Time-Varying Equity Premium with Noisy Consumption*

Alessandro Melone†

This version: August 28, 2025

Abstract

In the habit model, the surplus consumption ratio predicts returns with an R^2 of 11%, yet its structural empirical counterpart captures only about 40% of that. This paper shows that measurement error in consumption data largely explains the discrepancy. Using a simulated habit economy, I study how measurement error affects predictive regressions and propose an alternative measure—the consumption gap—designed to recover the surplus consumption signal even when consumption is mismeasured. Empirically, the consumption gap is a powerful return predictor, incremental to alternative business-cycle consumption measures, and it tracks time variation in the market price of risk in cross-sectional tests. These results highlight that measurement error—not model failure—largely accounts for the weak empirical performance of the surplus consumption ratio, supporting habit formation as a key mechanism for understanding time-varying risk premia.

Keywords: Consumption-Based Asset Pricing, Measurement Error, Time-Varying Equity Premium, Habit Formation Model, Conditioning Variable.
JEL codes: C22, E32, E44, G12.

*I am grateful to Federico Bandi, John Campbell, Zhanhui Chen, John Cochrane, Ian Dew-Becker, Jack Favilukis, Roberto Gomez-Cram, Andrei Gonçalves, Niels Gormsen, Robert Hodrick, Philipp Illeditsch, Ravi Jagannathan, Leonid Kogan, Tim Kroencke, Sydney Ludvigson, Roberto Marfè, Tom Meling, Florian Nagler, Dimitris Papanikolaou, and Andreas Stathopoulos for helpful comments. I also benefited from inputs of seminar participants at University of Bologna, WU Vienna, Northwestern University (Kellogg), City University of Hong Kong, Tilburg University, University of British Columbia (Sauder), Collegio Carlo Alberto, UNSW Business School, The Ohio State University (Fisher), Federal Reserve Board, and Indiana University. I am indebted to Carlo Favero, Robert Korajczyk, Andrea Tamoni, Christian Wagner, and Josef Zechner for insightful discussions and invaluable advice. Part of this paper was written while visiting Northwestern University; I thankfully acknowledge financial support from the OeAD (Marietta Blau Grant) during this time. Some of the results herein appeared in an earlier version of this paper titled "Consumption Disconnect Redux." All errors, of course, are my own.

†Fisher College of Business, The Ohio State University, 2100 Neil Avenue, Columbus, OH 43210. E-mail: melone.11@osu.edu. Personal website: Homepage.

"...one must take seriously measurement error and specification error in consumption data before estimating and testing any consumption-based model ..."

Campbell and Cochrane (2000)

1 Introduction

The consumption-based habit formation model of Campbell and Cochrane (1999) has become a benchmark for explaining time-varying risk premia. Its appeal lies in an intuitive economic mechanism: by embedding a slow-moving habit into a power utility function, the model generates countercyclical variation in risk aversion. As consumption approaches the habit level during business cycle troughs, the curvature of the utility function increases, raising expected returns and depressing risky asset prices. The surplus consumption ratio—the gap between consumption and habit—emerges as the key state variable driving fluctuations in the price-dividend ratio and expected returns.

In the data, however, the surplus consumption ratio is a much weaker predictor than the model implies. Figure 1 illustrates this point. The first bar reports the annual R^2 in the model; the second shows the performance of its structural empirical counterpart s, constructed following Cochrane (2011). The empirical R^2 is just 4.7%—roughly 40% of the model-implied value. The third bar offers a striking contrast. It reports the annual predictive R^2 of cyclical consumption (cc), a proxy for surplus consumption recently proposed by Atanasov, Møller, and Priestley (2020), which nearly twice the predictive power of s and reaches a level broadly consistent with the model's implications. Two questions naturally follow: Why does s dramatically underperform its theoretical counterpart? And why does cc perform so well? This paper shows that measurement error in consumption data offers a unified explanation for both the underperformance of s and the success of cc.

As preliminary evidence, the fourth bar in Figure 1 shows the empirical performance of s after correcting for measurement error in National Income and Product Accounts (NIPA) consumption data. I unfilter consumption using the approach of Kroencke (2017), and reconstruct the surplus consumption ratio following Cochrane (2011). The resulting variable, denoted s^{UNFIL}, achieves an R^2 comparable to that of cc and its model-implied value. These findings indicate that the weak empirical performance of s reflects measurement error rather

than model failure, thereby reinforcing habit formation as a central mechanism for understanding time-varying risk premia.

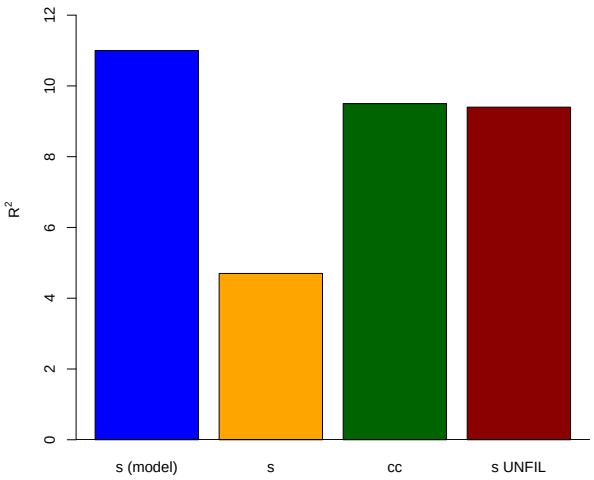


Figure 1: Surplus Consumption Predictability. This figure shows the annual R^2 from predictive regressions of market excess return using different measures for the surplus consumption ratio. The first bar reports the predictive R^2 in the Campbell–Cochrane model. Annual observations. The sample period is 1957 to 2022.

I begin by studying the effects of mismeasured, or noisy, consumption in the Campbell–Cochrane model. Noisy consumption induces more complex dynamics for habit, making the standard solution used to compute habit and detrend consumption potentially suboptimal. First, consistent with this argument and the empirical evidence in Figure 1, I show that the predictive power of the surplus consumption ratio deteriorates substantially when consumption is measured with error (second bar). Second, I find that the filter used to construct *cc* is calibrated to maximize the correlation between true surplus consumption—computed using consumption data without measurement error—and *cc* constructed from noisy consumption. That is, *cc* is implicitly designed to minimize the wedge between expected returns and surplus consumption in the presence of measurement error. These results provide an explanation for both the underperformance of s and the relative success of *cc* in the data.

Thus, I propose an alternative measure—the consumption gap (*cgap*)—explicitly constructed to maximize the correlation between true surplus consumption and its proxy under noisy consumption. I find that, while *cc* outperforms s in the presence of high levels of mis-

measurement, *cgap* exhibits an R^2 that is approximately 1.8 to 2.6 times higher than that of *cc*—a result that closely mirrors what I will find empirically. In the simulated habit economy, *cgap* emerges as the relevant measure when consumption is mismeasured, as its predictive power remains largely unaffected by the degree of measurement error. These findings both motivate and inform my empirical analysis.

Empirically, I study the predictability performance of the *cgap* for stock returns using U.S. quarterly data for the period 1957–2022. To compute the *cgap*, I apply a *one-sided* Hodrick-Prescott (HP) filter to the levels of log consumption. Importantly, I calibrate the filter's λ parameter using the value that, in the simulated data, yields the most robust correlation between *cgap* and the true surplus consumption ratio. I find that the consumption gap is a strong equity premium predictor: a one-standard-deviation decrease in the consumption gap is associated with an increase of about 1.6% (23.3%) in expected market returns over the next quarter (next five years). The R^2 from these predictive regressions is about 3% at a quarterly horizon and increases up to 48% at a five-year horizon. Importantly, return predictability holds out-of-sample, across different aggregation horizons and sub-samples, accounting for the small-sample Stambaugh bias, and using unrevised vintage consumption data. These results are particularly promising when compared to those reported by Welch, Goyal, and Zafirov (2024), who document disappointing performance for most equity premium predictors, characterized by weak in-sample predictability and negative out-of-sample R^2.

This time series predictability is both statistically and economically significant. A real-time mean-variance investor who allocates their wealth between equity and risk-free Treasury bills each quarter would be willing to pay an annualized fee of 4.27% to a portfolio manager who uses the consumption gap to predict aggregate market returns rather than the historical mean. This market timing strategy generates significant positive alphas with respect to several prominent factor models, and an out-of-sample Sharpe ratio of about 0.60, which is about 22 percentage points higher than the Sharpe ratio of a buy-and-hold investor. This outperformance is of similar magnitude to that recently documented in Gu, Kelly, and Xiu (2020) using state-of-the-art machine learning methods to predict the equity premium. Having a parsimonious, economically meaningful predictive model with a single predictor is particularly valuable in the age of big data, in which the curse of dimensionality is a severe

concern (e.g., Martin and Nagel, 2021).

Next, I investigate the predictive performance of the consumption gap in a multivariate setting. First, I study how the consumption gap compares to alternative empirical proxies for surplus consumption, including those constructed in Cochrane (2011), Wachter (2006), Atanasov, Møller, and Priestley (2020), and Bandi and Tamoni (2023). This analysis is important because the *cgap* should effectively identify the relevant habit-based variable given the dynamics of NIPA consumption data. The out-of-sample tests show that *cgap* is a superior equity predictor across different aggregation horizons. This predictability translates into *extra* utility gains for a mean-variance investor ranging between 118 and 378 basis points each year.

Additionally, I find that the predictive power of the consumption gap for stock returns is not captured by other standard predictors documented in the literature. In particular, I run multivariate predictive regressions using the consumption gap together with other well-known return predictors, including financial ratios, macro variables, and sentiment. The results indicate that the consumption gap captures most of the variation in expected excess market returns.

Is the documented predictability due to mispricing or a time-varying equity premium? My analysis supports the latter view for two main reasons. First, the consumption gap does not contain the level of stock prices, eliminating concerns that the predictability of returns is a result of mechanical mean-reversion in prices. Second, the cross-sectional analysis indicates that the consumption gap tracks the market price of risk over time, supporting the economic mechanism proposed in the habit model.

In particular, I estimate a conditional version of the Capital Asset Pricing Model (CAPM) in which the stochastic discount factor (SDF) is linear in the market portfolio and its coefficients are time-varying. The coefficients are a linear function of the consumption gap. The intuition for this specification is that the price of market risk is higher in bad times, when the consumption gap is low and marginal utility is high, than it is in good times, when the consumption gap is high and marginal utility is low. Conditioning the SDF coefficients using the information contained in the consumption gap leads to an unconditional two-factor model, in which the two factors are the market portfolio and the market portfolio scaled by the consumption gap. The market portfolio is the only fundamental factor. The scaled

market portfolio can be interpreted as a managed portfolio that invests more in the market when the consumption gap is high, and vice versa. Since the consumption gap negatively predicts returns, the managed portfolio represents a hedge against time-varying equity premia as it performs well when the stock market is low and marginal utility is high. Thus, the risk price associated with this managed portfolio should be negative. The cross-sectional analysis confirms this prediction.

Finally, I use the SDF coefficient estimates obtained from the cross-sectional analysis to construct an empirical SDF and study its dynamics. This empirical SDF fluctuates substantially over time, with an annualized conditional variance ranging between 0.4 and 3.3. Moreover, I find that the SDF variance is informative about the state of the economy. Indeed, the variance of the SDF correlates with measures of economic activity, and large spikes in the SDF coincide with periods of bad economic conditions and financial distress–the 1970s energy crisis, the early 1980s recession, the Great Financial Crisis and its aftermath, and the COVID-19 pandemic. In these situations, consumption declines toward habit, marginal utility is high, and individuals demand higher compensation for bearing risks, thus generating variation in the conditional risk premium in a way consistent with macro-finance theories featuring Abel (1990)'s "catching up with the Joneses" preferences (e.g., Campbell and Cochrane, 1999; Chan and Kogan, 2002).

This paper contributes to the literature showing that consumption-based asset pricing models can align theoretical predictions with empirical evidence once measurement error in consumption data is properly addressed. While prior work has focused on constructing alternative measures of consumption growth risk to explain the equity premium and the cross-section of returns (e.g., Parker and Julliard, 2005; Jagannathan and Wang, 2007; Savov, 2011; Kroencke, 2017; Bandi and Tamoni, 2023; Bryzgalova, Huang, and Julliard, 2024), this paper instead examines how mismeasured consumption affects return predictability. The analysis shows that measurement error distorts the information embedded in consumption-based state variables, significantly weakening their forecasting power.[1]

[1] Another strand of work includes Borup and Schütte (2021) and Guo and Pai (2024), who show that using unrevised consumption data improves asset pricing performance by enhancing predictability in the cross-section and time series, respectively. While this strategy addresses data revisions, my framework offers a general approach for analyzing the time-series effects of mismeasured consumption arising from broader economic mechanisms, including filtering and slow consumption adjustment.

This paper also contributes to the literature on market return predictability. Welch, Goyal, and Zafirov (2024) document that most return predictors exhibit unstable performance, particularly in out-of-sample tests. This includes key macro-based state variables such as the consumption–wealth ratio (*cay*) of Lettau and Ludvigson (2001a) and the cyclical consumption measure (*cc*) of Atanasov, Møller, and Priestley (2020). I show that a simple, economically motivated predictor generates out-of-sample performance comparable to that of state-of-the-art machine learning methods that leverage large datasets and non-linearities (e.g., Kelly and Pruitt, 2013; Gu, Kelly, and Xiu, 2020). These findings underscore the central role of business-cycle consumption dynamics in explaining countercyclical variation in risk premia.

2 Theoretical Motivation

The objective of this section is to discuss the conditions under which the consumption gap serves as a meaningful return predictor. Specifically, I show, using a simulated external habit model à la Campbell and Cochrane (1999), that i) the consumption gap forecasts long-horizon returns better than the surplus consumption ratio or an alternative cyclical consumption predictor when consumption is mismeasured, and ii) its predictability is largely unaffected by the severity of the measurement error.

I start by briefly describing the Campbell and Cochrane (1999) economy. Then, I discuss the implications of measurement error in consumption for the surplus consumption ratio and introduce the consumption gap. Finally, I describe the simulated external habit economy, both with and without mismeasured consumption, and discuss the resulting predictability outcomes.

2.1 The Campbell-Cochrane Economy

Identical agents maximize utility over consumption relative to a time-varying reference point X_t:

$$E\left[\sum_{t=0}^{\infty}\delta^t\frac{(C_t-X_t)^{1-\gamma}-1}{1-\gamma}\right],$$

where $\delta > 0$ is the time preference parameter and $\gamma > 0$ is a curvature parameter.[2] The level of habit, X_t, is defined indirectly through surplus consumption S_t

$$S_t = \frac{C_t - X_t}{C_t}. \tag{1}$$

Consumption growth is specified as an i.i.d. lognormal process:

$$\Delta c_{t+1} = g + v_{t+1}, \tag{2}$$

where g is the average log growth rate and v_{t+1} is an i.i.d. $\mathcal{N}(0, \sigma_v^2)$ shock. Dividend growth is also an i.i.d process and is imperfectly correlated with consumption growth:

$$\Delta d_{t+1} = g + w_{t+1}; \quad w_{t+1} \sim \text{i.i.d. } \mathcal{N}(0, \sigma_w^2), \quad \text{corr}(w_t, v_t) = \rho_{w,v}.$$

The logarithm of S_t in equation (1) is assumed to follow a heteroskedastic autoregressive process

$$s_{t+1} = (1 - \phi)\bar{s} + \phi s_t + \lambda(s_t)v_{t+1}, \tag{3}$$

where the parameters $\bar{s}$ and ϕ control the mean and persistence of s_t, while the function $\lambda(s_t)$ controls its sensitivity to the consumption shock and is specified as

$$\lambda(s_t) = \begin{cases} \bar{S}^{-1}\sqrt{1 - 2(s_t - \bar{s})} - 1, & s_t < s_{\max} \\ 0, & s_t \geq s_{\max} \end{cases} \tag{4}$$

where $\bar{S} = \sigma_v\sqrt{\frac{\gamma}{1-\phi}}$, $s_{\max} = \bar{s} + \frac{1}{2(1-\bar{S}^2)}$, and $\bar{s} = \ln \bar{S}$. This specification for s_t implies that (log) habit x_t adjusts slowly and geometrically to consumption c_t with coefficient ϕ (i.e., s_t is de-trended consumption).

