

In full-information estimates, long-run risks explain at most a quarter of p/d variance, and habit explains even less

Andrew Y. Chen

Federal Reserve Board

Fabian Winkler

Federal Reserve Board

Rebecca Wasyk

Federal Reserve Board

November 2019*

Abstract

Many consumption-based models succeed in matching long lists of asset price moments. We propose an alternative, full-information Bayesian evaluation that decomposes the price-dividend ratio (p/d) into contributions from long-run risks, habit, and a residual. We find that long-run risks account for less than 25% of the variance of p/d and that habit's contribution is negligible. This result is robust to the prior, including priors that assume long-run risks in consumption and highly persistent habit. However, the residual mostly tracks decades-long movements in p/d . At business cycle frequency, long-run risks explain about 70% of the movements of p/d while habit's contribution stays negligible.

*We thank the editor Ivo Welch, an anonymous referee, Francesco Bianchi, John Cochrane, Ed Herbst, Stijn Van Nieuwerburgh, Francisco Palomino, and Missaka Warusawitharana for advice and comments; and Carter Bryson and Chris Campano for excellent research assistance. Our emails are andrew.y.chen, rebecca.d.wasyk, and fabian.winkler, all @frb.gov. The views expressed herein are those of the authors and do not necessarily reflect the position of the Board of Governors of the Federal Reserve or the Federal Reserve System.

1. Introduction

In this paper, we examine the ability of long-run risks and habit to explain the time series of the price-dividend ratio. We estimate a model in which the log price-dividend ratio pd_t is linear in long run growth x_t , long run volatility σ_t^2 , surplus consumption s_t , and a residual e_t :

$$pd_t = \tilde{\mu}_{pd} + A_x x_t + A_V \sigma_t^2 + A_s s_t + A_e e_t, \quad (1)$$

where A_x , A_V , A_s , and A_e are coefficients that we estimate jointly with other parameters. Long-run risks explain only about 20% of the variance of pd_t . This unexplained portion of pd_t is both large and highly persistent. In fact, when we filter out low frequencies, we find long-run risks account for about 75% of business-cycle variation in pd_t . Habit, however, explains less than 5% of the variance of pd_t , regardless of the filtering.

The state variables on the right hand side of Equation (1) are latent, and require structure, parameters, and a statistical method for extraction from data. For structure, we assume that long-run risks and habit are related to consumption and dividend growth in the usual way (Campbell and Cochrane 1999, Bansal, Kiku, and Yaron 2012). For parameters, we estimate the model using Bayesian methods (Herbst and Schorfheide 2014). Finally, to extract the latent states, we apply a particle smoother (Godsill, Doucet, and West 2004) to annual data on the price-dividend ratio, consumption growth, and dividend growth. Combined with the estimated coefficients, these latent states decompose the time-series of pd_t into contributions from long-run risks and habit.

Long-run risks and habit were probably not designed explain the entire time-series of the price-dividend ratio. However, we are not asking their models to explain the entire time series either. Rather, we are asking more subtle questions: What share of the time-series do they explain? What features do they explain? What is the nature of the residual that they leave behind?

We find that long-run growth accounts for 13%, long-run volatility accounts for 11%, and surplus consumption accounts for -2% of the variance of the price-dividend ratio.¹ These baseline estimates come from priors that bias the esti-

¹This variance decomposition comes from taking the covariance of both sides of Equation (1) with pd_t , and thus can lead to negative variance shares if the latent states are negatively correlated with the price-dividend ratio.

mator in *favor* of long-run risks and habit. Our baseline estimate uses “believer” priors—priors that assume that long-run risks in consumption growth are real and that habit is highly persistent. Despite these priors, the estimation finds that a Bayesian believer would not only update to a low weight on long-run risks and habit, but she would become quite confident in these low weights. We find the posterior probability that long-run risks and habit account for more than 40% of the variance of the price-dividend ratio is negligible.

Though the consumption-based factors explain little of the variance of pd_t , they do explain important features of the time-series. In particular, they explain business-cycle variation in the price-dividend ratio quite well. This is clearly seen in crisis episodes. The stock market crashes of the Great Depression and 2008 Financial Crisis are accompanied by sharp declines in long-run growth, sharp spikes in long-run volatility, and also sharp declines in surplus consumption. More subtly, long-run growth booms and busts with valuations around the 1960s, and the late 1990’s technology “bubble” is marked by a boom and bust in long-run growth. These observations can be formalized with variance decompositions of either first-differenced or high-pass filtered versions of Equation (1). These filtered decompositions show that long-run growth accounts for about 50%, and long-run volatility accounts for about 20% of the business-cycle variation in pd_t .

Surplus consumption, however, accounts for less than 5% of the variance of p_t , even using filtered data. This poor explanatory power is consistent with the near-zero contemporaneous correlation between consumption growth and stock returns (Albuquerque et al. 2016), and the fact that the habit model implies a correlation close to 1. In contrast, the important role of long-run growth in business cycle variation in pd_t is consistent with the fact that stock returns are a strong predictor of future consumption growth.

With business-cycle frequencies accounted for, Equation (1) leaves a residual that is smooth and extremely persistent. Indeed, the residual extracted by particle smoother appears to be non-stationary. It rises dramatically during the 1980’s and 1990’s, before flattening out in a high plateau from 2000-2014. Thus, long-run risks and habit are missing a slow-moving, albeit nearly-permanent and very large component of valuations.

We believe these findings are important now that so many models match so many moments. Habit, long-run risks, ambiguity aversion, rare disasters,

financial frictions, near-rational learning, and time-preference shocks all can match the first and second moments of stock returns, the first and second moments of the risk-free rate, and equity premium predictability.² As a result, the traditional method of evaluating models by matching key moments cannot differentiate between these models.

This abundance of success means that we must push the models further. Adding more target moments is the natural next step, and is pursued by several papers that compare long-run risks and habit (Bansal, Gallant, and Tauchen 2007; Beeler and Campbell 2012; Bansal, Kiku, and Yaron 2012). This approach, however, leads to a muddled picture of model performance, as the preferred model depends on which moments one considers important. In fact, authors have continued to publish both models with habit (Bekaert and Engstrom 2017, Chen 2017) and long-run risks (Schorfheide, Song, and Yaron Forthcoming, Colacito et al. Forthcoming). Moreover, adding more moments by bringing in new data is arguably a step backward. An elegant parable that can only explain equity markets may be more useful than a complicated machine that explains many asset markets.

These motivations lead us to our empirical design. Our Bayesian method accounts for the entire likelihood of the data, and thus our estimation efficiently incorporates all moments of the price-dividend ratio, consumption growth, and dividend growth. At the same time, we do not use a formal likelihood-based model evaluation, as this would certainly lead to rejection of long-run risks and habit. Instead, our examination of the extracted latent states allows us to see in which ways these models succeed and in which ways they fail as parables for historical stock market events.

Our method also addresses the concern that long-run risks are difficult to

²These literatures are large, so we list only a selection of papers from each. Papers that show the strong performance of habit include Jermann (1998), Campbell and Cochrane (1999), Wachter (2006), Verdelhan (2010), Bekaert and Engstrom (2017), and Chen (2017). Papers that show the strong performance of long-run risks include Bansal and Yaron (2004), Drechsler and Yaron (2010), Bansal, Kiku, and Yaron (2012), Croce (2014), Nakamura, Sergeyev, and Steinsson (2017), Ai et al. (2018), Schorfheide, Song, and Yaron (Forthcoming), and Colacito et al. (Forthcoming). For rare disasters see Rietz (1988), Barro (2006), Gabaix (2012), Wachter (2013). For ambiguity aversion, see Ju and Miao (2012) and Jahan-Parvar and Liu (2014). For financial frictions see Mankiw and Zeldes (1991), Lustig and Van Nieuwerburgh (2006), Guvenen (2009), He and Krishnamurthy (2013), and Brunnermeier and Sannikov (2014). For near-rational learning, see Adam, Marcet, and Nicolini (2016), and Adam, Marcet, and Beutel (2017). Other quantitative explanations for asset prices include investment-specific technology shocks (Papanikolaou 2011) and preference shocks (Albuquerque et al. 2016).

identify. We explicitly use the price-dividend ratio to identify long-run risks, and impose priors that believe expected growth has a persistent component. As a result, we can identify subtle relationships between consumption growth and the price-dividend ratio that are lost in simpler regression-based methods. Indeed, we find that two-stage estimations that first estimate latent states without using asset prices and then estimate the price-dividend equation lead to no role for long-run risks, even at business cycle frequencies.

