A Non-Parametric Test For Representative Agent
Equilibrium Models

Luca Pezzo, Washington University in St. Louis *

September 15, 2017

Abstract

Popular consumption-based representative agent equilibrium models are shown to be re-
jected conditionally on highly uncertain and illiquid subperiods. These subperiods, selected
based on rules predicting low returns in a training sample, contain all the major financial
crises and recessions and are characterized by low consumption and GDP growth. Financial
uncertainty and illiquidity are found to be the most important rejection drivers. Interestingly,
the alternative intermediary-based pricing seems a more robust setup. These conclusions are
reached via a non-parametric conditional asset pricing test which exploits a new inequality
found in Martin [2017].

*Olin Business School, Washington University in St. Louis, Email: lpezzo@wustl.edu. I am thankful to my
advisor P.H. Dybvig, T.A. Maurer, F. Corielli, O. Kadan, M. Ringgenberg, G. Zhou, A. Manela and my fellow PhD
students for helpful comments and discussions. All remaining errors are my own.
1 Introduction

What is wrong with Consumption-based Representative Agent Equilibrium (CRAE) models? Since the seminal work of Lucas [1978], these models\(^1\) have represented an important and active area of research which tries to explain the behavior of aggregate stock market prices accounting for several empirical regularities. These include the puzzles of excess volatility (LeRoy and Porter [1981], and Shiller [1981]) and equity premium (Mehra and Prescott [1985]), as well as the predictability of stock market returns using the aggregate divided-price ratio (Campbell and Shiller [1988], Fama and French [1988]). Nonetheless, recent empirical work has questioned the validity of key implications of these models. CRAE models cannot disentangle the behavior of risk premia in financial crises and recessions (Muir [2017]), nor can they replicate the dynamics of aggregate uncertainty (Martin [2017]). A large fraction of investors do not form rational expectations (Greenwood and Shleifer [2014]) and profitable trading strategies exist that take less risk during bad times (Moreira and Muir [2017]). These issues call for a unified framework to analyze the CRAE pricing.

Exploiting a new inequality found in Martin [2017], I derive a non-parametric conditional asset pricing test for the CRAE models. I document how these models are rejected conditionally on highly uncertain and illiquid periods, which contain all the major financial crises and recessions, where consumption and GDP growths are low. In these periods the minimum risk premium implied by the models is at least 1.3\% monthly above the actual one, or 0.22 in risk-adjusted term.\(^2\) The inflated model-based predictions are consistent with higher disaster probabilities but not, in general, with higher risk aversion. I also find financial uncertainty and illiquidity to be the most important rejection drivers, contributing on average to 75\% of the magnitude of the rejection statistic: viewing financial uncertainty as a proxy for asymmetric information,\(^3\) this result is con-

\(^{1}\)Including Campbell and Cochrane [1999], Campbell and Viceria [1999], Latteu and Ludvigson [2001], Basal and Yaron [2004], Barro [2006], Bollerslev et. al [2009], Drechsler and Yaron [2011], Wachter [2013], Bansal et al. [2014], Campbell et al. [2017]

\(^{2}\)i.e. the average Sharpe ratio implied by the models is at least 0.22 higher than the actual one conditional on these periods.

\(^{3}\)Financial uncertainty, proxied by the Ludvigson et al. [2016] index, is highly correlated with classic asymmetric information proxies (see Moeller et al. [2007]) such as market volatility, the VIX index and analysts’ forecasts dispersion, and it positively correlates with price impact measures as asymmetric information models predict
sistent with the violations of the assumption of symmetric information and trading frictions in the tested models.

In intermediary-based models the marginal investor is a sophisticated intermediary rather than a household/consumer: my results suggest this alternative class of models to be in general a more robust alternative. In contrast to what happens in CRAE models, proxies of a key state variable in intermediary-based pricing are not sensitive to the periods originating the CRAE rejection, and the implied Stochastic Discount Factor (SDF) of a leading exponent of the class is never rejected, both conditionally on the periods originating the CRAE rejection and unconditionally. Moreover, the two most important drivers in the CRAE models, asymmetric information and illiquidity, are less likely to primarily concern intermediary-based models: requiring all intermediaries to have similar levels of information seems a much more reasonable assumption than extending the requirement to all investors. While, Frazzini et al. [2015], using proprietary data, show how transaction costs for this class of investors are an order of magnitude smaller than the average investor costs estimated in the literature (included those used in this article).

I reach these conclusions via a new and less restrictive framework to test the fundamental pricing equation, which delivers a non-parametric conditional asset pricing test for the CRAE models. The pricing equation states that the expectation of the SDF times the gross return of any asset is equal to one. Testing this equation is challenging because it involves the joint assessment of its linear structure as well as a particular asset pricing model, which specifies the SDF by imposing strong assumptions on its functional form and on its joint relation with the set of test asset returns. In general, any test for the pricing equation suffers from this joint hypothesis problem. Exploiting the novel inequality found by Martin [2017], I show how a test for violations of a lower bound on the market risk premium implies a joint non-parametric (SDF-free) test for the pricing equation in popular CRAE models, and ultimately a test for the models themselves. This is because such a bound only depends on a Negative Covariance Condition (an inequality involving the SDF

(Grossman and Stigliz [1980], Hellwig [1980], Diamond and Verracchia [1981], as well as Kyle [1985], and virtually all research on liquidity and market micro-structure, typically predict that asymmetric information decreases trading and destroy liquidity, defined as the ability to trade an asset without significantly changing its price)

and the market portfolio) and the pricing equation, and the structure of CRAE models satisfies the Covariance Condition. Since a test for the pricing equation is a test for the CRAE models’ First Order Conditions (FOCs), a test for the bound violations is a test for the CRAE models themselves. In particular, by defining lower bound violations as a subsample where on average the lower bound is above the realized excess market return (the sample analog of its conditional risk premium), I construct the non-parametric test for the pricing equation as a conditional test for the lower bound violations. An endogenous rule guides the selection of the subsample where a one sided t-test is run to check for lower bound violations. Specifically, in each period an econometric model, which is a function of variables proxying against the key CRAE models’ assumptions of symmetric information and real, frictionless, arbitrage-free markets, produces a one-step-ahead out-of-sample forecast for the market risk premium. Whenever such a forecast is below the lower bound an indicator function turns to one. The lower bound t-test is then performed on the subsample selected by the indicator function. This approach explicitly links the test to violations of key CRAE models’ assumptions and, in case of rejection, endogenously delivers the subsample originating it. To avoid any sample selection bias the econometric model specification and the set of actual proxies (regressors in the model) are selected in a training sample according to the best fit, while the actual test for the CRAE models is performed in a subsequent sample.

This paper indeed provides a formal unified framework to study the systematic CRAE models’ issues found by the extant literature. Martin [2017] finds that several CRAE models cannot replicate the dynamics of the lower bound through simulations. Using his lower bound measure, I set up a formal non-parametric test for all equilibrium models satisfying the Negative Covariance Condition. Muir [2017] shows how CRAE models’ have difficulty in matching the different behavior of risk premia in financial crises and economic recessions in light of the fact that consumption and its volatility respond very similarly to these events. The subsamples originating the CRAE models’ rejections in the present study also contain all the major financial crises and economic recessions and are sensitive to consumption growth. The difference is that in my test the selection is endogenous and guided by proxies against the key CRAE models’ assumptions.
Moreira and Muir [2017] find profitable trading strategies which, contrary to the predictions of CRAE models, take relatively less risk during recessions by timing the market according to lagged volatility. Due to the weak relation between lagged volatility and average returns and the strong relationship between lagged volatility and current volatility, Sharpe ratios decrease in the level of lagged volatility. This is because changes in volatility are not offset by proportional changes in expected returns. Thus, it is more profitable to lever up positions when the volatility is low and reduce the exposure in times when it is high. This is in contrast with the CRAE models’ key prediction of a representative agent demanding higher returns in periods of financial distress. Lower bound violations can be viewed as a formal test for this latter claim. First, the subsamples generating the violations contain lower Moreira-Muir exposures and all the major financial crises and recessions. Second, such subsamples are characterized by higher level of volatility and lower risk premia, implying strictly lower Sharpe ratios. Most importantly, lower bound violations imply Sharpe ratios from the perspective of CRAE models that are higher than the actual ones. This is exactly the concern Moreira and Muir raise. Finally, Greenwood and Shleifer [2014] and Amromin and Sharpe [2014], contrary to rational expectations, show how CRAE models’ required market returns disagree with actual expectations of a non-trivial fraction of investors. I show how their results are further exacerbated while conditional to the rejection subsamples.

From a methodological point of view, this paper parallels the work of Nagel and Singleton [2011]. As in my framework, they construct a test for the pricing equation guided by the selection of a set of pre-determined instruments (proxies in my wording), and reject several CRAE models with linear SDF. In their setup the instruments are non-linear functions of the first and second moments of models’ factors and test asset returns, and they are optimally selected to maximize the efficiency and the power of the test. In my approach the instruments are also selected in a nonlinear fashion, but with a different objective: they maximize the training sample fit of the forecasting model used to pin-down the rejection subsample, and they are not directly linked to the properties of the test. The upside is that my test applies to any arbitrary SDF as long as it satisfies the Negative Covariance Condition. The reason why Nagel and Singleton can tie the
selection of instruments to the efficiency and the power of their test is due to the fact that they have an explicit and linear SDF functional form, a finite number of parameters to estimate, and sharp pre-specified null and alternative hypothesis. If I were to tie a combination of instruments to a function that endogenously select the rejection subsample with the aim of maximizing the test efficiency and power, I would be dealing with an infinite dimensional parameter space defining such a function. Moreover, the null and the alternative hypothesis of this test would depend on the particular instruments’ combination and thus would not be pre-specified.

2 Framework

2.1 A new setup to test the pricing equation

The fundamental pricing equation states that the expected discounted return of any asset $i$ equals one

$$\mathbb{E}_t[M_{t+1} \times R_{t+1}^i] = 1$$

(1)

where $M_{t+1}$ is the Stochastic Discount Factor (SDF) for the period $[t : t + 1]$, $R_{t+1}^i$ is the gross return on asset $i$ over the same period, and $\mathbb{E}_t[\cdot]$ is the conditional expectation operator. The existence of equation (1) either requires: (i) the Law of One Price,$^5$ (ii) the absence of trading frictions,$^6$ and (iii) a potentially large but finite set of states of the world. Or alternatively: (i)' $M_{t+1} > 0$,$^7$ (ii) and (iii). In either context, an asset pricing model is, at minimum, a specification for $M_{t+1}$ and its joint distribution with $R_{t+1}^i$ for all $i \in \mathcal{I}$, where $\mathcal{I}$ represents the set of test assets.

Standard tests for the pricing equation$^8$ are joint assessments of all these components. In this paper I propose a new framework to test equation (1) focusing on the market return, $R_{t+1}$, by exploiting a recent asset pricing restriction linking the conditional risk premium on the market to observables in the investors’ information set. In particular, following Martin [2017], it is

$^5$The Law of One Price states that portfolios with equal payoffs have the same price.

$^6$Trading frictions are defined as any impediment to an immediate trade of any size.

$^7$This condition, under (ii) and (iii), implies no-arbitrage, defined as the absence of strategies delivering positive payoffs at zero (or negative) costs.

$^8$See footnote 1.
straightforward to derive the following proposition

**Proposition 1.** In any market which satisfies the pricing equation (1) and the Negative Covariance Condition (NCC)

$$\text{Cov}(M_{t+1} \times R_{t+1}, R_{t+1}) \leq 0, \text{ with } M_{t+1} > 0$$

(2)

it is possible to construct a real time conditional lower bound, \( LB_t \equiv \frac{\text{Var}^2(R_{t+1})}{R_{t,f}} \), on the market risk premium \( \mathbb{E}_t[\pi_{t+1}] \equiv \mathbb{E}_t[R_{t+1} - R_{t,f}] \) by

$$LB_t = 2 \left( \frac{DY_t}{S_t} \right)^2 \left( \int_0^\hat{F}_t \hat{p}u_t(k)dk + \int_\hat{F}_t^\infty \hat{c}a_l_t(k) \right) \geq 0$$

(3)

setting \( DY_t = 1 \) delivers the original Martin [2017] measure.

**Proof.** See Appendix □

quantities with hats are ex-dividend, \( DY_t \) is the gross dividend yield on the market portfolio with respect to the period \([t, t+1]\) assumed known\(^9\) at \( t \), \( \hat{S}_t \) is the closing market level at time \( t \), \( \hat{F}_t \) is the forward contract on the market with tenor \( 1 = (t+1) - t \) and finally \( \hat{p}u_t(k) \) and \( \hat{c}a_l_t(k) \) are European put and call option quotes on the market with unity tenor as a function of the common strike \( k \). By the Put-Call parity\(^{10}\) the forward contract \( \hat{F}_t \equiv \hat{F}_t(k^*) \) is the unique point \((k^*, \hat{F}_t(k^*))\) at which the call and put functions intersect so that \( LB_t \) is just a function of \( DY_t, \hat{S}_t, \{\hat{p}u_t(k_i), \hat{c}a_l_t(k_i)\}_{k_i \in K_t} \) where \( K_t \) is the set of observable strikes with unit tenor at time \( t \).

The most direct interpretation of the lower bound quantity obtains when the NCC equals zero: in this case \( LB_t \) measures the market risk premium itself from the perspective of any unconstrained myopic log-investor who is fully invested in the market.\(^{11}\)

Applying the logic of contraposition to Proposition 1 delivers a new framework to test for the pricing equation

---

\(^9\)Setting \( DY_t = 1 \) virtually delivers the same lower bound. More generally, the role played by dividends in the determination of the lower bound measure at horizon 1-month is irrelevant. See Appendix for details.

\(^{10}\)Adjusted for dividends if \( DY_t \neq 1 \), i.e. \( \hat{c}a_l_t(k) = \hat{p}u_t(k) + \hat{S}_t - PV(D_{t+1}) - \frac{k}{R_{t,f}} \).

\(^{11}\)See Example 2 in Martin [2007] Section III.
**Proposition 2.** *Given a violation of the lower bound measure in equation (3) if the Negative Covariance Condition in (2) holds the pricing equation is rejected.*

In this new setup, testing the pricing equation reduces to a joint test for the model-free lower bound measure (3) and the NCC.