All agents are identical so individual consumption and habit and aggregate consumption and habit are the same. Since habit is external, the intertemporal marginal rate of

[2] Agent's preferences exhibit the "catching up with the Joneses" feature of Abel (1990) so that habit formation is external. External habit formation is not an exclusive feature of the Campbell-Cochrane economy; it is also used in other theoretical models, such as the heterogeneous preferences economy of Chan and Kogan (2002).

substitution is

$$M_{t+1} = \delta e^{-\gamma(\Delta c_{t+1} + \Delta s_{t+1})}. \tag{5}$$

Using the fact that any asset return R_{t+1} must satisfy the Euler equation

$$1 = E_t[M_{t+1} R_{t+1}],$$

one can derive the price-dividend ratio of a claim to the consumption stream (or, equivalently, price-consumption ratio)

$$\frac{P_t}{C_t}(s_t) = E_t\left[M_{t+1}\frac{C_{t+1}}{C_t}\left(1 + \frac{P_{t+1}}{C_{t+1}}(s_{t+1})\right)\right]. \tag{6}$$

Similarly, the price-dividend ratio of a claim to the dividend stream is

$$\frac{P_t}{D_t}(s_t) = E_t\left[M_{t+1}\frac{D_{t+1}}{D_t}\left(1 + \frac{P_{t+1}}{D_{t+1}}(s_{t+1})\right)\right]. \tag{7}$$

The expressions (6) and (7) show that the price-consumption ratio and the price-dividend ratio depend only on the model's state variable s_t. By substituting M_{t+1} from (5) and consumption growth from (2), Campbell and Cochrane (1999) solve equation (6) numerically on a grid for the state variable s_t. Given the price-consumption ratio (or the price-dividend ratio) as a function of the state, they then calculate expected returns and other relevant quantities.

2.2 Measurement Error in Consumption, Noisy surplus consumption, and the Consumption Gap

The dynamics of aggregate consumption are crucial for the dynamics of surplus consumption, and, consequently, for the dynamics of the model-implied price-dividend ratio and expected returns. However, the *true* consumption growth dynamics—and thus the true surplus consumption dynamics—are unobservable. Indeed, *observed* National Income and Product Accounts (NIPA) consumption data are subject to several measurement issues, such as sampling error, imputation procedures, and filtering, which generate an artificially smooth con-

sumption time series (e.g., Wilcox, 1992; Daniel and Marshall, 1997; Parker, 2001; Kroencke, 2017).[3] Following Campbell and Cochrane (1999), I assume that consumption growth follows an i.i.d. process in the simulated economy I study.

To examine how measurement error in consumption affects the dynamics of the surplus consumption ratio, I consider the case in which observed consumption follows the specification:

$$\widetilde{\Delta} c_{t+1} = \Delta c_{t+1} + \varepsilon_{t+1} \ ,$$

where Δc_{t+1} is true consumption, as specified in (2), and ε_{t+1} is a mean-zero measurement error process. I refer to $\widetilde{\Delta} c_{t+1}$ as *noisy* consumption. Following Daniel and Marshall (1997) and Kroencke (2017), I adopt a tractable model of measurement error by specifying that current observed consumption is a weighted average of current true consumption and lagged reported consumption. This simple specification captures various technical factors (e.g., filtering) as well as economic mechanisms that generate an observed consumption series smoother than true consumption.[4] Specifically, observed consumption follows:

$$\widetilde{\Delta} c_{t+1} = \rho \widetilde{\Delta} c_t + (1-\rho) \Delta c_{t+1} \ , \tag{8}$$

where ρ lies between 0 and 1 and determines the degree of measurement error. Higher values of ρ result in observed consumption that is progressively smoother than true consumption Δc_{t+1}. Under (8), the error in consumption growth is given by

$$\varepsilon_t = \widetilde{\Delta} c_t - \Delta c_t = -\rho v_t + (1-\rho) \sum_{j=1}^{\infty} \rho^j v_{t-j} \ .$$

Consider an econometrician who observes $\widetilde{\Delta} c_{t+1}$ and wants to construct surplus consumption following its specification (3); I denote this new object as noisy surplus consumption

[3] Many papers argue that the deficiencies of consumption-based asset pricing models are possibly due to consumption mismeasurement, including Hansen and Singleton (1983); Breeden, Gibbons, and Litzenberger (1989); Cochrane (1996); Campbell and Cochrane (2000); Parker and Julliard (2005); Jagannathan and Wang (2007); Savov (2011); Kroencke (2017). Interestingly, even Campbell and Cochrane (1999) acknowledge that measurement error in consumption could explain their model's limited success in the late sample period.

[4] This specification is also widely used to unsmooth returns of illiquid assets; see, e.g., Geltner, 1991; Couts, Gonçalves, and Rossi, 2024.

$\widetilde{s}_t$. First, the consumption shock multiplying the sensitivity function $\lambda(s_t)$ will be different as a consequence of the fact that consumption is autocorrelated (see, e.g., Wachter, 2001; Campbell, Pflueger, and Viceira, 2020). Specifically, the consumption shock will be $\widetilde{v}_{t+1} = \widetilde{\Delta}c_{t+1} - E_t(\widetilde{\Delta}c_{t+1}) = \widetilde{\Delta}c_{t+1} - \rho\widetilde{\Delta}c_t - (1-\rho)g$. Second, $\bar{S}$ depends on the volatility of the consumption process, and the volatility of $\widetilde{\Delta}c_{t+1}$ is a function of the measurement error parameter ρ. Thus, also the sensitivity function specified in equation (4) will change accordingly.

As surplus consumption depends positively on recent past consumption, s_t is essentially interpreted as a recession state variable: the higher recent consumption growth, the further agents' consumption is from habit, the higher the marginal utility and the lower compensation they require for bearing aggregate risks. Under the joint assumption that log consumption levels are a random walk and surplus consumption evolves according to (3), habit x_t is approximately a weighted moving average of past consumption levels, and the surplus consumption ratio is basically detrended consumption (see, e.g., Cochrane, 2011). However, autocorrelated consumption implies more complex dynamics for habit than in the standard Campbell-Cochrane case.[5] In this context, the standard solution to compute habit and detrend consumption may be suboptimal.

Thus, I introduce a new measure that i) maintains the same economic interpretation of surplus consumption as a recession state variable and ii) better identifies the habit level given the dynamics of observed consumption. Specifically, I propose using a Hodrick-Prescott (HP) filter to identify the persistent component in consumption, i.e., the habit level. There are two main reasons why the HP filter is a natural method to identify habit and de-trend consumption levels. First, it assumes a highly persistent I(2) process for the consumption trend, consistent with both the economic interpretation of habit and the empirical evidence (e.g., Heaton, 1995; Li, 2001). Second, the HP filter is effective at capturing cyclical components of a time series when the underlying process is non-trivial (De Jong and Sakarya, 2016; Hodrick, 2020), as is the case for consumption (recent studies include Schorfheide, Song, and Yaron, 2018; Bandi and Tamoni, 2023; Bryzgalova, Huang, and Julliard, 2024).[6]

[5] For example, in the model used by Campbell, Pflueger, and Viceira (2020), log habit is also a function of past expected consumption growth; see their Appendix B for details.

[6] Hamilton (2018) discusses some shortcomings of the HP filter and proposes a different filter. However, (i) the main criticism is that the HP filter generates spurious dynamic relations, which I address by using a

I define the consumption gap as:

$$cgap_t = c_t - \tau_t \ , \tag{9}$$

where c_t are log consumption *levels* and τ_t is their *one-sided* HP trend. Using a one-sided filter ensures that *cgap* does not incorporate future information, thereby avoiding spurious predictability issues (i.e., look-ahead bias).

A critical input for applying the HP filter is the calibration of the smoothing parameter λ. While the textbook value of λ for quarterly data is 1,600, several studies have used significantly higher smoothing parameters, reaching up to $\lambda = 10^6$ (see, e.g., Harvey and Trimbur, 2008; Kaiser and Maravall, 2012; Flaig, 2015; Ramey and Zubairy, 2018; Coibion, Gorodnichenko, and Ulate, 2018; Barnichon, Debortoli, and Matthes, 2022).[7] I use simulated data to inform the calibration for λ in my asset pricing framework. Specifically, following Campbell and Cochrane (1999), I simulate a long history of artificial data from the external habit model described in Section 2.1. I calibrate the parameters as in the original paper; see Campbell and Cochrane (1999), Table 1. I run simulations for different values of λ, ranging from 10^1 to 10^{10}, and for different levels of measurement error ρ in (8), varying from 0.1 to 0.5.

Panel (a) of Figure 2 reports the correlation between the true surplus consumption ratio and the consumption gap constructed using observed noisy consumption, $\widetilde{cgap}$. It is evident that very high values of λ are required to generate a strong correlation between the true surplus consumption ratio and its noisy proxy, ranging from 64% to 74% for λ between 10^6 and 10^8. While $\lambda = 10^7$ maximizes the correlation between surplus consumption (sc) and the consumption gap ($cgap$), it also exhibits greater sensitivity to the intensity of measurement error (ρ). By contrast, $\lambda = 10^6$ achieves a comparably high correlation while maintaining

one-sided filter, (ii) the H filter of Hamilton (2018) is not optimal to compute a highly persistent component like habit as it defines the trend at time t as $\mathrm{E}_{t-8}(y_t)$, and (iii) the H filter displays large spurious dynamics due to base effects following an outlier shock such as the COVID-19 pandemic (e.g., Kamber, Morley, and Wong, 2024). Furthermore, Hodrick (2020) shows through simulations that while the H filter successfully identifies cycles when the underlying time series is a random walk, the HP filter more closely characterizes the underlying cycle for more complex time series, such as noisy consumption. I show that HP filtered-consumption uniformly dominates H filtered-consumption in, both in simulated data and empirically.

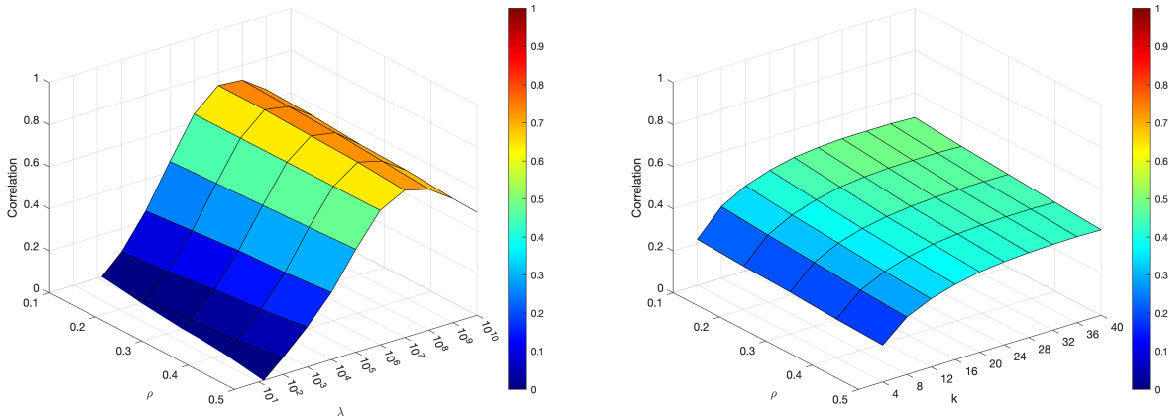
[7] Empirical estimates of potential output provided by the IMF, FED, and BIS also align with very high λ values; see, e.g., Coibion, Gorodnichenko, and Ulate (2018).

the lowest variance across different levels of measurement error.

More formally, since the actual form and severity of measurement error in the data are unobservable, I use a bias-variance tradeoff criterion to select the optimal λ by minimizing

$$(\rho_{\max} - \bar{\rho}(\lambda))^2 + \sigma_\rho^2(\lambda) \ ,$$

where $\rho_{\max}$ is the maximum correlation observed across different measurement error levels, $\bar{\rho}(\lambda)$ is the average correlation across measurement error levels, and $\sigma_\rho^2(\lambda)$ is the variance of the correlation across measurement error levels. This procedure selects $\lambda^* = 10^6$, which I use throughout the paper in *both* the simulation and empirical analyses.



(a) $\widetilde{cgap}$ **(b)** $\widetilde{cc}$

Figure 2: True surplus consumption Ratio vs. Noisy Measures. This figure shows the correlation between the true surplus consumption ratio and either the noisy consumption gap (Panel a) or noisy cyclical consumption (Panel b) as a function of the measurement error degree, ρ, and the filter parameters, λ and k, respectively.

Additionally, I consider an alternative proxy for surplus consumption, denoted as cyclical consumption (cc), recently proposed by Atanasov, Møller, and Priestley (2020). This variable is constructed using the linear projection method of Hamilton (2018) (i.e., the H filter) as the residual from the following regression:

$$c_t = b_0 + b_1 c_{t-k} + b_2 c_{t-k-1} + b_3 c_{t-k-2} + b_4 c_{t-k-3} + u_t \ .$$

Here, the key parameter is k.[8] While Hamilton (2018) suggests using $k = 8$ for quarterly data, Atanasov, Møller, and Priestley (2020) use a value as high as $k = 24$. To study the sensitivity of cc to this parameter, in a similar spirit to the calibration of λ above, I run simulations for different values of k, ranging from 4 to 40, and for different levels of measurement error.

Panel (b) of Figure 2 reports the correlation between the true surplus consumption ratio and cyclical consumption constructed using noisy consumption, $\widetilde{cc}$. Two key observations emerge. First, the simulation shows that cyclical consumption reaches its highest correlation (49%) with true surplus consumption precisely at $k = 24$. This finding suggests that, at least in part, the empirical success of cyclical consumption can be attributed to the presence of measurement error in consumption. Following Atanasov, Møller, and Priestley (2020) and in line with the simulation results, I use $k = 24$ to compute cyclical consumption throughout the paper. Second, $\widetilde{cc}$ never reaches the high levels of correlation with true surplus consumption that $\widetilde{cgap}$ achieves. This suggests that the HP filter more effectively identifies the relevant variable when consumption is mismeasured, compared to the H filter.

Ultimately, the ability of surplus consumption or alternative measures to forecast excess returns depends on the type and degree of measurement error in consumption. Next, I study the relative performance of these variables as equity premium predictors in a simulated economy with and without measurement error.[9]

[8] As for *cgap*, I construct this variable using an expanding window to avoid the look-ahead bias.

[9] An alternative to the approaches discussed in this section is to unfilter observed consumption in (8), for example, by following the methodology of Kroencke (2017), and use the resulting series to construct the surplus consumption ratio (s^{UNFIL}). This method is appealing because it aims to eliminate measurement error in simulations, yielding s^{UNFIL} with a high correlation to the true surplus consumption ratio s. However, it relies on the strong assumption that the econometrician knows the exact specification of the measurement error in observed consumption. Empirically, this approach has two key drawbacks. First, the nature and severity of measurement error are unobservable, making it crucial to adopt a method that is robust to this uncertainty. As I show, *cgap* meets this requirement. Second, applying the Kalman filter proposed by Kroencke (2017) requires calibrating several parameters and making specific assumptions about consumption volatility dynamics, introducing additional sources of uncertainty. Consistent with these considerations, I find that *cgap* is a better return predictor than s^{UNFIL} in the data. Specifically, Appendix Table B.1 shows that while constructing the surplus consumption ratio from unfiltered consumption improves the predictive R^2 by approximately 1.5 times relative to using observed NIPA consumption, incorporating *cgap* further increases the R^2 by more than 0.4 times. Moreover, Columns (4) and (5) indicate that in a multivariate predictive regression, only the *cgap* coefficient remains statistically and economically significant.

2.3 Return Predictability in a Habit Economy

(a) Return Predictability *without* Measurement Error in Consumption

I begin by studying model-implied predictability in the habit economy when consumption is observed without error. Following Campbell and Cochrane (1999), I simulate a long history of artificial data from the external habit model described in Section 2.1, and construct five variables (for both consumption and the dividend claim): log excess stock returns, the log price-dividend ratio (log pd), the surplus consumption ratio (s), the consumption gap ($cgap$), and the cyclical gap (cc). I then run long-horizon predictive regressions of returns on log pd, s, $cgap$, and cc.

Table 1: Predictive Regressions in the Campbell-Cochrane Economy *without* Measurement Error

This table reports OLS estimates and R^2 (in square brackets) from long-horizon predictive regressions of model-implied log excess stock returns on log price-dividend ratio, surplus consumption ratio (s), the consumption gap ($cgap$), or cyclical consumption (cc). The parameters are calibrated as in Campbell and Cochrane (1999) (see Table 1). All regressors are de-meaned and standardized.

Horizon (Years)	1	2	3	5	7
log pd	−1.93	−3.61	−5.07	−7.43	−9.18
	[11.84]	[21.96]	[30.47]	[43.71]	[52.84]
s	−0.51	−0.96	−1.35	−1.98	−2.45
	[11.00]	[20.38]	[28.37]	[40.72]	[49.24]
$cgap$	−0.36	−0.68	−0.95	−1.39	−1.72
	[5.44]	[10.10]	[14.00]	[20.05]	[24.18]
cc	−0.31	−0.58	−0.81	−1.20	−1.48
	[3.95]	[7.49]	[10.27]	[14.87]	[17.90]

Table 1 reports OLS estimates and R^2 for these predictive regressions.[10] Three comments are in order. First, the coefficients are negative for all four predictors: when marginal utility

[10] The first two rows replicate results reported in Table 5, Columns 2 to 5 in Campbell and Cochrane (1999). The results differ slightly from those reported in Campbell and Cochrane (1999) as I simulate quarterly rather than monthly data as in the original paper.

is low (s, $cgap$, and cc are high), expected returns are low, and prices are high. Second, the R^2 from the predictive regressions when using pd or s are essentially the same. This result follows from the fact that the price-dividend ratio is a function of surplus consumption only, as it is the only state variable in the economy. This evidence also allows me to focus on the relative performance of the consumption gap against s directly in the following analyses. Third, $cgap$ and cc have comparable predictive coefficients and R^2, but the proportion of explained variation in equity returns is substantially smaller for these variables compared to s. When consumption is observed without error, the surplus consumption ratio is the best-performing predictor.