Though our methodology is novel, the estimated latent states capture key properties of long-run risks and habit. Estimated long-run growth strongly predicts consumption growth in U.S. data. Estimated long-run volatility predicts simply proxies for consumption volatility and excess returns. Estimated surplus consumption strongly predicts excess returns, but has relatively little predictive power for consumption and dividend growth. This match with the literature is not only qualitative, but quantitative: The R^2 s using historical observables and estimated latent states are similar to those implied by simulations of Bansal, Kiku, and Yaron 2012 and Campbell and Cochrane (1999).

Our results are robust. Our baseline model assumes (1) a reduced form price-dividend equation without structural restrictions, (2) data from 1929-2014 can be used without accounting for structural breaks, and (3) “believer” priors that assume long-run risks in consumption growth are real. Using alternative estimations, we show that none of these three assumptions is important to our main result. Estimation of a model with structural restrictions, using only post-war quarterly data, or flat priors leads to the same main results: the residual explains the vast majority of the variance of the level of price/dividends, but expected growth and volatility explain the vast majority of business cycle variation in price/dividends.

Related Literature In a closely related paper, Schorfheide, Song, and Yaron (Forthcoming) (SSY) use Bayesian methods to estimate an extension of the long-run risks model with multiple volatility processes. Unlike us, they do not allow for factors other than long-run risks to drive stock prices, and essentially force long-run risks to fit the data.³ As a result, their Bayesian method provides no natural way to examine the explanatory power of long-run risks, and much of their empirical evaluation comes back to moment matching.

³SSY do allow for a time-preference state to drive asset prices, but the time-preference state is mostly identified with the risk-free rate.

Moreover, our results relate to an unusual feature of SSY. Unlike previous papers on long-run risks, SSY assume three independent volatility processes. This alternative assumption implies that long-run volatility is independent of the observed volatilities of consumption and dividend growth, and effectively disconnects long-run volatility from observables, making long-run volatility much like our residual.

Aldrich and Gallant (2011) and Gallant, Jahan-Parvar, and Liu (2018) use the Gallant and McCulloch (2009) Bayesian method to compare variants of ambiguity aversion, habit, long-run risks, and prospect theory. The Aldrich and Gallant method, however, requires the choice of an auxiliary model for computing the likelihood function. We avoid this potentially difficult choice by computing the likelihood function directly. Moreover, unlike these papers, we don't force a single model to fit the entirety of the data.

Our findings are consistent with papers that find that consumption and stock prices are disconnected at virtually all horizons (Lettau and Ludvigson 2004, Lettau and Ludvigson 2014). This disconnect is hard to reconcile in models that emphasize endowment shocks such as habit or long-run risks (Albuquerque et al. 2016). Our paper shows that this disconnect is critical for asset pricing models when the entire likelihood of the data is taken into account.

2. Model, Estimation, and Parameter Estimates

This section provides the background required for understanding our main results. Here, we describe the model (Section 2.1), estimation method (Section 2.3), baseline priors (Section 2.4) and parameter estimates (Section 2.5). We also cover model frequency and data (Section 2.2).

2.1. Semi-Structural Model with Multiple Sources of Risk

Our key variable of interest is the log price-dividend ratio pd_t . pd_t is linear in four state variables

$$pd_t = \mu_{pd} + A_x x_t + A_V \tilde{\sigma}_t^2 + A_s \tilde{s}_t + A_e e_t. \quad (2)$$

where x_t is long-run growth, $\tilde{\sigma}_t$ is long-run volatility, \tilde{s}_t is surplus consumption (habit), and e_t is a residual. The tildes over $\tilde{\sigma}_t$ and \tilde{s}_t indicate that these vari-

ables are demeaned ($\tilde{\sigma}_t^2 = \sigma_t^2 - \mathbb{E}(\sigma_t^2)$ and $\tilde{s}_t = s_t - \bar{s}$) and simplify the mapping to the data because they imply that μ_{pd} is the mean log price-dividend ratio.

Our goal is to estimate the coefficients A_x, A_V, A_s, A_e and find the smoothed paths of the latent states $x_t, \tilde{\sigma}_t^2, \tilde{s}_t$ and e_t . The coefficients and smoothed paths provide a simple description of the importance of each source of market volatility.

We estimate the reduced form (2) rather than a structural expression to let the data speak freely about cross-equation relationships. This agnostic view seems appropriate given the amount of disagreement in the literature about the right structural model for asset prices (e.g. Gabaix 2012; Cochrane 2017).

All states are latent, but they can be identified by their linkages with additional observables. In addition to pd_t (Equation (2)), the long-run risk states $x_t, \tilde{\sigma}_t^2$ are identified by their relationship with consumption and dividend growth

$$\begin{aligned}\Delta c_t &= \mu_c + x_{t-1} + \sigma_{t-1} \eta_{c,t} \\ \Delta d_t &= \mu_d + \phi_x x_{t-1} + \phi_{\eta_c} \sigma_{t-1} \eta_{c,t} + \varphi_d \sigma_{t-1} \eta_{d,t} \\ \eta_{c,t}, \eta_{d,t} &\sim N(0, 1) \text{ i.i.d.},\end{aligned}\tag{3}$$

where long-run growth x_t evolves according to the standard heteroskedastic AR(1)

$$\begin{aligned}x_t &= \rho_x x_{t-1} + \varphi_x \sigma_{t-1} \eta_{x,t} \\ \eta_{x,t} &\sim N(0, 1) \text{ i.i.d.},\end{aligned}\tag{4}$$

and long-run volatility σ_t is an AR(1) in logs

$$\begin{aligned}h_t &= \rho_h h_{t-1} + \sigma_h \eta_{h,t} \\ \sigma_t &= \bar{\sigma} \exp(h_t). \\ \eta_{h,t} &\sim N(0, 1) \text{ i.i.d.}\end{aligned}\tag{5}$$

This volatility specification borrows a technical fix from Schorfheide, Song, and Yaron (Forthcoming) (SSY), but otherwise the above consumption and dividends are equivalent to Bansal, Kiku, and Yaron (2012)'s processes.⁴ SSY's fix ensures that volatility is always positive.

⁴To see the mapping, note that $\sigma_t^2 - \bar{\sigma}^2 \approx 2\bar{\sigma}^2 h_t$ implies $\sigma_t^2 - \bar{\sigma}^2 \approx \rho_h (\sigma_{t-1}^2 - \bar{\sigma}^2) + 2\bar{\sigma}^2 \sigma_h \eta_{h,t}$.

Importantly, our specification does not include the multiple volatility processes of SSY. Using a single volatility process is consistent with the bulk of the long-run risk literature (Bansal and Yaron 2004; Bansal, Kiku, and Yaron 2012). Moreover, adopting SSY’s specification would lead to severe identification problems in our setting. This is because SSY’s specification disconnects the conditional volatility of x_t from Δc_{t+1} while preserving its connection with pd_t . In other words, SSY’s conditional volatility of x_t acts much like our residual, and identifying both would be impossible.

Surplus consumption \tilde{s}_t is also identified by consumption growth, but is more “backward looking.” \tilde{s}_t is an AR(1)-like process

$$\begin{aligned} \tilde{s}_t &= \rho_s \tilde{s}_{t-1} + \lambda(\tilde{s}_{t-1})(\Delta c_t - \mathbb{E}_{t-1} \Delta c_t) \\ \lambda(\tilde{s}_{t-1}) &= \begin{cases} \exp(-\bar{s})\sqrt{1 - 2\tilde{s}_{t-1}} - 1, & \tilde{s}_t \leq \frac{1}{2}[1 - \exp(2\bar{s})] \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (6)$$

which is equivalent to Campbell and Cochrane (1999)’s habit process. This process means that surplus consumption is the average of past consumption growth and that habit is the average of past consumption levels (Campbell 2003; Chen 2017). In the robustness section (Section 5), we examine a process in which \tilde{s}_t responds to consumption growth itself rather than innovations. This alternative assumption has little effect on our results.