### 2.2 Focusing on Representative Agent Equilibrium models

In this paper I focus on the popular context of Consumption-based Representative Agent Equilibrium models (CRAE models hereafter). Such models includes, among others, the setup of Lucas [1978] as well as the modern frameworks of external habit (e.g. Campbell and Cochrane [1999]), long-run risk (e.g. Bansal and Yaron [2004], Bansal et. al [2014], Campbell et al.[2017]), and rare disasters (e.g. Barro [2006], and Wachter [2013]).

From a structural point of view, in such frameworks the pricing equation (1) comes from the first order conditions and $M_{t+1}$ is the representative agent equilibrium inter-temporal marginal rate of substitution of consumption between $t$ and $t+1$. The coefficient of relative risk aversion is at least 1, if preferences are Epstein-Zin the elasticity of inter-temporal substitution is at least 1, and the market return as well as the (other) state variables are associated random variables.\(^{12}\)

From an environmental point of view, CRAE models operates under

**E1:**\(^{13}\) symmetric information and market completeness\(^{14}\)

**E2:** absence of trading frictions

**E3:** no-arbitrage

**E4:** (no trade) closed and real economies - single market with consumption good as numeraire

**E5:** rational expectations - representative agent knows the data generating process

\(^{12}\)The concept of associated random variables extends the concept of nonnegative correlation in a manner that can be extended to the multivariate setting. In particular, jointly normal random variables are associated if and only if they are nonnegatively correlated. (See Martin [2017] and references therein)

\(^{13}\)These conditions plus the homoteticity of utility functions (e.g. Constant Reative Risk Aversion or Epstein-Zin) are necessary conditions for the existence of a representative agent.

\(^{14}\)Defined as the ability of markets to replicate any payoff with tradeble assets.
The set of structural assumptions are sufficient for the NCC to hold, while the above environmental set of assumption E1-E5 will be used in the construction of the non-parametric test as illustrated in the Empirical Design section.

Because the NCC in (2) holds, following Proposition 2, a test for a violation of the lower bound in (3) is a non-parametric (SDF-free) test for the pricing equation (1) in CRAE models. This is a fairly general setup to study the performance of consumption-based representative agent models: if a lower bound violation is detected the First Order Conditions (FOCs) of all CRAE models are simultaneously violated via the failure of the pricing equation.

2.3 Designing the non-parametric test

This subsection explain how to construct the non-parametric test for the CRAE models. I first define the concept of a lower bound violation, then introduce the formal test

2.3.1 A lower bound violation

Let \( \pi_{t+1} \) be the realized excess market return \( R_{t+1} - R_{f,t} \) for the period \([t : t + 1]\) and recall that \( LB_t \) refers to the lower bound measure for \( E_t[\pi_{t+1}] \) computed through equation (3).

**Technical Assumption 1.**

**TA1:** the unconditional first moments of \( LB_t \) and \( \pi_{t+1} \) are bounded

TA1 guarantees\(^\text{15}\) the existence of the first conditional moments of \( LB_t \) and \( \pi_{t+1} \). It is a very weak assumption which for example is satisfied when the unconditional variance of \( LB_t \) and \( \pi_{t+1} \) is well-defined. Under TA1 I can define a lower bound violation as follows

**Definition 2.** A lower bound violation is a subsample \( I_t^v \) such that \( \mathbb{E}[\pi_{t+1}|I_t^v] < \mathbb{E}[LB_t|I_t^v] \), letting \( y_{t+1} \equiv \pi_{t+1} - LB_t \), this is equivalent to \( \mathbb{E}[y_{t+1}|I_t^v] < 0 \)

A lower bound violation is a subsample where on average \( LB_t \) fails to be a proper lower bound for the market risk premium: this is because the \( I_t^v \)-average excess market return, the subsample

\(^{15}\text{See Gut [2005], Chapter 10, definition 1.1.}\)
analog of the conditional market risk premium $\mathbb{E}[\pi_{t+1}|I_t^v]$, is below its average lower bound, the subsample analog of the conditional lower bound $\mathbb{E}[\pi_t|I_t^v]$.

2.3.2 A non-parametric test for consumption-based representative agent models

**Definition 3.** A non-parametric test for consumption-based representative agent equilibrium models is a one-sided $t$-test

$$H_0 : \mathbb{E}[y_{t+1}|I_t^v] \geq 0 \quad \text{vs.} \quad H_1 : \mathbb{E}[y_{t+1}|I_t^v] < 0$$

(4)

or equivalently,$^{16}$ for any nonnegative $I_t^v$

$$H_0 : \mathbb{E}[y_{t+1} \times I_t^v] \geq 0 \quad \text{vs.} \quad H_1 : \mathbb{E}[y_{t+1} \times I_t^v] < 0$$

(5)

the test is correctly specified as long as

**HP1:** the estimator for $\mathbb{E}[y_{t+1}|I_t^v]$ is well-defined

**HP2:** an objective rule to select $I_t^v$ exists

Given any conditioning set $I_t^v$, HP1 requires the Central Limit Theorem (CLT) to hold so that a proper limiting normal distribution for $\mathbb{E}[y_{t+1}|I_t^v]$ exists. The weakest assumptions under which the CLT hold impose (a) all up the $2 + \Delta$ moment of $y_{t+1}$ (for some $\Delta > 0$) to be bounded, and (b) the process for $y_{t+1}$ to be a strong mixing,$^{17}$ that is, a weakly dependent process in probability. A sufficient condition for strong mixing is temporal independence. For concreteness with respect to (a), I require the third moment of $y_{t+1}$ to be bounded, which means $y_{t+1}$ has to have well-defined skewness. This requirement is itself very mild and holds in general in pricing specifications, even in the presence of jumps: for example in a Black-Scholes world all moments are bounded, and the same remains true if we add a Jump diffusion component with constant intensity. In a framework like the time-varying rare disaster of Wachter 2013, where the intensities are time-varying according to a Cox-Ingeron-Ross [1985] model, it would be enough to impose a strictly

---

$^{16}$The equivalence assumes $P(I_t^v = 1) > 0$

$^{17}$See for example Thm. 5.20 of White [2001].
positive mean-reversion and long-run mean.

Another important assumption is HP2: $I_t^v$ cannot be selected had-hoc by the econometrician in order to maximize the chance of spotting a lower bound violation, otherwise the test is biased. To avoid sample-selection, $I_t^v$ should be selected endogenously and the selection criterion should not be directly linked to the test. With a little abuse of notation, define $I_t^v$ to be an objective rule. Following the stated rationale I specify the rule as

$$I_t^v = I_t^v(W_t, LB_t) \equiv 1_{[\hat{\pi}_t(W_t) < LB_t]} \quad (6)$$

where $\hat{\pi}_t(W_t)$ is the one-step-ahead out-of-sample forecast of $\mathbb{E}_t[\pi_{t+1}]$ according to an econometric model using a list $W_t$ of predictors. Specifically, I model $\pi_{t+1}$ as

$$\pi_{t+1} = f(W_t) + e_{t+1} \quad (7)$$

where $f(W)$ assumes one among the following forms:

**Linear** $f(W) = \beta_0 + \sum_{i=1}^{w} \beta_i W_i$

**Pure quadratic** $f(W) = \beta_0 + \sum_{i=1}^{w} \beta_i W_i + \sum_{j=1}^{w} \beta_{w+j} W_{w+j}^2$

**Interaction** $f(W) = \beta_0 + \sum_{i=1}^{w} \beta_i W_i + \sum_{k>l>w}^{w(w-1)} \beta_l W_l W_k$

**Quadratic** $f(W) = \beta_0 + \sum_{i=1}^{w} \beta_i W_i + \sum_{j=1}^{w} \beta_{w+j} W_{w+j}^2 + \sum_{k>l>2w}^{w(w-1)} \beta_{2w+l} W_l W_k$

with $w$ representing the number of elements in $W$. To avoid sample selection biases, the list of predictors $W_t$ are selected from a wider pool $Z_t$, and both the actual $W_t$ section as well as the econometric model specification are obtained according to the best in-sample fit in a training sample. While the non-parametric test is performed in a subsequent sample. Finally, in order to guide the selection of a subsample where CRAE models are rejected, I link the objective rule $I_t^v$ to the potential failure of the key CRAE assumptions $E1 - E5$. This step is achieved by identifying $Z_t$ with a pool of economic variables proxying against the subset of key testable assumptions
in $E1 − E5$. The details of this procedure are given in the Results section while discussing identification.

3 Data

The data used in this study is at the monthly frequency and covers the United States Financial Markets over the period Feb : 1973 − Dec : 2014. As already discussed, the sample is split into a training sample $TS = \{1, \ldots, T_s\}$ and a main sample $MS = \{T_s + 1, \ldots, T\}$ using the last 25 years. The choice of $T_s$ is due to the availability of option data (necessary for the construction of $LB_t$), that is, $T_s + 1$ is the first date for which $LB_t$ is computable. Choosing $T_s$ this way also allows to maximize the statistical power of the RAE models’ test: this is because to select the model for the risk premium forecasts via the proposed statistical method, I do not need to waste any single data point involving options and all available data is used to perform the main test. The time $t$ information set contain information up to the first business day of month $t$ included.\footnote{For all the ambiguous cases in which it is not clear what is the exact timing of an observation recorded at $t$ I lag it back one period (if anything this procedure only weakens the possibility for the test to find results)} In the same spirit of Jurado, Ludgivson and Ng [2015], because the focus of this study is to provide the most accurate description of the periods originating the RAE models’ violations rather then to come up with a profitable trading strategy, I take full advantage of all time $t$ information going beyond what is usually referred as “real-time” data.\footnote{This involves the usage of final revised historical data (e.g. the U.S. Dollar index and the M1 money supply measures from the FRED database) and ex-post best estimates (e.g. the Ludvigson et al. [2016] financial uncertainty index and the Baker-Wurgler [2006] sentiment index)} Data is divided into two categories: (i) the main variables, namely the market return $R_{t+1}$, the risk-free return $R_{t,f}$ and the lower bound $LB_t$ and (ii) the economic variables in $Z$, proxing against the CRAE models’ key assumptions and shaping the test conditioning set $I_t^v$.

3.1 Main Variables

The gross total market return is defined as $R_{t+1} \equiv \frac{\hat{S}_{t+1}}{\hat{S}_t} DY_t$ where $\hat{S}$ represents the daily closing level of the Standard & Poor’s 500 (SP500) index and $DY_t \equiv 1 + \frac{D_{t+1}}{\hat{S}_{t+1}}$ is the gross dividend yield.
with \( \{D_t\} \) being the \( SP500 \) dividend time series (divided by 12) available on Prof. Shiller website.\(^{20}\) The gross return on a risk-free investment, \( R_{t,f} \), is defined as the gross 1-month yield to maturity extracted from the Center for Research in Security Prices (CRSP) continuously compounded yield curve computed over liquid secondary market transactions on U.S. Treasuries.

The time-series of the market premium lower bound, \( \{LB_t\} \), is computed according to equation (3) in the most conservative way by linearly interpolating\(^{21}\) the Chicago Board Options Exchange (CBOE) SPX options closing bid prices; Data from January 1990 through December 1995 is provided by Optsum Data, while data from January 1996 through December 2014 is taken from OptionMetrics. For dates \( t \) in which the data is not sufficient/absent to deliver \( LB_t \) at the exact maturity of 1 month I linearly interpolate between the contemporaneous \( t \) lower bounds with the two closest maturities.

The following table summarizes the main variables

[ Table 1 goes about here ]

while Figure 1 plots in the top panel the time series of lower bound, \( LB_t \), computed according to equation (3) with \( DY = 1 \), linear interpolation and bid option quotes, against the next month excess market return \( \pi_{t+1} \),\(^{22}\) and zooms-in the lower bound dynamics in the bottom graph

[ Figure 1 goes about here ]

figures are in line with usual values, an annualized risk premium of 6.12% over the entire sample and a slightly higher estimates of 7.32% for the most recent period \( Jan1990-Dec2014 \). The time-series of conservative lower bounds \( LB_t \) features an annualized average of 3.96% and standard deviation of 3.84%.\(^{23}\) The lower bounds dynamics portrayed in Figure 1 are thoroughly described in Martin [2017].

\(^{21}\) In the Appendix I show how very similar results are obtained if we use a cubic spline interpolation instead.  
\(^{22}\) lagged back one period so to mach the time at which the expectations in \( LB_t \) are formed  
\(^{23}\) numbers that, once are restricted to the appropriate sample are very similar to those in Martin [2017]: the annualized sample mean and standard deviation of my lower bounds, computed using bid quotes, are 4.83% and 4.39%. Martin’s figures, which use mid rather than bid quotes, are 5% and 4.60%
3.2 Economic variables in $Z$

As already argued, the pool of economic variables in $Z$ should guide the selection of periods originating the CRAE models’ failure. In order to achieve this goal I propose a parsimonious list of 11 potential candidates proxying dimensions that go against assumption $E_1$-$E_5$ that are available since 1973 (beginning of $TS$). Their detailed description is reported in Appendix B.

$$Z = \{F, SII, GINI_{ch}, MDI, ILLIQ_{pi}, ILLIQ_{ts}, BM, Sent, TAX_{chg}, M1g, USDg\} \quad (8)$$

The set $Z$ carries a rich and diverse information content as shown by the average low correlation structure illustrated in panel (a) of Table 2

[Table 2 goes about here]

In what follows I briefly describe each variable and motivate its inclusion. To streamline the exposition I offer a characterization of such variables into proxies against each of the assumptions $E_1$ through $E_4$.\textsuperscript{24}

Assumption $E_1$ contains the key ingredients for the existence of a representative consumer: symmetric information,\textsuperscript{25} market completeness, and in general\textsuperscript{26} independence of preferences from wealth distribution. Because market completeness is not easy to proxy, I focus on proxies for asymmetric information and wealth distribution. In order to capture informational asymmetries I use the Ludvigson et al. [2016] financial uncertainty index $F$, designed to capture the latent degree of unpredictability in financial markets, and the Rapach et al. [2016] short interest index $SII$, capturing “the superior informational content of short seller in anticipating future aggregate cash flows and associated market returns”. I view $F$ as a proxy for asymmetric information because it nicely fits the typical description given by that literature: empirically Moeller et al. [2007] and the literature therein refer to uncertainty and analyst forecasts dispersion measures as asymmetric.

\textsuperscript{24}It is possible and not against the main goal of this paper that one or more variable simultaneously proxies for more then one dimension at a time.

\textsuperscript{25}This is true even in frameworks such as Basak [2005] or Bhamra and Uppal [2013] that allow aggregation under heterogeneity in beliefs.

