 1.34, which does not seem realistic.

Table 2 reports augmented Dickey-Fuller and Phillips-Perron p-values in the rightmost columns, which corroborate the point that log prices and log earnings are cointegrated. Tests confirm that both p and e are $I(1)$ processes. Visually, as seen in Figure 2, the VECM residual term, and the resulting counterfactual benchmark residual described in sub-section 3.1, has seen some sustained departures from the mean, though not by long enough to discard the notion of mean reversion.

Table 2: Estimation of the Cointegrating Relationship

	Model	Lags	Std. err. on earnings	R^2	RMSE	DFuller p-val	PPerron p-val
Real							
OLS	$p_{r,t} = 2.408 + 1.084 \times e_r$	0	0.012	0.815	0.370	0.000	0.000
OLS	$p_{r,t} = 2.101 + 1.203 \times e_r$	0	0.013	0.817	0.367	0.044	0.012
DOLS	$p_{r,t} = 2.320 + 1.113 \times e_r$	3	0.011	0.832	0.351	0.000	0.000
DOLS	$p_{r,t} = 2.168 + 1.164 \times e_r$	14	0.012	0.855	0.323	0.000	0.000
DOLS	$p_{r,t} = 2.139 + 1.175 \times e_r$	17	0.012	0.858	0.319	0.000	0.000
VEC	$p_{r,t} = 1.633 + 1.331 \times e_r$	3	0.081			0.000	0.000
VEC	$p_{r,t} = 1.646 + 1.336 \times e_r$	14	0.089			0.000	0.000
VEC	$p_{r,t} = 1.657 + 1.335 \times e_r$	17	0.105			0.000	0.000
Nominal							
OLS	$p_{n,t} = 2.615 + 1.059 \times e_n$	0	0.004	0.966	0.360	0.000	0.000
OLS	$p_{n,t} = 2.706 + 1.133 \times e_n$	0	0.004	0.977	0.295	0.005	0.001
DOLS	$p_{n,t} = 2.612 + 1.063 \times e_n$	3	0.005	0.969	0.341	0.000	0.000
DOLS	$p_{n,t} = 2.611 + 1.071 \times e_n$	14	0.004	0.973	0.317	0.000	0.000
DOLS	$p_{n,t} = 2.615 + 1.074 \times e_n$	21	0.004	0.974	0.311	0.000	0.000
VEC	$p_{n,t} = 2.666 + 1.104 \times e_n$	3	0.030			0.000	0.000
VEC	$p_{n,t} = 2.703 + 1.114 \times e_n$	14	0.031			0.000	0.000
VEC	$p_{n,t} = 2.769 + 1.122 \times e_n$	21	0.040			0.000	0.000

The resulting counterfactual valuation benchmark is shown in Figure 1. With the benefit of hindsight we can see that the counterfactual benchmark strays in some instances, notably in the early 1920s.[25] But for most of the sample period, the counterfactual benchmark does not look outrageously out of line. Some prominent boom episodes that were followed by crashes jut above the line, with periods of depressed market conditions falling below it. The residuals from the VECM and the counterfactual benchmark are shown in Figure 2.

construct the benchmark, implies a benchmark-to-smoothed-earnings ratio rising from about 15.1 in 1920 to about 17.8 in 2015.

[25]Mishkin and White (2002) describe a “very steep recession” where, from the business cycle peak in January 1920, the economy spiralled downwards until July 1921, exacerbated by the Fed’s raising of interest rates to combat inflation — 506 banks failed in 1921 owing to declining asset values, compared with 63 in 1919.

Figure 1: Counterfactual Valuation Benchmark

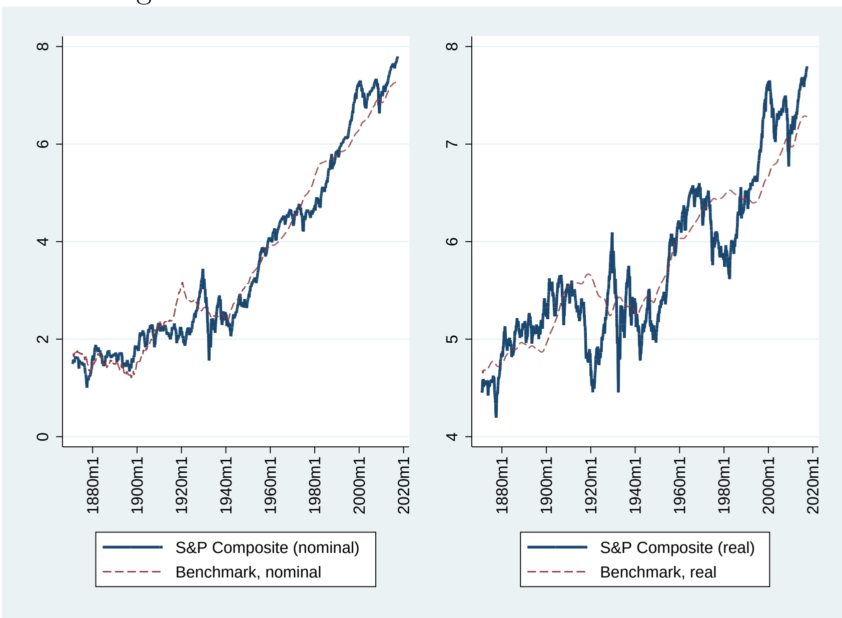
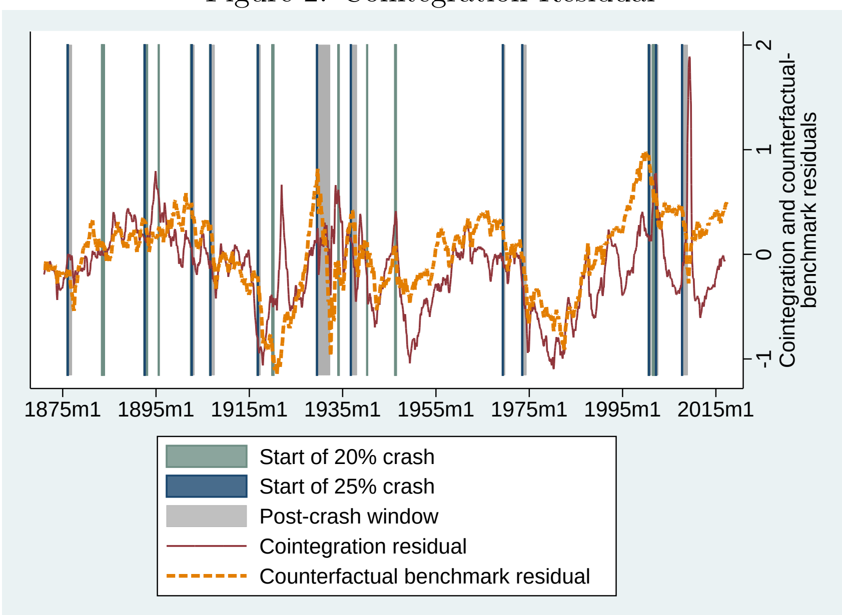


Figure 2: Cointegration Residual



4.2 Logit estimation of the probability of correction

The dummy variable for a forward-looking year-over-year crash shows some degree of autocorrelation, which is not surprising given the overlapping measurement of returns. OLS with Newey-West standard errors could be one way to address this issue, but it seems better to address this at the source.

With that goal in mind, I first-difference the year-over-year crash dummies, and also exclude the subsequent five months after a year-over-year drop exceeding the threshold is observed (e.g., following the start of a year-over-year crash in January, observations from at least February to June are excluded; if the year-over-year return starting in March is still more negative than the threshold, then July and August would also be excluded), because we would be ambivalent about evaluating model predictions while a crash could be in progress.[26] We do not want to penalize a high predicted crash probability when prices haven't fallen yet, and when the positive outcome variable has been first-differenced to 0. Thus, in effect, the included dummies are only the *initial* observation at the start of the year-over-year corrections, at least until the post-start-of-crash window is over. At the 25% threshold, the only dummies that are close together are for July and December 1929, and September 2000 and March 2002, which seems acceptable given that these relate to especially severe or protracted corrections.

The results are consistent over a broad range of specifications. Table 3 shows results for crashes ranging from 15% to 30% for the 1920 to 2015 sample period. The benchmark residual is consistently significant, with a positive effect on the likelihood of a crash — notably, the model fit improves as the crash threshold becomes more severe, peaking at the 25% threshold. This would seem to suggest that severe market corrections are more likely to be associated with a large divergence from fundamentals, whereas minor crashes of 15% or less could occur with or without any particular deviation from fundamentals. Appendix B shows similar results for the full 1881 to 2015 sample period.

The estimated crash probabilities are shown in Figure 3. Visually, the estimated crash probability seems to be relatively consistent in rising in advance of a crash, and ebbing subsequent to the crash. This alone is encouraging, in that there is a logical evolution to the model prediction. In addition, the predictions from using a 1920-2015, or 1881-2015 sample period, or a 20% or 25% threshold, look rather similar.

Table 4 adds additional regressors to the model, including five-year price growth, five-year volatility, and the de-trended dividend yield.[27] Except for the dividend yield, the coefficients take on the expected signs — but apart from the benchmark residual, none of them are statistically significant. One-year price growth and volatility have negative signs, reflecting that risk might be higher when growth has already slowed down or reversed.[28]

[26] Results are similar when excluding the subsequent 11 months.

[27] The dividend yield is de-trended with an HP filter.

[28] Analogously, Schularick and Taylor (2012) find that a negative second derivative of credit growth is a bad sign for financial stability.

Table 3: Logit Estimation of Likelihood of a Crash (No Added Pre-Crash Window): 1920-2015

	(1) Crash15	(2) Crash20	(3) Crash25	(4) Crash30
Benchmark residual	1.390*	2.581**	3.495***	3.056**
	(2.07)	(3.28)	(3.78)	(3.08)
Constant	-4.212***	-5.055***	-5.837***	-6.041***
	(-14.66)	(-12.06)	(-10.23)	(-9.82)
Observations	920	1021	1065	1090
Pseudo-R2	0.027	0.083	0.136	0.103
Chi-squared	4.278	10.780	14.284	9.477
P-value	0.039	0.001	0.000	0.002
Mfx	0.022	0.025	0.022	0.014
Mfx s.e.	0.012	0.011	0.010	0.008
AUROC	0.637	0.749	0.841	0.826
AUROC s.e.	0.069	0.069	0.069	0.096
Brier ratio	0.995	0.989	0.987	0.996
Brier ratio (crash=1)	0.990	0.976	0.968	0.986

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The crash dummy is defined by a forward-looking year-over-year drop in the nominal price index greater than the threshold, first-differenced to remove autocorrelation. The estimation excludes the five months immediately following a dummy value of 1 in the un-differenced series.

Figure 3: Estimated Probability of Crash Using Counterfactual Benchmark Residual

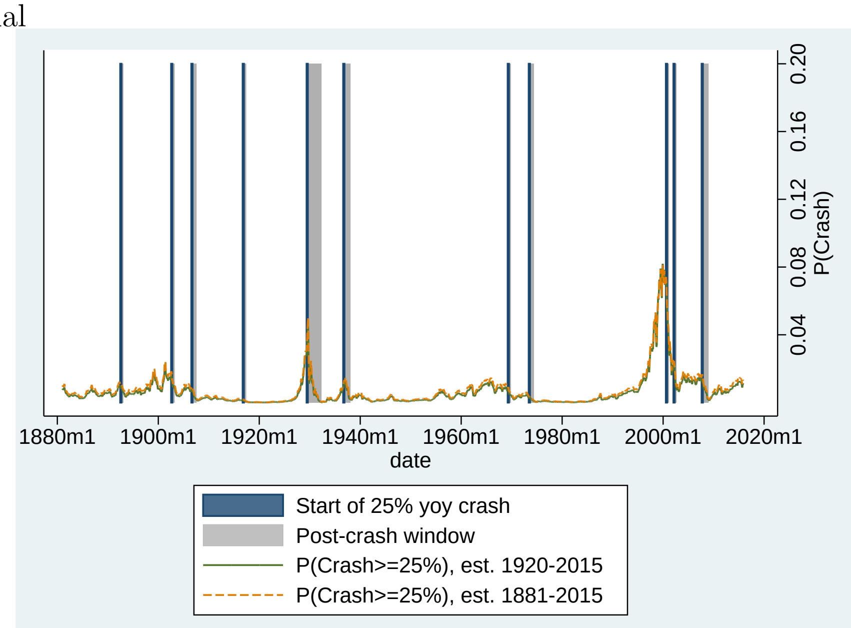


Table 4: Logit Estimation of Likelihood of a Crash (No Added Pre-Crash Window): 1920-2015

	(1) Crash25	(2) Crash25	(3) Crash25	(4) Crash25
Benchmark residual	3.495***	3.673**	3.494**	3.356**
	(3.78)	(3.25)	(3.18)	(2.79)
Price growth, 5yrs (annlzd)		3.893	7.481	6.601
		(0.52)	(1.08)	(1.02)
Volatility, 5yrs		5.789	7.535	8.611
		(1.21)	(1.48)	(1.65)
Div. yield, HP gap			0.467	0.446
			(1.07)	(0.68)
Price growth, 1yr				-0.598
				(-0.21)
Volatility, 1yr				-12.053
				(-1.47)
Constant	-5.837***	-7.244***	-7.697***	-6.916***
	(-10.23)	(-4.34)	(-4.52)	(-4.68)
Observations	1065	1065	1065	1065
Pseudo-R2	0.136	0.163	0.171	0.186
Chi-squared	14.284	14.710	15.811	19.974
P-value	0.000	0.002	0.003	0.003
Mfx	0.022	0.023	0.022	0.021
Mfx s.e.	0.010	0.011	0.011	0.011
AUROC	0.841	0.827	0.835	0.858
AUROC s.e.	0.069	0.090	0.088	0.074
Brier ratio	0.987	0.967	0.962	0.949
Brier ratio (crash=1)	0.968	0.943	0.936	0.922

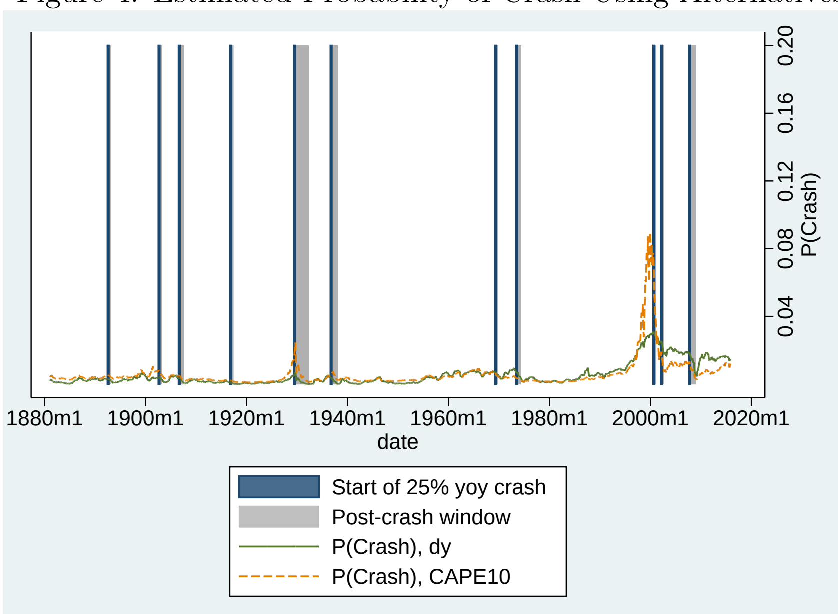
t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The crash dummy is defined by a forward-looking year-over-year drop in the nominal price index of ≥25%, first-differenced to remove autocorrelation. The estimation excludes the five months immediately following a dummy value of 1.