Unlike the other state variables, the residual is not identified by either consumption or dividends. It is simply an AR(1)

$$\begin{aligned} e_t &= \rho_e e_{t-1} + \eta_{e,t} \\ \eta_{e,t} &\sim N(0, 1) \quad \text{i.i.d.} \end{aligned} \quad (7)$$

e_t captures everything in market volatility that is *not* long-run growth, long-run volatility, or habit. The assumption of independent shocks in Equation (7) is important. This independence means that e_t captures a distinct element of market volatility that is orthogonal to long-run risks and habit. Thus, it is unlikely the e_t captures, say, an alternative specification for how surplus consumption responds to consumption growth, or non-linear effects.

2.2. Model Frequency and Data

For the baseline estimation, we assume the model frequency is annual, the same frequency as the data we use. This differs from the typical approach in the literature, which tests monthly models against annual data moments.

We choose this approach for two reasons. The first is that monthly consumption and dividends exhibit stark seasonality that is entirely unaccounted for by models. The enormous end-of-quarter boosts to dividend growth and spikes in consumption at the end of the year suggest that risk is properly understood at an annual horizon. Indeed, if monthly risk is relevant to agents in the economy, why would we observe such stark seasonality in equilibrium?

Moreover, modeling this seasonality is not a simple task. Simple deterministic month or quarter fixed effects do a poor job, leading to the Census Bureau's sophisticated X-13ARIMA-SEATS seasonal adjustment. As discussed in Ferson and Harvey (1992), the Census Bureau adjustments are forward-looking: They boost the current month's observation if the future months are high. This forward-looking and smoothed series is difficult to interpret in a model of consumption risk.

The second reason we use an annual model is that the robustness of asset pricing frameworks to changes in model frequency is an interesting question in itself. The annual frequency is particularly relevant, as annual data is far more accessible and uncontroversial. Indeed, monthly nondurable consumption is never directly observed, and instead is calculated by holding fixed shares observed every five years (Wilcox 1992). Time-aggregating a monthly model to the annual horizon is possible but dramatically increases the complexity of model evaluation (Schorfheide, Song, and Yaron Forthcoming).

Thus, we estimate the model using annual consumption, dividend, and stock price data from the Bureau of Economic Analysis (BEA) and the Center for Research on Security Prices (CRSP).⁵ Consumption is real non-durable and services consumption. Dividends and prices correspond to the CRSP index. The sample runs from 1929 to 2014.

⁵CRSP data is from Wharton Research Data Services (WRDS). wrds-web.wharton.upenn.edu/wrds/about/databaselist.cfm.

2.3. Bayesian Estimation

As our model contains several latent variables, it's important to fully utilize the data. Thus, we estimate the model using Bayesian MCMC methods. This approach also avoids the potentially contentious choice of moment conditions.

To evaluate the likelihood of our nonlinear model, we use a particle filter (Herbst and Schorfheide 2014). We also take advantage of the conditionally Gaussian nature of the model to adapt the filter, following Schorfheide, Song, and Yaron (Forthcoming). To estimate the model parameters, we embed the filter in a standard random-walk Metropolis-Hastings algorithm. Details of the particle filter and Metropolis-Hastings algorithms can be found in the Appendix, and code can be found at <https://sites.google.com/site/chenandrewy/>.

We fix some parameters outside of the estimation that are uninteresting or difficult to identify. The means of all observables μ_{pd}, μ_c, μ_d are fixed to be their sample means. \bar{s} and A_s are difficult to identify separately as they jointly determine the volatility of the habit contribution to the price-dividend ratio. Thus we chose $\bar{s} = \log 0.06$, close to the Campbell and Cochrane (1999) value. In unreported results, we estimate this parameter and find that it is poorly identified but does not affect the main results. Assuming alternative values of \bar{S} also did not have a significant impact on the main results.

2.4. Prior Parameters

To choose priors, we take the perspective of an econometrician who believes that long-run risks, habit, and other factors are real and contribute to the price-dividend ratio. That is, she believes there is a persistent component of expected consumption growth, that the volatility of long-run growth is persistent, and that surplus consumption is a persistent process that responds to consumption shocks. Moreover, she believes that these latent states help explain the price-dividend ratio.

She has no view, however, on the relative importance of these factors. Instead, she looks to the data to update her view. Thus, while her priors on the consumption process are normal and centered around values in the long-run risks literature (and similarly for the persistence of habit), her priors on the price-dividend coefficients are flat and diffuse.

These “open-minded believer” priors are important because of the ubiquity

of long-run risks and habits in the asset pricing literature. Moreover, these priors ensure that our results can be interpreted through the lens of long-run risks and habit.

The priors are laid out in detail in Table 1. The simple consumption and dividend priors are standard. The prior consumption volatility is on average 2.5% per year, dividends are expected to be 5 times more volatile than consumption, and consumption and dividend growth are expected to display some positive correlation.

Priors on long-run risk parameters are centered around annual versions of the monthly parameters in Bansal and Yaron (2004) (BY) and Bansal, Kiku, and Yaron (2012) (BKY). To annualize BY and BKY's parameters, we simulate their models, aggregate to the annual, and apply simple method of moments estimations to the aggregated simulations. Back of the envelope calculations also verify the conversions.

The prior persistence of long-run growth is centered around 0.85, close to the annualized persistences of 0.86 and 0.83 in BY and BKY, respectively. The dispersion of the persistence is rather small, at 0.05, as a high persistence is critical to the economics of the model, and our priors are those of a believer in long-run risks.

The prior relative volatility of long-run growth shocks φ_x is centered around 0.35, in the middle of the annualized values of 0.38 and 0.33 for BY and BKY, respectively. These values are roughly 12 times the monthly parameter value of about 0.04 in BY and BKY, due to the fact that long-run growth is highly persistent.⁶ The prior dispersion on φ_x of 0.10 is somewhat large, reflective of the disagreement in this parameter between BY and BKY.

Other long-run risk priors are chosen similarly. Prior distributions are centered around values in the literature, with dispersions that depend on the literature's agreement regarding the parameter.

Priors on the persistence of habit are based off of annualized values in Campbell and Cochrane (1999) and Bekaert and Engstrom (2017) of 0.87 and 0.96, respectively.

The persistence of the residual is uniform between -1 and 1, to be as diffuse

⁶Let $x_{m,t}$ be the monthly expected consumption growth and ϕ_m be its monthly AR(1) coefficient. Then expected consumption growth over the next year is approximately $x_t \approx (1 + \phi_m + \phi_m^2 + \dots + \phi_m^{11})x_{m,t}$.

as possible. Priors on the price-dividend coefficients impose the sign implied by theory, but are otherwise flat and diffuse, reflecting the uncertainty in the literature regarding the relative importance of the risk factors. The extreme bound on the coefficients are chosen so that each state variable can easily account for all of market volatility (assuming standard consumption growth parameters in the literature).

2.5. Posterior Parameter Estimates

The right half of Table 1 shows the posterior estimates. Full posterior distributions for each parameter are shown in Appendix A.6.

Beginning at the top of the table, the posteriors on simple consumption and dividend parameters are in line with estimates in the literature. The mean volatility of consumption innovations $\bar{\sigma}$ is 1.8% per year, and dividend innovations are roughly 6 times as volatile as consumption innovations.

The estimator finds that long-run risks persistence parameters are close to the values in the literature. The estimated persistence of long-run growth is 0.84, nearly identical to the prior mean parameter of 0.85. Indeed, the entire posterior distribution is very close to the prior (Appendix A.6), indicating that the data is not inconsistent with the belief in a persistent component of consumption growth. The estimated persistence of long-run volatility of 0.93 is also very close to its prior mean parameter of 0.92.