\textsuperscript{26}unless preference are homotetic and identical
information proxies, while from a theoretical point of view, standard predictions of asymmetric information models\textsuperscript{27} dictates a positive correlation with price impact measures. By looking at panel (b) of Table 2 we can see how $F$ is the measure that has the highest correlation with both standard analyst forecasts dispersion measures $\text{AnystForecastsDispIBES}$ (see for example Moeller et al. [2007]) and $\text{AnystForecastsDispYu}$ (Yu, [2011]) as well as a significant positive correlation of 0.3 with the Pastor-Stambaugh [2003] (il)liquidity index, $\text{ILLIQpi}$. Concerning the inclusion of $SII$, Rapach et al. [2016] argue the index “captures the superior informational content of short seller in anticipating future aggregate cash flows and associated market returns”. In order to capture wealth distribution I use the percentage changes in the $\text{GINI}$ index, $\text{GINIchg}$, from the United States Census Bureau.

Turning to trading frictions, the opposite of $E2$, I track this dimension either by looking at the impact of taxes, through the annual percentage changes in the aggregate dollar amount paid in capital gain taxes, $\text{TAXchg}$, or by employing two popular (il)liquidity indexes: the Pastor and Stambaugh [2003] index, $\text{ILLIQpi}$, designed to capture the price reaction to trading volume, and the mimicking portfolio for the Liu [2006] index, $\text{ILLIQts}$, constructed with the aim of seizing the trading speed dimension of liquidity.\textsuperscript{28}

The absence of arbitrage, assumption $E3$, is a sufficient condition for the Law of One Price (LoP) to hold thus if LoP fails there is arbitrage- Following the standard equilibrium framework I look at price relations in the financial rather than the commodity markets\textsuperscript{29} and track significant departures from the LoP through the Pasquariello [2014] market dislocation index $\text{MDI}$ which measures abnormal discrepancies between actual (mid-quote) and theoretical prices using three textbook arbitrage parities in stock, foreign exchange, and money markets. I also add two more general and popular mispricing proxies: the Berk and Wurgler [2006] sentiment index, $\text{Sent}$, designed to capture miss-pricing due to subjective valuations not reflecting rational risk com-

\textsuperscript{27}See footnote 5.

\textsuperscript{28}In subsection 4.4.1 I will add another, perhaps more direct, market friction measure, the $\text{SP500}$ ETF bid-ask spread, which is only available stating from 1993 and thus cannot be used to construct the information set here.

\textsuperscript{29}In the context of commodity markets Horvath et al. [2008], Pippenger and Phillips [2008] and Crucini and Shintani [2008] find contrasting results concerning the validity of the LoP.
pensation, and the Dow-Jones Industrial Average book-to-market ratio $BM$.\textsuperscript{30} Because $Sent$ is designed to reflect subjective valuations not reflecting rational risk compensation and it is a state variable in equilibrium models feature non-rational expectations,\textsuperscript{31} I view it also as proxy against rational expectations: i.e. against $E5$.

Finally $E4$ dictates CRAE models to be embedded in real and closed economies. I take into account the effect of nominal forces and the impact of foreign markets by including the growth rate of the U.S. money supply, $M1g$,\textsuperscript{32} and the rate at which the U.S. dollar appreciate, $USDg$, into the list of candidates.

\section{Results}

This section presents all the main finding of this paper: first the identification of the non-parametric test for the CRAE models is thoroughly discussed and then the test results are presented. The rest of the section broadly analyzes the test implications: section 4.3 discuss the main rejection drivers, the characteristics of the rejection subsample $I_t'$ and their link with the extant literature. Section 4.4 concludes by suggesting how intermediary-based pricing might be more robust than CRAE pricing.

If not otherwise stated, all figures are at the monthly frequency and standard errors are computed using the Newey and West [1997] heteroskedasticity and autocorrelation adjusted estimator following the Newey and West [1994] automatic lag selection procedure. Also in tables where figures have an associated significance level I highlight the boxes in dark green if the significance is at the 99\% level, mild green if the significance is at the 95\% level, and light green if the significance is at the 90\% level.

\textsuperscript{30}The other Goyal and Welch [2008] standard predictors are not included in that they are not in contrast with the key assumptions $E1$ through $E4$ or they proxy for dimensions already considered (e.g. inflation as a proxy for nominal forces as explained in the next paragraph).

\textsuperscript{31}For example Barberis et al. [2015].

\textsuperscript{32}For the sake of parsimony and due to the high correlation of 0.54 with inflation I do not include the latter.
4.1 Identification

The key results in this paper are the outcomes from the non-parametric test for the CRAE models described in Definition 3 of Section 2.3. This subsection contains a detailed analysis of the identification conditions required by such test: I first show how to construct actual objective rules $I_t^v \equiv [\hat{\pi}_t < LB_t]$ and then, given the rules, provide evidence in line with a well-defined estimator for $E[y_{t+1}|I_t^v]$, the test statistic.

4.1.1 Objective rule $I_t^v$

This subsection describes and discusses the actual construction of the rule $I_t^v \equiv [\hat{\pi}_t < LB_t]$. Following section 2.3, the first step is the identification of the best in-sample specification for the econometric model (7), the actual rule to be used in the non-parametric test is then computed in the main sample, a discussion on the appropriateness and effectiveness of such methodology concludes.

Selection of the best models in the training sample

The training sample $TS = [Feb1973, Dec1989]$ contains 191 monthly observations: this is the longest time period such that all the 11 candidate instruments $Z$ are available. Given 11 variables there are 2047 different non-empty subsets of instruments $W$ and for each of these subsets 4 different specifications of equation (7) for a total of 8188 models to predict the excess market return $\pi_{t+1}$. I rank models according to their in-sample adjusted $R^2$. To minimize a model selection purely driven by overfitting, I only keep for each possible set of instrument $W$, the best model specification.\footnote{This way I avoid comparisons only made in terms of functional form.} Figure 2 plots the first 100 out of the remaining 2047 models

Chow [1960] tests using linear, quadratic or cubic specifications unambiguously identify a brake in correspondence to model 6. Guided by this evidence, I consider for the main analysis the first 6 models which all roughly explain 30% of the one month ahead excess market return variation. The $R^2$ 95% confidence intervals, computed via bootstrap of size 10000, uniformly lies between
11% and 49%. While the 95% upper confidence bound for the ANOVA p-value on the regressions, also computed via bootstrap of size 10000, are never above 0.004.

The next table details each of the selected model in terms of instruments $W$ and functional form $f(\cdot)$

4 out of 6 models use the “Quadratic” functional form while the other 2 the “Interaction” one. In terms of selected instruments, $ILLIQts$, the Liu [2006] (il)liquidity measure, is never picked, $MDI$, the Pasquariello [2014] market dislocation index, is picked by half of the models and the Rapach et al. [2016] short interest index, $SII$, and the $GINIchg$ index are selected by 4 out of 6 models. Note how the proposed selection procedure is consistent with a ranking not guided by pure overfitting: none of the selected models contain all variables and within a given set of primitive variable $W$, the chosen functional form is not necessarily the one with the highest number of predictors. For instance, the best model is the one with the least overall number of variables. Also, all the in-sample $R^2$ are statistically significant and all selected functional form are relevant: the ANOVA p-value 95% bootstrapped upper bounds are never grater than 0.004.

**Out-of-sample predictions $\hat{\pi}_t$**

Given the selected specifications from $TS$, I estimate the out-of-sample time-series for model (7) recursively as one-period-ahead forecasts of the form $\hat{\pi}_t \equiv E_t[\hat{\pi}_{t+1}] = \hat{f}_t(W_t)$ where model parameters are re-estimated at each recursion using all information up to time $t$ included.

As shown by Goyal and Welch [2008], a naïve OLS regression of excess market returns on a large number of predictors will over-parametrize the model and lead to poor out-of-sample forecasts, I therefore combine the information from the set of predictors to obtain optimal forecasts using the Iterated Combination Method (ICM) of Lin, Wu and Zhou [2016]. First, predictive regressions are run on each predictor and a constant to obtain individual forecasts. Then a weighted average of the mean of all of the individual forecasts and the prevailing mean of the excess market return using all observations till time $t$, serves as the $t$ forecast. This methodology basically amounts to a weighted average of a shrinked OLS regression, in which the out-of-diagonal elements in the
regressors’ matrix are set to zero and the regressors’ coefficients are divided by the number of regressors, and the prevailing dependent variable mean.\textsuperscript{34}

The next table compares the performance of the ICM and OLS approaches for the 6 selected specifications

\[ \text{Table 4 goes about here} \]

for a given model (a specific panel among (1) through (6)) the in-sample (training) and out-of-sample (main) time series of next month excess return estimates\textsuperscript{35} \(\pi_{t+1}^M\) are produced with \(M \in \{\text{OLS, ICM}\}\). The performances of the two different methods are judged using the out-of-sample mean squared error statistic, MSE, and the following regression benchmark

\[ \pi_{t+1} = \alpha + \beta \pi_{t+1}^M + \varepsilon_{t+1} \quad (9) \]

in terms of the produced coefficient and \(R^2\). A good performance entails a (relatively) small MSE, a (relatively) high \(R^2\), \(\alpha = 0\) and \(\beta = 1\). No matter which model we look at OLS beats ICM in-sample (especially in terms of \(R^2\) and MSE\textsuperscript{36}) but ICM consistently out-perform OLS exactly where we care the most: out-of-sample in the period Jan1990 – Dec 2014. The ICM method, while producing similar out-of-sample \(R^2\), generates a MSE 1.5 smaller, non significant \(\alpha\), and \(\beta\)s which on average are much closer to 1 and 2.7 times bigger. In particular, according to the Diebold and Marino [1995] (DM) statistic, we always reject at the 95% the null of out-of-sample OLS MSE grater than the ICM ones. These results ex-post validate the choice of adopting the Lin et. [2016] ICM method to generate the out-of-sample forecasts for the market excess return.

A couple of econometric points on the validity of the training sample TS selection of models to be used in the main sample MS are in order at this time: the first is about the ex-ante versus ex-post performance of the predictive models. Despite the procedure adopted in Section 4.1, over-

\textsuperscript{34}The weights are designed to minimize the out-of-sample mean squared error and increase the out of sample \(R^2\). (See Lin, Wu and Zhou [2016].)

\textsuperscript{35}such estimates produced by either methods in either sample and for each model are not spurious: their first autocorrelation parameters safely lies at least 2 standard errors below 0.95

\textsuperscript{36}The Diebold and Marino [1995] (DM) statistic has to be interpreted as the usual t-statistic design to compare the OLS and ICM MSEs.
fitting might still play a role in the choice of selecting the first $n$ bests models in $TS$ to be used in $MS$. As a matter of facts, best performers in a given sample tend to under-perform when adopted in another sample and vice-versa with median models remaining more stable. One then might argue for a selection of the $n$ models around the median of the in-sample $R^2$ ranking distribution rather than in its tail. Figure 3 below shows how this is not a concern when constructing the forecasts using the ICM approach

[ Figure 3 goes about here ]

the graph plots the out-of sample ICM MSE on the in-sample counter-part for the 6 selected forecasting models. Models above and to the left of the 45 degree line passing through the origin performed better in-sample while those below and to the right performed better out-of-sample. As it is apparent from the graph the selected subsample of models is balanced.

The second point is about the validity of the specifications of the selected models in the main sample: specifically, is the selection of the subset $W_t \subseteq Z_t$ carried over in $TS$ according to the best fit valid in $MS$? to answer this question, for each $i$-th selected model, I compute the residuals $r^i_{t+1} \equiv \pi_t - \hat{\pi}^i_t$ and for every $z_t \in Z_t$ such that $z_t \notin W_t$ I run the following regression

$$r^i_{t+1} = \alpha + \beta z_t + u_{t+1} \quad (10)$$

if model $i$ is well-specified with respect to $Z_t$ in $MS$ $\beta$ should be statistically insignificant.

[ Table 5 goes about here ]

Panel A pf Table 5 reports for each model the $t$-statistics associated to $\beta$ in eq. (10). Except for the first best, all model specifications remain correct in $MS$. Fortunately the miss-specification in the first best model turns out to be negligible; The forth best model is exactly the correction needed to take such miss-specification into account, this is because the only difference between the first and the forth model is the inclusion of $SII$. Furthermore, and most importantly, the two rules are very similar$^{37}$ and, as it can be checked later, no result this paper finds is affected by

$^{37}$as shown in Table 8 the correlation between the first and fourth best rule is 0.615
the exclusion of $SII$: all findings brought by model 4 are brought by model 1 as well. A further and final encouraging observation on the validity of the $TS$ specifications to be used in the main sample $MS$ can be appreciated by jumping ahead to Table 11 and note that no variable that has been excluded from the $TS$ original specification is consistently able to discriminate between the rejection subsample and the rest of the data (i.e. no columns displaying a variable which is in $Z$ but not in all of the six specifications detailed in Table 3 has a statistically significant difference in the parameter $(\alpha_1 - \alpha_2)$ for each panel).

**Is the forecasting model too good?**

In the limiting case of a perfect model, $\hat{\pi}_t = \pi_{t+1}$, and the rule $I_{iv}$ would clearly bias the non-parametric test. Even if the maximum $MS$ ICM (out-of-sample) $R^2$ of the benchmark regression (8) displayed in Table 4 is 0.091, safely rejecting this extreme, such observation raises the question of how good should a model be to substantially bias the test. One way to quantify this bias is to look at the following probability $P(\pi_{t+1} < LB_t | \hat{\pi}_t < LB_t)$. If the selected models produce forecasts $\hat{\pi}_t$ such that the above probability is high the test is biased. Panel B of Table 5 computes $P(\pi_{t+1} < LB_t | \hat{\pi}_t < LB_t)$ for the 6 forecasting models and reports its point estimate as well as the lower and upper 95% confidence intervals: no model produces estimates statistically different from 50%. These values do not seem too high, especially if one realizes that the probability of $\pi_{t+1} < LB_t$ given the outcome of a fair coin toss is 50%.