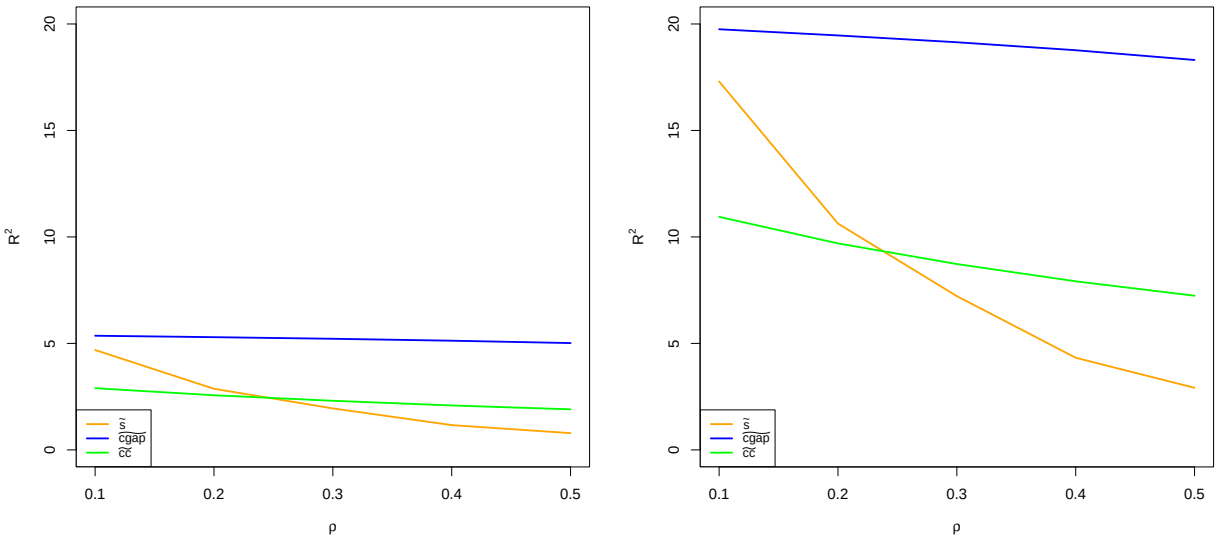
(b) Return Predictability *with* Measurement Error in Consumption

To study the implications of mismeasured consumption for return predictability, I assume that while the true consumption process is unobservable, one would observe the noisy consumption process $\widetilde{\Delta}c_t$, which follows the dynamics specified in equation (8). In this case, I construct the noisy surplus consumption ratio $\widetilde{s}_t$ as an econometrician who observes $\widetilde{\Delta}c_t$ and needs to infer the properties of the price-dividend ratio using noisy consumption. Specifically, as discussed in Section 2.2, the sensitivity function and the consumption shock will be a function of the volatility of noisy consumption and the degree of measurement error, and I calibrate the free parameters ϕ and γ to match the level and the volatility of the price-dividend ratio.[11] Similarly, noisy consumption gap ($\widetilde{cgap}_t$) and cyclical consumption ($\widetilde{cc}_t$) are extracted from mismeasured consumption levels.

Figure 3 shows the R^2 from regressing log excess stock returns on the model-implied $\widetilde{s}_t$, $\widetilde{cgap}_t$, and $\widetilde{cc}_t$ as a function of the degree of measurement error governed by the parameter ρ in (8). I display results for values of ρ between 0.1 and 0.5, corresponding to noisy consumption processes which have a correlation with true consumption in the model between about 83% and 54%. These correlations are approximately the correlations between consumption and unfiltered consumption in Kroencke (2017) and between consumption and garbage in Savov (2011), respectively.

Panel (a) of Figure 3 shows results from predicting 1-year ahead stock returns. Unlike

[11] Calibrated parameter values are reported in Appendix Table A.2.



(a) 1-year **(b)** 5-year

Figure 3: Predictive Regressions with Noisy Consumption. This figure shows the R^2 from long-horizon predictive regressions of model-implied log excess stock returns on noisy surplus consumption ratio (orange line), the consumption gap (blue line), or cyclical consumption (green line) as a function of the degree of measurement error ρ.

the predictive regression results without measurement error reported in Table 1, the surplus consumption ratio is a weaker equity premium predictor than the consumption gap when consumption is mismeasured, regardless of the level of measurement error. The ability of the surplus consumption ratio to predict returns appears to be particularly affected by the degree of mismeasurement. The severity of this issue is significantly amplified for longer horizons, as shown in Panel (b), which examines 5-year ahead returns. Furthermore, Figure 3 demonstrates that $\widetilde{cgap}_t$ is a remarkably stable excess return predictor for different degrees of mismeasuremet, compared to both $\widetilde{s}_t$ and $\widetilde{cc}_t$. Most interestingly, the predictive power of $\widetilde{cgap}_t$ is very similar to that of the true consumption gap reported in Table 1. This follows from the fact that, as true and noisy consumption levels are cointegrated, $\widetilde{cgap}_t$ and $cgap_t$ are highly correlated (at about 98% and 90% for $\rho = 0.10$ and 0.50, respectively), thus exhibiting similar return predictability.[12]

Overall, these results suggest that although surplus consumption is a powerful return predictor when consumption growth is i.i.d., the presence of even a small amount of measurement error in consumption substantially weakens the ability of this variable to forecast the equity premium.[13] Intuitively, the econometrician is unable to observe true consumption growth and, thus, unable to construct the proper surplus consumption ratio which is the state variable for expected returns. A novel measure emerges—the consumption gap—which appears to be robust to the severity of mismeasurement.

This section motivates the use of the consumption gap as a return predictor using a

[12] Appendix Figure A.1 shows results for different time horizons, while Appendix Figure A.2 demonstrates that the same results hold when using the dividend claim. Furthermore, Wachter (2005) finds that Campbell and Cochrane's numerical solution exhibits inaccuracies that bias the model's implications for unconditional asset pricing, while Schreindorfer (2023) shows that these inaccuracies extend to conditional asset price dynamics. To address potential concerns, Appendix Figure A.3 uses the numerically accurate solution based on a fine grid for the surplus consumption ratio proposed by Schreindorfer (2023) and shows that, while long-horizon predictability is marginally lower, the results remain qualitatively similar to those obtained using the original Campbell-Cochrane solution based on a sparse grid. I thank David Schreindorfer for making his code available on his webpage.

[13] While I follow Campbell and Cochrane (1999) in assuming that true consumption is i.i.d., some economic models—such as Kaltenbrunner and Lochstoer (2010); Blanchard, L'Huillier, and Lorenzoni (2013)—endogenously generate serially correlated consumption. Appendix Figure A.4 replicates the main simulation results from Figure 3, allowing for AR(1) consumption growth dynamics in the spirit of Li (2001); Wachter (2001); Campbell, Pflueger, and Viceira (2020), and shows that the results remain qualitatively unchanged. This finding generalizes the results presented in this section, indicating that *cgap* is robust to alternative assumptions about the true underlying data-generating process.

tractable specification for the measurement error process. Of course, the observed consumption dynamics can be characterized by more complexity. The rest of the paper documents empirically that the consumption gap i) represents a meaningful equity premium predictor and ii) explains a substantially larger portion of return variation compared to alternative macro-variables, especially at long horizons.

3 Time-Varying Equity Premium and the Consumption Gap

This section examines the time series properties of the consumption gap and its forecasting performance in a univariate setting. Specifically, I begin by identifying the consumption gap using consumption data and investigating its dynamics. Then, I show that the consumption gap is a robust predictor of the equity premium, both in-sample and out-of-sample, across various forecasting horizons ranging from 1 quarter to 5 years. Finally, I assess the economic magnitude of return predictability for a mean-variance investor who utilizes the consumption gap to time the market portfolio, and address potential measurement and sample issues.

3.1 Measuring the Consumption Gap

Following previous literature, I use aggregate quarterly seasonally adjusted consumption expenditures on non-durables and services from the National Income and Product Accounts (NIPA) constructed by the U.S. Bureau of Economic Analysis (BEA). The sample period ranges from 1947Q1 to 2022Q4; data are in real-per-capita terms, in chained 2017 U.S. dollars.[14] I compute the consumption gap (*cgap*) using a one-sided HP filter as in equation (9) to avoid any look-ahead bias. Specifically, I use a 10-year burn-in sample and compute the HP trend on an expanding window from 1957Q1 to the end of the sample.

Figure 4 shows the consumption gap over time for the period 1957–2022. The *cgap* has an unconditional average of -0.85% and a standard deviation of 2.69%. Fitting an AR(1) for

[14] I downloaded the two series from FRED on 6/24/2024, with codes A796RX0Q048SBEA and A797RX0Q048SBEA for non-durables and services, respectively.

the consumption gap yields a first-order autocorrelation parameter of 0.947, which implies an half-life of about three years ($\log(0.5)/\log(0.947) = 12.7$ quarters). Importantly, this is consistent with persistent expected stock returns as emphasized in, e.g., Campbell and Cochrane (1999); Pástor and Stambaugh (2000); Lettau and Ludvigson (2001a).

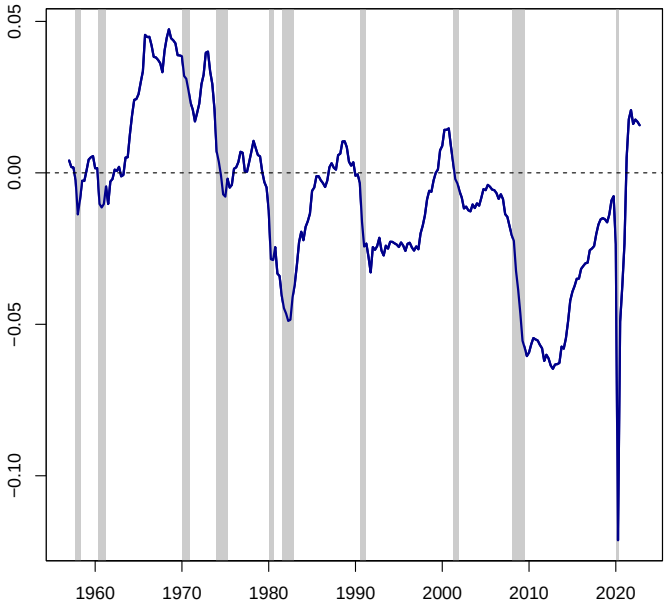


Figure 4: The Consumption Gap. This figure shows the consumption gap computed as in equation (9). Consumption data are NIPA non-durables and services seasonally-adjusted consumption expenditures for the period 1947–2022. Shaded areas are NBER recessions. Quarterly observations.

The consumption gap displays evident business cycle fluctuations: It reaches its highest values around the onset of recessions, drops throughout economic downturns, and rises after recessions, i.e. the *cgap* is procyclical. For instance, the correlation of *cgap* with real GDP and industrial production annual growth is, respectively, 44% and 35%. The consumption gap is negative for the majority of the sample—about two-thirds of the quarters—implying a positive equity premium on average (c.f. Table 1), in line with the empirical evidence.

3.2 Predicting Stock Returns using the Consumption Gap

I begin by studying the time series predictability properties of the consumption gap by running the univariate regression:

$$r_{m,t+h} = \gamma_0 + \gamma_1 cgap_t + \epsilon_{t+h} \ , \tag{10}$$

where $r_{m,t+h}$ are log market returns in excess of the 1-month Treasury bill rate compounded in the period $t : t + h$. Throughout the paper, I often refer to r_m as equity premium, excess market returns, or simply market returns.[15]

Table 2 reports results for the predictive regression (10) for different horizons from one-quarter ($h = 1$) to five years ($h = 20$). For $h > 1$, to account for potential inference issues related to overlapping observations, I follow Ang and Bekaert (2007) and use overlapping observations-corrected standard errors as in Hodrick (1992).

Panel A shows the in-sample results. Coefficient estimates increase linearly with the horizon, and the R^2 starts at about 3% and then rise almost to 50% at 5-year horizon, displaying the classic pattern documented in, e.g., Campbell and Shiller (1988) and Fama and French (1989). The predictive power of the consumption gap is both statistically and economically significant. For quarterly returns, the estimate for γ_1 reported in Table 2 is about -0.6, implying that a one-standard-deviation decrease in *cgap* is associated with an increase in expected returns of about 1.6%. For 5-year returns, a one-standard-deviation decrease leads to an increase in expected returns of about 23.3%. Most importantly, the predictive coefficients are negative across the different aggregation periods, suggesting that the *cgap* captures countercyclical variation in expected returns, consistent with the Campbell and Cochrane (1999) model. Periods when the consumption gap is low are, on average, bad times when marginal utility is high, and individuals require larger compensation for bearing aggregate risks, leading to a higher equity premium.

A common concern is that small-sample bias in long-horizon forecasting regressions can produce spurious predictability. To address this concern, I compute and report in square brackets the bias-adjusted coefficients using the analytical formula in Boudoukh, Israel, and

[15] Stock returns are value-weight return of Center for Research in Security Prices (CRSP) firms incorporated in the U.S. and listed on the NYSE, AMEX, or NASDAQ.

Table 2: Predicting Market Returns

This table reports OLS estimates and the R^2 from the predictive regression (10). Panel A shows the in-sample results for different aggregation periods. Panel B shows the out-of-sample results. In this case, I estimate the permanent component τ to compute *cgap* recursively from the latest available data using a HP filter (i.e., a one-sided HP filter). The R^2_{OOS} is computed as in Campbell and Thompson (2008); p-values for R^2_{OOS} are computed as in Clark and West (2007). I employ an expanding window for estimating the predictive regressions, using the first half of the sample as a training sample. The in-sample period starts in 1957Q1 and ends in 1989Q4, 1994Q4, and 1999Q4. In square brackets, I report the long-horizon bias-adjusted coefficients computed as in Boudoukh, Israel, and Richardson (2022), while in curly brackets, I report the bias-adjusted R^2. Values in parenthesis are heteroskedasticity and autocorrelation consistent (HAC) standard errors using Newey and West (1987) with automatic bandwidth selection procedure as described in Newey and West (1994); for $h > 1$, I use overlapping observations-corrected standard errors as in Hodrick (1992). ***, **, and * indicates respectively 1%, 5%, and 10% level of significance. Quarterly observations. The sample period is 1957Q1 to 2022Q4.

Panel A: In-Sample

	$r_{m,t+1}$	$r_{m,t+4}$	$r_{m,t+8}$	$r_{m,t+20}$
$cgap_t$	-0.588^{***}	-2.253^{***}	-3.611^{**}	-8.667^{**}
	(0.156)	(0.756)	(1.494)	(3.407)
	[−0.582]	[−2.015]	[−3.030]	[−7.257]
Constant	0.008^{*}	0.034	0.077^{*}	0.186
	(0.005)	(0.023)	(0.047)	(0.115)
Observations	263	260	256	244
R^2	0.032	0.123	0.182	0.478
	{0.032}	{0.121}	{0.176}	{0.464}

Panel B: Out-Of-Sample

R^2_{OOS}	From 1990	From 1995	From 2000
$r_{m,t+1}$	4.33^{***}	4.74^{***}	3.93^{***}
$r_{m,t+4}$	16.60^{***}	16.81^{***}	13.70^{***}
$r_{m,t+8}$	20.37^{***}	18.80^{***}	14.34^{***}
$r_{m,t+20}$	44.15^{***}	35.75^{***}	53.26^{***}

Richardson (2022), who extend the classic Stambaugh (1999) setting. The bias for the consumption gap is not severe. For example, at annual horizon, the coefficients (in absolute value) decrease by about 10%. This can be compared to a reduction of, e.g., about 60% (from 0.1 to 0.04, c.f. Figure 1 in Boudoukh, Israel, and Richardson, 2022) for the dividend-to-price ratio. This result is not surprising as the bias depends on the autocorrelation of the predictor, and *cgap* is significantly less persistent than the *dp* ratio, with a AR(1) parameter of about 0.94 versus 0.98. Importantly, all predictive coefficients remain highly statistically significant, with $t-$stats above two across different horizons. Additionally, in curly brackets, I report the bias-adjusted R^2 and show that the return variation explained by *cgap* remains largely unaffected by small-sample bias.