In contrast, the volatilities of long-run risks are revised up significantly. The mean posterior relative volatility of long-run growth is 0.47 compared to the prior value of 0.35. More notably, the relative volatility of long-run volatility's mean posterior of 0.23 is nearly twice the prior value of 0.10. We will return to these volatility estimates when we discuss the intuition behind our estimation (Section 3.1).

The estimated dividend loading on long-run growth of 1.8 is somewhat lower than the prior mean of 3.0, though there is a lot of posterior uncertainty in this parameter.

Moving down Table 1, habit persistence is estimated to be 0.88, close to the prior persistence of 0.90. Similar to the persistence of long-run growth, the posterior distribution is close to the prior, indicating that the data does not speak strongly against the belief that habit is persistent.

Table 1: Parameter Estimates

The table shows prior and posterior parameter estimates for the annual model

$$\begin{aligned}
 pd_t &= \mu_{pd} + A_x x_t + A_V \bar{\sigma}_t^2 + A_s \bar{s}_t + A_e e_t \\
 \Delta c_t &= \mu_c + x_{t-1} + \sigma_{t-1} \eta_{c,t} \\
 \Delta d_t &= \mu_d + \phi_x x_{t-1} + \phi_{\eta c} \sigma_{t-1} \eta_{c,t} + \varphi_d \sigma_{t-1} \eta_{d,t} \\
 x_t &= \rho_x x_{t-1} + \varphi_x \sigma_{t-1} \eta_{x,t} \\
 h_t &= \rho_h h_{t-1} + \sigma_h \eta_{h,t}, \quad \sigma_t = \bar{\sigma} \exp(h_t) \\
 \bar{s}_t &= \rho_s \bar{s}_{t-1} + \lambda(\bar{s}_{t-1})(\Delta c_t - \mathbb{E}_{t-1} \Delta c_t) \\
 e_t &= \rho_e e_{t-1} + \eta_{e,t}
 \end{aligned}$$

where pd_t is the log price-dividend ratio, Δc_t is consumption growth, Δd_t is dividend growth, and η_t 's are standard normal independent noise. Priors are truncated normal N_T (mean, std dev, lower bound, upper bound), normal N (mean, std dev), or uniform U (lower bound, upper bound). Posteriors are computed using annual consumption, dividend, and stock prices from 1929-2014, particle filter, and Metropolis Hastings. $\exp \bar{s} = 0.06$ is chosen outside of the estimation, as are μ_{pd} , μ_c , and μ_d , which are chosen to be their sample means. These parameter values are used in our price-dividend ratio decompositions (Figures 2-4).

Parameter	Prior	Posterior			
		Mean	5%	95%	
Simple Consumption and Dividends					
Vol of Cons Shock	$\bar{\sigma}$	$N_T(0.025, 0.010, 0, \infty)$	0.0181	0.0106	0.0288
Div Loading on Cons	$\phi_{\eta c}$	$N(2, 2)$	1.02	-0.721	2.63
Rel Vol of Dividends	φ_d	$N(5, 2)$	6.14	4.93	7.45
Long Run Risks					
Persist of LR Growth	ρ_x	$N_T(0.85, 0.05, 0, 1)$	0.843	0.767	0.914
Rel Vol of LR Growth	φ_x	$N_T(0.35, 0.10, 0, \infty)$	0.467	0.309	0.621
Div Load on LR Growth	ϕ_x	$N(3, 2)$	1.76	0.0167	3.81
Persist of log LR Vol	ρ_h	$N_T(0.92, 0.05, 0, 1)$	0.929	0.868	0.974
Rel Vol of LR Vol	σ_h	$N_T(0.10, 0.10, 0, \infty)$	0.233	0.152	0.324
Habit and Residual					
Persistence of Habit	ρ_s	$N_T(0.90, 0.10, 0, 1)$	0.882	0.733	0.994
Persistence of Residual	ρ_e	$U(-1, 1)$	0.962	0.912	0.994
Price-Dividend Coefficients					
LR Growth Coefficient	A_x	$U(0, 300)$	19.1	10.4	29.7
LR Vol Coefficient	A_V	$U(-3000, 0)$	-295	-751	-24.9
Habit Coefficient	A_s	$U(0, 10)$	0.194	0.0113	0.539
Residual Coefficient	A_e	$U(0, 2.3)$	0.134	0.102	0.169

The remaining parameter estimates are very different in nature, as they come from flat and diffuse priors. These diffuse priors let the data speak freely about the residual persistence and price-dividend coefficients.

The persistence of the residual is estimated to be very high at 0.96. This high persistence suggests that the residual is important: the portion of the price-dividend ratio that cannot be explained by the macro factors is very long-lived. Moreover, this persistence is very precisely estimated, with 95% of its posterior distribution above 0.91.

The remaining parameters are the price-dividend coefficients. The posterior coefficients for long-run risks are large and broadly consistent with the literature. The price-dividend ratio's loading on long-run growth $A_x = 19.1$ implies that a 1 percentage point increase in expected growth over the next year implies a notable boom in the price-dividend ratio of 19.1%. For comparison, we simulate the monthly Bansal and Yaron (2004) and Bansal, Kiku, and Yaron (2012) models and find that they imply an annual A_x of about 13 and 7, respectively. The posterior loading on long-run volatility $A_V = -295$ implies that a 1 percentage point increase in consumption volatility implies a 13.6% decline in the price-dividend ratio. For comparison, Bansal-Yaron and Bansal-Kiku-Yaron imply about a decline of about 23% and 32%, respectively.

The coefficient on surplus consumption $A_s = 0.194$, in contrast, is much smaller than implied by the literature. We find that an annual aggregation of the monthly Campbell and Cochrane (1999) model implies a much larger A_s of 2.0.

3. Estimation Diagnostics

Before we present the main results, we present a set of diagnostic checks. These diagnostics focus on the historical paths of our estimated latent states, as these paths determine the price-dividend ratio decompositions. Section 3.1 visually inspects the historical state paths. Section 3.2 quantitatively compares the forecasting power of our historical state paths with the power implied by classic long-run risks and habit models. Overall, the diagnostics show that our estimated latent states capture key properties of long-run risks, habit, and economic history.

3.1. Historical State Paths and Observables

We begin our diagnostic checks by comparing the estimated paths of latent states with observables in the data. This kind of analysis is rare in consumption based asset pricing, as standard moment-matching methods do not provide paths of latent states.⁷ We use a particle smoother (at mean posterior parameters) to extract latent paths. Details on the particle smoother can be found in Appendix A.4.

Figure 1 examines the estimated paths of our four states. The top left panel compares the estimated path of long-run growth (green line) with its observable counterpart, consumption growth (blue crosses). Long-run growth moves with business cycles, and also tracks long-term shifts in economic growth. Long-run growth rises as the economy recovers from the Great Depression, falls during the productivity slowdown of the 1970's, and once again falls in advance of the Great Recession.