**Is the forecasting model too bad?**

At this point one might wonder if these values are too low for the test to work. Fortunately for this paper, the answer is no. There is another force in place which enables the rule $I_{iv}$ to work without biasing the test: the persistence of the forecasts process $\hat{\pi}_t$. Figure 4 plots the autocorrelation functions of $\hat{\pi}_t$ and the excess return process $\pi_{t+1}$ as well as their time-series (in the bottom graph $\hat{\pi}_t$ is represented by the blue line while $\pi_{t+1}$, lagged back at $t$, by the dashed black line) along with the lower bound $LB_t$ (bottom graph, red line) for the representative case of forecasting model 1

[ Figure 4 goes about here ]
Note how the excess return process, in contrast to the forecasted time-series, does not display any noticeable temporal persistence. Inspecting the rejection subsample via the third graph of Figure 4 reveals three patterns: the rejection subsample, as highlighted by the vertical purple stripes, is made by 32 blocks. In each block one of the following three things happen: (1) all observations are such that \( \pi_{t+1} < LB_t \) and \( \hat{\pi}_{t+1} < LB_t \) - i.e. the blue and the dashed black lines are always below the red one - (2) while \( \hat{\pi}_{t+1} < LB_t \) at least one observation is such that \( \pi_{t+1} < LB_t \) but not all - i.e. the blue line is below the red and at least in one point the dashed black line is below the red line - (3) while \( \hat{\pi}_{t+1} < LB_t \) none of the observations are such that \( \pi_{t+1} < LB_t \). Case 1 represents the instances in which the forecasting model works exactly, for the present case 34.38% of the blocks. As I already argued, the fact that this number is low is reassuring against the sample selection bias. Case 2 is key: it shows how the simultaneous persistence in the forecast process and its absence in the excess market return process helps identifying rejection periods avoiding the sample selection bias issue. 34.38% of the time a negative shock captured by some regressor in \( \hat{\pi}_t \) (mostly coming from \( F \) or \( ILLIQpi \) as I will show in a later section) induces the forecast to go below the lower bound \( LB_t \), because the forecast process is persistent, \( \hat{\pi}_t \) remains low for some time giving the chance to the highly oscillatory excess return process \( \pi_{t+1} \) to go below the lower bound while also \( \hat{\pi}_t \) is below. The last and residual case 3 refers to situations (31.24%) in which the model for the excess return is not helpful at all: those are blocks in which \( \hat{\pi}_t < LB_t \) and not a single observation for the lagged excess return is below the lower bound \( LB_t \). These blocks work against the identification of a rejection, pushing the p-values of the non-parametric test up, but fortunately represents a minority. Panel C of Table 5 extend this analysis to the other 5 rules coming from their respective forecasting models: similarly to the discussed case, associated with forecasting model 1, on average 30.41% of the time the forecasting models work exactly, most often, 37.21%, the persistence of the forecasts help, and only in less than one-third (on average) the forecasting models do not work.

---

\(^{38}\)If we look at this percentage observation-wise instead of block-wise we are back to 52.94% as reported by the statistic \( Pr(\pi_{t+1} < LB_t | \hat{\pi}_t < LB_t) \) from Panel B of Table 5.
4.1.2 Well-defined estimator for $E[y_{t+1}|I_t^v]$

Given any objective rule $I_t^v$, the non-parametric test requires a well-defined estimator for $E[y_{t+1}|I_t^v]$. As I discussed in Section 2.3, the estimator is well-defined if (a) $y_{t+1}$ is a strongly mixing stochastic (weakly dependent) process, and if (b) its third moment is bounded. Also, $E[y_{t+1}|I_t^v]$ represents a potential lower bound violation only if the first moments of $\pi_{t+1}$ and $LB_t$ are bounded. This subsection provides corroborating evidence in favor (a) and the first moments of $\pi_{t+1}$ and $LB_t$ being bounded.

A necessary requirement for these conditions to hold is that $\pi_{t+1}, LB_t$ and $y_{t+1}$ are covariance stationary processes. Moreover, for the case of $\pi_{t+1}$ and $LB_t$ this requirement becomes sufficient. Panel A of Table 6 documents the results from the unit-root Augmented Dickey and Fuller [1979] test for the representative case of model 1. The test is applied against the alternative of an AR(1) process without intercept (AR), with intercept (ARD), and with intercept and deterministic trend (TS) to $LB_t, \pi_{t+1}$ and $y_{t+1}$ using the entire main sample (MS), multiplying each time series by $I_t^v$ (xIv), and conditioning on $I_t^v$ (Iv)

no process under no AR(1) specification contains a unit root process at the canonical 5% level. In particular, an AR(1) process is covariance stationary if contains no unit-root and if its residual are i.i.d. The next step is to ask whether the AR(1) specification is a good specification for $LB_t, \pi_{t+1}$ and $y_{t+1}$: i.e. if the residuals are i.i.d. To answer this question Figure 5 plot the autocorrelograms of $LB_t, \pi_{t+1}, y_{t+1}$ and the residuals from an AR(1)-ARD specification for $LB_t$

The flat autocorrelogram of the residuals from the AR(1)-ARD specification suggests the AR(1) structure imposed on $LB_t$ to be appropriate. The correlograms of $\pi_{t+1}, y_{t+1}$ are very similar and flat as well. Moreover, the Newey and West [1987] statistics on all the various $\pi_{t+1}$ and $y_{t+1}$ AR(1) specifications reported in Panel B of Table 6 are insignificant. This last two sets of evidence support an i.i.d. specification for $\pi_{t+1}$ and $y_{t+1}$ or a covariance-stationary AR(0) representation.
4.2 The non-parametric test

Recall from Definition 3 in Section 2.3.2 that, given an objective rule \( I_t^v \equiv 1_{\hat{\pi}_t < LB_t} \) with \( \hat{\pi}_t \) being the risk premium estimate from a specification of the forecasting model (7), and given a well-defined estimator for \( E[y_{t+1}|I_t^v] \) with \( y_{t+1} \equiv \pi_{t+1} - LB_t \), we want to test

\[
H_0 : E[y_{t+1}|I_t^v] \geq 0 \text{ vs. } H_1 : E[y_{t+1}|I_t^v] < 0
\]

or equivalently,\(^{39}\) for any nonnegative \( I_t^v \)

\[
H_0 : E[y_{t+1} \times I_t^v] \geq 0 \text{ vs. } H_1 : E[y_{t+1} \times I_t^v] < 0
\]

Table 7 reports the main results from the test

[ Table 7 go about here ]

each panel shows the results for a specific rule (model). For a given model 4 statistics are reported: \( E[y_{t+1}] \), telling us if unconditionally the lower bound holds, \( E[y_{t+1}|I_t^v] \), telling us if conditionally upon the subsamples selected by \( I_t^v \) the lower bound holds, \( E[y_{t+1}^{LB}|I_t^{vLB}] \), which is the same as the previous one except that the lower bound series is kept fixed at its unconditional mean \( LB \) with \( y_{t+1}^{LB} \equiv \pi_{t+1} - LB \), and finally \( E[y_{t+1}^0|I_t^{v0}] \), which instead set the lower bound series to zero with \( y_{t+1}^0 \equiv \pi_{t+1} \). The last two estimates capture the specific role played by the lower bound series \( LB_t \). The p-values of the discussed statistics are shown in the second and third rows in each panel: for the case of the conditional estimates I report the p-values against both the equivalent alternative hypothesis \( H_1 : E[y_{t+1}|I_t^v] < 0 \) and \( H_1 : E[y_{t+1}|I_t^v] < 0 \) respectively. The last row shows the number of observations included in each selected subsample: all available observations goes into the unconditional test. \( I_t^v, I_t^{vLB} \) and \( I_t^{v0} \) pin-down the number of observations reported by the next three figures, and the last two values refer to the overall number of shared observations and those only shared by the second and third statistic. The bar charts display the impact of statistics

\(^{39}\)The equivalence assumes \( P(I_t^v = 1) > 0 \)
3 and 4 on statistic 2.

First note how unconditionally, at any level of confidence, the lower bound holds: the lower bound is on average below the risk premium by 0.281\% in the period Jan1990 – Dec2014 and 0.101\% in the period Jan1990 – Dec2009, consistently with an unconditionally tight lower bound, such estimates are not different from 0. The bulk of this article resides in the next set of results, mainly revealed by the second statistic in the table: conditional on the objective rules the lower bound is on average always violated at the 10\% level and in 4 out of 6 cases at confidence 95\% \footnote{With model 2 being borderline between 10\% and 5\% and model 6 generating rejections that are at the 5\% if evaluated against \( H_1 : \mathbb{E}[y_{t+1}|I_t(W_t, LB_t)] < 0 \) and at the 7% if evaluated against \( H_1 : \mathbb{E}[y_{t+1}I_t(W_t, LB_t)] < 0 \).}. In particular the average lower bound is above its conditional risk premium by values that ranges from 1.262 to 1.654 monthly percentage points. The number of observations involved is on average 66, between 23\% and 29\% of the main sample, with a mean of 70 if we exclude model 6. According to Proposition 1 \( LB_t \equiv \frac{\text{Var}_Q(R_{t+1})}{R_{t,f}} \geq 0 \); the results associated to the third and forth statistic in the table speaks to the importance of its informational content in the test rejections. An informative lower bound is essential: \( I_t^{v0} \) where \( LB_t \equiv 0 \) is a proper subset of \( I_t^v \) where \( LB_t \equiv \frac{\text{Var}_Q(R_{t+1})}{R_{t,f}} \) and it is never able to pin-down rejection yielding test statistics which are on average 68\% of those in the second column and p-values grater than 0.10. A non-negative (informative) dynamic lower bound is needed: the rejection rule \( I_t^{vLB} \) associated to the unconditional but informative lower bound \( LB = E[\hat{LB}_t] > 0 \) shares an average of 91\% of the observations with \( I_t^v \) (a minimum of 86\%) and significantly improves the test performance, allowing one rejections and four marginal ones with statistics that are on average 78\% of those of column 2. This tells us that three-fourth of the main rejection magnitude (i.e. statistics 2) is due to the dynamics of the excess market return \( \pi_{t+1} \) rather than that of the lower bound \( LB_t \), nonetheless its dynamics it is not negligible yielding the extra quantum, the average residual 22\% gap, needed to consistently achieve the rejections.

Are the different rejection rules \( I_t^v \), byproduct of a particular subset of instruments \( W \subset Z \) and a given specification for the forecasting model (7), consistent with each another? As I show

\footnote{This point is furthered investigated in the Appendix where I show how, consistently with the documented violations, that when the lower bound is conditioned on \( I_t^v \) it becomes a much less tight more noisy measure.} \footnote{The reason why numbers are different for model 2, 5 and 6 is due to the inclusion of the MDI instrument which makes the sample end in Dec2009.
next, the answer is yes even if there is no \textit{a-priori} mechanism that guarantees a uniform selection. This feature, which is a strong robustness check on the internal consistency of the methodology, can be appreciated either by looking at Figure 6

[ Figure 6 goes about here ]

In particular, Table A summarizes the high correlations of the different rules adopted to reject the CRAE Models: the smallest correlation, the one between model 1 and and model 2, is 0.424, while the highest, the one between model 2 and model 3, is 0.927, the average correlation among all models is 0.626. Table B displays the pairwise percentage overlap across the six rules, the smallest percentage, the one between model 1 and and model 2, is nonetheless 50\%, while the highest, the one between model 2 and model 3, is 95.52\%, the average pairwise overlap among all models is 77.32\%. Turning to the graphs, the top one plots the time series of the six objective rules $I_t^v$ against the real GDP growth: in order to make the graph more readable I multiply each rule by its associated model, i.e. rule for model $j$ is plotted as $j \times I_t^v$ and assumes values 0 and $j$ in the rejection periods. The six rules clearly display a counter-cyclical pattern, the least correlated model, model 1, displays a negative 0.16 correlation with GDP growth while the most correlated model, model 2, shows a negative correlation of 0.40, and cluster around specific periods discussed next. The bottom graph shows in green the sample periods which are systematically detected by all the rejection rules against the real GDP growth for a total of 35 observations: these are times associated with negative GDP growth (correlation coefficient of -0.30), include all the major economic recessions of the last 25 years (the pink NBER recessions) as well as the 1997 Asian financial crises, the 1998 LTCM crises, the period (late 1999 to 2001) during which the dotcom bubble collapsed, a period in late 2002 when stock market was hitting new lows following the end of the dotcom boom, the quant meltdown in August 2007 and the European sovereign debt crises which accounts for the last two green stripes.
4.3 Analysis of the rejection periods

We already saw that these are turbulent times including all the major financial crises and economic recessions. In this subsection I further analyze such periods finding that they are characterized by high uncertainty, trading frictions, and crash probabilities, as well as low consumption and GDP growth. I also find that the two most important drivers in the selection of these periods are the high level of financial uncertainty and market illiquidity, while in general time-varying risk aversion cannot justify them. The last part of this subsection show how these results relate to the literature, complementing and extending, within a formal and unified framework, their critique of CRAE Models.

4.3.1 Characteristics of the rejection periods

Table 8 describes the rejections periods in terms of a set of variables $X$ containing the original pool of economic variables in $Z$, which contains proxies against the CRAE models’ assumptions $E1 – E5$, plus other variables that: (i) were excluded because redundant (inflation), (ii) where not available in the training sample ($PBA$: bid-ask spread on the SP500 SPDR ETF, or the $VIX$ index) or (iii) are not viewed as direct proxy against assumptions $E1 – E5$ but allow for a direct comparison with the literature (column 15 to the end).

Each column refers to a different variable in $X$ and each panel to a different rejection rule. For each variable and rule the table reports (1) the difference in conditional means between the rejection subsample, $Iv = 1$, and the rest of the sample, $Iv = 0$, (2) the mean conditional on a rejection, (3) whether such mean is greater than the unconditional median, (4) the mean conditional on the rest of the sample, and (3) whether such mean is greater than the unconditional median. As anticipated, note how only the following variables are consistently able to discriminate between the rejection subsample and the rest in terms of their conditional mean differential (their estimates are significantly different, and over the median only in rejections or in the rest of the sample, in at least 5 out 6 cases): the Ludvigson et al. [2016] financial uncertainty index $F$ (expressed in
VIX percentage point for ease of interpretation\footnote{Given the high correlation of 0.84 between VIX and $F$, I regress $F$ on VIX and use $\hat{F} = \hat{\beta}_0 + \hat{\beta}_1 VIX$. Results with respect to the raw $F$ are qualitatively identical.}, the (negative of the) of the Pastor-Stambaugh (2003) price impact liquidity index, $ILLIQpi$, the $SP500$ proportional bid-ask spread, $PBA$, the Moreira-Muir [2017] weights (described and discussed in a later section), $MMweights$, the real GDP and consumption growths, $RGDPg$ (directly available from FRED) and $Cg$ (measured as the sum of nondurables and services available from FRED) respectively and the conditional probabilities of a market crash, $Pt(marketcrash)$ (described and discussed in a later section).