Table 5 shows results for some alternative indicators, including the dividend yield, the CAPE, the gap from an HP-filtered price trend, price growth, and volatility. The benchmark residual shows better model fit and a higher area under the receiver-operator characteristic curve (AUROC).[29] The Brier score is marginally better, though none of the indicators improve much on the unconditional mean prediction (sub-section 4.3, which includes a pre-crash horizon, shows better Brier score results).[30] The dividend yield and CAPE are relatively better than the remaining ones, but visually, Figure 4 raises some concerns. For example, the dividend yield prediction over most years since 2000 is higher than at any other point in the sample period, even for 1929, and not far below the 2000 peak. Going forward, the simple fact of the dividend yield being lower than in the first half of the 20^{th} century would automatically imply a higher crash risk, which does not seem reasonable.[31] For the CAPE, the late 1990s crash predictions are much larger than anything else in the sample.

Figure 4: Estimated Probability of Crash Using Alternatives



More detail on the marginal effects is shown in Table 6. When the benchmark residual is 0, indicating that the price index equals the counterfactual valuation benchmark, an increase in the benchmark residual of 0.1 would be associated with an increase in the likelihood of a 25% crash from 0.3%, to 0.4%, a fairly small difference.[32] However, when the residual goes from 0.4 to

[29]The receiver operating characteristic (ROC) curve plots the true positive rate $P(Signal = 1|Event = 1)$ against the false-positive rate $P(Signal = 1|Event = 0)$, across thresholds for each value of the signalling variable in the sample, where $Signal = 1$ if the signalling variable exceeds the threshold, and 0 otherwise. An AUROC of 1 is a perfect classifier of the event, while an AUROC of 0.5 is uninformative.

[30]The Brier score is the mean squared error for the probabilistic prediction of a binary outcome variable.

[31]A de-trended dividend yield does not fare better.

[32]The unconditional probability of a $\geq$ 25% year-over-year crash starting in a particular

Table 5: Comparison to Alternative Indicators: 1920-2015

	(1) Crash25	(2) Crash25	(3) Crash25	(4) Crash25	(5) Crash25	(6) Crash25
Benchmark residual	3.495***					
	(3.78)					
Div. yield		-0.785*				
		(-2.51)				
Price growth, 5yrs (annlzd)			10.142			
			(1.79)			
Volatility, 5yrs				0.502		
				(0.10)		
Log price, HP gap					0.947	
					(0.51)	
CAPE10						0.118***
						(4.35)
Constant	-5.837***	-2.571**	-5.931***	-5.105***	-5.070***	-7.502***
	(-10.23)	(-2.97)	(-7.66)	(-5.35)	(-12.32)	(-9.37)
Observations	1065	1065	1065	1065	1065	1065
Pseudo-R2	0.136	0.083	0.049	0.000	0.005	0.108
Chi-squared	14.284	6.276	3.207	0.010	0.258	18.889
P-value	0.000	0.012	0.073	0.919	0.611	0.000
Mfx	0.022	-0.005	0.066	0.003	0.006	0.001
Mfx s.e.	0.010	0.003	0.044	0.032	0.012	0.000
AUROC	0.841	0.766	0.677	0.382	0.526	0.830
AUROC s.e.	0.069	0.076	0.102	0.143	0.124	0.068
Brier ratio	0.987	0.992	0.993	1.000	0.999	0.993
Brier ratio (crash=1)	0.968	0.985	0.989	1.000	0.999	0.974

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The crash dummy is defined by a forward-looking year-over-year drop in the nominal price index of $\geq$25%, first-differenced to remove autocorrelation. The estimation excludes the five months immediately following a dummy value of 1.

0.5, indicating that the price index goes from about 49% to 65% higher than the benchmark, the likelihood increases from about 1.2% to 1.7%. This reflects the intuition that the likelihood of a disorderly outcome increases non-linearly as prices move further from fundamentals.

Table 6: Marginal Effects of Benchmark Residual on Probability of Crash: 1920-2015

Residual	dy/dx	std.err.	z	$P \geq \|z\|$
-0.4	0.003	0.002	1.535	0.125
-0.2	0.005	0.003	1.950	0.051
-0.1	0.007	0.003	2.193	0.028
0	0.010	0.004	2.424	0.015
0.1	0.014	0.006	2.588	0.010
0.2	0.020	0.008	2.622	0.009
0.4	0.040	0.017	2.313	0.021
0.6	0.079	0.043	1.864	0.062
0.8	0.152	0.100	1.527	0.127

The unconditional probability of a $\geq$ 25% year-over-year crash starting in a particular month is about 0.7%, after first-differencing to remove serial correlation.

4.2.1 Historical returns relative to estimated crash probability

Although the focus of this paper is on crash risk rather than expected returns, one might wonder if higher crash risk is compensated by higher returns on the upside, as well. The anecdotal evidence provided by history does not appear to support this notion. Table 7 looks at the distribution of subsequent three-year returns relative to the estimated crash probability.

Returns for a three-year horizon at the mean, and 10^{th} and 25^{th} percentiles, mostly continue to worsen as the crash probability estimate increases.[33] For the 75^{th} and 90^{th} percentiles, a deterioration in returns is not as obvious, but it would be hard to make the case that there is any improvement. Caution is warranted in drawing any hard conclusions, since the overlapping of the returns observations means that there is less independence in the sample, and fewer distinct episodes, particularly at the higher estimates of crash risk.

4.2.2 Rolling out-of-sample estimation

The in-sample results have been encouraging, but translating that into out-of-sample performance proves to be more challenging. To assess the out-of-sample performance, I run a rolling regression starting in January 1920, estimating the counterfactual benchmark residual from a VECM with data from 1871 up until

month is about 0.7%, after first-differencing.

[33] For the three-year returns, the (approximate) 10^{th} percentile observation in the $[3.5, 8.5)$ crash risk category is -16.3%, and has the 4^{th}-worst rank, while the 3^{rd}-worst observation is -35.9%.

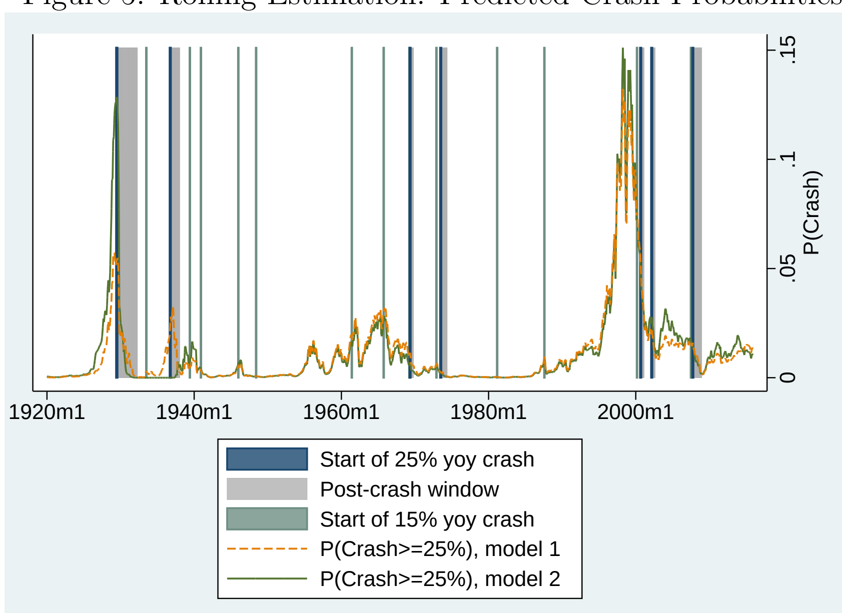
Table 7: Three-Year Returns (Annualized) by Estimated Probability of Crash: 1920-2013

Crash prob (%)	mean	p10	p25	p75	p90	N
$(0, 0.5)$	9.1	-2.6	4.0	14.5	20.2	708
$[0.5, 1.0)$	7.1	-4.1	1.8	12.6	16.8	223
$[1.0, 1.5)$	0.9	-12.9	-9.4	8.3	24.3	110
$[1.5, 2.0)$	0.0	-35.3	-22.0	22.8	24.1	17
$[2.0, 3.5)$	-4.5	-36.6	-29.6	13.6	18.6	32
$[3.5, 8.5)$	-9.4	-16.3	-13.2	-3.0	3.3	38

the end of the rolling window.[34] I then predict the probability of a crash using data from 1881 up until a year before the prediction, so as to not include observations that haven't yet been classified as a crash or no-crash. This is done for two models: Model 1 uses only the counterfactual residual and a constant, and Model 2 includes five-year volatility and five-year price growth. The results are shown in Figure 5.

At first glance, the rolling predictions in Figure 5 seem reasonable and relate to the incidence of crashes.

Figure 5: Rolling Estimation: Predicted Crash Probabilities



Model 1: Only the residual term from the counterfactual benchmark; Model 2: Residual term plus five-year price growth, and five-year volatility. The estimated probabilities are for a year-over-year crash $\geq 25\%$, using data up until a year before the prediction for the logit model, and up until the prediction, for the VECM estimation of price and earnings cointegration.

[34] I constrain the coefficient on log earnings to be at most -1, thus assuming that valuation ratios will not trend downwards exponentially over time. As can be seen in Figure 9 in Appendix B, this affects mostly just the 1920-1960 period of the rolling estimation.

Since the rolling estimation includes data only prior to each prediction observation, out-of-sample evaluation can be done directly on the rolling predictions — Table 8 shows a range of out-of-sample measures for the 1920-2015 period. The first column is the standard Brier score, and shows that most indicators match or almost match the null model with only a constant, with the exception of the CAPE, which does worse.

Column 2 shows the AUROC and its p-value is shown in column 3. It is important to note that the AUROC is not entirely valid as an out-of-sample measure — this is because the joint distribution of the reference and class variables will not necessarily be consistent from one sample to another.[35] Nevertheless, I include the AUROC since it is familiar in the early-warning literature and captures the general sense that a higher vulnerability measure is associated with clearer signals, at least within the given sample period. And it can also help to identify a negative case where an indicator clearly does *not* work, at least not in the expected sense of signals improving with a higher measure. In this respect, the null model, the HP gap for the nominal price index, and volatility do poorly since they all have AUROCs of less than 0.5.

Column 4 shows the out-of-sample pseudo-R^2; Model 1 does the best, although its out-of-sample pseudo-R^2 of .063 is rather modest.[36] Notably, Model 2 does poorly on this measure, along with volatility, which both have highly negative pseudo-R^2's.

Column 5 recalculates the Brier score while putting more weight (in this case, 10 times) on the crash outcomes. This seems like a reasonable adjustment, given the severity of the tail outcome in relation to other periods. The first-differencing of the crash return dummy variable also leaves a sparse number of crash observations. But these weighted quadratic prediction scores give only an almost-imperceptible edge to Model 1 and the CAPE. One reason for the only slight edge is that the predicted probabilities usually do not get too high — e.g., they reach 0.1 prior to the Internet crash; also, this measure is not using any pre-crash horizon, so higher predicted probabilities in the run-up to a crash weigh against the counterfactual benchmark and CAPE models. The next section will consider such a pre-crash horizon.

Closer examination of sub-periods can cast more light on what is driving the overall out-of-sample results for 1920-2015; additional tables by sub-period are provided in Appendix B. The out-of-sample pseudo-R^2 for Model 1 is attributable to the 1920-1949 period, where it and most other models actually do well, apart from Model 2 and volatility.

At the 25% threshold, there is relatively less happening in the way of crashes over the 1950-1989 period (with notable exceptions), and there is less differentiation across models. For the 1990-2015 period, Models 1 and 2 and the CAPE suffer somewhat — the reason is provided by Figure 5. They jump the

[35] For example, if the reference variable tends to take on higher values in the out-of-sample period, then the in-sample thresholds won't carry over well to the out-of-sample. The AUROC measures only signal reliability in relation to the distribution in the given sample.

[36] Without constraining the VECM coefficient for log smoothed earnings to be at most -1, the pseudo-R^2 would be .058

Table 8: Out-of-Sample, Rolling Estimation: 1920-2015

	Brier score	AUROC	AUROC p-val	Pseudo-R^2	QPS10
Null	0.0065	0.31	0.96	0.000	0.061
Div. yield	0.0066	0.75	0.01	0.021	0.060
HP gap	0.0066	0.49	0.54	-0.041	0.060
CAPE10	0.0080	0.82	0.00	-0.050	0.059
Volatility, 5yr	0.0065	0.37	0.88	-0.293	0.061
Price growth, 5yr	0.0065	0.61	0.16	0.017	0.061
CB, model 1	0.0067	0.81	0.00	0.067	0.059
CB, model 2	0.0068	0.67	0.06	-0.140	0.059

gun in predicting a higher crash probability too soon prior to the 2000 Internet crash — this higher predicted probability is eventually partially vindicated. As a result, all models except price growth and the null model have negative out-of-sample pseudo-R^2's in this period. The AUROC for Model 1 and the CAPE are 0.75, but this could give a misleading view of its helpfulness, given that valuation ratios were reaching unprecedented levels running up to the 2000 peak.