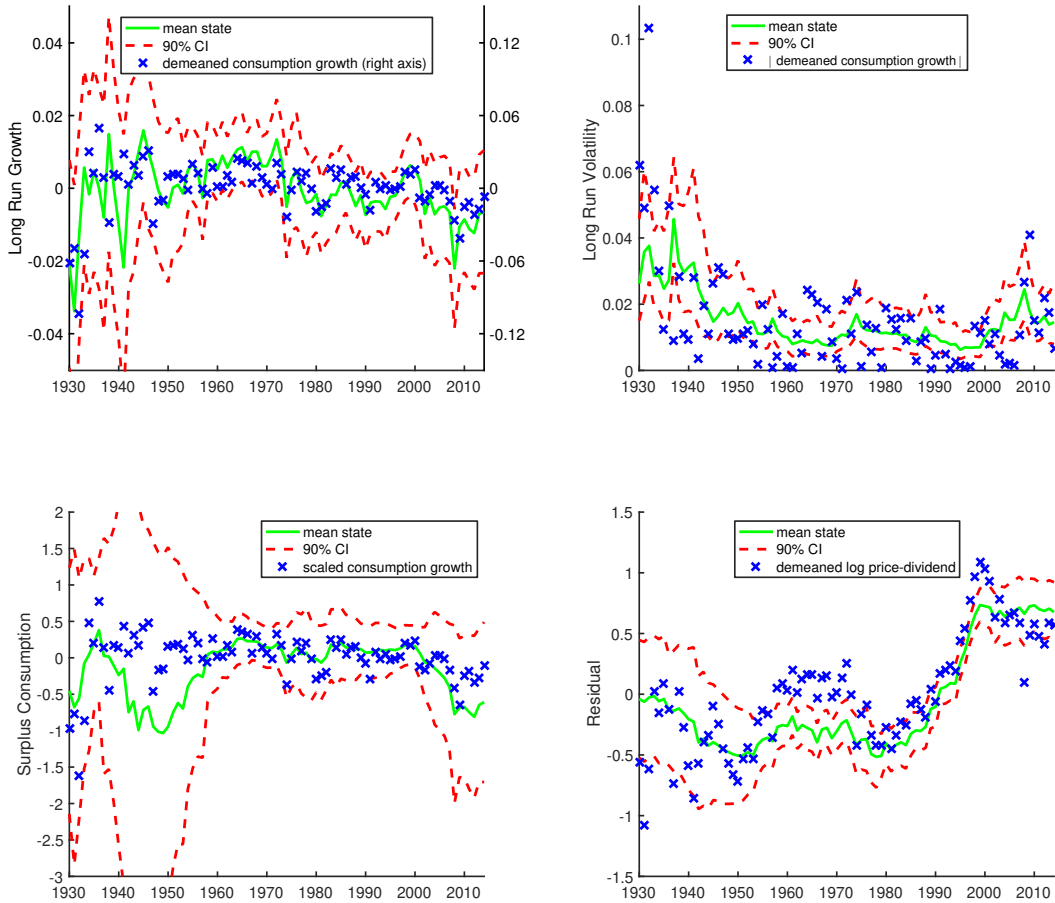
The top right panel of Figure 1 examines long-run volatility. For comparison, we plot the absolute value of demeaned consumption growth as a simple proxy for consumption volatility. The estimated volatility state tracks the history of consumption volatility very well. Consumption volatility is exceptionally high in the pre-war period, drops to unusual lows in the late 1950s and 1990s, and spikes during the crises of 1973 and 2009.

Figure 1 also shows why both long-run risks are estimated to be volatile. Large changes in expected growth and volatility help explain the extreme consumption growth observations of the Great Depression and the Great Recession. Indeed, these large changes also help explain the stock market crashes during these crises. The resulting long-run risks processes, however, are significantly more volatile than those in the long-run risks literature, leading to the choppy pattern in long-run growth alluded to earlier.

The bottom left panel of Figure 1 shows the historical path of surplus consumption along with scaled consumption growth. Consumption growth is scaled by subtracting out its mean, and then multiplying by the steady state $\lambda(s_t)$, to imitate the “shock” term in the habit process (6). The path of surplus consumption is intuitive. Surplus consumption is a persistent process that slowly responds to consumption growth. As such, it moves far from its steady

⁷Schorfheide, Song, and Yaron (Forthcoming) is a notable exception.

Figure 1: Smoothed States and Observables We apply a particle smoother to data on consumption, dividends, and stock prices, using mean posterior parameters (Table 1). Crosses plot observables for comparison. Scaled consumption growth in the lower left panel is demeaned consumption growth multiplied by the steady state of $\lambda(s_t)$ (see Equation (6)). The estimated latent states are intuitive given observables.



state during the large shocks of the Great Depression and Great Recession, but otherwise is relatively constant.

The residual, shown in the bottom right panel, tracks low-frequency movements in the price-dividend ratio. This behavior is natural, given that long-run growth, volatility, and surplus consumption tend to move with the business cycle. With business-cycle frequency components removed, the residual contains only low-frequency components.

3.2. A Comparison of Forecasting Power with the Literature

In our second diagnostic, we take a quantitative look at the estimated latent states. We examine the R^2 s from forecasting regressions using estimated states and historical data, and compare to R^2 s implied by simulations of Bansal and Yaron (2004) (BY), Bansal, Kiku, and Yaron (2012) (BKY), and Campbell and Cochrane (1999) (CC).⁸

The Forecasting Power of Long-Run Risks Table 2 examines the forecasting power of estimated long-run growth and volatility. Panel A examines the ability of long-run growth to forecast the following LHS variables: consumption growth, dividend growth, excess returns, and the risk-free rate. The “estimate” columns use the estimated path for long-run growth along with empirical data for the LHS variables. To help understand the identification, we examine an estimation which does not use asset prices (no P/D) in addition to our baseline (with P/D). For comparison, the “model implied” columns use simulated long-run growth and LHS variables from either the Bansal and Yaron (2004) (BY) or Bansal, Kiku, and Yaron (2012) (BKY) calibrations of the long-run risks model.

Panel A shows that the estimated long-run growth state displays the key property of long-run growth risk: long-run growth strongly predicts consumption growth. Indeed, the R^2 s are similar in magnitude to those implied by either BKY or BKY. Moreover, this predictive power is robust. Very large R^2 s are found regardless of whether or not asset prices are used in the estimation.

Estimated long-run growth also shows strong predictability for dividend growth, as long as the P/D is used in the estimation. This result is consistent with Beeler and Campbell (2012) and Bansal, Kiku, and Yaron (2012), who find

⁸We thank an anonymous referee for this helpful suggestion.

that dividend predictability is a subtle issue that is sensitive to the model specification. The BY and BKY models reflect different spectrums of this issue, with the BY model implying much stronger predictability. Our empirical results suggest that the BY model is closer to the data. The data show strong dividend predictability, but identifying it requires relatively sophisticated statistics (like our particle smoother).

Moving down the table, estimated long-run growth shows noticeable predictive power for excess returns. This differs from BY and BKY, which show little predictive power.

This apparent discrepancy is likely due to a simplifying assumption used in BY and BKY. For simplicity, BY and BKY both assume long-run independence between growth and volatility. As a result, times of high risk premiums in BY and BKY are entirely determined by volatility (see Panel B) and are disconnected from growth.

The empirical data, however, show a strong negative correlation between growth and volatility (Bloom 2009, for example). Our particle smoother cannot ignore this correlation, despite the fact that we also assume long-run independence. As a result, the small-sample paths of growth and volatility are negatively correlated, as can be seen in Figure 1, leading to forecasting power for both growth and volatility.

Finally, estimated long-run growth show little forecasting power for the risk free rate, in contrast to the extremely strong forecasting power implied by BY and BKY. The fact that the BY and BKY models imply excessive predictability of the risk-free rate is documented by Beeler and Campbell (2012). Thus, we view this discrepancy as an area for improvement in the classic models. Indeed, Schorfheide, Song, and Yaron (Forthcoming) introduce time-varying patience to help resolve this problem.

Panel B of Table 2 examines the properties of long-run volatility. We use long-run volatility to forecast observed volatility, where observed volatility is proxied using the absolute value of demeaned consumption growth ($|\Delta c_{t+j} - \mathbb{E}(\Delta c_{t+j})|$). Dividend volatility is measured analogously.

Panel B shows that estimated long-run volatility forecasts consumption and dividend volatility well. The R^2 s are large, at about 30-50%, and these large R^2 s are found regardless of whether asset prices are used in the estimation. These

Table 2: Estimation Diagnostics: Long-Run Risks

Regressions are of the form $\sum_{j=1}^H \text{LHS}_{t+j} = \alpha + \beta[\text{Latent State}]_t + \epsilon_t$. “Estimate” uses LHS variables from U.S. data and latent states extracted by particle smoother at mean posteriors. “Model Implied” uses simulated data from either the Bansal and Yaron (2004) (BY) calibration of the long-run risks model or the Bansal, Kiku, and Yaron (2012) (BKY) calibration. The consumption vol proxy is $|\Delta c_{t+j} - \mathbb{E}(\Delta c_{t+j})|$ and the dividend vol proxy is defined analogously. The estimated latent states capture key empirical properties of long-run risks.