In particular, the high values of $F$ and the VIX with respect to the rest of the sample characterized the rejections subsample as highly uncertain, while the high values of $ILLIQpi$ and $PBA$, a more direct measure of trading friction specific to the $SP500$ which is the market proxy used in this paper, characterized the rejections subsample, with respect to the rest of the sample, as highly illiquid.

\subsection*{4.3.2 What drives the rejection periods}

Recall that $I^r_t$ turns 1 when the time $t$ forecast $\pi_{t+1}^{ICM} \equiv \hat{\pi}_t$ is below $LB_t$. We can re-write the forecast as

$$\hat{\pi}_t = \hat{f}_t(W_t) \equiv g(W_t)\hat{\theta}_t$$

(13)

with $g(W_t)$ being the vector containing all the elements of $W_t$ plus potentially their interactions and/or squares (depending on the specification adopted by eq. (7)) and $\hat{\theta}_t$ being the vector of time-varying weights. This representation is useful because explicitly links the impact of each regressor in $g(W_t)$ to the estimated forecast $\hat{\pi}_t$. The four graphs on the left-side of Figure 7 and the bottom one on the right report a brake-down of the time-series of $\hat{\theta}_t$ for the representative case of the first forecasting model:

[ Figure 7 goes about here ]

the top-left graph gives an overall snapshot of all the weights, the second focuses on the intercept, the third plots the dynamics of the linear terms highlighting those of the two most important
regressors (because $ILLIQpi$ is not on the same scale as $F$ its dynamics is separately reported in the adjacent graph) while the bottom-left graph plots the dynamic weights for the interaction terms. A quick look at the dynamics of $F$ and $ILLIQpi$ and their first order impact on the forecast $\hat{\theta}_t$ (captured by the time-varying weights associated to their levels) gives the main intuition behind the mechanics of the selection rule $I^v_t$. First, notice how both weights $\hat{\theta}_t$ are always negative and their main negative spikes are inside the rejection subsample. Second, take a look at the regressors dynamics and note that all their main positive spikes are also inside the rejection subsample and mostly correspond to the negative spikes in their weights. These observations tells us that, up to a first order approximation, the impact of positive spikes in the main regressors translate to negative shocks to the contemporaneous forecast $\hat{\theta}_t$. If the impact of these regressors are dominant, negative and high in magnitude in the rejection subsample then it must be that high values of $F$ and $ILLIQpi$ cause the forecasts $\hat{\theta}_t$ to go below the the lower bound $LB_t$ thus turning the selection rule $I^v_t$ to one. Going one step forward, because we already know from the test results that such rule has selected a rejection subsample, we could conclude that high values of $F$ and $ILLIQpi$ cause the detected RAEMs rejection.

Using eq. (13) we can break $g(W_t)\hat{\theta}_t$ into the component due to the regressor $X$ and the rest as $g(W_t)\hat{\theta}_t \equiv g^X(W_t)\hat{\theta}^X_t + rest^X_t$. The next figure plots the exclusive$^{44}$ absolute contribution, $\frac{|g^X(W_t)\hat{\theta}^X_t|}{|g^X(W_t)\hat{\theta}^X_t|+|rest^X_t|}$, of each of the 7 variables in $Z$ which are shared by all of the six models in the rejection periods, $I^v = 1$, versus the rest of the sample, $I^v = 0$

[ Figure 8 goes about here ]

let us first concentrate on the rejections and note how $F$ dominates the scene contributing on average to 70.14% of the estimate for $\mathbb{E}[\tilde{\pi}_t|I^v = 1]$ followed at great distance by $ILLIQpi$ with a modest but not negligible 14.93%. The reason why the role of $ILLIQpi$ is non-negligible is because, while coupled with $F$, it contributes to an additional average 5%, as reported in the 8th column of the table displayed below the graphs in Figure 8, and it is of particular help in raising the joint contribution of model 5 and 6 up to approximately 50% and 60% respectively. The just mentioned

---

$^{44}$with $g^X(W_t)\hat{\theta}^X_t$ coming from the linear and potentially the squared terms containing $X$ but not the interactions
table additionally reports the conditional correlation between each exclusive contribution and the overall forecast $\hat{\pi}_t$: note again how the $F$ contribution is among the highest. In contrast, a whole different story emerges from the analysis of the contributions in the non-rejection periods, $I^v_t = 0$, where all the contributions are low and either negatively or insignificantly correlated with the forecasts $\hat{\pi}_t$.

What emerges is that the two most important drivers are $F$ and $ILLIQpi$ and their contribution is dominant only in rejection periods. The next table give a more detail analysis of the exclusive joint contribution of $F$ and $ILLIQpi$

[Table 9 goes about here]

by focusing on the figures on the right note how the joint contribution $g^{ILLIQpi}(W_t)\hat{\Gamma}^{ILLIQpi}_t$ is always statistically except in the first best model where it is marginally so different in rejections from the rest of the sample and how the residual contribution when $I^v_t = 1$, the red component of the bars in the left bar graph, is almost never significant. This last point implies the blue components of the last two bars are what drive the overall outcomes emphasizing the dominant role played by $F$ and $ILLIQpi$ even in these cases. Figure 9 concludes this subsection by plotting the actual exclusive joint contribution of $F$ and $ILLIQ$ directly on the rejections.

[Figure 9 goes about here]

Clearly the joint contribution pushes toward rejections: if it was only for it since $g^{ILLIQpi}(W_t)\hat{\Gamma}^{ILLIQpi}_t < LB_t$ (i.e. the blue line is always negative) the indicator $I^v_t$ would always be 1 and we would reject all the time. That is where the structure of the entire model helps: it counter-balance this effect but almost exclusively in the non-rejection periods leaving the impact of all the negative blue spikes intact in the dynamics of the actual rejections pattern (corresponding to the instances in which the red line go negative). This is both visually evident and soundly numerically backed by a correlation between the red and the blue lines which is 0.61 in rejection periods and $-0.30$ in the rest of the sample.

\footnote{with the only exception of model 3}
4.3.3 Comparisons with the literature

Muir, [2017] using annual data over 140 countries shows how three important CRAE Models exponents - Campbell and Cochrane [1999], Bansal and Yaron [2004] and Barro [2006] - are not able to match the actual risk premium behavior during financial crises, wars and recessions. The key issue is that those models are not able, using the properties of the consumption growth time series, to differentiate between financial crises and economic recessions and wars. Column 16 and 17 of Table 8 shows the impact of the (negative of) the monthly real log consumption\(^{46}\) growth rate, \(-Cg\), and the quarterly real GDP growth rate (divided by 4), \(-RGDPg\). They both consistently and significantly\(^{47}\) picture the rejection periods as low growth times, with average growth below the median only during rejections. The *endogenous* selection of the rejection periods of this paper matches the characteristics of the *exogenous* selection in Muir in that all the rejection subsamples contain both financial crises and economic recessions plus the salient wars’ characteristics of a low GDP and consumption growth. My framework extends the annual international framework of Muir (2017) to the entire CRAE Model class at the monthly frequency for the representative U.S. sample.

Moreira and Muir [2017] [to be completed]

Greenwood and Shlifer (2014), after showing the reliability of the sources\(^{48}\), compare expectations reported in surveys of market participants with expected return forecasts based on CRAE Models predictor variables: namely the dividend price ratio, \(DP\),\(^{49}\) the Lettau et al. (2001) \(CAY\) measure,\(^{50}\) and the (negative of the) Campbell et al. (1999) surplus consumption ratio, \(-SCR\). These CRAE Models variables all positively predict the market risk premium while survey data

\(^{46}\text{computed as the aggregate sum of nondurables and services}\)

\(^{47}\text{with the exception of consumption under the first rule where the rate is in any case smaller during rejections even if not statistically so.}\)

\(^{48}\text{they show how survey expectations are highly positively correlated with each other as well as with past stock returns and the level of the stock market}\)

\(^{49}\text{which dynamics is usually targeted in the CRAE Models calibrations}\)

\(^{50}\text{best understood as reflecting consumption behavior in CRAE Models frameworks under the permanent income hypothesis with time-series variation in risk premia and power utility}\)
directly report investors’ estimates of the market risk premium: if rational expectations in CRAE Models hold the model based time-series and survey data should theoretically have a correlation of one. The next table shows the actual correlations

\[ \text{Table 10 goes about here} \]

the first raw of Panel (a) associated to each of the three variables reports the unconditional correlation of the predictor and the estimates from the Graham-Harvey and (rescaled) Gallup surveys for the 1-year risk premium and replicates Greenwood-Shleifer (2014) striking results: the correlation coefficients are always negative and mostly statistically significant. Note that the Gallup is a monthly survey conducted among households holding an investment portfolio of $10,000 or more, while the Graham-Harvey is a quarterly survey interviewing CEOs of large U.S. corporations. Therefore at least a nontrivial fraction of investors do not have rational expectations. In the second and the third raw I report correlations for the non-rejection subsamples, \( I^v = 0 \), and the rejection ones, \( I^v = 1 \). Note how in general the correlations in rejections are higher: to better appreciate this feature Panel (b) reports the difference in correlations between rejection and non-rejections together with their p-values while Figure 10 plot them

\[ \text{Figure 10 goes about here} \]

periods of CRAE Models rejection significantly worsen the unconditional results against rational expectations especially when actual expectations are those of the households as reported by the Gallup survey which more closely reflect the characteristics of the model-based representative consumer. The \( CAY \) measure is the most affected, featuring an average correlation of 0.62 and 0.30 lower in the rejection periods as measure through the Graham-Harvey and the Gallup survey respectively and starting from average insignificant non-rejection levels of 0.02 and \(-0.17\) respectively. In-between in terms of rejection period sensitiveness the (negative) of the surplus consumption ratio, \(-SCR\), with correlation deterioration of the order of 0.10 and 0.25 from already significantly negative starting levels of \(-0.49\) and \(-0.52\) in non-rejection times. The dividend price ratio, \( DP \), is the least sensitive to CRAE Models rejection (especially in the Graham-Harvey case)
with non-significant difference for the Graham-Harvey survey and a (mostly marginal) deterioration in the order of 0.12 for the Gallup survey, but this is only because it already starts from very high and negative levels in the order of $-0.65$ for both surveys. This analysis extends the concern found in Greenwood and Shleifer [2014] to the entire CRAE Models class.

4.4 An alternative more robust framework?

[to be completed]

5 Conclusions

In this paper I propose a solution to the joint hypothesis problem plaguing the usual tests for the pricing equation by focusing on the market portfolio and the class of consumption-based representative agent equilibrium models. I derive a model-free test for the pricing equation which represents a fairly general setup to analyze the joint performance of such models.

I both reject the popular Consumption-based Representative Agent Equilibrium (CRAE) models as a whole and document how the intermediary-based framework, a close substitute, might be insensitive to this failure. The subsamples originating the CRAE models’ FOCs violations are mostly driven by proxies for financial uncertainty and market illiquidity, are characterized by low consumption and GDP growths as well as high uncertainty and illiquidity, and contain all the major financial crises and economic recessions. The documented conditional failures, which I argue mainly come from high levels of asymmetric information and frictions, complement and extend the CRAE models’ criticisms in the literature.

Interestingly, my setup suggests that intermediary-based pricing is more robust to the detected issues.
6 Appendix

6.1 Proof of Proposition 1

First I show why $LB_t$ is a lower bound for the market risk premium $\mathbb{E}_t[R_{t+1} - R_{t,f}]$ then I derive equation (3).

Suppose there exist a stochastic discount factor $M_{t+1} > 0$ satisfying the pricing equation (1), then by the Fundamental Theorem of Asset Pricing (FTAP, Ross [1973,1978], Harrison and Kreps [1979], Dybvig and Ross [1987]) there exist an equivalent risk-neutral measure $Q$ such that $R_f = \mathbb{E}[R^i]$ for any gross return $R^i$ (thus for the market return $R$ as well).

By definition the conditional risk neutral variance for the market return at horizon $t+1$ can be written as

$$Var^Q_t(R_{t+1}) \equiv E^Q_t[R^2_{t+1}] - E^Q_t[R_{t+1}]^2$$

where $R_{t+1}$ is the gross cum-dividend market return. Still from FTAP we can go back and forth from the physical probability measure and the risk-neutral one, thus $E^Q_t[R^2_{t+1}] = E_t[R_{t,f}M_{t+1}R^2_{t+1}]$ and by the definition of risk-neutral measure, $E^Q_t[R_{t+1}]^2 = R_{t,f}^2$, hence

$$Var^Q_t(R_{t+1}) = E_t[R_{t,f}M_{t+1}R^2_{t+1}] - R_{t,f}^2$$

dividing the above equation by the gross risk-free return $R_{t,f}$ and rearranging

$$\frac{Var^Q_t(R_{t+1})}{R_{t,f}} = E_t[R_{t+1} - R_{t,f}] + Cov_t(M_{t+1}R_{t+1}, R_{t+1})$$

if $Cov_t(M_{t+1}R_{t+1}, R_{t+1}) \leq 0$, which together with $M_{t+1} > 0$ defines the NCC, then $LB_t \equiv \frac{Var^Q_t(R_{t+1})}{R_{t,f}}$ is a lower bound for $RP_t \equiv E_t[R_{t+1} - R_{t,f}]$.