The rolling out-of-sample estimation illustrates the idiosyncratic nature of severe tail outcomes and how results can be swayed by the particular events that have occurred — this is explored further in Appendix B, where model coefficients can be seen to show some adjustment to each new event. Clearly, any given out-of-sample period should not be relied upon too heavily in dictating broader conclusions.

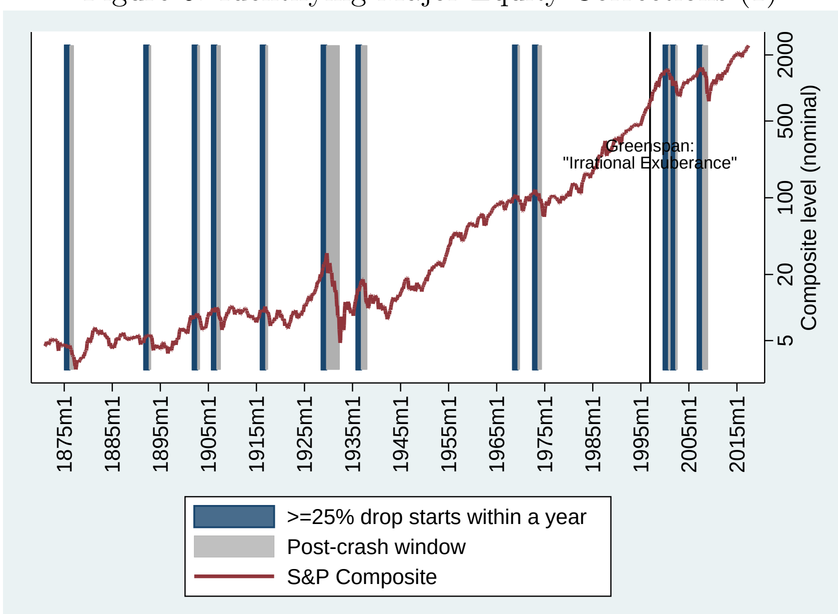
Overall, I would argue that the in-sample and out-of-sample results, along with relative coefficient stability (shown in Appendix B), visual inspection of predicted probabilities, and a stronger through-the-cycle rationale to the approach, suggest that the counterfactual benchmark residual is a more reliable indicator of downside risk for the equity market than the alternatives considered.

Interestingly, volatility does not do well by most criteria — this would cast some doubt on measures like Value-at-Risk, as applied to an aggregate equity market. Price growth and the HP price gap also do less well, yet such measures are often used in the early-warning literature. The dividend yield and the CAPE do relatively well according to some criteria, but the former seems to estimate risk in the recent period that is higher than much of the sample period, even observations for 1929, and not far below the 2000 Internet bubble episode, as seen in Figure 4. This could be at odds with common sense and could reflect the downwards trend in the dividend yield. Meanwhile, the CAPE does poorly using the Brier score and pseudo-R^2.

4.3 Pre-crash horizon of one year

Results so far haven't allowed for any forward-looking horizon relative to a crash — the estimations have attempted to fit the model to only the initial starts of these events. This is asking a lot — we might instead wish to include a forward-looking horizon whereby a crash will occur, if not imminently, then within an upcoming horizon of, say, one year. The estimated probabilities would then be evaluated against whether a crash starts over this near-term horizon, rather than against an exact timing of the event. This should also reduce a problem with first-differencing in that the "zero" outcomes — when no crash occurs — are over-weighted, since they exhibit the same overlapping year-over-year measurement as for the crash outcomes. Figure 6 shows the one-year pre-crash window for crashes exceeding 25%.

Figure 6: Identifying Major Equity Corrections (1)



If a crash occurs within the specified horizon, it is classified as part of a pre-crash window; this classification into pre-crash versus other periods becomes the dependent variable in the logit estimations. In other words,

$$C_t = \begin{cases} 1 & \text{if } \frac{P_{t+j}}{P_t} - 1 \leq thr \\ 0 & \text{if } \frac{P_{t+j}}{P_t} - 1 > thr \end{cases}, \tag{3}$$

where C_t is the dummy variable identifying the start of the crash events and j is set to 12 months, for the year-over-year return. To identify the pre-crash horizon,

$$Y_t = \begin{cases} 1 & \text{if } \sum_{k=0}^{K-1} C_{t+k} > 0 \\ 0 & \text{if } \sum_{k=0}^{K-1} C_{t+k} = 0 \end{cases}, \tag{4}$$

where for K, I focus on a 12-month horizon, and Y_t is the dummy variable identifying the pre-crash horizon.

Although the pre-crash horizon is intuitively appealing, it would introduce an even greater degree of autocorrelation in the model errors than with just the year-over-year definition (which we removed by differencing), arguably inducing a near-duplication of crash observations, despite there being only a limited number of distinct crash events. This would result in misleadingly smaller estimates of the standard errors.

A simple approach to address this is to collapse the data into a lower frequency, i.e., from monthly to annual, by estimating a model separately for each calendar month.[37] Since the dependent variable is set to true over a 12-month stretch before a crash, the annual frequency removes the statistical significance in the serial correlation of the model errors, and gives a more realistic picture of the number of independent observations. Thus the standard errors at an annual frequency should give a reasonable sense of the statistical reliability of the model.

Table 9 shows the results for the model with only the counterfactual benchmark residual. Columns 1 to 3 show the mean, minimum, and maximum, respectively, for each parameter or statistic, across the 12 estimations. Columns 4 to 7 show the estimations for the four quarter-end months — Appendix B provides the complete set of estimations. The benchmark residual continues to be both statistically significant and with an economically important marginal effect on the likelihood of a crash — a 0.1 increase in the residual, indicating a roughly 10% increase in prices relative to the counterfactual valuation benchmark, has been associated with a roughly 2.5 percentage point increase in the likelihood of a crash of $\geq$25% starting within the next year, compared with the unconditional probability of 8.0%, for the 1920 to 2015 sample period. In Appendix B, price growth and volatility have the expected signs, but are not statistically significant.

Because this approach does not over-weight the no-crash observations, as before, signal accuracy and model fit measures show an improvement, with an average AUROC of 0.85 and Brier score of 0.060 versus the unconditional Brier score of 0.073. The average pseudo-R^2 is 0.26.

Figure 7 shows the average of the one-year-ahead crash probability estimated across the annual data sets. It perhaps allows for a more intuitive interpretation than Figure 3, which corresponds to the probability of a crash starting in a particular month — an ambitious target, and not vital to the task of vulnerability assessment. With this annualized in-sample estimation, the predicted crash probability peaks at 55% prior to the 1929 crash, and about 71% prior to the Internet crash. The latter episode is distinguished by its sustained price run-up compared with the more ephemeral surge in valuations in 1929. Overall, the appearance of the estimated risk closely resembles the earlier figures.

Direct comparison of the economic significance of the results to the literature

[37] This is preferable to taking averages by year, or choosing one particular year-end, owing to the volatility in the data.

Table 9: Logit Estimation of Likelihood of a Crash, One-Year Pre-Crash Horizon: Model 1, 1920-2015

	(1) Mean PreCr	(2) Min PreCr	(3) Max PreCr	(4) March PreCr	(5) June PreCr	(6) Sept PreCr	(7) Dec PreCr
Residual	4.36	3.95	5.64	3.95**	4.16**	4.21**	4.50**
	(1.53)	(1.38)	(2.00)	(1.42)	(1.42)	(1.51)	(1.61)
Constant	-3.38	-3.70	-3.26	-3.32***	-3.36***	-3.31***	-3.40***
	(0.71)	(0.67)	(0.85)	(0.68)	(0.69)	(0.68)	(0.73)
Obs	88.75	87	90	90	90	89	87
Ps-R2	0.253	0.230	0.302	0.230	0.257	0.235	0.258
Chi2	12.42	11.30	14.77	11.30	12.62	11.53	12.58
P-val	0.000	0.000	0.001	0.001	0.000	0.001	0.000
Mfx	0.253	0.232	0.315	0.234	0.238	0.249	0.266
Mfx s.e.	0.086	0.075	0.103	0.082	0.077	0.088	0.091
AUROC	0.858	0.835	0.889	0.835	0.859	0.878	0.854
AUR s.e.	0.065	0.056	0.072	0.067	0.070	0.056	0.068
Brier	0.060	0.058	0.063	0.061	0.059	0.063	0.061
Brier_u	0.073	0.072	0.074	0.072	0.072	0.072	0.074
Month	n/a	n/a	n/a	3	6	9	12

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The crash dummy ("PreCr") is defined by a forward-looking year-over-year drop in the nominal price index of ≥25%, starting within one year. The estimation excludes the five months immediately following the start of the year-over-year correction. Regressions are run separately for each calendar month, thus collapsing the data to an annual frequency. Columns 1 to 3 show the mean, minimum, and maximum, respectively, for each parameter or statistic, across the twelve estimations. Columns 4 to 7 show the estimations for the four quarter-end months. Appendix B provides the complete set of estimations.

is not straightforward. Chen, Hong, and Stein (2001) translate the effects of stock turnover on conditional skewness into an effect on the prices of out-of-the-money put options. They find that a two-standard-deviation increase in turnover for the aggregate market would increase the price of a put option with a strike price 15% below the market price by about 25%. The effect of large movements in past returns are greater than what can be accommodated by the option-pricing model that they discuss (Chen, Hong, and Stein, 2001).[38]

As previously mentioned, Goetzmann (2015) and Greenwood et al. (2016) investigate returns conditional on a price run-up. Defining a crash as a 40% drawdown occurring within a two-year period, Greenwood et al. (2016) find that industry net-of-market returns of 50%, 100%, and 150% correspond to crash probabilities of 19%, 54%, and 81%, over the 1928-2014 period. Even though this relates to industry returns, which are more volatile than for the aggregate market, this does illustrate a comparable magnitude to the higher levels of risk estimated in Figure 7.[39] In contrast, Goetzmann (2015) finds that for a cross-section of 18 advanced economies over 1900-2014, following a boom of 100% or more over one year, the probability of a crash that at least halves the market goes from an unconditional 2% to 4% at a one-year horizon, and from an unconditional 6% to 15% at a five-year horizon. But he finds that a subsequent doubling is about twice as likely over the same horizons. These smaller probabilities are likely partly attributable to a steep threshold of -50% for the aggregate stock market of an advanced economy. More generally, this conditional rule about price run-ups does not differentiate between recoveries from depressed market levels, and other episodes.

4.4 Bootstrapped t-values

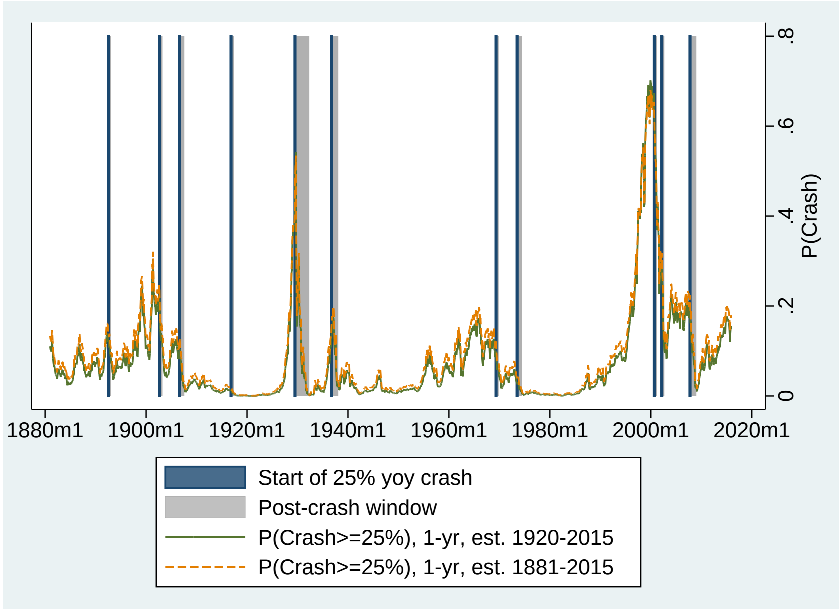
There could still remain the question of the sensitivity of the results to the particular set of crashes that happened to have occurred. The rolling out-of-sample estimation helps to answer this, though bootstrapping the standard errors and t-values might shed further light on the statistical robustness of the findings.

For the bootstrapping, I continue with the model from the previous section, with a pre-crash horizon of one year, and separate estimations by calendar month, so as to annualize the data. Because the annualization essentially eliminates the serial correlation in the model errors, I use a standard bootstrap,

[38]By comparison, if tail risk conditional on exceeding the threshold was assumed to be constant, the results in this paper imply that a deep out-of-the-money put option (e.g., a strike price 25% below the current market level) could vary in price by an order of magnitude, depending on the stage of the market cycle. For example, for an estimated crash risk of 10% by the exercise date of a European put option, the price should be 10 times greater than for an estimated crash risk 1%. Although the models in this paper are not estimated in a way that corresponds exactly to the mechanisms of a European option, because of the differencing of dependent variable, or the pre-crash horizon, the results convey the same pattern of relative risk.

[39]The one-year pre-crash horizon with a crash defined by the year-over-year change results in a similar time horizon as a maximum drawdown over two years.

Figure 7: Estimated Probability of Crash Starting Over One-Year Horizon



generating the average t statistic from 1,000 replications, and repeating this procedure 250 times for each of the 12 annual data sets by calendar month. Since price growth and volatility have not reliably been significant up until now, I focus on the benchmark residual, with results shown in Figure 8.