Then, I test the ability of the consumption gap to predict returns out-of-sample (OOS). The OOS exercise is particularly important because if the consumption gap would not perform well OOS, it would be of little use to long-term investors who want to time the market in real-time. Furthermore, the OOS analysis is a robust way to address concerns about spurious predictability (e.g., Welch and Goyal, 2008). I test for out-of-sample predictability using the metric introduced by Campbell and Thompson (2008). Specifically, I compute the R^2_{OOS}

$$R^2_{OOS} = 1 - \frac{\sum_{t=1}^{T} \left(r_{m,t+h} - \widehat{r}_{m,t+h}\right)^2}{\sum_{t=1}^{T} \left(r_{m,t+h} - \bar{r}_{m,t+h}\right)^2}$$

where $\widehat{r}_{m,t+h}$ is the fitted value from my predictive regression estimated through period $t-1$ and $\bar{r}_{m,t+h}$ is the historical average return estimated thorough period $t-1$. If the R^2_{OOS} is positive, then the predictive regression has lower average mean squared prediction error than the historical average return.

Panel B in Table 2 shows results for the out-of-sample tests. I find that the consumption gap significantly predicts expected excess market returns for different aggregation horizons ranging from one quarter to five years. This result is not associated to any particular sample, as the R^2_{OOS} is positive and significant regardless the OOS period starts in 1990, 1995, or 2000. Furthermore, Appendix Figure B.1 plots the OOS predictive performance of the consumption gap benchmarked against the historical mean over time. Panel (a) displays

results for an aggregation horizon of one quarter, while Panel (b) displays results for 5-year returns. The graphs show the cumulative squared prediction errors of the historical mean minus the cumulative squared prediction error of the predictive regression using the *cgap*. Whenever a line increases, it means that the *cgap* predicts better; whenever it decreases, the historical mean predicts better. This analysis clearly shows that the consumption gap is a powerful predictor, displaying superior performance compared to the historical mean for most periods over more than 30 years.

Overall, the empirical evidence shows that the consumption gap is a strong, robust predictor for the U.S. equity premium. These findings are particularly promising when compared to those recently reported by Welch, Goyal, and Zafirov (2024), who document disappointing performance for most equity premium predictors, with weak in-sample predictability and negative R^2_{OOS}.

3.3 The Economic Value of Market Timing

I quantify the economic gains for a *real-time* investor who uses the consumption gap to predict the market portfolio. Specifically, I consider a mean-variance investor who faces the following objective function at the end of quarter t:

$$\arg\max_{w_{t+1|t}} w_{t+1|t}\hat{r}_{m,t+1|t} - \frac{1}{2}\gamma w^2_{t+1|t}\hat{\sigma}^2_{m,t+1|t} \ , \tag{11}$$

where $w_{t+1|t}$ is the allocation to the market portfolio in period $t+1$ given its forecast at time t, $\hat{r}_{m,t+1|t}$ is the investor's equity premium forecast, $\hat{r}_{m,t+1|t}$ is the investor's forecast of the variance of the market excess return, and γ represents the coefficient of relative risk aversion. Following the literature (e.g., Campbell and Thompson, 2008; Rapach, Strauss, and Zhou, 2010; Dong, Li, Rapach, and Zhou, 2022), I set $\gamma = 3$. I also assume that the investor uses the sample variance computed over a 20-quarter (i.e., five years) rolling estimation window to predict the variance. From (11), one can solve for the optimal weights:

$$w^*_{t+1|t} = \frac{1}{\gamma}\frac{\hat{r}_{m,t+1|t}}{\hat{\sigma}^2_{m,t+1|t}} \ . \tag{12}$$

To keep the equity allocation within reasonable ranges, I impose a maximum leverage constraint of 100% and exclude short sales, i.e., $0 \leq w_{t+1|t} \leq 2$.[16]

The average utility realized by the investor is given by

$$\overline{U}_j = \hat{r}_j - \frac{1}{2}\gamma\hat{\sigma}_j^2 \ , \quad \text{for } j = 0, 1,$$

where a subscript of 0 or 1 indicates the mean and variance for the portfolio return over the *out-of-sample* period when the investor uses, respectively, the historical mean or the consumption gap to predict $r_{m,t+1}$. I use half of the sample (from 1957Q1 to 1989Q4) as a training sample for the predictive regressions, thus the out-of-sample period is 1990Q1–2022Q4. Finally, I compute the average utility gain (or increase in certainty equivalent return) when the investor uses the competing forecast in lieu of the historical mean benchmark as $\Delta\overline{U} = \overline{U}_1 - \overline{U}_0$.

The average utility gain has the same unit of measure as returns and, multiplied by four, can be interpreted as the annualized portfolio management fee (as a proportion of wealth) that the investor would be willing to pay to have access to the information in the competing forecast relative to that in the historical mean benchmark. I find that the annualized utility gain reaches 4.27% for the market timing, which is well above the 2% threshold for economic significance discussed in Pástor and Stambaugh (2000).

Such a real-time market timing strategy generates an annualized performance of 10.5% with a $t-$stat of 3.4 and an associated out-of-sample Sharpe Ratio of 0.60, which is about 58% higher than the unconditional Sharpe Ratio of the market portfolio for the 1990–2022 period. This outperformance is of similar magnitude to that documented in Gu, Kelly, and Xiu (2020), who use state-of-the-art machine learning methods to predict the equity premium. To better understand how this market timing strategy performs over time, Figure 5 plots the cumulative gains of a buy-and-hold strategy that invests in the market portfolio, alongside the performance of the market timing strategy based on both the historical mean and the consumption gap. To ensure that the buy-and-hold strategy has the same average exposure to stocks compared to the dynamic strategy, I scale the buy-and-hold return by the average portfolio weight of the market timing strategy (see Appendix Figure B.2). The

[16] Appendix Figure B.2 shows the weights over time for the market timing strategy that uses the *cgap*.

managed portfolio that utilizes information from the *cgap* significantly outperforms both the portfolio based on the historical mean benchmark forecast and the excess market portfolio.

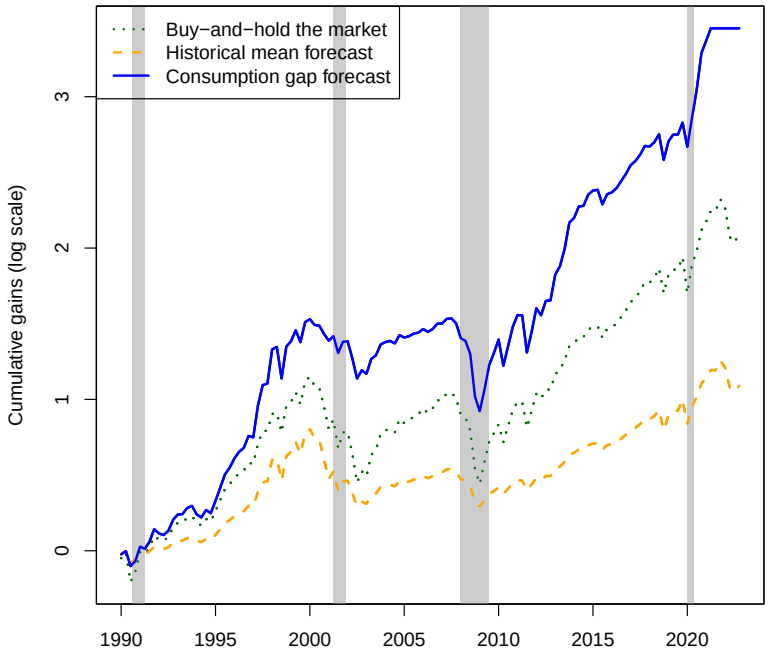


Figure 5: Market Timing Performance Over Time. This figure shows the log cumulative excess return for a portfolio constructed using the market excess return forecast with the *cgap* (blue solid line) and the historical mean (orange dashed line), as well as the log cumulative excess return for the CRSP value-weighted market portfolio (dark green dotted line). Shaded areas are NBER recessions. Quarterly observations. The sample period is 1957Q1 to 2022Q4; the sample period 1957Q1–1989Q4 is used as a training sample for the predictive regressions.

To further support that the *cgap* effectively times the market, rather than increasing risk exposure, I compute the alphas from regressing the market timing strategy returns on several prominent factor models. Specifically, Table 3 reports alphas from the regression

$$r_t^{cgap} = \alpha^{cgap} + \beta X_t + \varepsilon_t \ ,$$

where r_t^{cgap} is the performance resulting from the market timing strategy which uses the consumption gap to predict the market and X_t is a matrix containing factor returns for different characteristics-based factor models. In particular, I consider the CAPM, the (Fama

and French, 1993) three-factor model, the Carhart (1997) four-factor model, the (Fama and French, 2015) five-factor model, the q-factor model proposed by Hou, Xue, and Zhang (2015); Hou, Mo, Xue, and Zhang (2018), and the CAPM plus "betting against the beta" factor proposed by Frazzini and Pedersen (2014) or a volatility-managed market portfolio constructed as in Moreira and Muir (2017). I find that, while about 80% of total variation in the market timing strategy is explained by the market portfolio itself, the market timing strategy generates a positive and significant alpha which ranges between 3.76 to 4.50% per year across different factor models, with $t-$stats above two.

Table 3: Disconnect Between Market Timing Performance and Other Factors

This table reports the results from regressing returns generated using the market timing strategy which uses the consumption gap to predict the market (r^{cgap}) on several characteristics-based risk factors. Column (1) is the CAPM. Column (2) is the Fama and French (1993) three-factor model (FF3). Column (3) is the Carhart (1997) four-factor model. Column (4) is the Fama and French (2015) five-factor model (FF5). Column (5) is the Hou, Xue, and Zhang (2015); Hou et al. (2018) five-factor model (q5). Column (6) is CAPM plus "betting against the beta" (BAB) factor proposed by Frazzini and Pedersen (2014). Column (7) is CAPM plus the volatility-managed market portfolio (VMP) constructed as in Moreira and Muir (2017). Alphas are annualized and in percentage. Values in parenthesis are heteroskedasticity and autocorrelation consistent (HAC) standard errors using Newey and West (1987) with automatic bandwidth selection procedure as described in Newey and West (1994). ***, **, and * indicates respectively 1%, 5%, and 10% level of significance. Quarterly observations. The sample period is 1990Q1 to 2022Q4.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
α^{cgap}	4.46***	4.33***	3.86**	4.36**	4.50**	3.76**	4.44***
	(1.56)	(1.52)	(1.61)	(1.85)	(2.19)	(1.50)	(1.56)
Observations	132	132	132	132	132	132	132
R^2	0.82	0.83	0.83	0.83	0.84	0.83	0.82

3.4 Addressing Potential Measurement and Sample Issues

(a) Vintage Data

As I use revised NIPA consumption data to compute the consumption gap, a natural question is whether using only vintage (i.e. unrevised) data for these series would affect the forecasting performance of the consumption gap. Thus, I construct the *cgap* using vintage data for aggregate consumption expenditures on non-durables and services from the Real-Time Data

Set for Macroeconomists provided by the Federal Reserve Bank of Philadelphia.[17] Vintage data for consumption are available from 1965Q3.

Table B.2 shows OLS estimates and the R^2 from the predictive regressions. Panel A reports the in-sample results, while Panel B reports results for the out-of-sample forecasting regressions. The in-sample tests show that the coefficients on the consumption gap are negative for different aggregation horizons, confirming the economic mechanism behind the predictability generated by the consumption gap. Furthermore, the $t-$stat is well above 3 for quarterly returns with an associated R^2 of 4.7%. Return predictability increases at longer horizons, consistent with what I find using revised consumption data.

Finally, Panel B shows the out-of-sample results and documents that the forecasts of market returns based on the consumption gap are more accurate than forecasts based on the historical mean. This result holds for different aggregation horizons and different subsamples, confirming that the *cgap* is a valuable return predictor for a real-time investor.

(b) Return Predictability Excluding COVID-19

Figure 4 shows that consumption expenditures fell dramatically below their trend during the COVID-19 pandemic—2020Q2. Is the consumption gap predictable power generated by this large negative consumption shock? I address this question by replicating results in Table 2 excluding the pandemic year. As the World Health Organization (WHO) declares COVID-19 a pandemic in March 2020, I use as ending sample date 2019Q4.

Appendix Table B.3 reports results for the predictive regression (10) for different horizons from $h = 1$ to $h = 20$ (i.e. five years). Results—both in-sample and out-of-sample—are qualitatively and quantitatively similar to the results reported in Table 2, suggesting that the documented predictability is not an artifact of the large consumption drop in 2020Q2.

(c) Subsamples Analysis

Welch and Goyal (2008) find that the forecasting performance of many predictors is substantially weaker after the oil price crisis in the mid-1970s. If this observation applies to

[17] Data are available at FRB Philadephia.

the *cgap*, it could impair the usefulness of the proposed predictor going forward. Table B.4 addresses this concern by investigating the predictability over different subsamples starting in 1980, 1990, and 2000. The results are very similar (and even better for the late sample in Panel C) to those reported in Panel A of Table 2. This evidence is complementary the out-of-sample results reported in Panel B of Table 2. Overall, the predictability documented does not appear to be generated by a specific subsample.

(d) Macroeconomic Data Release Lag

Macroeconomic data are typically released with a lag. Specifically, consumption data for quarter t are available in the middle of quarter $t+1$. Thus, to address potential concerns that this lag generates predictability in my sample, I replicate the results documented in Table 2 using the second lag of *cgap* when predicting stock returns. That is, stock return forecasts at time $t+h$ are based on the consumption gap in time $t-1$. Appendix Table B.5 reports results for these predictive regressions. Both the in-sample and out-of-sample analyses deliver similar conclusions when compared to results reported in Table 2, suggesting that the lag in releasing consumption data is not driving the results.

(e) Alternative Consumption Trends

A key result in real business cycle models is that output and consumption share the same stochastic trend (see, e.g., King, Plosser, Stock, and Watson, 1991). Recently, Campbell, Pflueger, and Viceira (2020) used this result within a habit-based economy with consumption and output to study stock-bond correlation dynamics. Following Campbell, Pflueger, and Viceira (2020), I construct their measure of stochastically detrended consumption ($\hat{c}$) by detrending consumption using an exponentially declining moving average.

Appendix Table B.6 reports results from predictive regressions using $\hat{c}$. Panel A reports in-sample test results, while Panel B shows out-of-sample test results. The findings are consistent with those in Table 2. In particular, $\hat{c}$ significantly predicts excess returns in-sample with a negative sign, with an R^2 ranging from approximately 3% to nearly 44% at a 5-year horizon. This predictability persists out-of-sample across different aggregation horizons and sample periods.

These findings are important as they further support the economic mechanism underlying return predictability: when consumption is above its trend, marginal utility is low, leading to lower future equity premia. At the same time, Table 2 shows that the predictability associated with *cgap* is consistently stronger than that of $\hat{c}$. Using the mean-variance framework described in Section 3.3, I find that a real-time investor would be willing to pay an additional 106 basis points per year to access the information in *cgap* rather than in $\hat{c}$. This result indicates that *cgap* is the relevant detrended consumption measure in the data.

Finally, Leser (1961) shows that as the HP smoothing parameter $\lambda \to \infty$, the fitted trend becomes nearly linear. Thus, given that the simulations select $\lambda^* = 10^6$, a reasonable question is whether I am effectively using a deterministic time trend. First, the asymptotic result in Leser (1961) pertains to a *two-sided* HP filter, whereas I use a real-time consumption trend estimate, i.e., a one-sided filter. Second, Appendix Table B.7 presents predictability results using a deterministic time trend to de-trend consumption levels. The predictive performance is poor, characterized by low statistical power, low in-sample R^2, and predominantly negative R^2_{OOS}.

4 The Consumption Gap and Other Equity Premium Predictors

This section examines the ability of the consumption gap to predict returns in a multivariate setting. I begin by studying the predictive performance of the consumption gap when compared to alternative empirical proxies for the surplus consumption ratio. Next, I investigate whether the predictability of the consumption gap remains robust after accounting for other common return predictors proposed in the literature, including valuation ratios, sentiment, and other macroeconomic predictors.

4.1 The Consumption Gap and Alternative Habit-Based Measures

It is valuable to show that the consumption gap proposed in this paper predicts the equity premium at least as well as alternative empirical proxies for surplus consumption. This analysis is important because the *cgap* should identify the relevant habit-based measures given the dynamics of NIPA consumption data.

I start by running the predictive multivariate regression:

$$r_{m,t+1} = \gamma_0 + \gamma_1 cgap_t + \gamma_2 x_t + \epsilon_{t+1} \ , \tag{13}$$

where $r_{m,t+h}$ are market excess returns at time $t+1$, *cgap* is the consumption gap computed as in equation (9), and x is an alternative surplus consumption ratio proxy. I consider four different measures proposed in the literature. First, I use the empirical proxy for the surplus consumption ratio filtered from the Campbell and Cochrane (1999) constructed in Cochrane (2011)'s Presidential Address, denoted as s_t.Second, I use the weighted average of past consumption growth as constructed in Wachter (2006), which I denote as s_t^W. Then, I include cyclical consumption (cc_t), recently introduced by Atanasov, Møller, and Priestley (2020). Finally, I follow Bandi and Tamoni (2023), who construct a proxy for the business-cycle component of consumption as the 4-year aggregate of consumption growth, which I denote as bc_t. Since Bandi and Tamoni (2023) show that bc_t is strongly related to both the surplus consumption ratio and cyclical consumption, I include it as an additional measure. As expected, the four variables are highly correlated: the $cgap_t$ has a correlation of 65.5%, 73.3%, 75.5%, and 90.6%, with s_t, s_t^W, cc_t, and bc_t, respectively.