Panel A: R^2 from Forecasting LHS with Long-Run Growth (%)									
H	LHS = Consumption Growth				LHS = Dividend Growth				
	Estimate		Model Implied		Estimate		Model Implied		
	With P/D	No P/D	BY	BKY	With P/D	No P/D	BY	BKY	
1 Y	46.3	26.6	26.8	18.6	24.2	4.0	23.9	3.5	
3 Y	31.3	18.7	34.3	24.6	17.6	2.2	28.6	6.0	
5 Y	19.2	10.1	29.2	20.8	10.1	1.4	24.2	5.9	

H	LHS = Excess Returns				LHS = Risk-Free Rate			
	Estimate		Model Implied		Estimate		Model Implied	
	With P/D	No P/D	BY	BKY	With P/D	No P/D	BY	BKY
1 Y	2.0	0.2	0.6	0.9	3.2	1.9	77.9	64.8
3 Y	15.9	6.0	1.6	1.6	2.2	0.2	53.3	40.2
5 Y	23.7	8.8	2.4	1.9	0.4	0.0	37.1	25.8

Panel B: R^2 from Forecasting LHS with Long-Run Volatility (%)									
H	LHS = Cons Vol Proxy				LHS = Div Vol Proxy				
	Estimate		Model Implied		Estimate		Model Implied		
	With P/D	No P/D	BY	BKY	With P/D	No P/D	BY	BKY	
1 Y	34.0	29.2	1.5	9.4	47.5	33.7	2.5	10.0	
3 Y	37.1	37.3	3.6	22.1	50.1	55.4	4.1	23.4	
5 Y	46.1	42.7	4.4	30.2	55.5	62.4	5.6	30.5	

H	LHS = Excess Returns				LHS = Risk-Free Rate			
	Estimate		Model Implied		Estimate		Model Implied	
	With P/D	No P/D	BY	BKY	With P/D	No P/D	BY	BKY
1 Y	0.37	0.04	0.83	1.81	3.98	3.88	4.59	19.43
3 Y	5.27	3.87	2.18	5.28	16.65	14.33	4.24	19.91
5 Y	13.09	8.83	3.73	7.85	33.38	25.12	5.07	19.57

large R^2 s support the BKY calibration of long-run risks, which produces similarly large R^2 s. In contrast, the BY model implies that volatility is largely not predictable.

Moving down the panel, forecasts of asset prices also support the BKY calibration over the BY calibration. Estimated long-run volatility has strong predictive power for both excess returns and the risk-free rate. A similarly large R^2 s are produced by the BKY model, while the BY model produces little predictability.

The Forecasting Power of Surplus Consumption Table 3 examines the forecasting power of surplus consumption. Overall, the table shows that estimated surplus consumption has similar properties to the classic Campbell and Cochrane (1999) (CC) model, as long as the price-dividend ratio is used in the estimation. Using all available information, estimated surplus consumption has little forecasting power for consumption growth, dividend growth, or the risk-free rate, but has strong power for forecasting excess returns. These are precisely the key properties of the CC model. Indeed, the CC model was essentially designed to have these properties.

Interestingly, if asset prices are omitted from the estimation, estimated surplus consumption is very unlike the CC model. If asset prices are ignored, surplus consumption has far too much forecasting power for consumption growth and too little power for excess returns compared to the CC model.

These results are consistent with other empirical facts about consumption growth and stock returns. If asset prices are ignored, surplus consumption becomes an average of historical consumption growth. Historical consumption growth has forecasting power, as consumption growth has an annual AR1 coefficient of 0.45 (Bansal, Kiku, and Yaron 2012, for example). Similarly, historical consumption growth is not known to have significant forecasting power for excess returns.

Overall, Tables 2 and 3 show that our estimation captures key properties of long-run risks and habit. Taken with our other diagnostic checks, the estimation overall does a good job of synthesizing a variety of consumption-based models in a single framework, and extracting their latent states from historical data.

Table 3: Estimation Diagnostics: Surplus Consumption

Regressions are of the form $LHS_t = \alpha + \beta[\text{Surplus Consumption}]_t + \epsilon_t$. “Estimate” uses LHS variables from U.S. data and surplus consumption is extracted by particle smoother at mean posteriors. “Model Implied” uses simulated data from the Campbell and Cochrane (1999) model (CC). “NA” entries are due to the fact that the risk-free rate is constant in the CC model. The estimated surplus consumption state captures key properties of habit, as long as asset prices are used in the estimation.

R^2 from Forecasting LHS with Surplus Consumption (%)						
H	LHS = Consumption Growth			LHS = Dividend Growth		
	Estimate		Model Implied	Estimate		Model Implied
	With P/D	No P/D	CC	With P/D	No P/D	CC
1 Y	5.2	25.9	0.7	0.2	6.8	0.6
3 Y	5.6	16.1	2.1	1.3	0.9	1.6
5 Y	8.2	10.4	3.1	2.7	0.4	2.3

LHS = Excess Returns							LHS = Risk-Free Rate		
H	Estimate		Model Implied	Estimate		Model Implied			
	With P/D	No P/D	CC	With P/D	No P/D	CC			
	1 Y	4.67	0.03	16.90	0.05	1.94	NA		
3 Y	12.77	3.25	37.97	0.44	0.13	NA			
5 Y	24.44	1.31	50.21	2.85	0.14	NA			

4. Results: Price-Dividend Ratio Decompositions

Having explained our parameter estimates and checked that our latent states capture key properties of the classic models, we can now address the main questions of the paper: How well do long run-risks and habit explain the time-series of price/dividends? What features do they explain? What is the nature of the residual that they leave behind?

4.1. Price-Dividend Ratio Decompositions

Figure 2 shows our first main result. It decomposes the historical path of the price-dividend ratio into contributions from long-run risks, habit, and a residual. The decomposition follows naturally from the linear price-dividend ratio (Equation (2)) in our semi-structural model (Section 2.1).

The figure shows that the residual (yellow bars) plays a dominant role in market volatility between 1929 and 2014. The residual is responsible for the relatively low asset prices of the 1940s, the boom of the 1950s, the bear market of the 1970s, the epic rise of valuations between 1980 and 2000, and the sharp crash in the early 2000s. Indeed, the residual closely tracks the price-dividend ratio (blue line) for the vast majority of the sample.

Long-run growth (green bars) plays a non-trivial role. It depresses asset prices during the Great Depression, boosts them during World War 2 and the 1960's, and drags prices down in advance of the Great Recession.