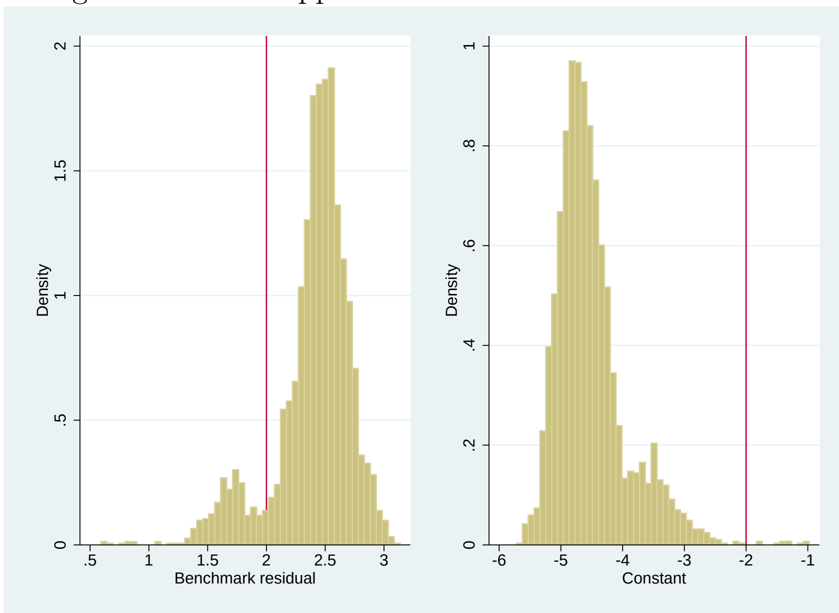
The t statistic for the benchmark residual at the 10^{th} percentile is about 1.91, suggesting that for almost 90% of the replicated samples, the residual would be significant at a standard significance level. Arguably, this supports the argument that a large deviation from the counterfactual valuation benchmark can be a useful predictor of negative tail risk in equity returns.

5 Discussion

Mean reversion is a well-known phenomenon of financial markets, and in this sense the results should not be surprising. Yet it can be an elusive goal to provide any sensible commentary about the likelihood of there being a bubble, where debate often continues after the fact (DeLong and Magin, 2006; Fama, 2013; Greenwood et al. 2016), and prospects for a correction. Media chatter and speculation about the next correction will continue to be a hallmark of each cycle. Rather than becoming resigned to an absence of any rigorous analysis, a balanced, probabilistic approach should be preferable — with emphasis on the probabilistic aspect, since the goal is not to "call" a correction or absence of one.[40] Instead, the point is to illustrate the increase in risk that occurs when prices become unhinged from fundamentals, and to validate the

[40]On the flip side of worrying about poor returns, recognizing when pessimism has been overdone could be equally helpful. Quantification of the magnitude of uncertainty can also be helpful in its own right, i.e., even when high valuations prompt doomsday prognostications, a quantitative perspective could lead to a more modest level of conviction.

Figure 8: Bootstrapped t Statistics of Model Coefficients



Bootstrapped t-statistics for the logit model of 25% year-over-year nominal price crashes starting within a one-year horizon, with the counterfactual benchmark residual as a regressor. Each bar is the average t statistic from a bootstrap with 1,000 replications. This procedure was repeated 250 times, separately for each annual data set ending on a given calendar month, to generate the distribution of bars.

use of through-the-cycle, cyclically smoothed earnings as a way to proxy for these fundamentals.

As stated earlier, the parsimonious approach in this paper abstracts from types of events that might trigger corrections, as well as reasons for the general rise in valuation ratios over time, evidenced by the coefficient on earnings of greater than 1 in Table 2. A trigger could be an easily identifiable event like the Kobe earthquake of January 17, 1995, and its effect on Japanese markets, or something less immediately noticeable, like the lock-up expirations and insider selling seen with the bursting of the Internet bubble (Ofek and Richardson, 2003). The latter factor could conceivably have been foreseen in a general sense — but for the most part, foresight into the timing of any specific type of trigger will be unattainable.

A great deal of worthwhile historical context is beyond the scope of this paper, with Table 24 in Appendix C providing only a brief summary. But the recent financial crisis of 2008-09, and its comparison to previous episodes, is of particular interest, and illustrates both the potential and limitations of this paper's approach.

The Financial Crisis Inquiry Commission (2011) recounts that home refinancing following rate cuts by the Federal Reserve led to $2 trillion in equity extraction over 2000 to 2007, at a rate up to seven times greater than a decade previously. A surge in house prices ensued, accompanied by burgeoning securitization and a collapse in lending standards. The effects did not escape other asset classes, with share prices of "the most aggressive financial firms" reaching all-time highs (FCIC, 2011) and risky assets in general benefiting from the repricing of risk by investors (Hennessy et al., 2011). Consumption fuelled by the expansion of credit, in spite of stagnant wages, also presumably benefited the economy on a short-term basis, and thus corporate earnings and stock prices.

Though the conduit for speculation differed from 1929, parallels are apparent. The rapid rise in house prices led to a "gold rush" mentality and increased speculative activity (FCIC, 2011), reminiscent of descriptions (e.g., Galbraith, 1979; Chancellor, 1999) of the mania that became more earnest by 1928.[41] The mid-2000s enthusiasm over financial innovation and unwavering faith in housing markets could be argued to have parallels with Chancellor's (1999) account of the new-era mentality in the 1920s, inspired by the earlier creation of the Federal Reserve (hailed as a remedy to "booms, slumps, and panics"), the extension of free trade, declining inflation, and more "scientific" corporate management.[42] Leverage played an important role in both 1929 and 2007, with brokers' loans in the 1920s viewed as being safe due to cash margin and stock collateralization (Galbraith, 1979), mirroring the complacency toward risk in mortgage-backed securities, in that case helped along by their triple-A ratings.

With the 2008-2009 crisis, it is often emphasized that vulnerability in the financial system was exacerbated by short-term funding, extraordinary leverage, and the opaqueness and interconnectedness of major financial institutions. Such aspects might lead one to conclude that the stock market is a sideshow for financial stability concerns, but Galbraith (1979, 1994) argues that the 1929 stock market debacle can be blamed for shaken consumer and business confidence, leading directly to falling business investment. Chancellor (1999) asserts that the stock market decline had a profound effect on expectations and caused the failure of certain banks engaged in the securities business, in turn leading to a crisis of confidence in the banking system.[43]

In its quantitative abstraction of these events, Chart 7 shows a fairly elevated estimated risk of correction starting within a year, of about 21% in mid-2007, exceeding the mid-1960s estimates of close to 18%, but well below the Great Crash of 1929, which peaked at 55% in September 1929, and even further dwarfed by the Internet bubble, which peaked at 71% in December 1999. As already pointed out, the latter two events were much more focused

[41] Both these booms were accompanied by a mushrooming of morally dubious conduct. For instance, a Treasury Department analysis in 2006 found a 20-fold increase in mortgage fraud reports between 1996 and 2005 (FCIC, 2011). Hennessy et al. (2011) state that fraud was not an essential cause of the recent crisis, but that it should have been a leading indicator of deeper structural problems in the market. In the late 1920s, over 100 stocks on the New York Stock Exchange in 1929 were subject to price manipulation (Galbraith, 1979), while investment banks "frequently dumped" unwanted stocks or included related-party issues in the investment trusts that they promoted to the public (Chancellor, 1999).

[42] New technologies were also a popular focus of stock speculation in the 1920s (Chancellor, 1999).

[43] There is sharp disagreement about the role of the crash in the ensuing depression, a debate that is not likely to be resolved owing to its politicization (Chancellor, 1999).

on the stock market. Other risk channels in the economy, and contagion stemming from that, will not be fully reflected — risk assessment, even if it pertains to equity markets, clearly can be complemented with other types of analysis. Nevertheless, high valuation levels were a symptom that could have reinforced a diagnosis of permissive risk-taking.

Insight could undoubtedly be improved if these and other factors explored in the literature could judiciously be captured, although long time series for the sample period considered in this paper in most cases would not be available. In terms of sentiment measures, examples include investor surveys, mutual fund flows (reflecting retail investor activity), trading volume, closed-end fund discounts, option implied volatility, IPO activity, and equity issuance (Baker and Wurgler, 2007). While one could expect general optimism to be already reflected in prices to a large extent, heterogeneous sentiment or actions of specific groups seem likelier to help, such as the aforementioned insider selling described in Ofek and Richardson (2003).

Concerning the broader question of financial stability, stock markets could provide one indication of potential underpricing of risk, even if they are not themselves a core issue. Although the estimation of equity market crash risk might be considered unfeasible by many, their volatility and presumable mean reversion should make this task *more* feasible rather than less, compared with the challenging post-crisis mission undertaken by central bankers to predict financial crises. The rigorous estimation of relevant vulnerabilities, one by one, could provide a more fruitful approach than the ad-hoc collection of indicators that are often assembled to gauge the risk of a financial crisis.

This paper might prompt the notion that such models could be adapted as market timing tools, though the aim is not to advocate for this. The obvious issue is the general noisiness of signals of crash risk (and implied overvaluation), the fact that overvaluation can persist for a long time, and insufficient out-of-sample evidence. Furthermore, historical data are a hazy guide to what might materialize in the future. Based on a hypothetical portfolio that reinvests dividends and switches to long-term bonds whenever the estimated overvaluation exceeds a given percentage threshold, for the 1871-2015 period, market timing in this manner is detrimental to total returns for thresholds below roughly 60% above the benchmark. Above this threshold, it would appear to have been beneficial; but thresholds this high amount to cherry-picking only a couple of periods that appear to have been the most extremely overvalued, in particular from 1928m10 to 1929m10 and 1996m10 to 2002m5.[44] Which is perhaps the point: truly frenzied manias are rare occurrences, but if they could be avoided, obviously it would help. Over-concern about more garden-variety levels of market enthusiasm comes at the cost of reinvested dividends and the difficulty of re-timing entry.

Having said all of this, caution is warranted in interpreting results. Even a century of data provides an uncertain guide to the future, and there are few major booms and busts from which to draw empirical conclusions.

[44]This abstracts from tax considerations; in the presence of capital gains taxes, the threshold would be even higher.

6 Conclusion

This paper suggests an approach for assessing downside risk — namely, crashes — in equity markets. In pursuit of this goal, I first construct a counterfactual valuation benchmark by estimating the cointegration of log prices and log earnings, and then replacing log earnings with a smoothed measure to approximate through-the-cycle fundamentals.

Inevitably, there is considerable noise in this endeavour; however, the residual between the counterfactual benchmark and the S&P Composite Index is a statistically significant, but also an economically meaningful predictor of large equity price corrections in the 15% to 30% range. This is a novel result, given that the general view has been that there might be only some token statistical significance to shorter-term return predictions. The counterfactual benchmark residual out-performs measures such as the dividend yield, the CAPE, equity price growth, volatility, and an HP-filtered gap.

Out-of-sample results hold some promise, but show that the models should be taken with a grain of salt. Depending on how one measures the results, an unconditional mean might be the "best" out-of-sample prediction, but the idiosyncrasy of rare events warrants caution with regard to any given out-of-sample exercise. Apart from the statistical issue of having only a small number of major corrections to work with, factors such as secular trends in demographics, investor attitudes, and economic growth could mean a permanent departure from relationships estimated with past data. Nevertheless, a reasoned analysis of past data is a relevant contribution to such questions.

Even if one is more concerned with the long-term equity risk premium, market dislocations can last for extended periods, a fact that should be relevant to any investor with a finite horizon or liquidity requirements. From a financial stability point of view, equity markets are not the focal point, but can provide a lens into risk-taking behaviour. And, finally, market crashes will continue to be the subject of much prognostication, regardless of how one views the feasibility of estimating this risk.

References

Baker, M., and J. Wurgler (2000): "The Equity Share in New Issues and Aggregate Stock Returns," *Journal of Finance*, 55(5), 2219–57.

——— (2007): "Investor Sentiment in the Stock Market," *Journal of Economic Perspectives*, 21(2), 129–151.

Barberis, N., and R. Thaler (2003): "A Survey of Behavioral Finance," in *Handbook of the Economics of Finance*, ed. by G. Constantinides, M. Harris, and R. Stulz, vol. 1, chap. 18, pp. 1051–121. Elsevier, London/New York.

Baron, M., and W. Xiong (2017): "Credit Expansion and Neglected Crash Risk," *Quarterly Journal of Economics*, pp. 713–764.

Bauer, G. (2014): "International House Price Cycles, Monetary Policy and Risk Premiums," *Bank of Canada Staff Working Paper* (2014-54).

Bilmes, L. J., and J. E. Stiglitz (2008): "The Iraq War Will Cost Us $3 Trillion, and Much More," http://www.washingtonpost.com/wp-dyn/content/article/2008/03/07/AR2008030702846.html, Accessed: 2017-06-22.

Borgy, V., L. Clerc, and J.-P. Renne (2009): "Asset-Price Boom-Bust Cycles and Credit: What is the Scope of Macro-Prudential Regulation," *Banque du France Document de Travail*, (263).

Borio, C., and M. Drehmann (2009): "Assessing the Risk of Banking Crises Revisited," *BIS Quarterly Review*, (March).

Borio, C., and P. Lowe (2002): "Asset Prices, Financial and Monetary Stability: Exploring the Nexus," *Bank of International Settlements Working Paper*, (114).

Brady, N. F. (1988): "Report on the Presidential Task Force on Market Mechanisms," *U.S. Government Printing Office*.

Burr, J. B. (1938): *The Theory of Investment Value*. Harvard University Press.

Bussiere, and M. Fratzscher (2006): "Towards a New Early Warning System of Financial Crises," *Journal of International Money and Finance*, 25(6), 953–973.

Campbell, J., A. Lo, and A. MacKinlay (1997): *The Econometrics of Financial Markets*. Princeton University Press, Princeton, New Jersey.

Campbell, J., and R. Shiller (1987): "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95, 1062–1088.

——— (1998): "Valuation Ratios and the Long-Run Stock Market Outlook," *Journal of Portfolio Management*, (Winter).

——— (2001): "Valuation Ratios and the Long-Run Stock Market Outlook: An Update," *NBER Working Paper*, (No. 8221).

Cardarelli, R., S. Elekdag, and S. Lall (2009): "Financial Stress, Downturns, and Recoveries," *IMF Working Paper*, (09/100).

Chancellor, E. (1999): *Devil Take the Hindmost*. Farrar, Straus and Giroux, New York.

Chen, J., H. Hong, and J. C. Stein (2001): "Forecasting Crashes: Trading Volume, Past Returns and Conditional Skewness in Stock Prices," *Journal of Financial Economics*, 61, 345–391.

Cochrane, J. (2008): "The Dog That Did Not Bark: A Defense of Return Predictability," *Review of Financial Studies*, 21(4), 1533–1575.

Davis, E., and D. Karim (2008): "Comparing Early Warning Systems for Banking Crises," *Journal of Financial Stability*, pp. 89–120.

Davis, E. P. (2003): "Comparing Bear Markets - 1973 and 2000," *National Institute Economic Review*, (183), 78–89.

DeLong, J., and K. Magin (2006): "A Short Note on the Size of the Dot-Com Bubble," *National Bureau of Economic Research*.