Table 4 reports the results. For ease of interpretation, I standardize all predictors to have unit standard deviations in the full sample. Columns (1) to (4) investigates the predictability of the alternative variables in a univariate setting. All variables have a negative predictive coefficients, in line with a habit interpretation, and are highly statistical significant, suggesting that they are meaningful equity premium predictors. However, the portion of return variation explained by these variables in the period 1957–2022 is substantially lower than that explained by *cgap*, that has a R^2 of 3.2% (c.f. Table 2). Furthermore, Columns (5) to (8) show that once I include both the consumption gap and one of the alternative macro

Table 4: The Consumption Gap and Other Habit-Based Measures: IS Predictability

This table reports OLS estimates and the R^2 from the multivariate predictive regression (13). The alternative habit-based measures includes: the surplus consumption ratio constructed in Cochrane (2011) (s_t), the surplus consumption ratio constructed in Wachter (2006) (s_t^W), cyclical consumption constructed in Atanasov, Møller, and Priestley (2020) (cc_t), and business-cycle consumption constructed in Bandi and Tamoni (2023) (bc_t). In Columns (1) to (4), I study the univariate predictive regressions, i.e., $\gamma_1 = 0$. Values in parenthesis are heteroskedasticity and autocorrelation consistent (HAC) standard errors using Newey and West (1987) with automatic bandwidth selection procedure as described in Newey and West (1994). ***, **, and * indicates respectively 1%, 5%, and 10% level of significance. Regressors are de-meaned and standardized. Quarterly observations. The sample period is 1957Q1 to 2022Q4.

	$r_{m,t+1}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$cgap_t$					-0.014^{**}	-0.014^{**}	-0.016^{**}	-0.025^{**}
					(0.006)	(0.007)	(0.007)	(0.010)
s_t	-0.012^{***}				-0.003			
	(0.004)				(0.004)			
s_t^W		-0.013^{***}				-0.003		
		(0.005)				(0.007)		
cc_t			-0.012^{***}				0.000	
			(0.004)				(0.007)	
bc_t				-0.013^{***}				0.010
				(0.004)				(0.010)
Observations	263	262	263	263	263	262	263	263
Adjusted R^2	0.015	0.018	0.014	0.017	0.026	0.026	0.025	0.027

predictor, the *cgap* appears to subsume the return predictive power, with $t-$stats above two.

While it is still useful to understand how well different variables fit the data in-sample, out-of-sample tests are ultimately the most relevant, as they provide the most rigorous evidence on stock return predictability (e.g., Welch and Goyal, 2008; Martin and Nagel, 2021; Dong, Li, Rapach, and Zhou, 2022). Table 5 shows results for two out-of-sample tests. First, Panel A reports the improvements in R^2_{OOS} when using *cgap* as a return predictor benchmarked against each of the four alternative measures. A positive value indicates that the *cgap* predicts market returns better than the alternative predictor. Panel A of Table 5 shows that the consumption gap is a consistently superior equity predictor across all variables and virtually all aggregation horizons, from one quarter to five years. Finally, following Campbell and Thompson (2008), I quantify the economic magnitude of return predictability with a market timing strategy (c.f. Section 3.3) that uses each of the variables as the equity premium predictor. Panel B in Table 5 reports the average utility gains (in %). A mean-variance investor would be willing to pay an *additional* 118 to 378 basis points in annualized portfolio management fees to access the equity premium forecasts based on the information in the *cgap*.

Table 5: The Consumption Gap and Other Habit-Based Measures: OOS Tests

Panel A reports the improvements in R^2_{OOS} when using *cgap* as a return predictor benchmarked against each of the three alternative habit-based measures. Panel B reports average utility gains (in %) when using each of the variables. I employ an expanding window for estimating the predictive regressions, using the first half of the sample as a training sample. The sample period is 1957Q1 to 2022Q4; the in-sample period starts in 1957Q1 and ends in 1989Q4..

Panel A: R^2_{OOS} Improvements

$R^2_{OOS^{cgap}}$	s_t	s^W_t	cc_t	bc_t
$r_{m,t+1}$	4.42^{***}	3.99^{***}	1.32^{***}	1.34^{***}
$r_{m,t+4}$	15.35^{***}	14.84^{***}	7.15^{***}	3.75^{***}
$r_{m,t+8}$	24.41^{***}	17.33^{***}	13.06^{***}	-0.89
$r_{m,t+20}$	56.16^{***}	38.53^{***}	17.94^{***}	7.39^{***}

Panel B: Economic Gains

	$cgap_t$	s_t	s^W_t	cc_t	bc_t
$\Delta\overline{U}$	4.27	0.91	0.49	3.09	2.87

Overall, this evidence demonstrates that the consumption gap provides substantial additional predictive content for returns, suggesting that it effectively identifies consumption deviations from habit in NIPA consumption data.

4.2 Additional Return Predictors

(a) The Goyal-Welch Dataset

Is the return predictability of the consumption gap captured by standard predictors proposed in the literature? To study this question, I run the predictive multivariate regression (13) controlling for a comprehensive set of control variables. Specifically, I consider the original sixteen variables from Welch and Goyal (2008) available at a quarterly frequency. From this set, I exclude all regressors that do not display significant return predictability in my sample period. The final set of control variables includes: the dividend-price ratio (dp), the dividend-payout ratio (de), long-term government bond returns (ltr), and the investment-to-capital ratio (ik).[18]

Table 6 reports results for regression (13). In Columns (1) to (4), I consider the case $\gamma_1 = 0$ to study the in-sample predictability for the alternative predictors individually. Market returns are predictable, with an R^2 ranging from 0.2% to 2.3% using quarterly data; the investment-to-capital ratio displays the stronger predictive performance, in line with with results in Welch, Goyal, and Zafirov (2024). In Column (5) to (8), I include the consumption gap as an additional predictor. The $cgap$ is always statistically significant at 5% level of confidence. Furthermore, controlling for the consumption gap, the evidence of predictability for the other predictors is typically weaker, suggesting that $cgap$ captures most of the time variation in returns.

(b) The Output Gap

Cooper and Priestley (2009) find that the output gap ($ygap$)—a key production-based macro

[18] I use the repurchase-adjusted dividend-price ratio constructed in Nagel and Xu (2021), as it displays superior predictive performance in my sample when compared to the standard dp ratio. I thank Zhengyang Xu for making the time series available on his website.

Table 6: The Consumption Gap and Other Market Predictors

This table reports OLS estimates and the R^2 from the multivariate predictive regression (13). Control variables includes: the (repurchase-adjusted) dividend-price ratio constructed in Nagel and Xu (2021) (dp_t), the dividend-payout ratio (de), long-term government bond returns (ltr), and the investment-to-capital ratio (ik). In Columns (1) to (4), I study the univariate predictive regressions, i.e., $\gamma_1 = 0$. Values in parenthesis are heteroskedasticity and autocorrelation consistent (HAC) standard errors using Newey and West (1987) with automatic bandwidth selection procedure as described in Newey and West (1994). ***, **, and * indicates respectively 1%, 5%, and 10% level of significance. Regressors are de-meaned and standardized. Quarterly observations. The sample period is 1957Q1 to 2022Q4.

	$r_{m,t+1}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$cgap_t$					-0.013^{***}	-0.016^{***}	-0.015^{***}	-0.011^{**}
					(0.005)	(0.004)	(0.004)	(0.005)
dp_t	0.013^{**}				0.009			
	(0.006)				(0.006)			
de_t		0.007^{*}				0.007^{**}		
		(0.004)				(0.004)		
ltr_t			0.012^{*}				0.011^{*}	
			(0.007)				(0.006)	
ik_t				-0.014^{***}				-0.009
				(0.005)				(0.006)
Observations	263	263	263	263	263	263	263	263
Adjusted R^2	0.019	0.002	0.017	0.023	0.034	0.032	0.041	0.033

variable—predicts stock returns. This result follows naturally from the fact that *cgap* predicts returns and that consumption and output share the same stochastic trend (see, e.g., King, Plosser, Stock, and Watson, 1991).[19] Since consumption is the main component of output, accounting for approximately two-thirds of total output, I examine the extent to which the predictability generated by the output gap can be attributed to its consumption component.

Specifically, I decompose the output gap into two components: one related to consumption and another orthogonal to consumption, which captures the contributions of other output components (i.e., private domestic and government expenditures, and net exports of goods and services). To do so, I estimate the following regression:

$$ygap_t = a + b \cdot cgap_t + v_t \ , \tag{14}$$

where, for consistency, I estimate *ygap* using the same one-sided HP filter applied to real GDP that I use to construct *cgap*.[20] I denote the fitted value from regression (14) as $ygap^{cons}$, representing the consumption-driven component of the output gap, while the residual component is denoted as $ygap^{\perp cons}$.

Table 7 reports results from the predictive regression (10), where I consider the two distinct components of *ygap* as separate predictors of the equity premium. Panel A shows that $ygap^{cons}$ significantly predicts the equity premium, with the same t-statistics and R^2 values as those reported in Table 2. In contrast, Panel B demonstrates that $ygap^{\perp cons}$ does not predict market returns at any aggregation horizon.

These results indicate that the predictive power of the output gap is entirely driven by its consumption component.

(c) Consumer Sentiment

The consumption gap is higher during good times and lower during bad times, suggesting a potential correlation with measures of consumer sentiment. Thus, a question is whether

[19] In Appendix Section C, I formally explore this idea using Johansen (1991)'s procedure and provide additional supporting evidence for this result.

[20] I obtain similar results when computing *ygap* using a quadratic trend as in Cooper and Priestley (2009).

Table 7: Output and Consumption Gap as Return Predictors

This table reports OLS estimates and the R^2 from the predictive regression: $r_{m,t+h} = \gamma_0 + \gamma_1 ygap^i{}_t + \epsilon_{t+h}$, where $r_{m,t+h}$ are log market excess returns compounded in the period $t : t+h$ and $ygap^i$ represents one of the two components of $ygap$ constructed using (14). Specifically, $ygap^{cons}$ corresponds to the fitted value, while $ygap^{\perp cons}$ represents the residual component. Panel A shows results when using $ygap^{cons}$ in the predicting regressions. Panel B shows results when using $ygap^{\perp cons}$ in the predicting regressions. Values in parenthesis are heteroskedasticity and autocorrelation consistent (HAC) standard errors using Newey and West (1987) with automatic bandwidth selection procedure as described in Newey and West (1994); for $h > 1$, I use overlapping observations-corrected standard errors as in Hodrick (1992). ***, **, and * indicates respectively 1%, 5%, and 10% level of significance. Quarterly observations. The sample period is 1957Q1 to 2022Q4.

Panel A: $ygap^{cons}$

	$r_{m,t+1}$	$r_{m,t+4}$	$r_{m,t+8}$	$r_{m,t+20}$
$ygap_t^{cons}$	-0.650^{***}	-2.489^{***}	-3.990^{**}	-9.578^{**}
	(0.172)	(0.835)	(1.651)	(3.765)
Constant	0.002	0.011	0.041	0.099
	(0.006)	(0.027)	(0.055)	(0.134)
Observations	263	260	256	244
R^2	0.032	0.123	0.182	0.478

Panel B: $ygap^{\perp cons}$

	$r_{m,t+1}$	$r_{m,t+4}$	$r_{m,t+8}$	$r_{m,t+20}$
$ygap_t^{\perp cons}$	-0.132	-0.303	-0.132	-2.152
	(0.268)	(1.031)	(1.895)	(4.817)
Constant	0.013^{***}	0.054^{**}	0.110^{***}	0.257^{**}
	(0.005)	(0.021)	(0.043)	(0.106)
Observations	263	260	256	244
R^2	0.001	0.001	0.000	0.017

consumer sentiment generates the documented predictability. Indeed, e.g., Lemmon and Portniaguina (2006) find that consumer sentiment predicts stock returns. Is the sentiment component of *cgap* that is driving the documented predictability?

To address this question, I disentangle two parts of the consumption gap: one part related to consumer sentiment and one part orthogonal to sentiment and related to economic fundamentals. In particular, I run the regression:

$$cgap_t = a + bS_t + v_t \ , \tag{15}$$

where S_t is a matrix containing the two survey measures of consumer sentiment: the Consumer Sentiment Index (CSI) from the University of Michigan Surveys of Consumers and the Consumer Confidence Index (CCI) from the OECD. Following Lemmon and Portniaguina (2006), I orthogonalize both CSI and CCI with respect to current and lagged fundamentals, as captured by the principal components constructed in Ludvigson and Ng (2009). This procedure isolates the component of sentiment that is not explained by fundamental factors. I denote the fitted value from regression (15) as $cgap^{sent}$, which represents the sentiment-part of the consumption gap. Similarly, the residual component represents the fundamentals-part of the consumption gap, which I denote as $cgap^{\perp sent}$.

Table 8 reports results from the predictive regression (10) where I consider the two different components of *cgap* as separate market predictors. Panel A shows that the sentiment-*cgap* does not predict market returns for any aggregation horizons. On the contrary, Panel B shows that the portion of *cgap* related to fundamentals predicts returns, with $t-$stats and R^2 similar to the ones reported in Table 2. Overall, these results suggest that the predictability generated by the consumption gap relates to economic fundamentals rather than to some behavioral element.

Table 8: The Role of Consumer Sentiment

This table reports OLS estimates and the R^2 from the predictive regression: $r_{m,t+h} = \gamma_0 + \gamma_1 cgap^i{}_t + \epsilon_{t+h}$, where $r_{m,t+h}$ are log market excess returns compounded in the period $t : t+h$ and $cgap^i$ represents one of two components of $cgap$ components using (15). Specifically, $cgap^{sent}$ corresponds to the fitted value, while $cgap^{\perp sent}$ represents the residual component. Panel A shows results when using $cgap^{sent}$ in the predicting regressions. Panel B shows results when using $cgap^{\perp sent}$ in the predicting regressions. Values in parenthesis are heteroskedasticity and autocorrelation consistent (HAC) standard errors using Newey and West (1987) with automatic bandwidth selection procedure as described in Newey and West (1994); for $h > 1$, I use overlapping observations-corrected standard errors as in Hodrick (1992). ***, **, and * indicates respectively 1%, 5%, and 10% level of significance. Quarterly observations. The sample period is 1957Q1 to 2020Q4.

Panel A: $cgap^{sent}$

	$r_{m,t+1}$	$r_{m,t+4}$	$r_{m,t+8}$	$r_{m,t+20}$
$cgap_t^{sent}$	0.431	−3.195	−3.471	−33.866
	(3.090)	(9.609)	(18.259)	(37.590)
Constant	1.742	2.660	8.048	−3.285
	(2.743)	(9.181)	(17.691)	(36.637)
Observations	250	247	243	231
R^2	0.000	0.001	0.001	0.043

Panel B: $cgap^{\perp sent}$

	$r_{m,t+1}$	$r_{m,t+4}$	$r_{m,t+8}$	$r_{m,t+20}$
$cgap_t^{\perp sent}$	−0.591***	−2.253***	−3.671**	−8.469**
	(0.155)	(0.754)	(1.484)	(3.409)
Constant	1.350***	5.375**	10.814**	26.932**
	(0.452)	(2.198)	(4.391)	(10.731)
Observations	250	247	243	231
R^2	0.034	0.127	0.193	0.476

5 The Consumption Gap as an Instrument for Conditional Models

This section illustrates the use of the consumption gap as a conditioning variable. First, I construct an empirical SDF using the consumption gap as an instrument in a conditional version of the Capital Asset Pricing Model (CAPM). Next, I examine the properties of this heuristic SDF and its relation to economic conditions.

5.1 Estimating the SDF Coefficients

Following Lettau and Ludvigson (2001b), I use a conditional version of the CAPM to study the behavior of the *cgap* as a conditioning variable. A general SDF representation for the CAPM is (Hansen and Jagannathan, 1991):

$$M_{t+1} = 1 - b_t \cdot \widetilde{R_{m,t+1}} \ , \tag{16}$$

where $R_{m,t+1}$ is the excess market return and $\widetilde{R_{m,t+1}} = R_{m,t+1} - \mathrm{E}_t[R_{m,t+1}]$.[21] In the absence of arbitrage, the SDF in equation (16) satisfies the fundamental pricing equation $0 = \mathrm{E}_t[M_{t+1}R_{i,t+1}]$ for any excess asset return $R_{i,t+1}$. b_t represents the SDF loadings and can be interpreted as the time-varying price of risk on the market portfolio.