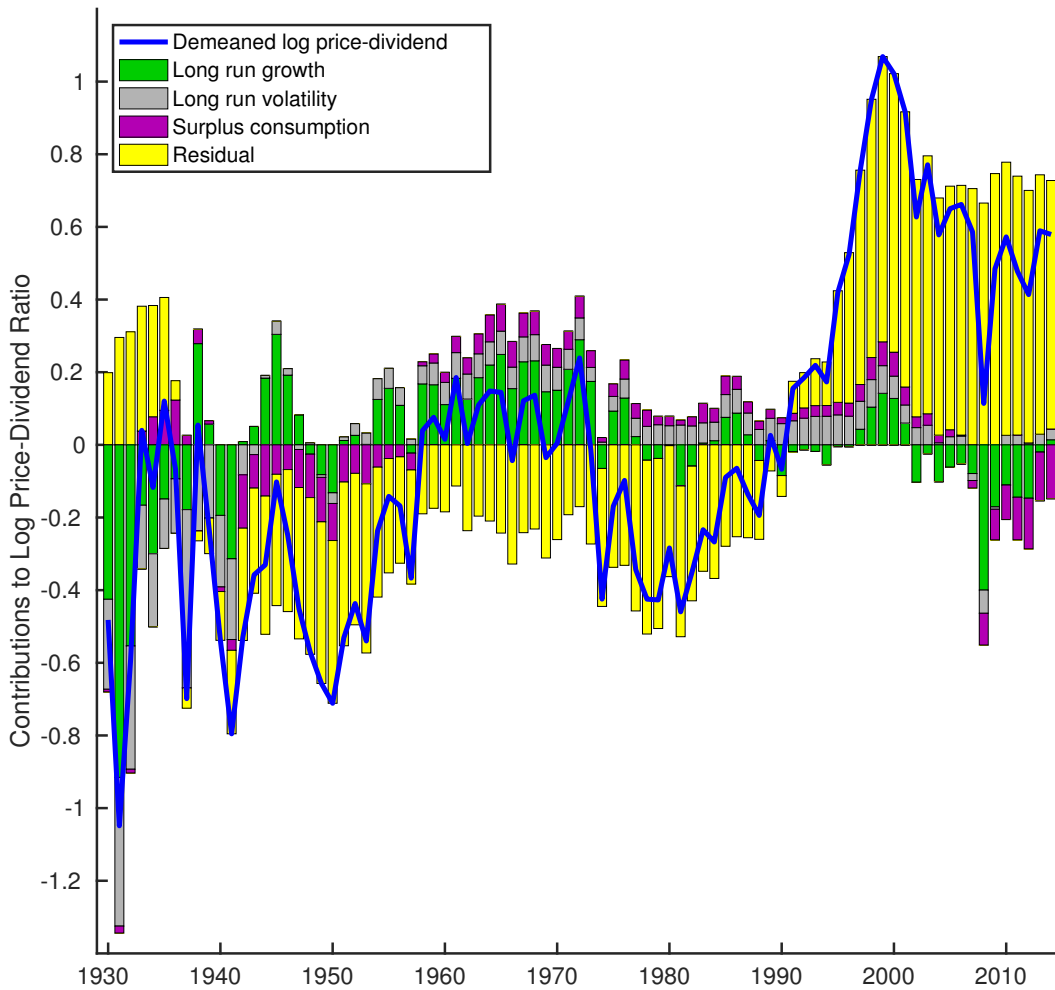
Long-run growth's contributions are smaller than the residual's, however. The residual's range is huge, spanning nearly 2 log points, roughly twice the range of long-run growth. Moreover, long-run growth often moves in the opposite direction of price/dividends, disconnecting it from the narrative of stock market history.

Long-run volatility (gray bars) has only a limited effect on asset prices. Its contribution tracks well-known events in economic history, exerting a downward force in the volatile early part of the sample before boosting asset prices during the Great Moderation. The magnitudes are small, however, and long-run volatility has little ability to explain major features of the price-dividend ratio such as the boom of the 1980's and 1990's.

Figure 2: Decomposition of the Historical Price-Dividend Ratio. The figure plots the decomposition implied by equation (2)

$$pd_t = \mu_{pd} + A_x x_t + A_v \tilde{\sigma}_t^2 + A_s \tilde{s}_t + A_e e_t$$

where pd_t is log price/dividends, x_t is long-run growth, $\tilde{\sigma}_t$ is long-run volatility, \tilde{s}_t is surplus consumption, and e_t is the residual. The states are the mean state found by particle smoother using mean posterior parameters from a Bayesian estimation (Table 1). Coefficients are the mean posterior coefficients. All states are adjusted for their in-sample means. The residual contribution (yellow) is dominant and closely tracks the price-dividend ratio.



This result sheds light on why the most recent version of the long-run risks model (Schorfheide, Song, and Yaron Forthcoming) assumes independent volatility processes for short run consumption growth shocks and shocks to long-run growth. Without this complication, long-run volatility would be cleanly identified by consumption growth, leading to a minor role for time-varying risk, a large role for long-run growth, and excessive consumption growth predictability. Schorfheide, Song, and Yaron (Forthcoming) rescue the predictability implications of their model by adding volatility processes, but this complication disconnects their key driver of asset prices from fundamental observables. Indeed, their volatility of long-run growth behaves much like an inverted version of our residual, rising and falling with the booms and busts of the dividend yield, independent of the path of consumption volatility.

Habit's (purple bars) impact is similarly small. Its effect is just barely noticeable in a few portions of the sample (before 1960, and after the 2008 Financial Crisis) and has a negligible effect otherwise.

To understand these results, it helps to examine large increases valuations during the 1950s and 1980s. These two episodes provide a nice illustration of how our particle smoother identifies the roles of the latent states using the observable data seen in Figure 1.

During the 1950s boom, consumption growth displays just a hint of a similar increase (Figure 1, top left panel). Indeed, consumption growth falls sharply in the early 1970s, just as the boom ends. These changes in consumption growth are quite subtle, but taken with evidence from asset prices our estimator then concludes that long-run growth experienced a boom during this episode, as seen in the green bars of Figure 2.

In contrast, the dramatic boom of that begins in the 1980s does not come with even a hint of a rise in observed consumption growth. Weighing the sharp rise in valuations against the stoically flat path of consumption growth, the estimator concludes that long-run growth played no role in the 1980's stock market boom. Indeed, growth generally declined over the sample period while the price-dividend ratio trended upward, making it hard for the estimator to give long-run growth a large role in market volatility.

Movements in consumption volatility (Figure 1, top right panel) played a relatively minor role in either the 1950s or 1980s. Indeed, observed consumption volatility was relatively high during the 1960s, when the boom of the 1950s

reached its peak, and overall volatility steadily declines from 1950-2000 while asset prices experienced dramatic booms and busts. Instead, consumption volatility displays most of its dynamics during the end of the Great Depression. As a result, the estimator finds that long-run volatility largely plays a large role during the early part of the sample (Figure 2, grey bars).

In Section 2.5 we found that the price-dividend loadings on long-run risks are similar in magnitude to the literature. How, then, do long-run risks play a minor role in the price-dividend ratio? Despite the strong sensitivity of price/dividends to long-run risks, long-run risks have not, in general, moved in the direction of asset prices over the historical sample. Indeed, over much of the sample asset prices moved in the opposite direction of those implied by long-run risks, leading to the large residual seen in Figure 2.

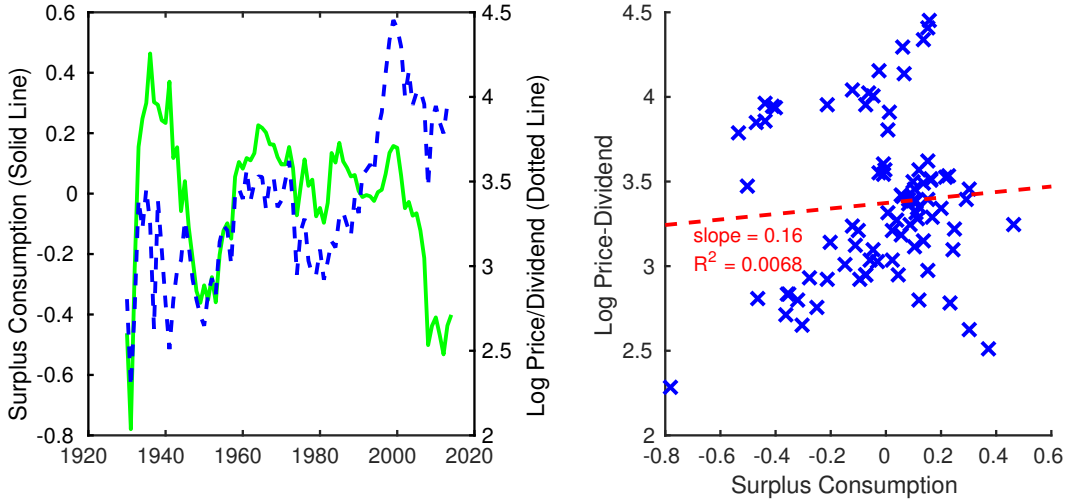
The small role for surplus consumption comes from its relatively strict definition. While innovations to long-run risks are unobserved, innovations in surplus consumption are perfectly correlated with innovations in consumption growth (Equation (6)). As a result, surplus consumption is simply a persistent process that slowly responds to consumption growth, and thus significantly departs from its steady state only during crisis periods (Figure 1, bottom left panel). As a result, it explains a portion of price-dividend fluctuations during the pre-war period and the Great Recession, but plays little role otherwise.