Duprey, T., B. Klaus, and T. Peltonen (forthcoming): "Dating Systemic Financial Stress Episodes in the EU Countries," *Journal of Financial Stability*.

Edmans, A., I. Goldstein, and W. Jiang (2012): "The Real Effects of Financial Markets: The Impact of Prices on Takeovers," *Journal of Finance*, 67(3), 933–971.

——— (2015): "Feedback Effects, Asymmetric Trading, and the Limits to Arbitrage," *American Economic Review*, 105(2), 3766–3797.

Engle, R., and C. Granger (1987): "Cointegration and Error-Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276.

Fama, E. (2013): "Prize Lecture: Two Pillars of Asset Pricing," `https://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2013/fama-lecture.pdf`, Accessed: 2016-09-29.

Fama, E., and K. French (1988): "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–25.

FCIC (2011): "The Financial Crisis Inquiry Report," `https://fcic.law.stanford.edu/report`, Accessed: 2017-03-19.

Frankel, J., and G. Saravelos (2012): "Can Leading Indicators Assess Country Vulnerability? Evidence from the 2008–09 Global Financial Crisis," *Journal of International Economics*, 87(2), 216–231.

Friedman, M., and A. J. Schwartz (1963): *A Monetary History of the United States, 1867-1960*. Princeton University Press, Princeton.

Galbraith, J. K. (1979): *The Great Crash, 1929*. Houghton Mifflin Company, Boston.

——— (1994): *A Short History of Financial Euphoria*. Penguin Books, New York, NY.

GFD (2017): "Global Financial Data," `https://www.globalfinancialdata.com`.

Goetzmann, W. (2015): "Bubble Investing: Learning from History," *NBER working paper*, (21693).

Gordon, M. (1962): *The Investment, Financing, and Valuation of the Corporation*. Irwin, Homewood, Ill.

Graham, B., and D. Dodd (2009): *Security Analysis*. McGraw-Hill, New York, NY, 6 edn.

Greenwood, R., A.Schleifer, and Y.You (2016): "Bubbles for Fama," *NBER Working Paper Series*, (23191).

Gromb, D., and D. Vayanos (2010): "Limits of Arbitrage: The State of the Theory," *NBER Working Paper* (15821).

Habib, A., H. Jiang, and M. M. Hasan (2017): "Stock Price Crash Risk: Review of the Empirical Literature," *Accounting and Finance*.

Hamilton, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.

Hennessy, K., D. Holtz-Eakin, and B. Thomas (2011): "Causes of the Financial Crisis," `https://fcic.law.stanford.edu/report`, Dissenting Statement. Accessed: 2017-03-19.

Illing, M., and Y. Liu (2006): "Measuring Financial Stress in a Developed Country: An Application to Canada," *Journal of Financial Stability*, 3(2), 243–265.

International Monetary Fund (2010): "The IMF-FSB Early Warning Exercise," *International Monetary Fund*.

——— (2015): "Global Financial Stability Report," *International Monetary Fund*.

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

Kaminsky, G., and C. Reinhart (1999): "The Twin Causes of Crises: The Causes of Banking and Balance of Payments Problems," *American Economic Review*, 89, 473–500.

Kim, J.-B., Y. Li, and L. Zhang (2011a): "CFOs Versus CEOs: Equity Incentives and Crashes," *Journal of Financial Economics*, 101, 713–730.

——— (2011b): "Corporate Tax Avoidance and Stock Price Crash Risk: Firm-Level Analysis," *Journal of Financial Economics*, 100, 639–662.

Lander, J., A. Orphanides, and M. Douvogiannis (1997): "Earnings Forecasts and the Predictability of Stock Returns: Evidence from Trading the S&P," *Journal of Portfolio Management*, 23, 24–35.

LeRoy, S., and R. Porter (1981): "The Present-Value Relation: Tests Based on Implied Variance Bounds," *Econometrica*, 49, 97–113.

Makinen, G. (2002): "The Economic Effects of 9/11: A Retrospective Assessment," *Congressional Research Service*.

Mishkin, F. S., and E. N. White (2002): "U.S. Stock Market Crashes and Their Aftermath: Implications for Monetary Policy," *NBER Working Paper Series*, (8992).

Misina, M., and G. Tkacz (2009): "Credit, Asset Prices, and Financial Stress," *International Journal of Central Banking*, (5), 95–122.

Ofek, E., and M. Richardson (2003): "DotCom Mania: The Rise and Fall of Internet Stock Prices," *Journal of Finance*, 58(3), 1113–1137.

Pasricha, G., T. Roberts, I. Christensen, and B. Howell (2013): "Assessing Financial System Vulnerabilities: An Early Warning Approach," *Bank of Canada Review* (Autumn): 10–19.

Phillips, P. (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.

Scheinkman, J., and W. Xiong (2003): "Overconfidence and Speculative Bubbles," *Journal of Political Economy*, 111, 1183–1219.

Schularick, M., and A. Taylor (2012): "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008," *American Economic Review*, 102(2), 1029–1061.

SEC (2002): "Acceleration of Periodic Report Filing Dates and Disclosure Concerning Website Access to Reports," `https://www.sec.gov/rules/final/33-8128.htm`, Accessed: 2017-03-19.

Shiller, R. (1981): "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in dividends?," *American Economic Review*, 71(3), 421–436.

Shiller, R. (1989): *Market Volatility*. MIT Press, Cambridge, MA.

——— (2015): *Irrational Exuberance*. Princeton University Press, Princeton, New Jersey, 3 edn.

Shleifer, A., and R. Vishny (2011): "Fire Sales in Finance and Macroeconomics," *Journal of Economic Perspectives*, 25(1), 29–48.

Siegel, J. (2014): *Stocks for the Long Run*. McGraw-Hill, New York, 5 edn.

Stock, J. (1987): "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica*, 55(5), 1035–1056.

Stock, J., and M. Watson (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61(4), 783–820.

Summers, L. (2016): "Heed the Fears of the Financial Markets," Financial Times, January 11.

Appendix A Dynamic Gordon Growth Model and the Cointegration of (d-p) and (e-p)

The dynamic Gordon growth model, following Campbell et al. (1997), starts with defining the log stock return r_{t+1}:

$$\begin{aligned} r_{t+1} &\equiv \log(P_{t+1} + D_{t+1}) - \log(P_t) \\ &= p_{t+1} - p_t + \log(1 + \exp(d_{t+1} - p_{t+1})) \end{aligned}$$

Campbell et al. (1997) then use a first-order Taylor expansion to approximate the right-hand side term:

$$r_{t+1} \approx k + \rho p_{t+1} - p_t + (1-\rho) d_{t+1}$$

where $\rho \equiv 1/(1 + exp(\mu_{d-p}))$, μ_{d-p} is the mean log dividend-price ratio, and $k \equiv \log(\rho) - (1-\rho) log(1/\rho - 1)$. The log-price formula is then

$$p_t = \frac{k}{1-\rho} + E_t \left[\sum_{j=0}^{\infty} \rho^j [(1-\rho) d_{t+1+j} - r_{t+1+j}] \right],$$

subject to the condition that $\lim_{j\to\infty} \rho^j p_{t+j} = 0$, to rule out that the log price could grow forever at a rate faster than $1/\rho$.

Subtracting this equation from d_t, and decomposing future dividends into summations of the changes, gives the log dividend-price ratio in terms of changes in log dividends:

$$\begin{aligned} d_t - p_t &= d_t - \frac{k}{1-\rho} + E_t \left[\sum_{j=0}^{\infty} \rho^j [r_{t+1+j} - (1-\rho)[d_t + \sum_{m=j}^{\infty} \rho^{m-j} \Delta d_{t+1+j}]] \right] \\ &= -\frac{k}{1-\rho} + E_t \left[\sum_{j=0}^{\infty} \rho^j [-\Delta d_{t+1+j} + r_{t+1+j}] \right] \end{aligned}$$

Thus, if changes in log dividends and expected price returns are stationary (as tests confirm), then the log dividend-price ratio is likewise stationary (Campbell et al., 1997). Since $d_t = \psi_t + e_t$ — i.e., dividends equal the payout ratio times earnings, with ψ_t being the log of the payout ratio — it follows that the earnings yield $e_t - p_t$ is also stationary — tests also confirm that changes in the payout ratio are stationary.

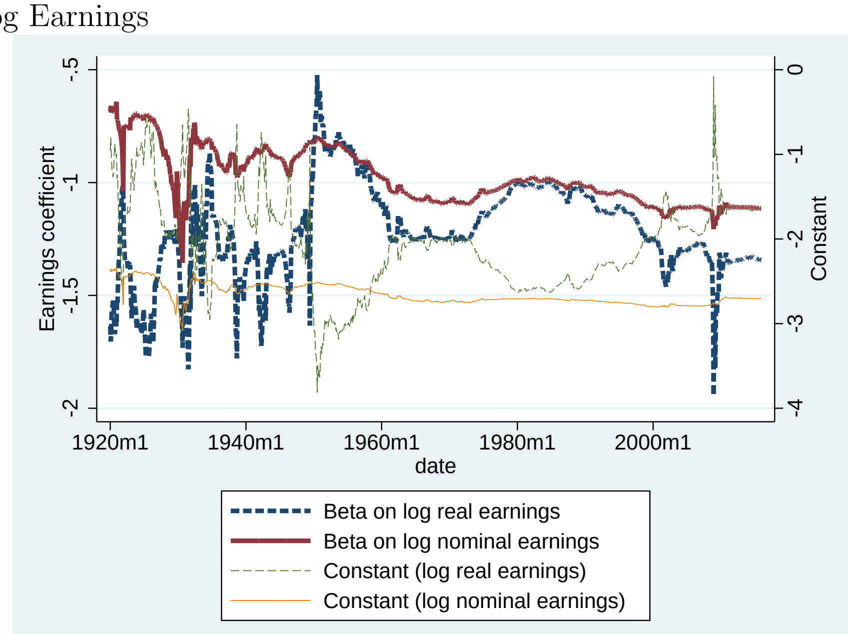
Appendix B Robustness Checks

This appendix presents additional robustness-related results, starting with the cointegration of log prices and log earnings. In addition to varying the sample period, definition of crash, and other regressors, subsequent sub-sections look at other robustness checks: (1) alternative measures of through-the-cycle earnings, (2) different assumptions for the cointegration of log prices and log earnings, and further results for (3) the rolling out-of-sample estimation and (4) a one-year pre-crash horizon. The last sub-section also extends the analysis to S&P 500 Composite sector indices.

For the cointegration of log prices and log earnings, the difference between nominal and real results is further examined in Figures 9 and 10. In Figure 9 the sample period starts in 1871m1 and the end observation is rolled forward, from 1920m1 up to 2016m6; conversely in Figure 10, the last observation is 2016m6 and the sample period start is rolled forward from 1871m1 to 1990m12. The main message from these figures is that the estimates with nominal prices and earnings are much more stable than with real prices and earnings — this gives another reason to opt for the nominal specification.

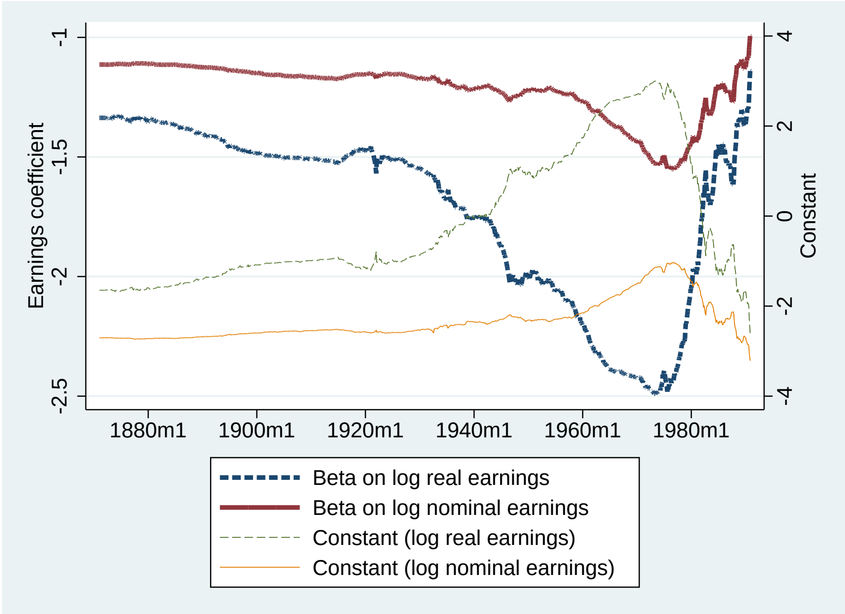
B.1 Additional figures and tables for the S&P Composite

Figure 9: Parameter Stability in Vector Error-Correction Model of Log Prices and Log Earnings



A given observation shows the estimated coefficient for the sample period from 1871m1 up until that date.

Figure 10: Parameter Stability in Vector Error-Correction Model of Log Prices and Log Earnings



A given observation shows the estimated coefficient for the sample period from that date until 2016m6.