Next, I derive equation (3). From the definition of variance, using hats to denotes ex-dividend
quantities and letting \( S \) be the cum-dividend market level

\[
Var_t^Q(R_{t+1}) = E_t^Q \left( \left( \frac{S_{t+1}}{S_t} \right)^2 \right) - E_t^Q \left( \frac{S_{t+1}}{S_t} \right)^2
\]

\[
= E_t^Q \left( \left( \frac{\hat{S}_{t+1}}{\hat{S}_t} \right)^2 \right) - R_{t,f}^2
\]

\[
= \left( \frac{(DY_t)^2}{(\hat{S}_t)^2} \right) E_t^Q \left( \frac{\hat{S}_{t+1}^2}{R_{t,f}} \right) - R_{t,f}^2
\]

by no arbitrage (see Martin 2017), since the options are written on \( \hat{S}_t \)

\[
E_t^Q \left[ \frac{\hat{S}_{t+1}^2}{R_{t,f}} \right] = 2 \int_0^\infty \text{call}_t(k) dK = 2 \left( \int_{F_t}^\infty \text{call}_t(k) dK + \int_{F_t}^\infty \text{call}_t(k) dK \right)
\]

since deep-in-the-money call options are neither liquid in practice nor intuitive to think about, it is convenient to split the range of integration for \( E_t^Q \left[ \frac{\hat{S}_{t+1}^2}{R_{t,f}} \right] \) into two and use the put-call parity to replace in-the-money call prices with out-of-the-money put prices. Assume that Market Dividends are paid as lump sums \( D_{t+1} \) at the end of the period \( [t : t + 1] \) but before \( t + 1 \), then the following is true

\[
max(S_{t+1} - D_{t+1} - k, 0) = max(k - S_{t+1} + D_{t+1}, 0) + (S_{t+1} - D_{t+1}) - k
\]

since \( \hat{S}_{t+1} = S_{t+1} - D_{t+1} \)

\[
max(\hat{S}_{t+1} - k, 0) = max(k - \hat{S}_{t+1}, 0) + (S_{t+1} - D_{t+1}) - k
\]

by linearity of the pricing equation

\[
\text{call}_t(k) = \text{put}_t(k) + \hat{S}_t - PV(D_{t+1}) - \frac{k}{R_{t,f}}
\]
where \( PV(D_{t+1}) = \mathbb{E}_t^Q \left[ \frac{D_{t+1}}{R_{t+1}} \right] = (1 - DY_t) \mathbb{E}_t^Q \left[ \frac{S_{t+1}}{R_{t+1}} \right] = \frac{DY_t}{DY_t} \tilde{S}_t \) and the last equality comes from \( R_{t,f} = \mathbb{E}_t^Q \left[ \frac{S_{t+1}}{S_t} \right] \). Applying the put-call parity

\[
\int_0^{\hat{F}_t} \text{call}_t(k)dK = \int_0^{\hat{F}_t} \text{put}_t(k)dK + \hat{F}_t \left( \frac{\hat{S}_t}{DY_t} - \frac{\hat{F}_t}{2R_{t,f}} \right) = \int_0^{\hat{F}_t} \text{put}_t(k)dK + \hat{F}_t \left( \frac{\hat{S}_t}{DY_t} - \frac{\hat{F}_t}{2R_{t,f}} \right) \]

which implies

\[
\mathbb{E}_t^Q \left[ \frac{\hat{S}_{t+1}^2}{R_{t,f}} \right] = 2 \left[ \int_0^{\hat{F}_t} \text{put}_t(k)dK + \hat{F}_t \left( \frac{\hat{S}_t}{DY_t} - \frac{\hat{F}_t}{2R_{t,f}} \right) + \int_0^{\infty} \text{call}_t(k)dK \right] \]

plugging \( \mathbb{E}_t^Q \left[ \frac{\hat{S}_{t+1}^2}{R_{t,f}} \right] \) in \( \text{Var}_t^Q(R_{t+1}) = \frac{(DY_t)^2R_{t,f}}{(S_t)^2} \mathbb{E}_t^Q \left[ \frac{S_{t+1}}{R_{t,f}} \right] - R_{t,f}^{-2} \) delivers equation (3)

\[
LB_t = 2 \frac{(DY_t)^2}{(S_t)^2} \left( \int_0^{\hat{F}_t} \text{put}_t(k)dK + \text{call}_t(k)dK \right) \]

### 6.2 The impact of dividends on the lower bound measure

Martin [2017] derives a lower bound for the risk premium, \( LB_t^M \), which is an implicit function of the market dividends. In his main formulation dividends are assumed known and part of the Standard & Poor’s 500 (SP500) index.\(^{51}\) Following this assumption all the contracts on the SP500 are to be considered as if written on the total value of the index rather than the ex-dividend one, an expedient which simplify the derivations and it is equivalent to the assumption that there are no dividends at all: as a matter of fact in my derivation \( LB_t \equiv LB_t^M \) if and only if the gross dividend yield \( DY_t \) is equal to 1. I argue that, more realistically, one should account for the fact that such contracts are written on the ex-dividend level of the SP500 so that dividends (or divided yields), even if assumed known, should become an explicit input in the lower bound derivation. Empirically whether they are a function of the dividends or not and whether dividends are indeed to be considered deterministic or stochastic turns out to be irrelevant in the current analysis.

\(^{51}\)this way the stochastic component of \( S \) only comes from the ex-dividend level \( \hat{S} \)
However, the realization of such a convenient simplification, would have been otherwise impossible to detect if no such formula, namely equation (3), for the bound as a function of dividend had been derived. The next table makes this point evident by comparing the key moments of the lower bound empirical distribution without dividends, $LB^n_M$, and with dividends, $LB_m$, computed both using a linear $m \equiv l$ and a cubic-spline $m \equiv cs$ interpolation to approximate the integral in (3)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>q25</th>
<th>q50</th>
<th>q75</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>$LB^n_t$</td>
<td>0.328</td>
<td>0.319</td>
<td>0.07</td>
<td>0.152</td>
<td>0.243</td>
<td>0.395</td>
<td>3.481</td>
</tr>
<tr>
<td>$LB_t$</td>
<td>0.330</td>
<td>0.321</td>
<td>0.07</td>
<td>0.153</td>
<td>0.249</td>
<td>0.396</td>
<td>3.502</td>
</tr>
<tr>
<td>$LB^n_M$</td>
<td>0.327</td>
<td>0.316</td>
<td>0.07</td>
<td>0.147</td>
<td>0.249</td>
<td>0.391</td>
<td>3.450</td>
</tr>
<tr>
<td>$LB_M$</td>
<td>0.329</td>
<td>0.318</td>
<td>0.07</td>
<td>0.148</td>
<td>0.251</td>
<td>0.394</td>
<td>3.471</td>
</tr>
</tbody>
</table>

the four distributions are virtually the same: the empirical role of deterministic dividends is negligible. Nonetheless, the conclusion in the current framework is even more general: if dividends were stochastic and the correlation between the gross dividend yield and the ex-dividend market return was zero, $\rho \equiv corr(DY_t, \hat{R}_t) = 0$, then $Var^Q(R_{t+1}) \approx Var^Q(\hat{R}_{t+1})^{52}$ so $LB^n_M$ would still be a good overall measure. The overall in-sample correlation is $\hat{\rho} = -0.0515$ with a p-value of 0.2334. I thus conclude that the impact of dividends is empirically irrelevant.

6.3 Linear versus cubic spline lower bound

In order to compute the lower bound measure at time $t$, $LB_t$, according to equation (3) I use the SPX options (Put and Call) bid quotes at horizon 1 month for the different available strikes as at the of the first business day of month $t$ from Optsum and Optionmetrics. In order to compute the integral in (3) we first need to interpolate the functions $\hat{put}(k)$ and $\hat{call}(k)$ over a continuum of strikes. In the study, following Martin [2017], I have used a linear interpolation. Another popular interpolant option is the cubic spline. The figure shows the time-series of lower bounds in the main sample $MS$ computed with the linear as well as the cubic-spline method with and without dividends (i.e. with $DY \geq 1$ as well $DY = 1$)

---

52 This relation follows from the fact that if $\rho \equiv corr(DY_t, \hat{R}_t) \geq 0$ then $Var^Q(R_{t+1}) \geq Var^Q(\hat{R}_{t+1})$ as shown in footnote 10 of Martin (2017)
the top panel uses the full available data: note how, independently from the presence of dividends, the spline and the linear interpolation almost perfectly overlap except for isolated points in the pre-1996 period. I adopt the most conservative of the approaches by excluding from the main sample all instances in which the spline and the linear interpolation differ proportionally (with respect to the linear scheme) by more than 50%. The new resulting sample (which is the one used in the main analysis), along with the lower bound estimates, is shown in the middle panel. All the estimates are now very close; The same point can be more precisely appreciated by looking at the bottom graph which plots the absolute percentage difference between the lower bound measures (with respect to the linear scheme) when computed using the linear as opposed to the spline approximation for the case the bound features dividends and for the case it does not. Again, it is impossible to distinguish between the case in which dividends are included from the case in which they are not, furthermore, the maximum discrepancy is now around 30% and on average the two scheme only differ by 2%.
6.4 Properties of the lower bound

Consistently with a lower bound violation, $LB_t$ becomes a less tight more noisy risk premium bound conditional on $I_t^v$ (given any selected forecasting model). The following tables display the estimates from the benchmark regressions of the form $\pi_{t+1} = \alpha^u + \beta^u LB_t + e_{t+1}$ and $\pi_{t+1}|I_t^v = \alpha^c + \beta^c LB_t|I_t^v + u_{t+1}$.

$$\pi_{t+1} = \alpha + \beta \times LB_t + e_{t+1}$$

<table>
<thead>
<tr>
<th>$I_t^v$</th>
<th>$\hat{\alpha}$</th>
<th>Std. err.</th>
<th>$\hat{\beta}$</th>
<th>Std. err.</th>
<th>$R^2$(%)</th>
<th>$R_{05}^2$(%)</th>
<th>p-val: $\alpha = 0, \beta = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unc.</td>
<td>0.000</td>
<td>0.004</td>
<td>1.774</td>
<td>1.240</td>
<td>1.56</td>
<td>3.76</td>
<td>0.63923</td>
</tr>
</tbody>
</table>

$$\pi_{t+1}|I_t^v = \alpha + \beta \times (LB_t|I_t^v) + e_{t+1}$$

<table>
<thead>
<tr>
<th>$I_t^v$</th>
<th>$\hat{\alpha}$</th>
<th>Std. err.</th>
<th>$\hat{\beta}$</th>
<th>Std. err.</th>
<th>$R^2$(%)</th>
<th>$R_{05}^2$(%)</th>
<th>p-val: $\alpha = 0, \beta = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>-0.021**</td>
<td>0.010</td>
<td>2.729*</td>
<td>1.663</td>
<td>5.09</td>
<td>-</td>
<td>0.094506</td>
</tr>
<tr>
<td>Model 2</td>
<td>-0.021*</td>
<td>0.011</td>
<td>2.399</td>
<td>1.650</td>
<td>3.59</td>
<td>-</td>
<td>0.16478</td>
</tr>
<tr>
<td>Model 3</td>
<td>-0.025**</td>
<td>0.012</td>
<td>2.890*</td>
<td>1.667</td>
<td>4.84</td>
<td>-</td>
<td>0.10092</td>
</tr>
<tr>
<td>Model 4</td>
<td>-0.024**</td>
<td>0.010</td>
<td>2.920*</td>
<td>1.682</td>
<td>5.60</td>
<td>-</td>
<td>0.070878</td>
</tr>
<tr>
<td>Model 5</td>
<td>-0.022**</td>
<td>0.010</td>
<td>2.415</td>
<td>1.601</td>
<td>3.82</td>
<td>-</td>
<td>0.078862</td>
</tr>
<tr>
<td>Model 6</td>
<td>-0.028**</td>
<td>0.012</td>
<td>2.831*</td>
<td>1.621</td>
<td>5.35</td>
<td>-</td>
<td>0.084104</td>
</tr>
</tbody>
</table>

The unconditional results are in line with Martin [2017], as the last column shows we cannot reject the null of $\alpha^u = 0$ and $\beta^u = 1$. The average annualized lower bound in $MS$ is 3.96%, while the Fama and French [2002, Table IV] estimates for the unconditional average equity premium over the period 1951-2000 are 3.83% or 4.78%. Also the out-of-sample Goyal-Welch [2008] $R^2$ ranges from 3.76% and 4.22%, compares favorably with the literature.\(^{54}\) Conditionally on the rules $I_t^v$, $LB_t$ can still be considered a proxy for the market risk premium. The null of $\alpha^c = 0$ and $\beta^c = 1$ is still not rejected at the conventional 0.05 level. Nonetheless, the generally statistically significant $\hat{\alpha}$s, the higher $\hat{\beta}$s, and the much lower p-values makes $LB_t|I_t^v$ a less tight more noisy conditional risk premium bound.

\(^{53}\)Computed as $1 - \sum \frac{\epsilon_i^2}{\nu_i^2}$ where $\epsilon$ is the error when $LB_t$ is used to forecast the equity premium and $\nu$ is the error when the historical mean equity premium (computed on a rolling basis) is used to forecast the equity premium.

\(^{54}\)In Campbell and Thompson [2008]'s most recent monthly subsample, from 1980-2003, the $R^2$ ranges from −0.0027 to 0.0003, while Goyal and Welch [2008] $R^2$ performs even worse. For a more comprehensive discussion see Martin [2017], Section V
### Appendix B

#### List of economic variables in Z

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial Uncertainty Ft</td>
<td>The cross sectional time average of the conditional Root Mean Squared Errors of more than 150 financial time-series. It captures the conditional volatility of a disturbance that is unforecastable from the perspective of economic agents.</td>
<td>The index introduced in Ludvigson et al. (2016) can be downloaded from C. Ludvigson website at <a href="https://www.sydneyludvigson.com/data-and-appendixes/">https://www.sydneyludvigson.com/data-and-appendixes/</a></td>
</tr>
<tr>
<td>Short Interest Index Silt</td>
<td>The log of the equal-weighted mean of short interest (as a percentage of share outstanding) across all publicly listed stocks on U.S. exchanges. It captures the superior informational content of short sellers.</td>
<td>The index introduced in Ludvigson et al. (2016) can be downloaded from C. Ludvigson website at <a href="https://www.sydneyludvigson.com/data-and-appendixes/">https://www.sydneyludvigson.com/data-and-appendixes/</a></td>
</tr>
<tr>
<td>Change in GINI Index GINIchgt</td>
<td>Annual percentage change in GINI measure of U.S. Household Income Dispersion</td>
<td>U.S. Census Bureau</td>
</tr>
<tr>
<td>Market Dislocation Index MDIt</td>
<td>The monthly average of hundreds of abnormal (with respect to the conditional mean of the distribution of that month) absolute violations (mid-quotes minus theoretical prices) of the Covered Interest Rate Parity, the Triangular Arbitrage Parity and the American Depository Receipt Parity.</td>
<td>The index is introduced in Pasquariello (2014). Data is available upon request writing to P. Pasquariello at <a href="mailto:pasquar@umich.edu">pasquar@umich.edu</a></td>
</tr>
<tr>
<td>Illiquidity (price impact) ILLIQpit</td>
<td>The (negative of the) aggregate average (over a month) daily response of signed volume to next day return for all individual stocks on the New York Stock Exchange and the American Stock Exchange. It represents the % cost incurred in a 1 million USD trade in the market. Similarly to the Amihud 2002 measure, it is a price impact proxy.</td>
<td>The index is taken from Pastor and Stambaugh (2003) and can be downloaded at <a href="http://finance.wharton.upenn.edu/~stambaugh/liq_data_1962_2016.txt">http://finance.wharton.upenn.edu/~stambaugh/liq_data_1962_2016.txt</a></td>
</tr>
<tr>
<td>Illiquidity (trading speed) ILLIQts1</td>
<td>Defined as the standardized turnover-adjusted number of zero daily trading volumes over the prior 12 months. Similarly to the Hou and Moskowitz (2005) measure, captures the trading speed dimension of liquidity.</td>
<td>The index is taken from Liu (2005) and it is available through W. Liu website at <a href="http://www.nottingham.edu.cn/en/business/people/academic/weimin-liu.aspx">http://www.nottingham.edu.cn/en/business/people/academic/weimin-liu.aspx</a></td>
</tr>
</tbody>
</table>
8 References


Barro R.J., (2006) Rare Disasters And Asset Markets In The Twentieth Century, The Quarterly Journal of Eco-
nomics, 121, no. 3: 823- 866.