Following, e.g., Cochrane (1996); Jagannathan and Wang (1996), I parameterize the conditional model assuming that the time-varying coefficients are a linear function of a conditioning variable z_t. This characterization allows the conditional model (16) to be specified as:

$$M_{t+1} = 1 - b_0\widetilde{R_{m,t+1}} - b_1 z_t \widetilde{R_{m,t+1}} \ . \tag{17}$$

This is a multifactor unconditional model in which the factors are the excess market return and the excess market return scaled by the instrument. The market portfolio is the only fundamental factor. The scaled excess return can be interpreted as a managed portfolio

[21] The Campbell and Cochrane (1999)'s external habit economy described in Section 2.1 implies an SDF of the form $M_{t+1} = a_t - b_t \cdot (\Delta c_{t+1} - \mathrm{E}_t[\Delta c_{t+1}])$. Appendix D.1 shows that the conditional CAPM in (16) can be derived as an approximation of this SDF specification.

which invests more aggressively in the market when z_t is high. Following, e.g., Ferson, Sarkissian, and Simin (2003); Lettau and Ludvigson (2001b), I demean the instrument z_t.

In general, $b = \Sigma_f^{-1}\mu_f$ (with an $\mathbb{E}[M] = 1$ normalization), where Σ_f and μ_f represent the variance-covariance matrix and the vector of risk premia for the factors, respectively (see, e.g., Ludvigson, 2013). This formulation implies that the SDF coefficients can be directly estimated by projecting the SDF onto the factors. To compute the standard errors of $\hat{b}$, I use the Generalized Method of Moments (GMM) framework developed in Chabi-Yo, Gonçalves, and Loudis (2024).[22] There are two key advantages to using this methodology rather than estimating SDF coefficients based on a cross-section of test assets. First, this approach yields $\hat{b}$ estimates that are asymptotically efficient, as efficient GMM disregards all test assets except for the factors themselves (see Nagel, 2013). This directly addresses the issue of test asset selection in estimating the cross-sectional model (see, e.g., Lewellen, Nagel, and Shanken, 2010; Bryzgalova, Pelger, and Zhu, 2023; Giglio, Xiu, and Zhang, 2025). Second, Chabi-Yo, Gonçalves, and Loudis (2024) demonstrates that the estimates are robust, as $\hat{b}$ converges in probability to the projection of the SDF onto the factors even if the SDF is misspecified.

As the previous analyses show that the consumption gap is a relevant conditioning variable, carrying important information about the time-varying equity premium, I estimate (17) using $z_t = cgap_t$:

$$M_{t+1} = 1 - \underset{[2.81]}{0.201} \cdot \widetilde{R_{m,t+1}} + \underset{[-2.97]}{0.188} \cdot cgap_t \widetilde{R_{m,t+1}} \ . \tag{18}$$

I report the SDF coefficient estimates with GMM-VARHAC t-statistics in square brackets below the coefficients. The *cgap* SDF coefficient is significant at the 1% level and has a negative sign, as expected. A high consumption gap is associated with low marginal utility and, consequently, low expected returns. Thus, b_1 should be negative (e.g., Barroso, Boons, and Karehnke, 2021). Moreover, its magnitude is economically meaningful. As the SDF coefficients are not directly comparable, I multiply each $\hat{b}$ by the standard deviation of its corresponding factor. The estimates in (18) show that a one-standard-deviation change in the managed portfolio leads to approximately a 0.2 change in the SDF, which is substantial given that $\mathbb{E}[M] = 1$. These results suggest that the consumption gap plays a significant

[22] For more details and complete derivations, see Appendix B of Chabi-Yo, Gonçalves, and Loudis (2024).

role in explaining the time variation in the SDF's exposure to the market portfolio—i.e., it effectively tracks the price of market risk.

Appendix Table D.1 reports estimates of $\hat{b}$ for different specifications of (17), including *cc* as an alternative conditioning variable. The results show that *cgap* remains the relevant measure, with an absolute t-statistic above 2 and an economic magnitude comparable to that in (18), even when controlling for *cc*.

5.2 Properties of the SDF and Economic Conditions

Estimating (17) enables the computation of the SDF implied by the empirical equilibrium model and facilitates the study of its properties. Specifically, I use the SDF coefficient estimates from (18) to construct the heuristic SDF as: $M_{t+1} = 1 - 0.024 \times R_{m,t+1} + 0.830 \times cgap_t R_{m,t+1}$.[23] Figure 6 shows the conditional variance for the SDF calculated using a standard GARCH$(1,1)$. The annualized conditional variance of the SDF ranges between 0.46 and 3.65, with an average of 1.00. Its yearly autocorrelation is 0.86, reflecting slow, long-term changes in economic conditions.

This empirical SDF displays properties consistent with leading macro-finance theories. First, the variance of the SDF fluctuates substantially over time. When marginal utility is more volatile, investors ask higher compensation for bearing risks, in a line with, for example, habit (Campbell and Cochrane, 1999) and heterogeneous agents models (Chan and Kogan, 2002). Furthermore, the SDF variance is related, on average, to the state of the business cycle. Indeed, the large spikes in the SDF in Figure 6 coincide with periods of intense economic and financial stress–the 1970s energy crisis, the early 1980s recession, the Great Financial Crisis 2007-2009, and the COVID-19 pandemic.

6 Conclusion

This paper explores the return predictability implications of measurement error in NIPA consumption data. I introduce a novel measure, denoted as the consumption gap (*cgap*),

[23] These values correspond to $\hat{b}_0$ and $\hat{b}_1^{cgap}$ without scaling the coefficients by the standard deviation of their respective factors.

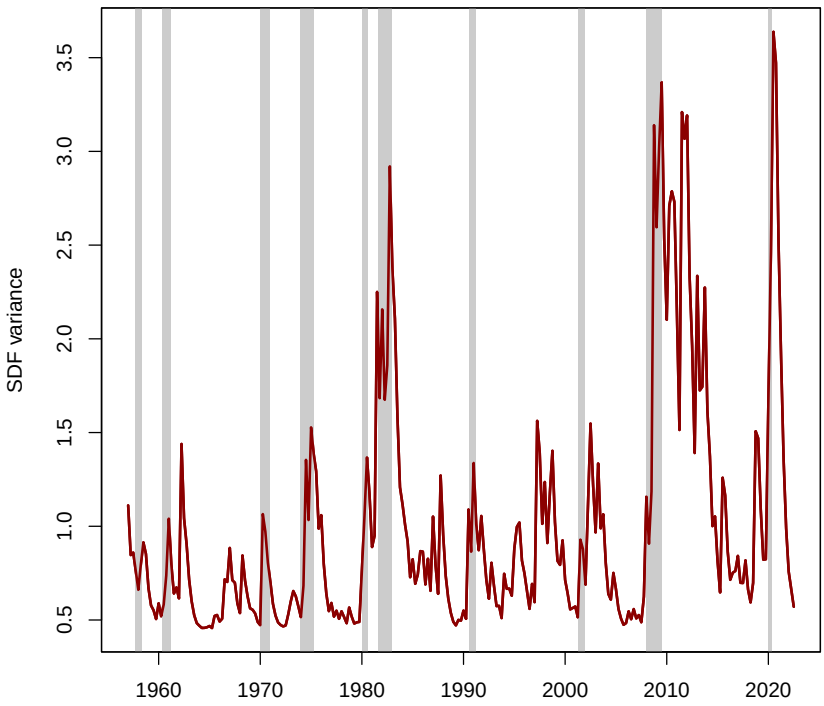


Figure 6: SDF Conditional Variance. This figure shows the annualized conditional variance of the empirical SDF obtained using the consumption gap as a conditioning variable for the market price of risk in (17). I calculate the conditional variance using a GARCH$(1,1)$ for the SDF computed as $M_{t+1} = 1 - 0.024 \times R_{m,t+1} + 0.830 \times cgap_t R_{m,t+1}$. Shaded areas are NBER recessions. Quarterly observations. The sample period is 1957Q1 to 2022Q4.

that robustly identifies consumption deviations from a time-varying subsistence level (i.e., habit) when consumption is mismeasured.

I start by showing, in a simulated Campbell and Cochrane (1999) economy, that the consumption gap predicts long-horizon returns better than the surplus consumption ratio or an alternative cyclical consumption measure when consumption is mismeasured and that its predictability is largely unaffected by the severity of the measurement error. I use these results to motivate my empirical analysis.

The empirical evidence shows that the consumption gap is a powerful equity premium predictor. At horizons from one quarter to five years, the consumption gap generates accurate estimates of countercyclical expected returns and out-of-sample forecasts. This predictability is both statistically and economically significant, and cannot be captured by existing alternative return predictors. Furthermore, in the cross-section of stocks, I use the consumption gap as an instrument to construct an empirical SDF and study its properties. I find that this heuristic SDF varies with economic conditions and is heteroskedastic, consistent with benchmark macro-finance theories.

As discussed in, e.g., Campbell and Cochrane (2000); Ludvigson (2013), empirical tests of consumption-based models must confront the fact that macroeconomic data feature a number of measurement issues. However, these models should not be dismissed. Instead, we need methodologies that can effectively address the challenges related to mismeasurement. The evidence documented in this paper is important as it clearly supports the economic prediction that consumption is crucial for understanding asset prices fluctuations. These findings indicate that developing empirical methods to address measurement and specification error in macroeconomic data is a promising direction for future research.

References

Abel, A. B. 1990. Asset prices under habit formation and catching up with the joneses. *American Economic Review: Papers & Proceedings* 80:38–42.

Ang, A., and G. Bekaert. 2007. Stock return predictability: Is it there? *Review of Financial Studies* 20:651–707.

Atanasov, V., S. V. Møller, and R. Priestley. 2020. Consumption fluctuations and expected returns. *Journal of Finance* 75:1677–713.

Bandi, F., and A. Tamoni. 2023. Business-cycle consumption risk and asset prices. *Journal of Econometrics* 237:105447–.

Barnichon, R., D. Debortoli, and C. Matthes. 2022. Understanding the size of the government spending multiplier: It's in the sign. *Review of Economic Studies* 89:87–117.

Barroso, P., M. Boons, and P. Karehnke. 2021. Time-varying state variable risk premia in the icapm. *Journal of Financial Economics* 139:428–51.

Beeler, J., and J. Y. Campbell. 2012. The long-run risks model and aggregate asset prices: An empirical assessment. *Critical Finance Review* .

Blanchard, O. J., J.-P. L'Huillier, and G. Lorenzoni. 2013. News, noise, and fluctuations: An empirical exploration. *American Economic Review* 103:3045–70.

Borup, D., and E. C. M. Schütte. 2021. Asset pricing with data revisions. *Journal of Financial Markets* 100620.

Boudoukh, J., R. Israel, and M. Richardson. 2022. Biases in long-horizon predictive regressions. *Journal of Financial Economics* 145:937–69.

Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger. 1989. Empirical tests of the consumption-oriented capm. *Journal of Finance* 44:231–62.

Bryzgalova, S., J. Huang, and C. Julliard. 2024. Consumption in asset returns. *Journal of Finance (Forthcoming)* –:–.

Bryzgalova, S., M. Pelger, and J. Zhu. 2023. Forest through the trees: Building cross-sections of stock returns. *Journal of Finance (Forthcoming)* .

Campbell, J. Y., and J. H. Cochrane. 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy* 107:205–51.

———. 2000. Explaining the poor performance of consumption-based asset pricing models. *Journal of Finance* 55:2863–78.

Campbell, J. Y., C. Pflueger, and L. M. Viceira. 2020. Macroeconomic drivers of bond and equity risks. *Journal of Political Economy* 128:3148–85.

Campbell, J. Y., and R. J. Shiller. 1988. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1:195–228.

Campbell, J. Y., and S. B. Thompson. 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21:1509–31.

Carhart, M. M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52:57–82.

Chabi-Yo, F., A. S. Gonçalves, and J. A. Loudis. 2024. An intertemporal risk factor model. *Management Science* .

Chan, Y. L., and L. Kogan. 2002. Catching up with the joneses: Heterogeneous preferences and the dynamics of asset prices. *Journal of Political Economy* 110:1255–85.

Clark, T., and K. West. 2007. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* 138:291–311.

Cochrane, J. H. 1996. A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy* 104:572–621.

———. 2011. Presidential address: Discount rates. *Journal of Finance* 66:1047–108.

Coibion, O., Y. Gorodnichenko, and M. Ulate. 2018. The cyclical sensitivity in estimates of potential output. *Brookings Papers on Economic Activity* 343.

Cooper, I., and R. Priestley. 2009. Time-varying risk premiums and the output gap. *Review of Financial Studies* 22:2801–33.

Couts, S. J., A. S. Gonçalves, and A. Rossi. 2024. Unsmoothing returns of illiquid funds. *Review of Financial Studies* hhae006.

Daniel, K., and D. Marshall. 1997. Equity-premium and risk-free-rate puzzles at long horizons. *Macroeconomic Dynamics* 1:452–84.

De Jong, R. M., and N. Sakarya. 2016. The econometrics of the hodrick-prescott filter. *Review of Economics and Statistics* 98:310–7.

Dong, X., Y. Li, D. E. Rapach, and G. Zhou. 2022. Anomalies and the expected market return. *Journal of Finance* 77:639–81.

Fama, E. F., and K. R. French. 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25:23–49.

———. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* .

———. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116:1–22.

Ferson, W. E., S. Sarkissian, and T. T. Simin. 2003. Spurious regressions in financial economics? *Journal of Finance* 58:1393–413.

Flaig, G. 2015. Why we should use high values for the smoothing parameter of the hodrick-prescott filter. *Jahrbücher für Nationalökonomie und Statistik* 235:518–38.

Frazzini, A., and L. H. Pedersen. 2014. Betting against beta. *Journal of Financial Economics* 111:1–25.

Geltner, D. M. 1991. Smoothing in appraisal-based returns. *Journal of Real Estate Finance and Economics* 4:327–45.

Giglio, S., D. Xiu, and D. Zhang. 2025. Test assets and weak factors. *The Journal of Finance* 80:259–319.

Gonçalves, A. S. 2021. Reinvestment risk and the equity term structure. *Journal of Finance* 76:2153–97.

Gu, S., B. Kelly, and D. Xiu. 2020. Empirical asset pricing via machine learning. *Review of Financial Studies* 33:2223–73.

Guo, H., and Y.-J. Pai. 2024. Dissecting nipa data revisions through habit. *Available at SSRN 3914344* .

Hamilton, J. D. 2018. Why you should never use the hodrick-prescott filter. *Review of Economics and Statistics* 100:831–43.

Hansen, L. P., and R. Jagannathan. 1991. Implications of security market data for models of dynamic economies. *Journal of Political Economy* 99:225–62.

Hansen, L. P., and K. J. Singleton. 1983. Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of Political Economy* 91:249–65.

Harvey, A., and T. Trimbur. 2008. Trend estimation and the hodrick-prescott filter. *Journal of the Japan Statistical Society* 38:41–9.

Heaton, J. 1995. An empirical investigation of asset pricing with temporally dependent preference specifications. *Econometrica: Journal of the Econometric Society* 681–717.

Hodrick, R. J. 1992. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies* 5:357–86.

———. 2020. An exploration of trend-cycle decomposition methodologies in simulated data. Working Paper 26750, National Bureau of Economic Research.

Hou, K., H. Mo, C. Xue, and L. Zhang. 2018. Which factors? *Review of Finance* 23:1–35.

Hou, K., C. Xue, and L. Zhang. 2015. Digesting anomalies: An investment approach. *Review of Financial Studies* 28:650–705.

Jagannathan, R., and Y. Wang. 2007. Lazy investors, discretionary consumption, and the cross-section of stock returns. *Journal of Finance* 62:1623–61.

Jagannathan, R., and Z. Wang. 1996. The conditional capm and the cross-section of expected returns. *Journal of Finance* 51:3–53.

Johansen, S. 1991. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica* 1551–80.

Kaiser, R., and A. Maravall. 2012. *Measuring business cycles in economic time series*, vol. 154. Springer Science & Business Media.

Kaltenbrunner, G., and L. A. Lochstoer. 2010. Long-run risk through consumption smoothing. *Review of Financial Studies* 23:3190–224.

Kamber, G., J. Morley, and B. Wong. 2024. Trend-cycle decomposition after covid. Working Paper, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy.

Kelly, B., and S. Pruitt. 2013. Market expectations in the cross-section of present values. *Journal of Finance* 68:1721–56.

King, R. G., C. I. Plosser, J. H. Stock, and M. W. Watson. 1991. Stochastic trends and economic fluctuations. *American Economic Review* 819–40.

Kroencke, T. A. 2017. Asset pricing without garbage. *Journal of Finance* 72:47–98.

Lemmon, M., and E. Portniaguina. 2006. Consumer confidence and asset prices: Some empirical evidence. *Review of Financial Studies* 19:1499–529.

Leser, C. E. V. 1961. A simple method of trend construction. *Journal of the Royal Statistical Society: Statistical Methodology* 23:91–107.

Lettau, M., and S. C. Ludvigson. 2001a. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance* 56:815–49.

———. 2001b. Resurrecting the (c) capm: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy* 109:1238–87.

Lewellen, J., S. Nagel, and J. Shanken. 2010. A skeptical appraisal of asset pricing tests. *Journal of Financial Economics* 96:175–94.

Li, Y. 2001. Expected returns and habit persistence. *Review of Financial Studies* 14:861–99.