Figure 11: Estimated Probability of Crash Using Counterfactual Benchmark Residual

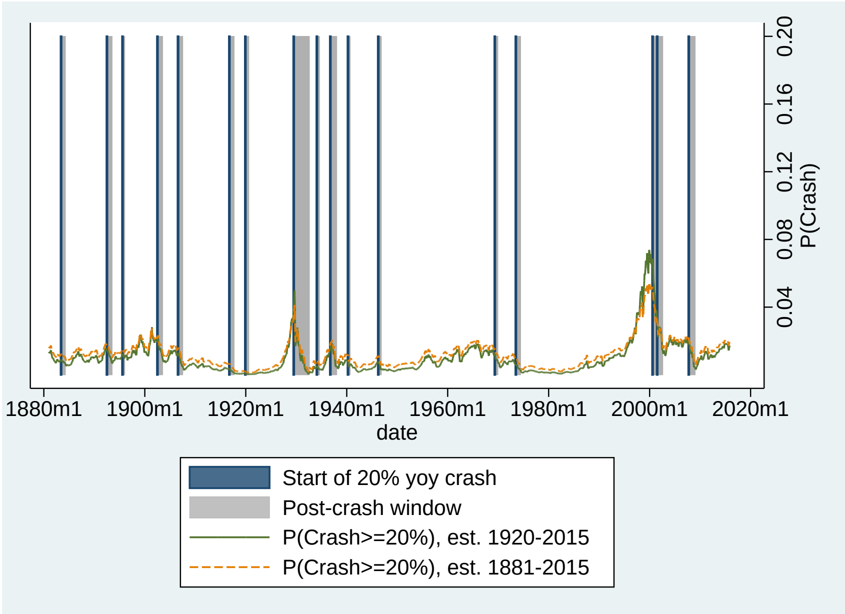


Table 10: Logit Estimation of Likelihood of a Crash (No Added Pre-Crash Window): 1881-2015

	(1) Crash15	(2) Crash20	(3) Crash25	(4) Crash30
Benchmark residual	1.024	1.814*	3.170***	2.438*
	(1.57)	(2.14)	(3.74)	(2.25)
Constant	-4.085***	-4.626***	-5.529***	-5.788***
	(-17.97)	(-14.95)	(-12.65)	(-11.47)
Observations	1297	1428	1506	1545
Pseudo-R2	0.013	0.037	0.098	0.056
Chi-squared	2.468	4.589	13.961	5.074
P-value	0.116	0.032	0.000	0.024
Mfx	0.018	0.021	0.023	0.011
Mfx s.e.	0.012	0.011	0.009	0.006
AUROC	0.598	0.681	0.796	0.730
AUROC s.e.	0.061	0.069	0.071	0.108
Brier ratio	0.997	0.993	0.990	0.996
Brier ratio (crash=1)	0.995	0.988	0.976	0.992

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The crash dummy is defined by a forward-looking year-over-year drop in the nominal price index greater than the threshold, first-differenced to remove autocorrelation. The estimation excludes the five months immediately following a dummy value of 1 in the un-differenced series.

Table 11: One-Year Returns by Estimated Probability of Crash: 1920-2015

Crash prob (%)	mean	p10	p25	p75	p90	N
$(0, 0.5)$	10.3	-14.7	-3.1	23.4	32.7	708
$[0.5, 1.0)$	6.3	-14.5	0.0	17.2	24.0	229
$[1.0, 1.5)$	3.3	-27.5	-5.6	12.3	23.5	128
$[1.5, 2.0)$	0.3	-26.8	-20.6	27.6	35.6	17
$[2.0, 3.5)$	0.1	-25.0	-18.8	24.3	29.2	32
$[3.5, 8.5)$	-1.0	-26.1	-17.8	14.2	20.2	38

B.2 Alternative through-the-cycle measures

The 10-year trailing moving average was a fairly arbitrary choice for the smoothing function of earnings. To evaluate the robustness of the results with respect to the smoothing method, Table 12 estimates the logit model for crash probability using counterfactual residuals constructed with the log of the 7-, 10-, 12-, and 15-year trailing moving averages of earnings, as well as the log of a Hodrick-Prescott (HP) smoothed trend of earnings.[45]

The results in Table 12 show marginally less model fit for the 7- and 15-year moving averages. But they do not change any conclusions.

B.3 Range of cointegration assumptions

As seen in Figures 9 and 10, the sample period can affect the cointegration estimates. This could be a concern with the results of the cointegration estimate being an input into the logit model, compounding model uncertainty. Table 13 examines this issue by constraining the coefficient on log earnings for the counterfactual benchmark, to values of 1.0, 1.05, 1.1, 1.15, and 1.2, and then re-estimating the main logit model with the log of the 10-year moving average of earnings. The results suggest that within a reasonable range, the log earnings coefficient appears not to matter a great deal in terms of the general result that crash probability increases as prices diverge from fundamentals.

However, the log earnings coefficient will still have a bearing on the current estimate of the counterfactual valuation benchmark. That is to say, a higher coefficient implies a higher current counterfactual valuation that would be supported by smoothed earnings, and less crash risk. The goal of this paper is not to provide a current assessment of downside risk, but if this were the objective, then uncertainty about the coefficient would certainly complicate things. Rolling out-of-sample estimates, in the next sub-section, simulate the case of estimating the crash probability in real time, and thus contend with a similar challenge.

[45] For the HP filter, I set lambda to 400000×3^4. This follows Borio and Drehmann (2009) and much of the early-warning literature, which often uses a lambda value of 400000 with quarterly data. I multiply by 3^4 to account for the monthly frequency of my data set.

Table 12: Logit Estimation of Likelihood of a Crash, by Measure of Smoothed Earnings: 1920-2015

	(1) Crash25	(2) Crash25	(3) Crash25	(4) Crash25	(5) Crash25
Residual, 7yma	3.543***				
	(3.73)				
Residual, 10yma		3.495***			
		(3.78)			
Residual, 12yma			3.366***		
			(3.91)		
Residual, 15yma				3.193***	
				(4.03)	
Residual, HP trend					3.762***
					(3.44)
Constant	-5.810***	-5.837***	-5.799***	-5.782***	-5.883***
	(-10.49)	(-10.23)	(-10.63)	(-10.92)	(-9.67)
Observations	1065	1065	1065	1065	1065
Pseudo-R2	0.124	0.136	0.129	0.125	0.132
Chi-squared	13.937	14.284	15.263	16.226	11.851
P-value	0.000	0.000	0.000	0.000	0.001
Mfx	0.023	0.022	0.022	0.020	0.024
Mfx s.e.	0.010	0.010	0.010	0.009	0.011
AUROC	0.840	0.841	0.844	0.841	0.825
AUROC s.e.	0.073	0.069	0.064	0.060	0.079
Brier ratio	0.991	0.987	0.990	0.991	0.987
Brier ratio (crash=1)	0.975	0.968	0.972	0.974	0.969

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The crash dummy is defined by a forward-looking year-over-year drop in the nominal price index of ≥25%, first-differenced to remove autocorrelation. The estimation excludes the five months immediately following a dummy value of 1.

Table 13: Logit Estimation of Likelihood of a Crash, by Earnings Coefficient Assumption: 1920-2015

	(1) Crash25	(2) Crash25	(3) Crash25	(4) Crash25	(5) Crash25
Coeff=1.00	3.120***				
	(3.69)				
Coeff=1.05		3.288***			
		(3.75)			
Coeff=1.10			3.452***		
			(3.78)		
Coeff=1.15				3.607***	
				(3.76)	
Coeff=1.20					3.747***
					(3.70)
Constant	-5.883***	-5.867***	-5.844***	-5.814***	-5.774***
	(-10.23)	(-10.25)	(-10.24)	(-10.17)	(-10.07)
Observations	1065	1065	1065	1065	1065
Pseudo-R2	0.125	0.130	0.135	0.139	0.143
Chi-squared	13.625	14.092	14.285	14.161	13.727
P-value	0.000	0.000	0.000	0.000	0.000
Mfx	0.020	0.021	0.022	0.023	0.024
Mfx s.e.	0.009	0.009	0.010	0.010	0.011
AUROC	0.830	0.836	0.841	0.845	0.848
AUROC s.e.	0.068	0.068	0.069	0.070	0.072
Brier ratio	0.989	0.988	0.987	0.986	0.984
Brier ratio (crash=1)	0.972	0.970	0.968	0.966	0.963

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

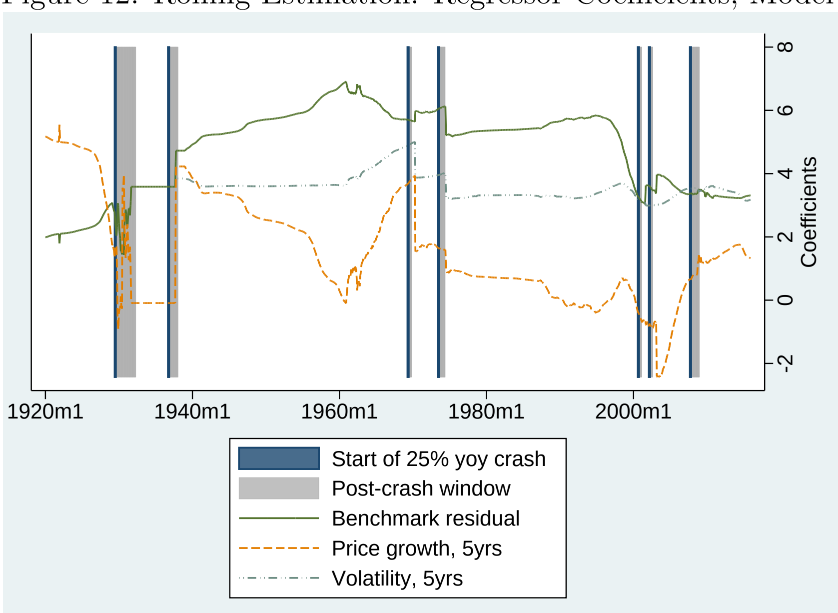
The crash dummy is defined by a forward-looking year-over-year drop in the nominal price index of ≥25%, first-differenced to remove autocorrelation. The estimation excludes the five months immediately following a dummy value of 1.

B.4 Rolling out-of-sample evaluation

Figures 12 and 13 describe the rolling in-sample results in more detail. Figure 12 shows coefficient estimates for Model 2. The coefficient on price growth, in particular, ranges from negative to positive depending on the sample period. This suggests that momentum over some given horizon is not a reliable indicator of crash risk, and corroborates the view in Greenwood et al. (2016) and Goetzmann (2015) that not all price run-ups are bad — this could be a more idiosyncratic feature that varies by crash episode. The coefficient on the benchmark residual has come down since the 2000 Internet bubble, but would appear to be telling a consistent story that a high valuation relative to smoothed fundamentals portends greater downside risk.

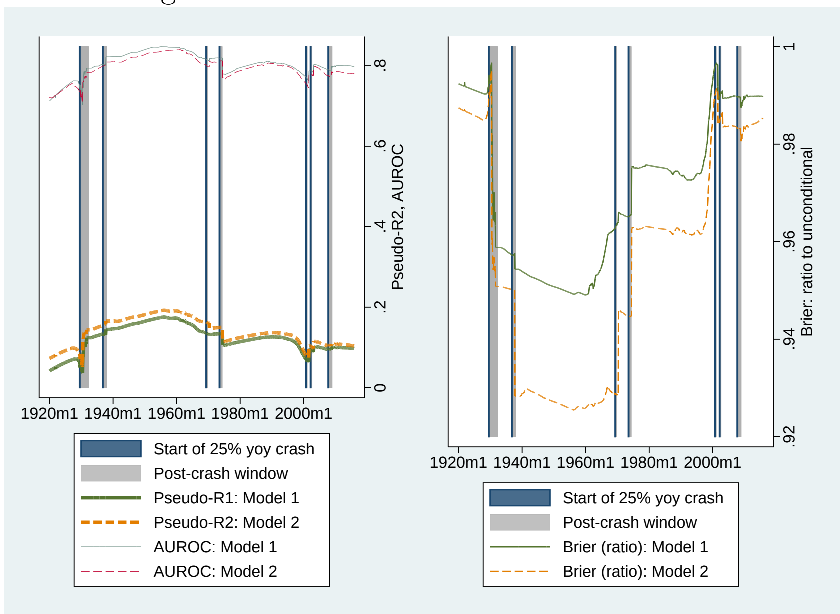
Figure 13 shows the in-sample pseudo-R^2's, AUROC, and Brier scores as a ratio to the unconditional Brier score. The first two measures consistently corroborate the full-sample finding that the model predictions can help explain crash probabilities. The Brier score shows only a modest improvement relative to the unconditional; however, this is a stringent measure since there is far more weight on the much more numerous non-crash observations.

Figure 12: Rolling Estimation: Regressor Coefficients, Model 2



The estimated coefficients are for Model 2, using data up until that point for both the VECM estimation of price and earnings cointegration, and the logit model.

Figure 13: Rolling Estimation: Evaluation with AUROC and Brier Scores



Each observation describes the in-sample measure up until that date.

Table 14: Out-of-Sample, Rolling Estimation: 1920-1949

	Brier score	AUROC	AUROC p-val	Pseudo-R^2	QPS10
Null	0.0064	0.14	0.96	0.000	0.060
Div. yield	0.0062	0.95	0.01	0.250	0.058
HP gap	0.0062	0.89	0.03	0.163	0.056
CAPE10	0.0056	0.98	0.01	0.400	0.050
Volatility, 5yr	0.0063	0.51	0.49	-1.005	0.059
Price growth, 5yr	0.0063	0.89	0.03	0.152	0.058
CB, model 1	0.0060	0.97	0.01	0.356	0.056
CB, model 2	0.0061	0.59	0.34	-0.302	0.054

Table 15: Out-of-Sample, Rolling Estimation: 1950-1989

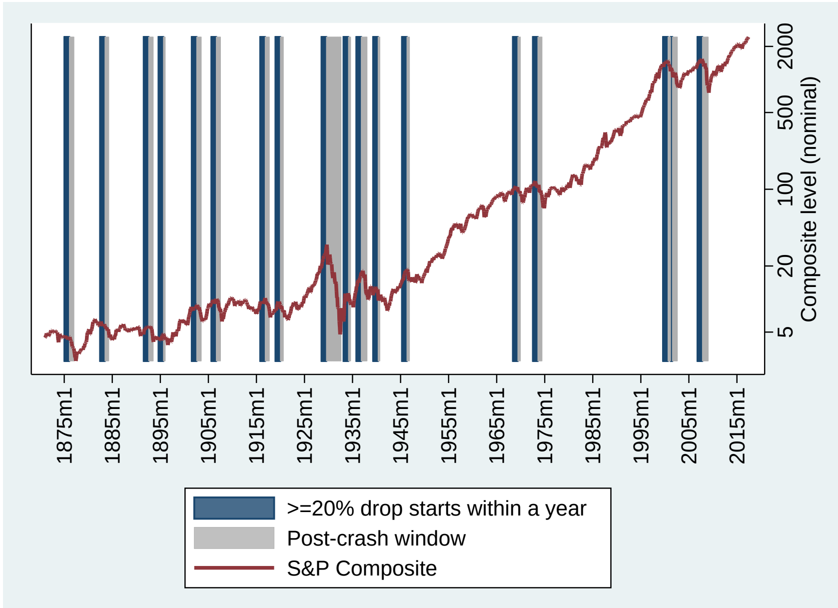
	Brier score	AUROC	AUROC p-val	Pseudo-R^2	QPS10
Null	0.0043	0.17	0.95	0.000	0.041
Div. yield	0.0045	0.65	0.24	-0.125	0.040
HP gap	0.0044	0.11	0.97	-0.227	0.041
CAPE10	0.0044	0.68	0.18	0.006	0.041
Volatility, 5yr	0.0043	0.17	0.94	-0.004	0.041
Price growth, 5yr	0.0044	0.15	0.96	-0.149	0.041
CB, model 1	0.0043	0.63	0.26	-0.003	0.041
CB, model 2	0.0043	0.60	0.32	-0.024	0.041

Table 16: Out-of-Sample, Rolling Estimation: 1990-2015

	Brier score	AUROC	AUROC p-val	Pseudo-R^2	QPS10
Null	0.0103	0.34	0.83	0.000	0.094
Div. yield	0.0104	0.67	0.16	-0.030	0.091
HP gap	0.0104	0.38	0.76	-0.044	0.094
CAPE10	0.0166	0.75	0.07	-0.422	0.096
Volatility, 5yr	0.0103	0.37	0.78	-0.002	0.094
Price growth, 5yr	0.0103	0.68	0.15	0.048	0.093
CB, model 1	0.0114	0.75	0.07	-0.088	0.091
CB, model 2	0.0116	0.72	0.09	-0.112	0.091

B.5 Pre-crash horizon of one year

Figure 14: Identifying Major Equity Corrections (2)



B.6 Sector indices

A similar methodology can be extended to sector indices, rather than just the aggregate market as a whole. Toward this end, I use the 10 economic sectors from S&P's and Morgan Stanley Capital International's Global Industry Classification System (MSCI GICS), available through Global Financial Data. These sectors consist of consumer durables, consumer staples, energy, financials, health care, industrials, information technology, materials, telecoms, and utilities. Table 19 summarizes the data availability for this sector breakdown. Price history is more limited for several of the sectors, in particular financials, health care, information technology, and materials, and even more so for the price-earnings ratio.