Goyal, A., Welch, I., (2008), A Comprehensive Look at The Empirical Performance of Equity Premium Predic-


Table 1: Statistics on Main Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std.Dev.</th>
<th>Min</th>
<th>Max</th>
<th>N. Obs.</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>((R_{t+1} - 1) \times 100)</td>
<td>0.93</td>
<td>4.57</td>
<td>-21.62</td>
<td>17.05</td>
<td>492</td>
<td>All</td>
</tr>
<tr>
<td>((R_{t,f} - 1) \times 100)</td>
<td>0.42</td>
<td>0.29</td>
<td>0.000</td>
<td>1.38</td>
<td>492</td>
<td>All</td>
</tr>
<tr>
<td>(\pi_{t+1} \times 100)</td>
<td>0.61</td>
<td>4.53</td>
<td>-16.62</td>
<td>16.04</td>
<td>289</td>
<td>Main</td>
</tr>
<tr>
<td>(LB_t \times 100)</td>
<td>0.33</td>
<td>0.32</td>
<td>0.07</td>
<td>3.48</td>
<td>289</td>
<td>Main</td>
</tr>
</tbody>
</table>

The table summarizes the main variables: \(R_{t+1} - 1\) is the total net return on the SP500, \(R_{t,f} - 1\) is the 1-month yield to maturity on U.S. Treasuries, \(\pi_{t+1} = R_{t+1} - R_{t,f}\) is the excess market return and \(LB_t\) is the market premium lower bound measure computed through (3) using linear interpolation, \(DY = 1\) and bid quotes. Observations are at the monthly frequency (not annualized). The lower bound and excess market return statistics are computed in the main sample \(Jan : 1990 – Dec : 2014\) while the market and the risk-free return are computed over the entire sample \(Feb : 1973 – Dec : 2014\).
Table 2: Panel (a): Pearson correlation matrix for the candidate instruments $Z$

<table>
<thead>
<tr>
<th></th>
<th>$F$</th>
<th>$SII$</th>
<th>$TAXchg$</th>
<th>$ILLIQpi$</th>
<th>$ILLIQts$</th>
<th>$MDI$</th>
<th>$USDg$</th>
<th>$BM$</th>
<th>$M1g$</th>
<th>$Sent$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SII$</td>
<td>-0.03</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$TAXchg$</td>
<td>-0.16</td>
<td>0.03</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$ILLIQpi$</td>
<td>0.35</td>
<td>0.06</td>
<td>-0.04</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$ILLIQts$</td>
<td>-0.05</td>
<td>0.01</td>
<td>0.05</td>
<td>0.10</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$MDI$</td>
<td>0.44</td>
<td>0.04</td>
<td>-0.10</td>
<td>0.23</td>
<td>-0.04</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$USDg$</td>
<td>0.00</td>
<td>-0.14</td>
<td>0.02</td>
<td>-0.02</td>
<td>0.09</td>
<td>0.10</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$BM$</td>
<td>0.02</td>
<td>-0.41</td>
<td>0.05</td>
<td>0.06</td>
<td>0.00</td>
<td>0.05</td>
<td>-0.11</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M1g$</td>
<td>0.15</td>
<td>-0.03</td>
<td>-0.09</td>
<td>-0.07</td>
<td>-0.05</td>
<td>0.19</td>
<td>-0.05</td>
<td>0.20</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$Sent$</td>
<td>-0.08</td>
<td>0.08</td>
<td>0.23</td>
<td>-0.15</td>
<td>0.13</td>
<td>-0.19</td>
<td>0.10</td>
<td>-0.42</td>
<td>-0.06</td>
<td>1</td>
</tr>
<tr>
<td>$GINIchg$</td>
<td>-0.13</td>
<td>-0.01</td>
<td>0.24</td>
<td>-0.04</td>
<td>0.06</td>
<td>-0.10</td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.05</td>
<td>0.22</td>
</tr>
</tbody>
</table>

The table displays Pearson correlation coefficients for the candidate instruments $Z$, described in Section 3.2, over the entire sample $Feb: 1973 − Dec: 2014$, the overall average absolute correlation is 0.10.

Panel (b): Asymmetric information variables - correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>$ILLIQpi$</th>
<th>AnalystForecastsDispIBES</th>
<th>AnalystForecastsDispYu</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F$</td>
<td>0.30</td>
<td>0.67</td>
<td>0.63</td>
</tr>
<tr>
<td>$GARCH$</td>
<td>0.17</td>
<td>0.72</td>
<td>0.35</td>
</tr>
<tr>
<td>$VIX$</td>
<td>0.31</td>
<td>0.62</td>
<td>0.44</td>
</tr>
</tbody>
</table>

The table displays the Pearson correlation coefficients for competing asymmetric information candidates to be included in $Z$ over the portion of the main sample $Jan1990 − Dec2014$ where they overlap. $F$ is the Ludvigson et al. [2016] financial uncertainty index, $GARCH$ is a the Standard and Poor’s 500 volatility calculated via a $GARCH(1,1)$ model on monthly returns, $VIX$ is the Chicago Board of Exchange implied volatility index, $ILLIQpi$ is the Pastor-Stambaugh [2003] (il)liquidity index, $AnalystForecastsDispIBES$ is the I/b/e/s market analyst forecasts dispersion measure adopted in Moeller et al. [2007] and $AnalystForecastsDispYu$ is the bottom-up market analyst forecasts dispersion measure computed in Yu [2011].
Table 3: First 6 best selected models for the excess market return

<table>
<thead>
<tr>
<th>W</th>
<th>F</th>
<th>SII</th>
<th>TAXchg</th>
<th>ILLIQpi</th>
<th>ILLIQts</th>
<th>MDI</th>
<th>USDg</th>
<th>BM</th>
<th>M1g</th>
<th>Sent</th>
<th>GINIchg</th>
<th>f(·)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6th</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table describes the characteristics of the first 6 models to predict the excess market return $\pi$, ranked by their adjusted in-sample $R^2$ in the training sample Feb : 1973 – Dec : 1989.
Table 4: Out-of-sample forecasts $\hat{\pi}_t$: OLS vs. ICM

$\pi_{t+1} = \alpha + \beta \pi_{t+1} + \tau_{t+1}, \ M \in \{\text{OLS}, \text{ICM}\}$ (*)

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$R^2$</th>
<th>MSE</th>
<th>DMstat</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training (in-sample)</td>
<td>OLS</td>
<td>0.0000</td>
<td>1.000***</td>
<td>1.000***</td>
<td>0.440</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>ICM</td>
<td>0.0000</td>
<td>1.000***</td>
<td>1.000***</td>
<td>0.077</td>
<td>0.002</td>
</tr>
<tr>
<td>Main (out-of-sample)</td>
<td>OLS</td>
<td>0.0000</td>
<td>0.301***</td>
<td>0.301***</td>
<td>0.054</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>ICM</td>
<td>0.0000</td>
<td>1.002***</td>
<td>1.002***</td>
<td>0.072</td>
<td>0.002</td>
</tr>
<tr>
<td>Training (in-sample)</td>
<td>OLS</td>
<td>0.0000</td>
<td>1.000***</td>
<td>1.000***</td>
<td>0.423</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>ICM</td>
<td>0.0000</td>
<td>1.000***</td>
<td>1.000***</td>
<td>0.061</td>
<td>0.002</td>
</tr>
<tr>
<td>Main (out-of-sample)</td>
<td>OLS</td>
<td>0.0000</td>
<td>0.305***</td>
<td>0.305***</td>
<td>0.086</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>ICM</td>
<td>0.0000</td>
<td>0.666***</td>
<td>0.666***</td>
<td>0.080</td>
<td>0.002</td>
</tr>
<tr>
<td>Training (in-sample)</td>
<td>OLS</td>
<td>0.0000</td>
<td>1.000***</td>
<td>1.000***</td>
<td>0.383</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>ICM</td>
<td>0.0000</td>
<td>1.000***</td>
<td>1.000***</td>
<td>0.061</td>
<td>0.002</td>
</tr>
<tr>
<td>Main (out-of-sample)</td>
<td>OLS</td>
<td>0.005**</td>
<td>0.299***</td>
<td>0.299***</td>
<td>0.069</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>ICM</td>
<td>0.0000</td>
<td>0.930***</td>
<td>0.930***</td>
<td>0.083</td>
<td>0.002</td>
</tr>
<tr>
<td>Training (in-sample)</td>
<td>OLS</td>
<td>0.0000</td>
<td>1.000***</td>
<td>1.000***</td>
<td>0.464</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>ICM</td>
<td>0.0000</td>
<td>1.000***</td>
<td>1.000***</td>
<td>0.073</td>
<td>0.002</td>
</tr>
<tr>
<td>Main (out-of-sample)</td>
<td>OLS</td>
<td>0.004</td>
<td>0.346***</td>
<td>0.346***</td>
<td>0.083</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>ICM</td>
<td>0.003</td>
<td>1.155***</td>
<td>1.155***</td>
<td>0.091</td>
<td>0.002</td>
</tr>
<tr>
<td>Training (in-sample)</td>
<td>OLS</td>
<td>0.0000</td>
<td>1.000***</td>
<td>1.000***</td>
<td>0.538</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>ICM</td>
<td>0.0000</td>
<td>1.000***</td>
<td>1.000***</td>
<td>0.074</td>
<td>0.002</td>
</tr>
<tr>
<td>Main (out-of-sample)</td>
<td>OLS</td>
<td>0.005*</td>
<td>0.267***</td>
<td>0.267***</td>
<td>0.115</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>ICM</td>
<td>0.002</td>
<td>0.614***</td>
<td>0.614***</td>
<td>0.082</td>
<td>0.002</td>
</tr>
<tr>
<td>Training (in-sample)</td>
<td>OLS</td>
<td>0.0000</td>
<td>1.000***</td>
<td>1.000***</td>
<td>0.497</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>ICM</td>
<td>0.0000</td>
<td>1.000***</td>
<td>1.000***</td>
<td>0.077</td>
<td>0.002</td>
</tr>
<tr>
<td>Main (out-of-sample)</td>
<td>OLS</td>
<td>0.004</td>
<td>0.313***</td>
<td>0.313***</td>
<td>0.122</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>ICM</td>
<td>0.002</td>
<td>0.518***</td>
<td>0.518***</td>
<td>0.074</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Among all possible combinations using functions of set of regressors in $Z$ to forecast the excess market return, $\pi_{t+1}$, the 6 best predictive specifications in terms of adjusted $R^2$ as measured in the training sample [Feb1972-Dec1989] are selected and their performance analyzed in panel (1) through (6). For each of the given specifications two different estimation methodologies are implemented: OLS and the Lin, Wu and Zhou [2016] Iterated Combination Method (ICM). In-sample (training) the models are estimated once while out-of-sample (main, [Jan1990-Dec2014]) the predictions are generated as iterative one-step-ahead conditional forecasts. To assess the models’ performances, the actual market excess return time series is regressed on the time series of in-sample and out-of sample predictions generated by the two different models according to (*). The regression coefficients as well as the $R^2$ and the out-of-sample Mean Squared Error (MSE) are reported in column 2 to 5. Column 6 reports the Diebold and Mariano [1995] statistic, which should be read as the usual t-statistic, to compare the OLS and ICM MSEs. The last column reports the number of observations in the training and the main sample across the different scenarios.
Table 5: Identification of the objective rule $I_t^v$

Panel A reports for each selected model in the training sample [Feb1972 – Dec1989] the t-statistics associated to $\beta$ in equation (10) with $r_{t+1}^i = \pi_{t+1} - \hat{\pi}_t$ where $\hat{\pi}_t$ is the forecast of $\pi_{t+1}$ given one of the six specifications of model (7) performed in the main sample [Jan1990 – Dec2014]. Panel B computes $P(\pi_{t+1} < LB_t | \hat{\pi}_t < LB_t)$ for the 6 models and reports its point estimate as well as the lower and upper 95% confidence intervals: values much higher than 50% lead to biased estimates for the RAEMs tests. Panel C shows the number of buckets selected by the rule $I_t^v$ (in % terms) according to the following categorization: case (1) all observation $t$ in that bucket are such that $\pi_{t+1} < LB_t | \hat{\pi}_t < LB_t$, case (2) given $\hat{\pi}_t < LB_t$ there is at least 1 observation such that $\pi_{t+1} < LB_t$, and case (3): given $\hat{\pi}_t < LB_t$ there is no observation such that $\pi_{t+1} < LB_t$.