Ludvigson, S. C. 2013. Advances in consumption-based asset pricing: Empirical tests. *Handbook of the Economics of Finance* 2:799–906.

Ludvigson, S. C., and S. Ng. 2009. Macro factors in bond risk premia. *Review of Financial Studies* 22:5027–67.

Martin, I., and S. Nagel. 2021. Market efficiency in the age of big data. *Journal of Financial Economics (Forthcoming)* –:–.

Moreira, A., and T. Muir. 2017. Volatility-managed portfolios. *Journal of Finance* 72:1611–44.

Nagel, S. 2013. Empirical cross-sectional asset pricing. *Annu. Rev. Financ. Econ.* 5:167–99.

Nagel, S., and Z. Xu. 2021. Asset Pricing with Fading Memory. *Review of Financial Studies* .

Newey, W. K., and K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–8.

———. 1994. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61:631–53.

Parker, J. A. 2001. The consumption risk of the stock market. *Brookings Papers on Economic Activity* 2001:279–333.

Parker, J. A., and C. Julliard. 2005. Consumption risk and the cross section of expected returns. *Journal of Political Economy* 113:185–222.

Pástor, L., and R. F. Stambaugh. 2000. Comparing asset pricing models: an investment perspective. *Journal of Financial Economics* 56:335–81.

Phillips, P. C., and S. Ouliaris. 1990. Asymptotic properties of residual based tests for cointegration. *Econometrica: journal of the Econometric Society* 165–93.

Ramey, V. A., and S. Zubairy. 2018. Government spending multipliers in good times and in bad: evidence from us historical data. *Journal of Political Economy* 126:850–901.

Rapach, D. E., J. K. Strauss, and G. Zhou. 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies* 23:821–62.

Savov, A. 2011. Asset pricing with garbage. *Journal of Finance* 66:177–201.

Schorfheide, F., D. Song, and A. Yaron. 2018. Identifying long-run risks: A bayesian mixed-frequency approach. *Econometrica* 86:617–54.

Schreindorfer, D. 2023. By force of habit and cyclical leverage. *Available at SSRN 4532711* .

Stambaugh, R. F. 1999. Predictive regressions. *Journal of Financial Economics* 54:375–421.

Wachter, J. A. 2001. Habit formation and returns on bonds and stocks. *Available at SSRN 263031* .

———. 2005. Solving models with external habit. *Finance Research Letters* 2:210–26.

———. 2006. A consumption-based model of the term structure of interest rates. *Journal of Financial Economics* 79:365–99.

Welch, I., and A. Goyal. 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21:1455–508.

Welch, I., A. Goyal, and A. Zafirov. 2024. A comprehensive 2022 look at the empirical performance of equity premium prediction. *Review of Financial Studies (Forthcoming)* –:–.

Wilcox, D. W. 1992. The construction of us consumption data: some facts and their implications for empirical work. *American Economic Review* 922–41.

Internet Appendix

"Time-Varying Equity Premium with Noisy Consumption"

By Alessandro Melone

This Internet Appendix is organized as follows. Section A presents additional results from the simulated external habit economy. Section B provides further findings that supplement the main results in the paper. Section C describes a cointegration analysis of non-durable and services NIPA consumption and real GDP. Section D contains technical details showing that a conditional CAPM can be derived as an approximation of the stochastic discount factor (SDF) implied by the model in Campbell and Cochrane (1999), as well as additional results for the cross-sectional analysis in Section 5.

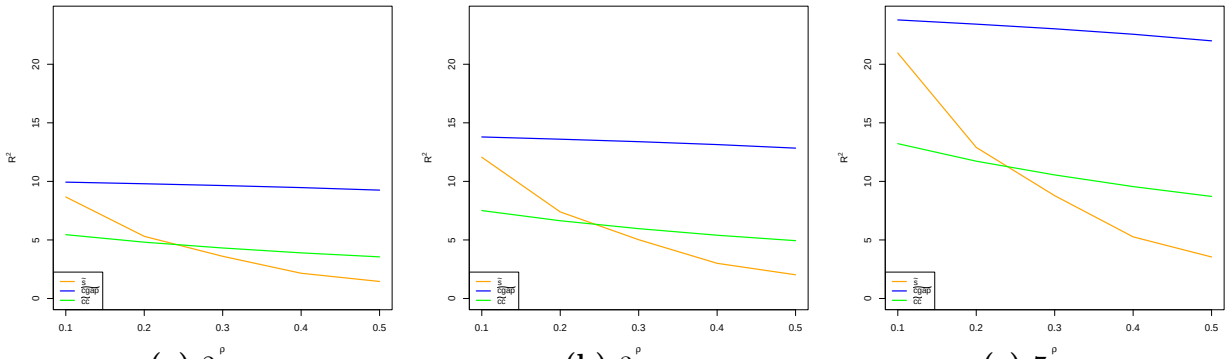
A Simulated Campbell-Cochrane Economy: Further Results

Table A.1: This table replicates Table 1 for the dividend claim.

Horizon (Years)	1	2	3	5	7
log pd	−1.83	−3.45	−4.84	−7.07	−8.71
	[7.04]	[12.85]	[17.50]	[24.14]	[28.03]
s	−0.51	−0.96	−1.34	−1.96	−2.41
	[6.87]	[12.47]	[16.92]	[23.23]	[26.97]
$cgap$	−0.36	−0.68	−0.95	−1.39	−1.71
	[3.42]	[6.27]	[8.51]	[11.67]	[13.52]
cc	−0.31	−0.59	−0.81	−1.19	−1.46
	[2.54]	[4.65]	[6.17]	[8.60]	[9.88]

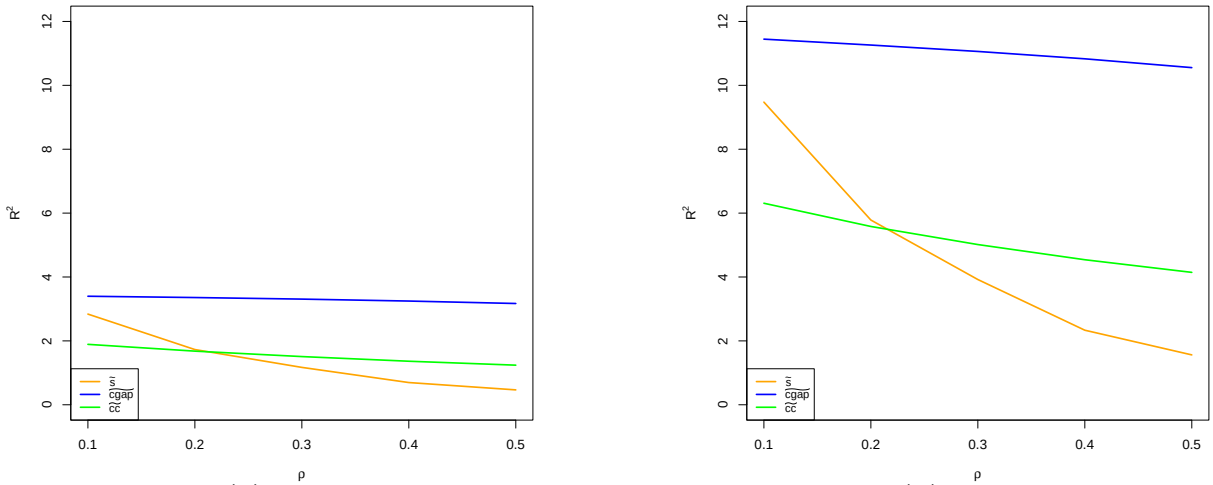
Table A.2: This table reports the calibrated parameters values for γ and ϕ for different values of ρ. The last row reports the correlation between true and noisy consumption in the simulated economy.

ρ	0.1	0.2	0.3	0.4	0.5
γ	1.45	1.15	1.10	0.85	0.70
ϕ	0.975	0.985	0.990	0.994	0.996
$\text{corr}(\Delta c_t, \widetilde{\Delta} c_t)$	0.83	0.74	0.67	0.61	0.54



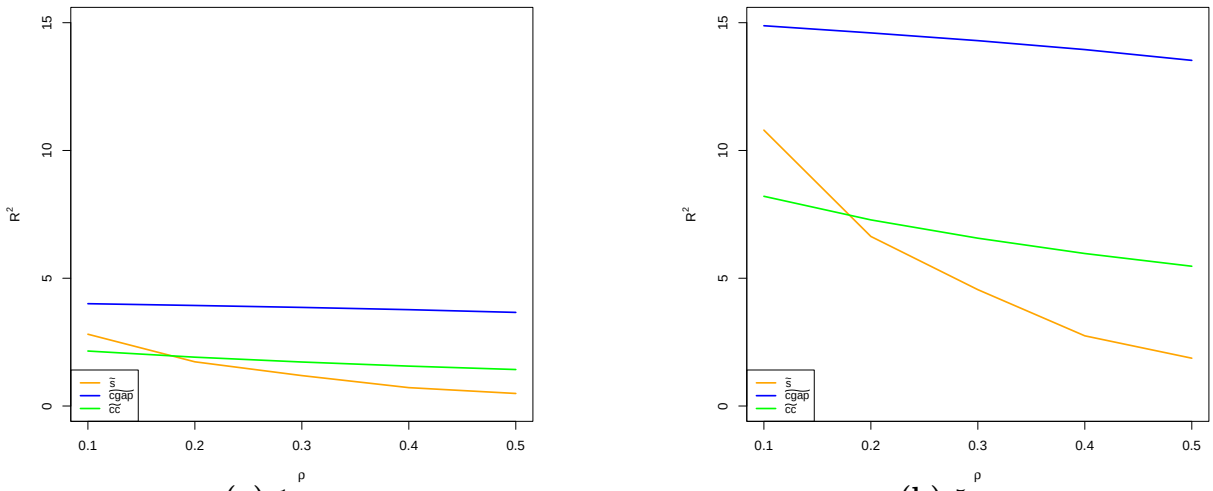
(a) 2-year **(b)** 3-year **(c)** 7-year

Figure A.1: This figure replicates Figure 3 for different time horizons.



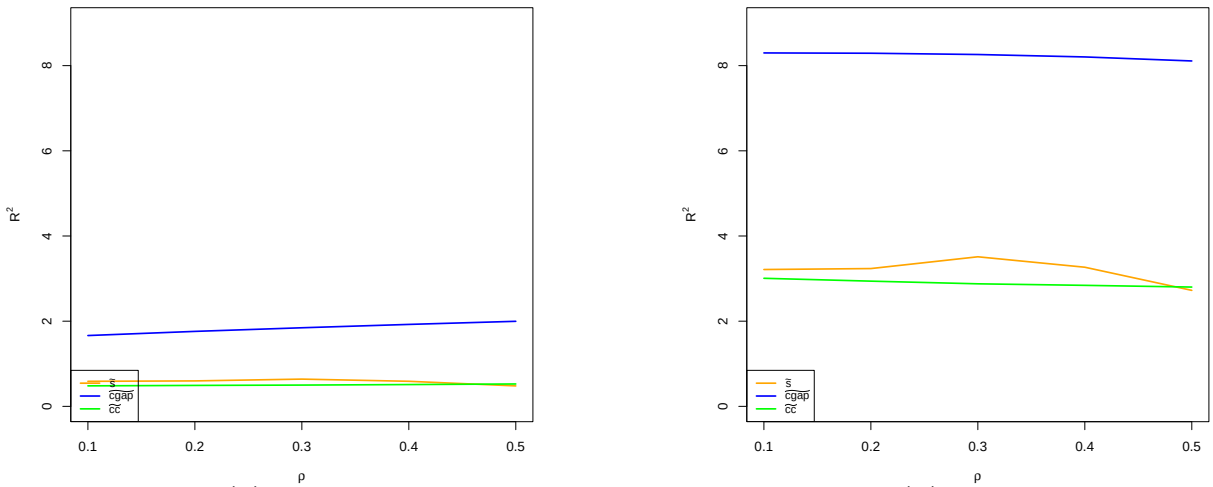
(a) 1-year **(b)** 5-year

Figure A.2: This figure replicates Figure 3 for the dividend claim.



(a) 1-year **(b)** 5-year

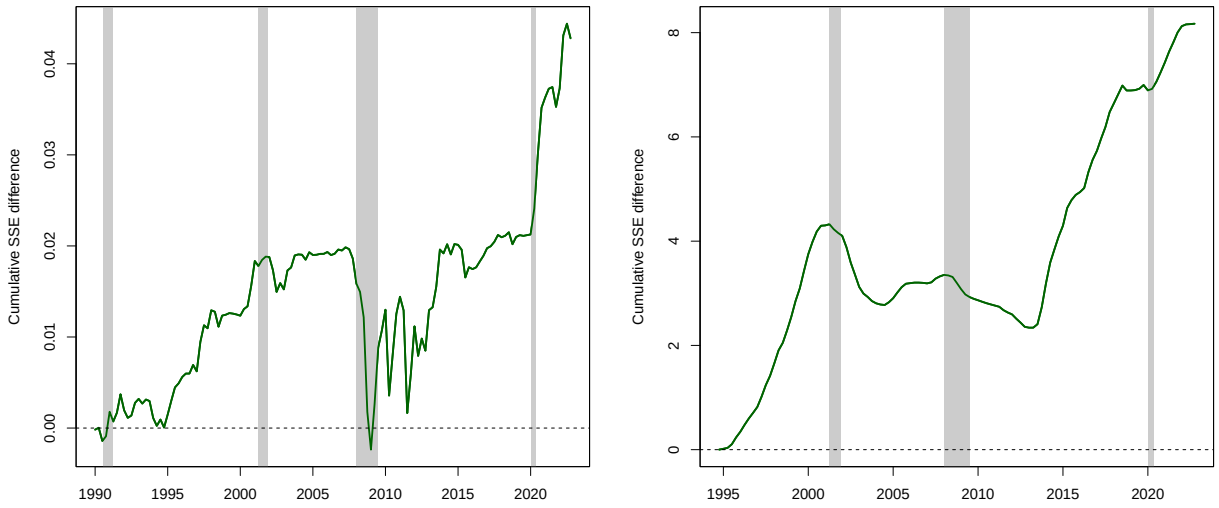
Figure A.3: This figure replicates Figure 3 using the numerically accurate solution for the Campbell-Cochrane economy proposed by Schreindorfer (2023).



(a) 1-year **(b)** 5-year

Figure A.4: This figure replicates Figure 3 when consumption growth is an AR(1) process. Consistent with the estimates reported in, e.g., Beeler and Campbell (2012); Schorfheide, Song, and Yaron (2018), I calibrate the (annualized) AR(1) parameter to 0.45.

B Return Predictability: Further Results



(a) 1-quarter **(b)** 5-year

Figure B.1: OOS Predictive Performance. This figure shows the OOS performance of predictive regressions for 1-quarter (Panel (a)) and 5-year returns (Panel (b)). The graphs show the cumulative squared prediction errors of the historical mean minus the cumulative squared prediction error of the predictive regression using the *cgap*. Whenever a line increases, it means that the *cgap* predicts better; whenever it decreases, the prevailing mean predicts better.

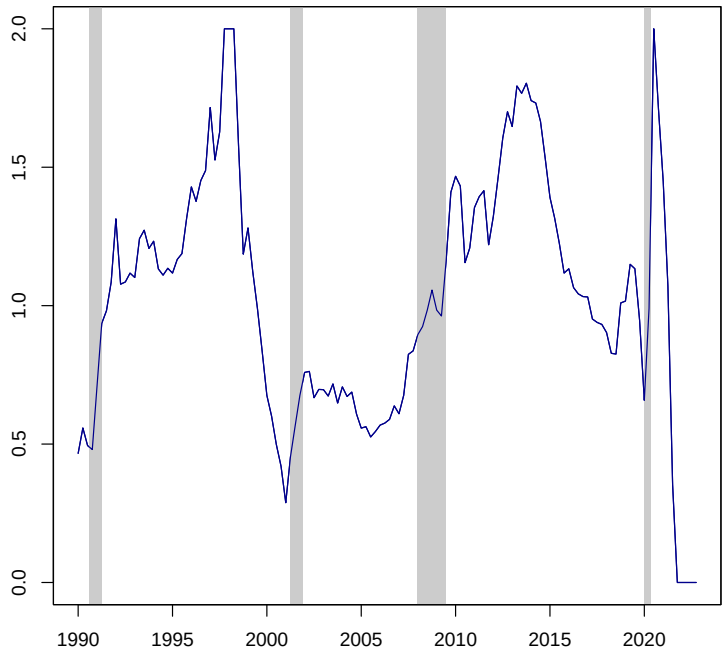


Figure B.2: Market Timing: Weights. This figure shows the portfolio equity weights for the market timing strategy using the *cgap* described in Section 3.3.