Similar to the aggregate market models, I use a dummy indicator of year-over-year crashes exceeding a given threshold, differenced to remove serial correlation, and excluding the five-month post-crash window. Tables 20 and 21 show the results without and with sector fixed effects, for year-over-year crash magnitudes of 30%, 35%, and 40%, with robust standard errors clustered by sector. In the sector case, the more severe crash definitions of 35% and 40% give a better model fit than thresholds of 30% and below.

The sector-specific benchmark residuals are statistically significant and of a similar magnitude as for the aggregate models.[46] The sectoral models also appear to offer some vindication for price growth and volatility measures as

[46] I use a seven-year moving average owing to the shorter sample size for some of the sector indices.

Table 17: Logit Estimation of Likelihood of a Crash, One-Year Pre-Crash Horizon: Model 1, 1920-2015

	(1) PreCr	(2) PreCr	(3) PreCr	(4) PreCr	(5) PreCr	(6) PreCr	(7) PreCr	(8) PreCr	(9) PreCr	(10) PreCr	(11) PreCr	(12) PreCr
Residual	4.50**	4.35**	3.95**	4.12**	4.06**	4.16**	4.07**	4.09**	4.21**	5.64**	4.63**	4.50**
	(1.58)	(1.53)	(1.42)	(1.45)	(1.41)	(1.42)	(1.38)	(1.44)	(1.51)	(2.00)	(1.65)	(1.61)
Constant	-3.42***	-3.38***	-3.32***	-3.36***	-3.34***	-3.36***	-3.37***	-3.26***	-3.31***	-3.70***	-3.39***	-3.40***
	(0.73)	(0.71)	(0.68)	(0.70)	(0.69)	(0.69)	(0.69)	(0.67)	(0.68)	(0.85)	(0.74)	(0.73)
Obs	88	89	90	89	90	90	90	88	89	88	87	87
Ps-R2	0.264	0.251	0.230	0.246	0.244	0.257	0.254	0.238	0.235	0.302	0.262	0.258
Chi2	12.91	12.33	11.30	12.06	11.99	12.62	12.50	11.64	11.53	14.77	12.76	12.58
P-val	0.000	0.000	0.001	0.001	0.001	0.000	0.000	0.001	0.001	0.000	0.000	0.000
Mfx	0.260	0.253	0.234	0.241	0.236	0.238	0.232	0.243	0.249	0.315	0.272	0.266
Mfx s.e.	0.087	0.086	0.082	0.082	0.080	0.077	0.075	0.083	0.088	0.103	0.093	0.091
AUROC	0.855	0.861	0.835	0.841	0.842	0.859	0.861	0.868	0.878	0.889	0.857	0.854
AUR s.e.	0.065	0.061	0.067	0.071	0.072	0.070	0.068	0.057	0.056	0.057	0.067	0.068
Brier	0.060	0.061	0.061	0.060	0.059	0.059	0.059	0.063	0.063	0.058	0.060	0.061
Brier_u	0.073	0.072	0.072	0.072	0.072	0.072	0.072	0.073	0.072	0.073	0.074	0.074
Month	1	2	3	4	5	6	7	8	9	10	11	12

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The crash dummy ("PreCr") is defined by a forward-looking year-over-year drop in the nominal price index of ≥25%, starting within one year. The estimation excludes the five months immediately following the start of the year-over-year correction. Regressions are run separately for each calendar month, thus collapsing the data to an annual frequency.

Table 18: Logit Estimation of Likelihood of a Crash, One-Year Pre-Crash Horizon: Model 2, 1920-2015

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	PreCr	PreCr	PreCr	PreCr	PreCr	PreCr	PreCr	PreCr	PreCr	PreCr	PreCr	PreCr
Residual	5.60*	5.36*	4.98*	4.94*	4.89*	4.85*	4.59*	4.37*	4.33*	5.33*	6.02*	5.76*
	(2.23)	(2.09)	(1.99)	(2.03)	(2.04)	(2.01)	(1.97)	(1.88)	(1.95)	(2.39)	(2.41)	(2.30)
Growth, 5yrs	1.77	0.73	-0.28	0.24	1.45	2.40	2.34	4.29	5.37	7.53	2.09	2.61
	(7.71)	(7.73)	(7.64)	(7.61)	(7.74)	(7.74)	(7.62)	(7.95)	(7.90)	(8.36)	(7.74)	(7.66)
Volatility, 5yrs	7.56	6.37	5.83	5.24	6.21	6.34	5.81	6.13	6.17	5.98	8.58	8.66
	(4.86)	(4.78)	(4.68)	(4.54)	(4.61)	(4.47)	(4.30)	(4.30)	(4.15)	(4.16)	(5.06)	(4.98)
Constant	-5.23***	-4.87**	-4.62**	-4.51**	-4.76***	-4.85***	-4.74***	-4.76***	-4.86***	-5.30***	-5.47**	-5.50***
	(1.58)	(1.53)	(1.45)	(1.39)	(1.44)	(1.44)	(1.39)	(1.39)	(1.37)	(1.44)	(1.68)	(1.65)
Obs	88	89	90	89	90	90	90	88	89	88	87	87
Ps-R2	0.320	0.299	0.276	0.281	0.285	0.300	0.293	0.280	0.279	0.352	0.328	0.325
Chi2	15.65	14.66	13.60	13.77	14.02	14.75	14.42	13.69	13.70	17.21	15.96	15.81
P-val	0.001	0.002	0.004	0.003	0.003	0.002	0.002	0.003	0.003	0.001	0.001	0.001
Mfx	0.305	0.299	0.283	0.279	0.272	0.263	0.250	0.244	0.240	0.271	0.329	0.314
Mfx s.e.	0.117	0.113	0.112	0.114	0.113	0.107	0.106	0.104	0.108	0.119	0.125	0.120
AUROC	0.877	0.878	0.857	0.866	0.857	0.862	0.861	0.855	0.868	0.884	0.882	0.880
AUR s.e.	0.068	0.067	0.075	0.075	0.083	0.083	0.083	0.077	0.074	0.072	0.066	0.066
Brier	0.054	0.055	0.056	0.056	0.054	0.053	0.054	0.057	0.057	0.050	0.054	0.054
Brier_u	0.073	0.072	0.072	0.072	0.072	0.072	0.072	0.073	0.072	0.073	0.074	0.074
Month	1	2	3	4	5	6	7	8	9	10	11	12

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The crash dummy ("PreCr") is defined by a forward-looking year-over-year drop in the nominal price index of $\geq$25%, starting within one year. The estimation excludes the five months immediately following the start of the year-over-year correction. Regressions are run separately for each calendar month, thus collapsing the data to an annual frequency.

Table 19: Data Availability for Sector Indices

Sector	Price index	Price-earnings
Consumer durables	1925m12	1991m3
Consumer staples	1925m12	1991m3
Energy	1915m1	1946m12
Financials	1970m1	1976m7
Health care	1987m1	1987m12
Industrials	1925m12	1915m1*
Information technology	1986m1	1994m3
Materials	1989m9	1992m3
Telecoms	1915m1	1957m3
Utilities	1915m1	1915m1*

*For industrials and utilities, I extend the price-earnings ratio backwards by using data from industrials and utilities composites from prior to the S&P and MSCI GICS classification.

indicators of risk — in particular, volatility over the past five years consistently appears as a statistically significant indicator of heightened crash risk. This is in contrast to what was found for the aggregate market. It could reflect that with higher volatility in the more detailed breakdown by sector, prices could be more subject to reversals in sentiment even when not conditioned on a divergence from fundamentals. Price growth is also statistically significant and with a positive coefficient, with sector fixed effects.

Since the sector price-earnings series are not available for the full sample period, I also estimate models where I include both the aggregate S&P Composite benchmark residual, as well as a sector-specific measure. For the sector-specific measure, I first construct the sector benchmark residual in the same way, and then regress this on the aggregate benchmark residual, such that the residual from this last equation gives the final sector-specific measure. This is to isolate any independent contribution of sector log price divergence from log smoothed earnings, above and beyond what is happening in the aggregate market.

Table 22 shows the results with a 40% crash definition for the six sectors with price data going back to roughly 1920 (columns 1 and 2), all 10 sectors (columns 3 and 4), as well as with a two-standard-deviation definition of correction, based on annual sector returns (columns 5 and 6), for all 10 sectors. Overall, Table 22 corroborates findings for the aggregate market. The aggregate residual is statistically significant, and of a similar magnitude as that seen for the aggregate-only model: downside risk increases as prices diverge above the through-the-cycle fundamentals.

The estimated crash probability is shown in Figure 15. Telecoms, as well as information technology (not shown), stand out in the 2000 Internet bubble as being particularly at risk.

One can debate the statistical inference for the panel of sector indexes — given that the sector indices are highly correlated with the aggregate market (0.59 for monthly log differences), there are fewer independent observations than there would appear to be, from the total count. This would require more

Table 20: Panel Logit Estimation of Likelihood of -40% Crash, without Sector Fixed Effects: 1920-2015

	(1) SecCrash30	(2) SecCrash35	(3) SecCrash40
Sector residual	1.703*	2.025***	1.750***
	(2.54)	(4.19)	(4.43)
Price growth, 5yrs	4.861*	3.957	4.575*
	(2.51)	(1.56)	(2.03)
Volatility, 5yrs	5.960**	6.306***	6.470***
	(2.83)	(3.84)	(6.18)
Constant	-7.110***	-7.418***	-7.604***
	(-18.38)	(-17.41)	(-17.60)
Observations	5205	5319	5398
Pseudo-R2	0.104	0.114	0.107
Chi-squared	75.145	142.690	119.243
P-value	0.000	0.000	0.000

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

work to better ascertain. In addition, the challenges of out-of-sample estimation and identification of vulnerabilities no doubt extend to the sector case.

Table 21: Panel Logit Estimation of Likelihood of -40% Crash, with Sector Fixed Effects: 1920-2015

	(1) SecCrash30	(2) SecCrash35	(3) SecCrash40
Sector residual	1.597*	1.824**	1.652***
	(2.31)	(3.00)	(3.86)
Price growth, 5yrs	6.455***	5.714*	5.550***
	(3.95)	(2.26)	(4.15)
Volatility, 5yrs	7.447*	8.112*	8.060***
	(2.14)	(2.55)	(4.88)
sector==cd	-1.221	-0.941*	-0.663***
	.	(-2.49)	(-3.32)
sector==cs	.	.	.
	.	.	.
sector==en	-0.799	-1.513***	-1.186***
	.	(-9.75)	(-13.38)
sector==fn	-1.529***	-1.340***	-0.986***
	(-12.87)	(-13.79)	(-10.06)
sector==hc	-1.685***	.	.
	(-8.51)	.	.
sector==in	-1.702**	-1.989*	-2.084***
	(-2.72)	(-2.53)	(-4.24)
sector==it	-1.595	-1.369**	-0.690
	.	(-2.67)	.
sector==mt	-0.519	-0.247	0.011
	.	.	(0.05)
sector==tc	-1.238***	-0.667**	-0.799
	(-5.94)	(-3.16)	.
sector==ut	.	.	.
	.	.	.
Constant	-6.571***	-6.905***	-7.143***
	(-8.67)	(-10.60)	(-22.70)
Observations	4907	4733	4812
Pseudo-R2	0.131	0.147	0.136
Chi-squared	-23.534	426.197	703.898
P-value	1.000	0.000	0.000

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Sectors are as follows: consumer discretionary (cd), consumer staples (cs), energy (en), financials (fn), health care (hc), industrials (in), information technology (it), materials (mt), telecommunications (tc), and utilities (ut).