<table>
<thead>
<tr>
<th>Panel A</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SI</td>
<td>ILIQts</td>
<td>MD1</td>
<td>GINichg</td>
</tr>
<tr>
<td>Model 1</td>
<td>-3.15***</td>
<td>-0.762</td>
<td>-0.5559</td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>-</td>
<td>-1.2339</td>
<td>-</td>
<td>-1.81*</td>
</tr>
<tr>
<td>Model 3</td>
<td>-</td>
<td>-1.0336</td>
<td>0.4791</td>
<td>-1.91*</td>
</tr>
<tr>
<td>Model 4</td>
<td>-</td>
<td>-0.8182</td>
<td>-0.539</td>
<td>-</td>
</tr>
<tr>
<td>Model 5</td>
<td>-</td>
<td>-1.0486</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Model 6</td>
<td>-1.72*</td>
<td>-0.8274</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>95% LB</td>
<td>95% UB</td>
</tr>
<tr>
<td>Model 1</td>
<td>52.94</td>
<td>41.43</td>
<td>64.45</td>
</tr>
<tr>
<td>Model 2</td>
<td>54.41</td>
<td>41.20</td>
<td>67.63</td>
</tr>
<tr>
<td>Model 3</td>
<td>55.41</td>
<td>44.40</td>
<td>67.41</td>
</tr>
<tr>
<td>Model 4</td>
<td>54.93</td>
<td>44.00</td>
<td>65.85</td>
</tr>
<tr>
<td>Model 5</td>
<td>58.21</td>
<td>47.62</td>
<td>68.80</td>
</tr>
<tr>
<td>Model 6</td>
<td>58.33</td>
<td>46.01</td>
<td>70.66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>If works</td>
<td>ACF(h1) helps</td>
<td>If not work</td>
</tr>
<tr>
<td>(%)</td>
<td>case 1</td>
<td>case 2</td>
</tr>
<tr>
<td>Model 1</td>
<td>34.28</td>
<td>34.38</td>
</tr>
<tr>
<td>Model 2</td>
<td>20.83</td>
<td>33.33</td>
</tr>
<tr>
<td>Model 3</td>
<td>24.00</td>
<td>40.00</td>
</tr>
<tr>
<td>Model 4</td>
<td>35.29</td>
<td>28.41</td>
</tr>
<tr>
<td>Model 5</td>
<td>29.17</td>
<td>41.67</td>
</tr>
<tr>
<td>Model 6</td>
<td>38.89</td>
<td>44.44</td>
</tr>
<tr>
<td>Avg</td>
<td>30.41</td>
<td>37.21</td>
</tr>
</tbody>
</table>
Panel A shows the results from an Augmented Dickey and Fuller [1979] test for the representative case of model 1. The test is applied against the alternative of an AR(1) process without intercept (AR), with intercept (ARD), and with intercept and deterministic trend (TS) to $LB_t$, $\pi_{t+1}$, and $y_{t+1}$ using the entire main sample (MS), multiplying each time series by $I_v$ (xIV), and conditioning on $I_v$ (IV). Panel B uses the same categorization as Panel A only it computes the Newey and West [1987] statistics on the AR(1) coefficients. 

Table 6: Necessary condition for a well-defined estimator for $\mathbb{E}_t[y_{t+1}|I_v']$
Table 7: Results from the non-parametric test

| mom (%) | 1 \( E[y_{t+1}] \) | 2 \( E[y_{t+1}|I_v^t] \) | 3 \( E[y_{t+1}|I_v^{LB}] \) | 4 \( E[y_{t+1}|I_v^0] \) |
|---------|----------------|----------------|----------------|----------------|
| **Estimate** | | | | |
| p-val for \( H_0; \text{mom} < 0 \) | | | | |
| Obs. in \( I_v^t \) | | | | |
| Panel (a): model 1 | 0.281 | -1.283** | -0.697 | -0.895 |
| 0.835 | 0.049 | 0.173 | 0.193 | |
| 289 | 68 | 70 | 44 | |
| Panel (b): model 2 | 0.101 | -1.262* | -1.150** | -0.403 |
| 0.618 | 0.069 | 0.076 | 0.351 | |
| 230 | 68 | 68 | 44 | |
| Panel (c): model 3 | 0.281 | -1.399** | -1.068** | -0.565 |
| 0.835 | 0.043 | 0.081 | 0.300 | |
| 289 | 74 | 71 | 46 | |
| Panel (d): model 4 | 0.281 | -1.364** | -1.108** | -1.107 |
| 0.835 | 0.035 | 0.041 | 0.111 | |
| 289 | 71 | 80 | 50 | |
| Panel (e): model 5 | 0.101 | -1.479** | -1.219** | -0.781 |
| 0.018 | 0.044 | 0.080 | 0.196 | |
| 230 | 67 | 73 | 50 | |
| Panel (f): model 6 | 0.101 | -1.654* | -1.363* | -1.565 |
| 0.618 | 0.070 | 0.095 | 0.143 | |
| 230 | 49 | 49 | 32 | |

Each panel shows the results for a specific rule (model). For a given model 4 statistics are reported: \( E[y_{t+1}] \), telling us if unconditionally the lower bound holds, \( E[y_{t+1}|I_v^t] \), telling us if conditionally upon the subsamples selected by \( I_v^t \) the lower bound holds, \( E[y_{t+1}|I_v^{LB}] \), which is the same as the previous one except that the lower bound series is kept fixed at its unconditional mean \( LB \) with \( y_{t+1}^{LB} = \pi_{t+1} - LB \), and finally \( E[y_{t+1}|I_v^0] \), which instead set the lower bound series to zero with \( y_{t+1}^0 = \pi_{t+1} \). The last two estimates capture the specific role played by the lower bound series \( LB_t \). The p-values of the discussed statistics are shown in the second and third rows in each panel: for the case of the conditional estimates I report the p-values against both the equivalent alternative hypothesis \( H_1 : E[y_{t+1}|I_v^t] < 0 \) and \( H_1 : E[y_{t+1}|I_v^0] < 0 \) respectively. The last row shows the number of observations included in each selected subsample: all available observations goes into the unconditional test. \( I_v^t, I_v^{LB} \) and \( I_v^0 \) pin-down the number of observations reported by the next three figures, and the last two values refer to the overall number of shared observations and those only shared by the second and third statistic. The bar charts display the impact of statistics 3 and 4 on statistic 2.
Each column refers to a different variable $X$ and each panel to a different rejection rule. For each variable and rule the table reports (1) the difference in conditional means between the rejection subsample, $I^v = 1$, and the rest of the sample, $I^v = 0$, (2) the mean conditional on a rejection, (3) whether such mean is greater than the unconditional median, (4) the mean conditional on the rest of the sample, and (5) whether such mean is greater than the unconditional median.
The table breaks down, numerically and graphically, the forecasts for the excess market return $\hat{\pi}_t$ as the sum of the exclusive joint contribution of the Ludvigson et al. [2016] financial uncertainty $F$ and the Pastor-Stambaugh [2003] $ILLIQpi$, $g^{F,LILLIQpi}(W_t)\theta_t^{F,LILLIQpi}$, and the rest of the model, $rest_t$. 

$$X_{contribution} = \frac{|g^{F}(W_t)\theta_t^F|}{|g^{F}(W_t)\theta_t^F| + |rest_t|}$$
Table 10: Greenwood and Shleifer [2014] rational expectation test

<table>
<thead>
<tr>
<th></th>
<th>model 1</th>
<th>model 2</th>
<th>model 3</th>
<th>model 4</th>
<th>model 5</th>
<th>model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP Uncond. pixel</td>
<td>GH 0.619</td>
<td>GH 0.673</td>
<td>GH 0.666</td>
<td>GH 0.619</td>
<td>GH 0.673</td>
<td>GH -0.619</td>
</tr>
<tr>
<td></td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
</tr>
<tr>
<td>DP</td>
<td>v</td>
<td>= 1 pixel</td>
<td>GH -0.641</td>
<td>GH -0.641</td>
<td>GH 0.059</td>
<td>GH 0.031</td>
</tr>
<tr>
<td></td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
</tr>
<tr>
<td>DP Uncond. pixel</td>
<td>GH 0.603</td>
<td>GH 0.635</td>
<td>GH -0.741</td>
<td>GH -0.641</td>
<td>GH 0.059</td>
<td>GH 0.031</td>
</tr>
<tr>
<td></td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
</tr>
<tr>
<td>CAY Uncond. pixel</td>
<td>GH -0.198</td>
<td>GH -0.287</td>
<td>GH -0.174</td>
<td>GH -0.279</td>
<td>GH -0.198</td>
<td>GH -0.198</td>
</tr>
<tr>
<td></td>
<td>Gall 0.210</td>
<td>Gall 0.001</td>
<td>Gall 0.884</td>
<td>Gall 0.001</td>
<td>Gall 0.210</td>
<td>Gall 0.001</td>
</tr>
<tr>
<td>CAY</td>
<td>v</td>
<td>= 0 pixel</td>
<td>GH -0.076</td>
<td>GH -0.203</td>
<td>GH 0.132</td>
<td>GH -0.131</td>
</tr>
<tr>
<td></td>
<td>Gall 0.210</td>
<td>Gall 0.001</td>
<td>Gall 0.884</td>
<td>Gall 0.001</td>
<td>Gall 0.210</td>
<td>Gall 0.001</td>
</tr>
<tr>
<td>CAY</td>
<td>v</td>
<td>= 1 pixel</td>
<td>GH 0.099</td>
<td>GH 0.504</td>
<td>GH 0.259</td>
<td>GH 0.224</td>
</tr>
<tr>
<td></td>
<td>Gall 0.790</td>
<td>Gall 0.270</td>
<td>Gall 0.645</td>
<td>Gall 0.002</td>
<td>Gall 0.955</td>
<td>Gall 0.005</td>
</tr>
<tr>
<td>Panel (b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corr. lv1 - Corr lv0</td>
<td>model 1</td>
<td>model 2</td>
<td>model 3</td>
<td>model 4</td>
<td>model 5</td>
<td>model 6</td>
</tr>
<tr>
<td>DP P(est</td>
<td>0)</td>
<td>GH -0.521</td>
<td>GH 0.567</td>
<td>GH 0.467</td>
<td>GH 0.562</td>
<td>GH 0.521</td>
</tr>
<tr>
<td></td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
</tr>
<tr>
<td>CAY P(est</td>
<td>0)</td>
<td>GH -0.535</td>
<td>GH 0.494</td>
<td>GH 0.433</td>
<td>GH 0.559</td>
<td>GH 0.575</td>
</tr>
<tr>
<td></td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
</tr>
<tr>
<td>-SCR P(est</td>
<td>0)</td>
<td>GH -0.429</td>
<td>GH 0.802</td>
<td>GH -0.590</td>
<td>GH 0.757</td>
<td>GH -0.672</td>
</tr>
<tr>
<td></td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
<td>Gall 0.000</td>
</tr>
</tbody>
</table>

Table 12: Panel (a) compares the conditional (upon rejection, I_v = 1 and the rest of the sample, I_v = 0) and unconditional correlations of RAEMs proxies for the expected market returns versus actual investors’ expectations from survey data: the model-based measures are the dividend price ratio, DP, the Lettau and Ludvigson [2001] CAY measure, and the (negative of the) Campbell and Cochrane [1999] surplus consumption ratio, −SCR. Panel (b) reports the difference in conditional correlations between the rejection periods and the rest of the sample for the variables introduced in Panel (a).
10 Figures

Figure 1: Main variables

The figure plots the excess market return $\pi_{t+1} = R_{t+1} - R_{t,f}$ and the lower bound measure $LB_t$ computed according to (3) setting $DY = 1$, using linear interpolation and bid quotes.
Figure 2 shows the first 100 models ranked by their adjusted in-sample $R^2$ along with a third order polynomial fit. Chow tests using linear, quadratic or cubic specifications unambiguously identify a break in correspondence of model 6. The sample is $TS = [Feb1973 : Dec1989]$. 
The figure plots the out-of-sample ICM MSE on the in-sample counter-part for the 6 selected forecasting models. Models above and to the left of the 45 degree line passing through the origin performed better in-sample while those below and to the right performed better out-of-sample. As it is apparent from the graph the selected subsample of models is balanced.
Figure 4: why $I_t^*$ works

Figure 4 plots the autocorrelation functions of the forecast process $\hat{\pi}_t$ and the excess return process $\pi_{t+1}$ as well as their time-series (in the bottom graph $\hat{\pi}_t$ is represented by the blue line while $\pi_{t+1}$, lagged back at $t$, by the dashed black line) along with the lower bound $LB_t$ (bottom graph, red line) for the representative case of the first best forecasting model.
Figure 5: ACFs of $\pi_{t+1}, LB_t, y_{t+1} \equiv \pi_{t+1} - LB_t$ and residuals from AR(1) model on $LB_t$.
The top graph plots the time series of the six objective rules $I_t^v(W_t, LB_t)$ against the real GDP growth: in order to make the graph more readable I multiply each rule by its associated model, i.e. rule for model $j$ is plotted as $j \times I_t^v(W_t, LB_t)$ and assumes values 0 and $j$ in the rejection periods. The bottom graph shows in green the sample periods which are systematically detected by all the rejection rules against the real GDP growth for a total of 35 observations. Tables A and B show the correlation as well as the % overlapping across the different rules respectively.
Figure 7: RAEMs Rejection mechanics

The first column shows the dynamics for the time-varying weights $\hat{\theta}_t$ in $\hat{\pi}_t = \hat{f}_t(W_t) \equiv g(W_t)\hat{\theta}_t$ with $g(W_t)$ being the vector containing all the elements of $W_t$ plus their interactions (according to the first best “interaction” specification adopted by eq. (7)). The top graph depicts all the $\hat{\theta}_t$ dynamics, the second focuses on the intercept, the third plots the dynamics of the linear terms highlighting those of the two most important regressors - the Ludvigson et al. [2016] uncertainty index $F$ and the Pastor-Stambaugh [2003] (il)liquidity index $ILLIQ_{pi}$ (because $ILLIQ_{pi}$ is not on the same scale as $F$ its dynamics is separately reported in the adjacent graph) and the bottom graph plots the dynamics for the interaction terms. Finally, to better visualize the mechanics, the time-series for $F$ and $ILLIQ_{pi}$ are reported in the first and second graph in the second column. In all these graphs the vertical purple stripes highlights the rejection subsample.
The figure plots the exclusive contribution to the excess return forecasts $\hat{\pi}_t$ of each of the seven economic variable $W \in Z$ that are common to all six chosen rule $I_v$ plus the joint exclusive contribution of the first two conditional on the rejection subsample $I_v = 1$ and on the rest of the sample $I_v = 0$. The table in the bottom contains the mean of the contributions across the six different model specification.
Figure 9: (exclusive) joint contribution of $F$ and $ILLIQpi$

The figure uses the representative forecasting model 1 to plot the dynamics for the exclusive joint contribution $g^{ILLIQpi}(W_{t})\hat{\theta}_{t}^{ILLIQpi}$ net of the lower bound $LB_t$ and the dynamics for the difference in the excess return forecasts $\hat{\pi}_t$ and the lower bound $LB_t$ also reporting their conditional correlations in rejections as well as in the rest of the sample.
Figure 10: Greenwood and Shleifer [2014] rational expectation test

The figure plots Panel (b) of Table 10, refer to it for the descriptions.