Table B.1: Predicting Market Returns with Unfiltered Consumption

This table OLS estimates and the R^2 from the predictive regression $r_{m,t+1} = \gamma_0 + \gamma_1 cgap_t + \gamma_2 s_t + \epsilon_{t+1}$, where $r_{m,t+h}$ are market excess returns at time $t+1$, $cgap$ is the consumption gap computed as in equation (9), and s is the surplus consumption ratio constructed as in Cochrane (2011). The surplus consumption ratio is derived using either observed NIPA non-durable and services consumption (s) or the unfiltered non-durable and services consumption (s^{UNFIL}), constructed as in Kroencke (2017). Values in parenthesis are heteroskedasticity and autocorrelation consistent (HAC) standard errors using Newey and West (1987) with automatic bandwidth selection procedure as described in Newey and West (1994). ***, **, and * indicates respectively 1%, 5%, and 10% level of significance. Regressors are de-meaned and standardized. Annual observations. The sample period is 1957 to 2022.

	(1)	(2)	(3)	(4)	(5)
$cgap_t$	-0.062^{***}			-0.063^{***}	-0.050^{**}
	(0.017)			(0.017)	(0.020)
s_t		-0.038^{***}		0.000	
		(0.007)		(0.009)	
s_t^{UNFIL}			-0.054^{***}		-0.017
			(0.010)		(0.015)
Observations	65	65	65	65	65
Adjusted R^2	0.114	0.032	0.080	0.100	0.105

Table B.2: Predicting Market Returns with Vintage Consumption

This table replicates Table 2 using vintage NIPA consumption data. Quarterly observations. The sample period is 1965Q3 to 2022Q4.

Panel A: In-Sample

	$r_{m,t+1}$	$r_{m,t+4}$	$r_{m,t+8}$	$r_{m,t+20}$
$cgap_t^{vintage}$	-0.396^{***}	-1.527^{***}	-2.499^{***}	-4.655^{**}
	(0.115)	(0.479)	(0.894)	(1.969)
Constant	0.940	3.852	8.873	26.007^{*}
	(0.722)	(3.063)	(5.967)	(13.900)
Observations	172	169	165	153
R^2	0.047	0.174	0.266	0.443

Panel B: Out-Of-Sample

R^2_{OOS}	From 1990	From 1995	From 2000
$r_{m,t+1}$	4.89^{***}	6.92^{***}	6.46^{***}
$r_{m,t+4}$	22.92^{***}	24.25^{***}	22.09^{***}
$r_{m,t+8}$	29.54^{***}	25.88^{***}	24.57^{***}
$r_{m,t+20}$	40.91^{***}	26.33^{***}	24.44^{***}

Table B.3: Predicting Market Returns Excluding COVID-19

This table replicates Table 2 excluding the sample period after the COVID-19 pandemic. Quarterly observations. The sample period is 1957Q1 to 2019Q4.

Panel A: In-Sample

	$r_{m,t+1}$	$r_{m,t+4}$	$r_{m,t+8}$	$r_{m,t+20}$
$cgap_t$	-0.506^{***}	-2.055^{**}	-3.861^{**}	-8.507^{**}
	(0.148)	(0.814)	(1.640)	(3.712)
Constant	0.009^{*}	0.036	0.072	0.182
	(0.005)	(0.023)	(0.048)	(0.118)
Observations	251	248	244	232
R^2	0.024	0.101	0.201	0.469

Panel B: Out-Of-Sample

R^2_{OOS}	From 1990	From 1995	From 2000
$r_{m,t+1}$	2.65^{***}	2.97^{***}	1.49^{***}
$r_{m,t+4}$	12.77^{***}	12.77^{***}	8.33^{***}
$r_{m,t+8}$	23.98^{***}	22.45^{***}	18.35^{***}
$r_{m,t+20}$	41.61^{***}	31.20^{***}	50.04^{***}

Table B.4: Predicting Market Returns for Different Subsamples

This table replicates Panel A in Table (2) for different subsamples. Constant estimates are not tabulated. Quarterly observations. The sample period ends in 2022Q4.

Panel A: Sample starts in 1980

	$r_{m,t+1}$	$r_{m,t+4}$	$r_{m,t+8}$	$r_{m,t+20}$
$cgap_t$	-0.744^{***}	-2.906^{**}	-4.202^{**}	-9.605^{***}
	(0.216)	(1.140)	(1.990)	(3.411)
Observations	171	168	164	152
R^2	0.036	0.135	0.161	0.364

Panel B: Sample starts in 1990

	$r_{m,t+1}$	$r_{m,t+4}$	$r_{m,t+8}$	$r_{m,t+20}$
$cgap_t$	-0.933^{***}	-3.485^{***}	-5.027^{**}	-11.653^{***}
	(0.209)	(1.257)	(2.134)	(3.755)
Observations	131	128	124	112
R^2	0.061	0.204	0.204	0.428

Panel C: Sample starts in 2000

	$r_{m,t+1}$	$r_{m,t+4}$	$r_{m,t+8}$	$r_{m,t+20}$
$cgap_t$	-1.046^{***}	-3.977^{***}	-5.126^{***}	-11.218^{***}
	(0.214)	(1.306)	(1.909)	(2.894)
Observations	91	88	84	72
R^2	0.090	0.281	0.241	0.624

Table B.5: Predicting Market Returns Using the Second Lag

This table replicates Table 2 lagging the *cgap* by two periods in the predictive regression. Quarterly observations. The sample period is 1957Q1 to 2022Q4.

Panel A: In-Sample

	$r_{m,t+1}$	$r_{m,t+4}$	$r_{m,t+8}$	$r_{m,t+20}$
$cgap_{t-1}$	-0.603^{***}	-2.057^{***}	-3.431^{**}	-8.437^{**}
	(0.160)	(0.756)	(1.494)	(3.407)
Constant	0.008	0.036	0.079^{*}	0.188
	(0.005)	(0.023)	(0.047)	(0.115)
Observations	262	259	255	243
R^2	0.034	0.102	0.164	0.454

Panel B: Out-Of-Sample

R^2_{OOS}	From 1990	From 1995	From 2000
$r_{m,t+1}$	4.71^{***}	5.15^{***}	4.63^{***}
$r_{m,t+4}$	17.86^{***}	17.66^{***}	14.05^{***}
$r_{m,t+8}$	22.62^{***}	20.82^{***}	16.72^{***}
$r_{m,t+20}$	45.12^{***}	36.08^{***}	55.87^{***}

Table B.6: Predicting Market Returns with Stochastically Detrended Consumption

This table replicates Table 2 using stochastically detrended consumption ($\hat{c}$) constructed in Campbell, Pflueger, and Viceira (2020). Quarterly observations. The sample period is 1957Q1 to 2022Q4.

Panel A: In-Sample

	$r_{m,t+1}$	$r_{m,t+4}$	$r_{m,t+8}$	$r_{m,t+20}$
$\hat{c}_t$	-0.627^{***}	-2.417^{***}	-3.860^{**}	-9.505^{**}
	(0.168)	(0.842)	(1.703)	(3.720)
Constant	0.053^{***}	0.207^{***}	0.355^{***}	0.877^{***}
	(0.011)	(0.055)	(0.109)	(0.236)
Observations	263	260	256	244
R^2	0.031	0.119	0.175	0.439

Panel B: Out-Of-Sample

R^2_{OOS}	From 1990	From 1995	From 2000
$r_{m,t+1}$	3.24^{***}	3.38^{***}	2.38^{***}
$r_{m,t+4}$	12.79^{***}	12.73^{***}	9.26^{***}
$r_{m,t+8}$	15.46^{***}	13.91^{***}	8.24^{***}
$r_{m,t+20}$	36.57^{***}	28.96^{***}	40.59^{***}

Table B.7: Predicting Market Returns with a Deterministic Trend

This table replicates Table 2 using a deterministic time trend to detrend consumption expenditures. Quarterly observations. The sample period is 1957Q1 to 2022Q4.

Panel A: In-Sample

	$r_{m,t+1}$	$r_{m,t+4}$	$r_{m,t+8}$	$r_{m,t+20}$
$cgap_t^{time}$	-0.155^{*}	-0.602^{*}	-1.026	-1.462
	(0.084)	(0.347)	(0.695)	(1.550)
Constant	0.013^{**}	0.055^{**}	0.113^{***}	0.270^{**}
	(0.005)	(0.021)	(0.043)	(0.106)
Observations	263	260	256	244
R^2	0.011	0.043	0.069	0.057

Panel B: Out-Of-Sample

R^2_{OOS}	From 1990	From 1995	From 2000
$r_{m,t+1}$	-0.53	-0.2	-2.86
$r_{m,t+4}$	-0.08	0	-7.97
$r_{m,t+8}$	3.61^{***}	1.35^{***}	-7.45
$r_{m,t+20}$	-3.74	-31.71	-4.89

C Cointegration Analysis

I test whether log aggregate NIPA non-durable and service expenditures and log real GDP are cointegrated using the testing procedure proposed by Johansen (1991). This procedure presumes a p-dimensional vector autoregressive (VAR) model with k lags, where p corresponds to the number of variables among which the investigator wishes to test for cointegration. For my application, I choose the number of lags in the VAR according to the Akaike information criterion (AIC): $p = 2$. The Johansen procedure provides two tests for cointegration: Under the null hypothesis, H_0, that there are exactly r cointegrating relations, the "Trace" statistic supplies a likelihood ratio test of H_0 against the alternative, H_1, that there are p cointegrating relations. A second approach uses the "L-max" statistic to test the null hypothesis of r cointegrating relations against the alternative of $r + 1$ cointegrating relations.

Table C.1 reports the test statistics along with the 90, 95, and 99 percent critical values. Both tests reject at 99% level of confidence the null of no cointegration between consumption expenditures and real GDP. Furthermore, I perform the Phillips and Ouliaris (1990) (PO) test for the null hypothesis that consumption and real output are not cointegrated. I obtain a test statistic of -28.34 with an associated p-value of 0.01. Overall, the empirical evidence uniformly suggests that log NIPA consumption and log real GDP for the sample period 1947 to 2022 share a common stochastic trend.

Table C.1: Johansen Cointegration Test

This table reports results for the Johansen (1991) procedure to test for cointegration between the I(1) series consumption expenditures and real GDP. Panel A reports results for the "Trace" statistics. Panel B reports results for the "L-Max" statistics. The null hypothesis is that the number of cointegration vectors is at most r. Testing is sequential for $r^* = 0, 1, ..., k-1$ and the first non-rejection of the null represents an estimate of r. The number of lags for the Johansen test is chosen according to the Akaike information criterion (AIC) and is equal to 2; I include a linear trend in the long-run regression. Quarterly observations. The sample period is 1947Q1 to 2022Q4.

Panel A: Trace

	Test statistics	90% CV	95% CV	99% CV
r <= 1 |	2.80	10.49	12.25	16.26
r = 0 |	27.91	22.76	25.32	30.45

Panel B: L-Max

	Test statistics	90% CV	95% CV	99% CV
r <= 1 |	2.80	10.49	12.25	16.26
r = 0 |	25.11	16.85	18.96	23.65

D SDF Analysis

First, I describe some technical details showing that the conditional CAPM in (16) can be derived as an approximation of the Campbell and Cochrane (1999)'s SDF. I then present additional results that complement those reported in Section 5.

D.1 Technical Derivations

(a) Normalization Approach

In the Campbell and Cochrane (1999) external habit economy described in Section 2.1, marginal utility is given by

$$u(C_t, X_t) = (C_t - X_t)^{-\gamma} = S_t^{-\gamma} C_t^{-\gamma}$$

and the log marginal rate of substitution is

$$\begin{aligned} m_{t+1} &= \log(\delta) - \gamma(\Delta s_{t+1} + \Delta c_{t+1}) \\ &= \log(\delta) - \gamma g - \gamma(\phi - 1) f(s_t) - \gamma(1 + \lambda(s_t)) v_{t+1} \ , \end{aligned} \tag{D.1}$$

where $f(s_t) = s_t - \bar{s}$ and $v_{t+1} = \Delta c_{t+1} - \mathrm{E}_t[\Delta c_{t+1}]$.

Given the sensitivity function in (4) and using the fact that s_t ventures above s_{max} only rarely (e.g., Wachter, 2006), the term multiplying the consumption shock v_{t+1} in (D.1) simplifies to

$$k\sqrt{1 - 2f(s_t)} \approx k(1 - f(s_t)) \ , \tag{D.2}$$

where the approximation follows from a first-order Taylor expansion around the steady state $\bar{S}$, and $k = \gamma/\bar{S}$ represents risk aversion at the steady state.

Finally, substituting (D.2) into (D.1), I obtain

$$m_{t+1} = a_t - b_t \cdot v_{t+1} \tag{D.3}$$

where $a_t = a_0 + a_1 \cdot f(s_t)$, with

$$\begin{aligned} a_0 &= \log(\delta) - \gamma g \\ a_1 &= -\gamma(\phi - 1) \ , \end{aligned}$$

and $b_t = b_0 + b_1 \cdot f(s_t)$, with

$$b_0 = k$$
$$b_1 = -k \ .$$

The SDF in (D.3) can be normalized such that $\mathbb{E}[M] = 1$. To see this, start from

$$E_t[M_{t+1} R^e_{t+1}] = 0$$

where R^e_{t+1} denotes excess returns. Then,

$$E_t \left[\exp(a_0 + a_1 \cdot f(s_t) - b_t \cdot v_{t+1}) R^e_{t+1}\right] = 0 \ .$$

Factoring out $a_0 + a_1 \cdot f(s_t)$ and adding and subtracting a constant, I obtain

$$\exp(a_0 + a_1 \cdot f(s_t) - \text{const}) \cdot E_t \left[\exp(\text{const} - b_t \cdot v_{t+1}) R^e_{t+1}\right] = 0 \ .$$

Thus, it follows that

$$E_t \left[\exp(\text{const} - b_t \cdot v_{t+1}) R^e_{t+1}\right] = 0 \ ,$$

and, taking the unconditional expectation, I get

$$E \left[\exp(\text{const} - b_t \cdot v_{t+1}) R^e_{t+1}\right] = 0 \ .$$

Finally, using the normalization condition

$$E \left[\exp(\text{const} - b_t \cdot v_{t+1})\right] = 1 \ ,$$

I obtain the SDF:

$$M_{t+1} = 1 - b_t \cdot v_{t+1} \ . \tag{D.4}$$

(b) Substituting Consumption with Returns

Gonçalves (2021) demonstrates that a conditional ICAPM holds with the external habit model (see Internet Appendix A). This implies that the ICAPM budget constraint is satisfied, allowing me to substitute the consumption shocks in (D.4) with the sum of the shocks to returns on the wealth portfolio and the shocks to the log consumption-wealth ratio:

$$\widetilde{\Delta c_{t+1}} = \widetilde{R_{w,t+1}} + \widetilde{cw_{t+1}} \ , \tag{D.5}$$

where $\widetilde{x_{t+1}}$ denotes the shock to the variable x_{t+1}; for example, $\widetilde{\Delta c_{t+1}} = v_{t+1}$.

As in the Campbell and Cochrane (1999) economy surplus consumption is the only state variable, cw_t is a (decreasing) function of s_t only. Thus, (D.4) can be rewritten as

$$M_{t+1} = 1 - b_t^R \cdot \widetilde{\dot{R}_{w,t+1}} + b_t^s \cdot g(\widetilde{\dot{s}_{t+1}}), \tag{D.6}$$

which shows that the habit economy is consistent with a conditional model featuring two fundamental factors: the market portfolio (as in the CAPM) plus shocks to surplus consumption.

There are three main reasons why I focus on the conditional CAPM (17), effectively imposing the constraint $b_t^s = 0$ in (D.6). First, testing a conditional CAPM naturally follows from the evidence that *cgap* predicts the equity premium (e.g., Lettau and Ludvigson, 2001b). Second, using only traded factors allows me to apply the robust and efficient GMM methodology discussed in Section 5.1, mitigating concerns related to the cross-sectional estimation of factor models with non-traded factors (see, e.g., Lewellen, Nagel, and Shanken, 2010). Finally, in my sample, an AR(1) surplus consumption shock has a standard deviation of 0.009, suggesting that the last factor in (D.6) has limited economic relevance.

D.2 Additional Results

Table D.1: SDF Coefficients

This table reports SDF coefficients for the model (17). Column (1) presents the CAPM. Columns (2) to (4) use *cgap*, *cc*, or both variables as the conditioning variable z_t in (17). I use the generalized method of moments (GMM) framework described in Chabi-Yo, Gonçalves, and Loudis (2024) to estimate $\hat{b}$. Since the SDF coefficients are not directly comparable, I multiply each $\hat{b}$ by the standard deviation of its corresponding factor. This transformation allows the reported values to be interpreted as the change in the SDF induced by a one-standard-deviation change in the respective factor, holding other factors constant. Values in parenthesis $t-$stat computed according to GMM asymptotic theory with Newey and West (1987, 1994) for the spectral density matrix. Quarterly observations. The sample period is 1963Q3 to 2022Q4.

	(1)	(2)	(3)	(4)
$\hat{b}_0$	0.199 (2.73)	0.201 (2.81)	0.145 (2.00)	0.196 (2.51)
$\hat{b}_1^{cgap}$		−0.188 (−2.97)		−0.180 (−2.00)
$\hat{b}_1^{cc}$			−0.144 (−2.02)	−0.014 (−0.15)