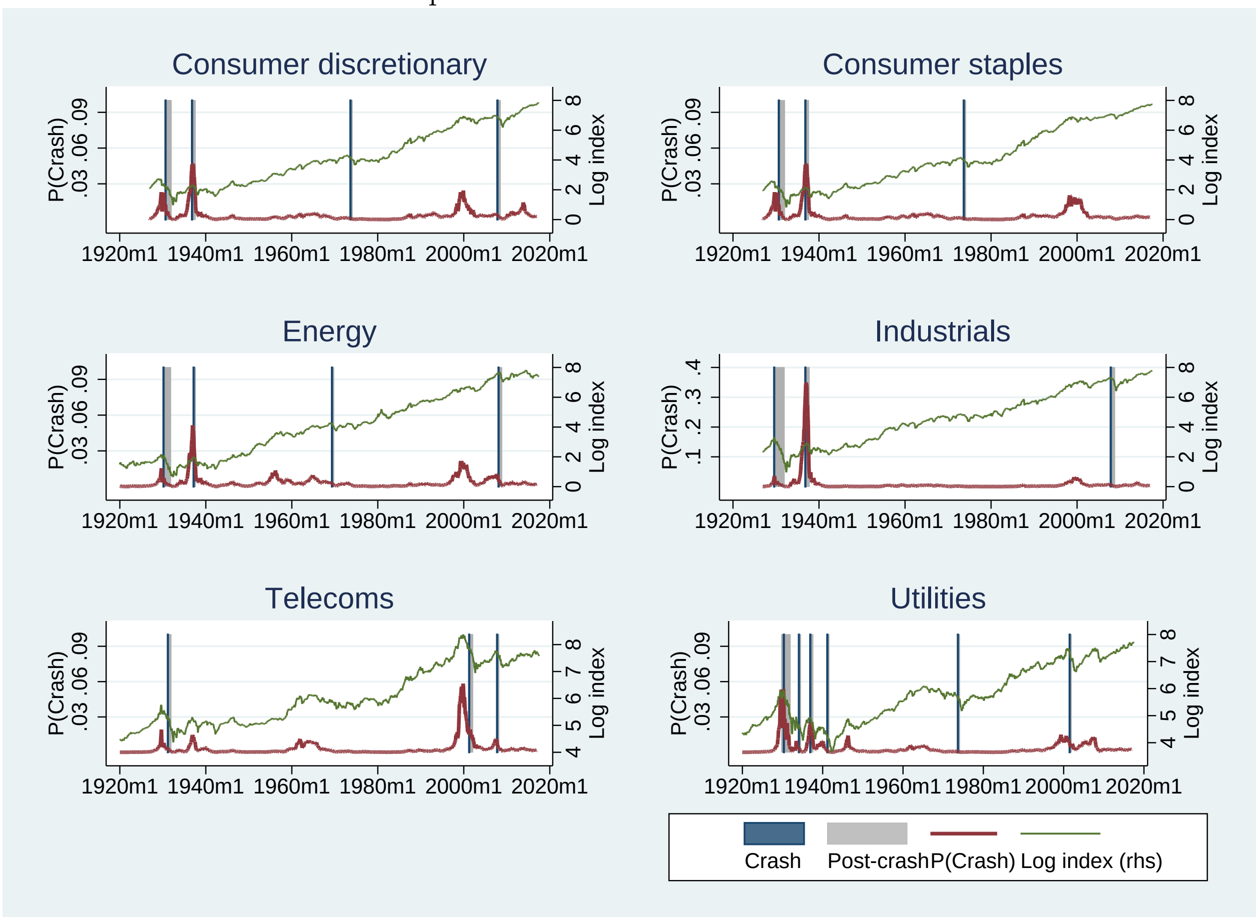
Table 22: Panel Logit Estimation of Likelihood of -40% Crash, with Aggregate Residual: 1920-2015

	(1) SecCrash40	(2) SecCrash40	(3) SecCrash40	(4) SecCrash40	(5) SecCrash2sd	(6) SecCrash2sd
Residual, 10yma	2.823***	2.851***	2.585***	2.559***	2.679***	2.747***
	(5.43)	(5.88)	(4.96)	(5.37)	(6.44)	(5.96)
Sector res., ex-agg	0.869*	0.847*	1.552**	1.045*	1.785**	1.411*
	(2.12)	(2.37)	(3.10)	(2.11)	(2.98)	(2.26)
Sec. pr.grwth, 5yrs	3.922	4.791*	3.521	5.055**	3.642	6.059*
	(1.19)	(1.97)	(1.71)	(2.96)	(1.65)	(2.23)
Sec. vol., 5yrs	8.056***	9.886***	5.894***	7.575***	4.841**	7.198**
	(4.39)	(3.95)	(4.70)	(3.70)	(3.18)	(3.06)
sector==cd		-0.655		-0.594		-0.309**
		.		.		(-2.97)
sector==cs		-0.857		-0.816		-0.425
		.		.		(-1.95)
sector==en		-0.740***		-0.730***		-1.032***
		(-6.38)		(-5.60)		(-7.11)
sector==fn		.		-1.481***		-1.789***
		.		(-6.58)		(-5.48)
sector==hc		.		.		-2.755***
		.		.		(-3.86)
sector==in		-1.955***		-1.540***		-1.834***
		(-4.45)		(-6.51)		(-6.12)
sector==it		.		-1.728***		-2.678***
		.		(-6.34)		(-3.41)
sector==mt		.		-2.102***		-2.261***
		.		(-3.45)		(-5.03)
sector==tc		-0.871		-0.856***		-0.463***
		.		(-6.71)		(-10.65)
sector==ut		.		.		.
		.		.		.
Constant	-8.171***	-7.900***	-7.956***	-7.369***	-7.535***	-7.221***
	(-11.66)	(-10.08)	(-14.06)	(-11.37)	(-17.42)	(-11.11)
Observations	6679	6679	11219	10067	11064	11064
Sectors	6	6	10	10	10	10
Pseudo-R2	0.146	0.167	0.128	0.154	0.135	0.184
Chi-squared	51.483	9.170	39.099	-183.965	71.083	107.810
P-value	0.000	0.027	0.000	1.000	0.000	0.000

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Sectors are as follows: consumer discretionary (cd), consumer staples (cs), energy (en), financials (fn), health care (hc), industrials (in), information technology (it), materials (mt), telecommunications (tc), and utilities (ut).

Figure 15: Estimated Probability of Correction for Sector Indices: with Residual Term from Main S&P Composite



Appendix C Additional Background Tables

Table 23: Detailed Timing of Crash Events

Start		Peak		Trough		Returns From Start (%)					Peak to
Date	Index	Date	Index	Date	Index	1m	3m	6m	1y	3y	Trough (%)
1883m6	5.82	1881m6	6.58	1885m1	4.24	-1.5	-5.0	-8.2	-23.4	-3.4	-35.6
1892m8	5.62	1887m5	5.90	1893m8	4.08	-2.5	-0.9	-2.0	-27.4	-5.2	-30.8
1895m8	4.79	1895m9	4.82	1896m8	3.81	.6	-4.2	-7.1	-20.5	3.2	-21.0
1902m9	8.85	1902m9	8.85	1903m10	6.26	-3.2	-9.0	-8.7	-26.9	1.4	-29.3
1906m9	10.03	1906m9	10.03	1907m11	6.25	-3.0	-1.9	-16.7	-25.7	0.5	-37.7
1916m11	10.21	1916m11	10.21	1917m12	6.80	-4.0	-11.6	-13.2	-31.0	-3.4	-33.4
1919m12	8.92	1919m7	9.51	1921m8	6.45	-1.0	-2.8	-11.2	-23.7	-0.5	-32.2
1929m7	28.48	1929m9	31.3	1932m6	4.77	5.7	-1.7	-23.8	-26.1	-44.0	-84.8
1934m2	11.32	1934m2	11.32	1935m3	8.41	-5.1	-13.3	-19.6	-20.7	17.0	-25.7
1936m10	16.89	1937m2	18.11	1938m4	9.89	2.8	4.1	0.7	-27.3	-8.6	-45.4
1940m4	12.27	1938m11	13.07	1942m4	7.84	-13.8	-18.6	-12.6	-21.4	-2.3	-40.0
1946m4	18.66	1946m5	18.7	1948m2	14.1	0.2	-3.3	-21.0	-21.8	-7.2	-24.6
1969m5	104.6	1968m12	106.5	1970m6	75.6	-5.2	-10.0	-8.0	-27.3	1.0	-29.0
1973m7	105.8	1973m1	118.4	1974m12	67.1	-1.9	3.8	9.2	-25.0	-0.5	-43.4
1987m8	329.4	1987m8	329.4	1987m12	241.0	-3.2	-25.6	-21.6	19.9	0.1	-26.8
2000m9	1468.1	2000m8	1485.5	2003m2	837.0	-5.3	-9.3	-19.2	-28.8	-11.4	-43.7
2007m10	1539.7	2007m10	1539.7	2009m3	757.1	-5.0	-10.5	-11.0	-37.1	-8.7	-50.8

Index values and corresponding timing are based on the average closing value in the month.

Table 24: Historical Context of Crashes (Source: Mishkin and White (2002), except where otherwise indicated.)

Year	**Historical Context**	**Triggers/Coinciding Events**	**Outcomes**
1902-1903	Financial system was sound (Mishkin and White, 2002). Stock market boom as United States overtook Britain as the world's leading industrial power. There was a sense of a "new era" of American greatness (Chancellor, 1999).	Coincided with a relatively mild economic contraction, though relative timing is hard to pin down. Banks called in loans made by syndicates, leading to the syndicates liquidating securities.	Mild effect on spreads. Secretary of the Treasury bought bonds and increased government deposits at banks.
1906-1907	Business cycle peak in May 1907; crash preceded severe recession, with fall in real GDP of 6.5 over second half of 2017. Banks vulnerable due to risky ventures.	Failed manipulation of the copper market led to failure of a trust, and subsequent banking panic.	Crash may have exacerbated widening of credit spreads. Late and limited response by the Treasury. The crisis led to creation of the Federal Reserve.
1916-1917	Massive volatility caused by World War I. Economy was booming and peaked in August 1918.	Partly generally rising interest rates and controls on new capital issues. Diversion of financial resources to government bonds to finance the war.	Rising spreads, and large market decline, though crash has been "virtually ignored by historians."
1929	Production, employment, and corporate earnings were high and rising in the 1920s (Galbraith, 1979), with stock prices rising for most of 1924-1929. Relatively weak banking system in 1929. Galbraith (1994) highlights the role of leverage and self-fulfilling expectations in the run-up, and endorsements of the boom from academic and financial circles. In the lead-up, there was a weakening in industrial production and some other economic indices.	The market was already performing badly in the week of October 21, with margin calls and brief attempts to restore confidence (Galbraith, 1994). Tuesday, October 29, was then the worst day in the exchange's history. Galbraith (1994) puts forth that a specific trigger is unimportant and that the "supply of intellectually vulnerable buyers" simply could have been exhausted.	After initial October 1929 crash, lender-of-last resort actions by New York Fed helped prevent panic and freeze-up in broker loans. Spreads only soared from mid-1930s onwards. Fed maintained tight monetary policy, contributing to the severe depression. The Glass-Steagall Act followed in 1933 and the Securities Exchange Commission was created in 1934. The role of the crash in the ensuing depression has been vigorously debated.

Year	Historical Context	Triggers/Coinciding Events	Outcomes
1937-1938	By early 1937, economy had expanded steadily and stock prices had doubled for past two years. Following banking crises of 1930-1933, banks had vastly increased their liquidity.	To tighten monetary policy, Fed doubled reserve requirements and increased margin requirements for stock purchases, leading banks to cut lending.	Decline in asset values may have contributed to widening spreads, and thus severity of recession. In the recession, real GDP fell 10% from May 1937 to June 1938.
1969-1970	The market peaked in advance of a mild recession. Galbraith (1994) writes that the 1960s were marked by "speculative surges and ensuing breaks," and economic growth, low unemployment, and low inflation, with accompanying investment optimism.	Abrupt decline in May 1970 coincided with Penn Central Railroad being on verge of bankruptcy.	Fall in firm valuations may have induced some adverse selection in credit markets and widening of spreads, though not as severely as in other crises.
1973-1974	The early 1970s saw a rapid growth in output (Davis, 2003). Banking system was vulnerable, with troubled loans and market rates rising above regulatory ceilings for deposit rates. Davis (2003) makes note of the exchange rate volatility, inflation, and industrial unrest of the period.	Move into recession at the same time as the market crash, with rising inflation and oil price shock. Alignment of market troughs across G-7 countries; write-down of capital that was poorly suited for higher oil prices (Davis, 2003).	Large decline in asset values may have aggravated access to credit of low-quality borrowers. But other factors were instrumental in the financial instability. With high inflation, fall in real prices was even more severe and protracted.
1987	Monetary aggregates had been increasing, but slowed in first half of 1987. Fed was preoccupied with inflation and there were market worries about a weak trade balance and dollar, and rising interest rates. Largest one-day stock market decline in U.S. history. Galbraith (1994) points out the re-emergence of leverage, and financial "innovation" in the form of junk bonds.	The Brady Commission (Brady, 1988) reported that an initial decline led to "mechanical, price-insensitive selling" of portfolio insurance strategies. Some market participants then sold in anticipation of further market declines, prompting more portfolio insurance selling.	Brokers needed to "extend huge amounts of credit" to customers facing margin calls. Fear about collapse of securities firms and breakdown of clearing and settlement systems led to Fed supplying liquidity of more than 25% of bank reserves. The intervention prevented financial stability repercussions.

Year	Historical Context	Triggers/Coinciding Events	Outcomes
2000-2003	Twin deficits in current account and private sector (Davis, 2003). The economy was slowing and moved into recession in March 2001. Banking system was strong. Makinen (2002) posits that timely action contained the short-run adverse economic effects of the September 11 terrorist attacks.*	The crash was most concentrated in high-tech stocks, which had reached steep valuations. Ofek and Richardson (2003) highlight the expiration of lock-up periods as a possible catalyst.	Sharp increase in spreads in December 2001 reflected Enron scandal, and the revelation of greater-than-expected information asymmetries, rather than the crash (Mishkin and White, 2002). Investment in telecoms may have been excessive at the time, leading to a write-down of capital (Davis, 2003).
2007-2009	A relaxation of credit standards accompanied the expansion of credit in the 2000s that contributed to higher asset valuations, while exposure to opaque and illiquid securitized products helped bring about the fragility and failure of a number of financial institutions (FCIC, 2011).	The failure of Lehman Brothers, placement of Fannie Mae and Freddie Mac into conservatorship, impending collapse of AIG, and opaqueness and interconnectedness of financial institutions led to a freeze-up of credit markets and a plummeting stock market (FCIC, 2011).	With declining asset prices and unemployment soaring above 10%, households cut back on consumption. Business financing and investment suffered, and over 2009 and 2010, 297 mostly small and medium-sized U.S. banks failed. $17 trillion of household net worth was lost between 2007 and 2009q1, compared with GDP of $14.4 trillion in 2008 (FCIC, 2011).

This summary relies heavily on Mishkin and White (2002) for the 1900 to 2002 period, with additional sources as indicated. Mishkin and White (2002) in turn draw on Friedman and Schwartz (1963) and others.

* Bilmes and Stiglitz (2008) emphasize, though, that the military responses and their ensuing fiscal implications would be costly in the long run.