Some readers may be surprised by the poor performance of surplus consumption. In particular, Campbell and Cochrane (1999) and Cochrane (2017) show historical time-series plots of surplus consumption and pd_t that visually suggest that surplus consumption explains pd_t well, at least during some subsamples of the data.

The left panel of Figure 3 provides similar visual evidence. The figure plots the smoothed surplus consumption state extracted by our estimation (solid line, left axis), overlaid with pd_t (dotted line, right axis). There are episodes in which the two series line up nicely, particularly during the rise in valuations in the beginning of the sample, and during the rise in the 1950s. It also appears that surplus consumption would fit the late 1990s Tech Boom and Bust and the 2008 Financial Crisis, if the surplus consumption could be shifted up in the latter half of the sample.

These visual suggestions of a good fit, however, are not found in more systematic evaluations. The right panel of Figure 3 illustrates an OLS regression of

Figure 3: A Closer Look at Surplus Consumption. The left panel overlays surplus consumption (solid line) and the price-dividend ratio (dotted line). Surplus consumption is the smoothed state estimated at posterior mean parameters, as in Figure 1. The right panel shows a scatter-plot of the two time-series. The dotted line shows an OLS regression. Though the time-series plot suggests some relationship during some periods, a scatterplot shows that the relationship is overall weak.



pd_t on surplus consumption. Taking all of the data together, the best fit line is relatively flat, and explains little variation in pd_t . Alternative specifications for surplus consumption also lead to a low R^2 , as do regressions of filtered pd_t on filtered surplus consumption (Appendix A.7).

Overall, Figure 2 shows that the residual (something other than long-run risks and habit) is the dominant factor in the history of the price-dividend ratio from 1929-2014. Indeed, the residual dominates despite the fact that the prior places a strong belief in the existence of long-run risks in consumption growth. Indeed, Section 5 shows that the residual dominates under a wide variety of alternative priors.

To place our results in numerical terms, we calculate a variance decomposition. This decomposition is found by taking the covariance of Equation (2) with pd_t

$$\begin{aligned} \text{Var}(pd_t) = & \text{Cov}(A_x x_t, pd_t) + \text{Cov}(A_V \tilde{\sigma}_t^2, pd_t) \\ & + \text{Cov}(A_s \tilde{s}_t, pd_t) + \text{Cov}(A_e e_t, pd_t). \end{aligned} \quad (8)$$

Equation (8) boils Figure 2 down to just 4 numbers. Note that negative variance shares, as well as shares in excess of 100%, are possible if a state variable has a negative sample correlation with pd_t .

In addition to boiling down the price-dividend decomposition, Equation (8) allows us to address the issue of estimation uncertainty—that is, how confident can we be about our decomposition given the short sample length and dispersion in the posterior parameters? To account for this dispersion, we repeatedly draw parameters from the posterior distribution and recalculate Equation (8) using these draws.

Figure 4 plots the resulting distribution of the variance shares. The residual's share is huge: on average 78%. There is some dispersion in this share, but we can be very confident that the residual share is larger than 60%.

Long-run risks and habit have small mean shares, accounting for the remaining 23% of market variance all together. Long-run growth's small 13% share is due to its tendency to move in the opposite direction of asset prices. Growth declined for most of the post-war period, while the price-dividend ratio increased, leading to a small covariance contribution. Long-run volatility has a similarly small share of 11%, though its weakness is simply due to the small magnitude of the estimated coefficient A_V . Surplus consumption's share is negligible.

It's important to note that long-run risks are identified using the price-dividend ratio. The price-dividend equation (2) and long-run risks equations (3)-(5) are estimated simultaneously. Estimating the long-run risks equations without using pd_t and then estimating the price-dividend equation in a second stage leads to no role at all for long-run risks (Section 4.3).

Alternative priors and model specifications lead to a similarly dominant residual share (Section 5). Indeed, this robustness can be traced to simple facts about consumption growth and asset prices, as we discuss next.

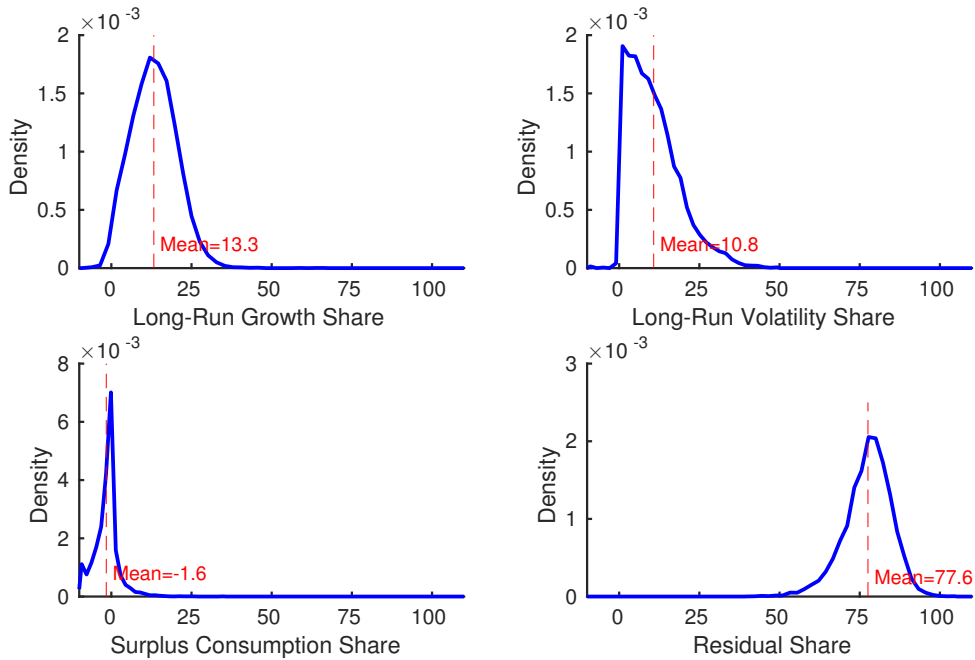
4.2. Decomposition of Changes in the Price-Dividend Ratio

Section 4.1 showed that long-run risks and habit explain very little of the variance of the price-dividend ratio. The intuition is straightforward: the historical booms and busts in the price-dividend ratio are largely not visible in the historical paths of consumption growth or consumption volatility. But we also

Figure 4: Price-Dividend Ratio Variance Decomposition and Estimation Uncertainty. Shares are in percent and implied by the decomposition

$$\begin{aligned} \text{Var}(pd_t) = & \text{Cov}(A_x x_t, pd_t) + \text{Cov}(A_V \tilde{\sigma}_t^2, pd_t) \\ & + \text{Cov}(A_s \tilde{s}_t, pd_t) + \text{Cov}(A_e e_t, pd_t). \end{aligned}$$

where pd_t is log price/dividends, x_t is long-run growth, $\tilde{\sigma}_t$ is long-run volatility, \tilde{s}_t is surplus consumption, e_t is the residual, and A 's are the coefficients from Equation (2). The densities are computed by drawing parameters from the posterior (Table 1), using the draw to find smoothed mean states, and calculating variance contributions according to the above equation. The plots show that the residual's dominant role is robust to estimation uncertainty.



have seen that the residual track mostly the lower-frequency movements of the price-dividend ratio (Figure 1).

In this section, we take a closer look at higher frequency movements. We show that expected growth plays a large role in explaining first differences of pd_t . This result can be interpreted as saying that expected growth is important for business-cycle frequency fluctuations in stock prices. Indeed, we find that very similar results hold using high-pass filtered pd_t (Appendix A.7).

Taking first differences of the price-dividend equation (2) leads to the following decomposition:

$$\Delta pd_t = A_x \Delta x_t + A_V \Delta \tilde{\sigma}_t^2 + A_s \Delta \tilde{s}_t + A_e \Delta e_t. \quad (9)$$

To motivate this decomposition, one can think of the price-dividend ratio is the sum of a long series of first differences Δpd_t . Thus, Δpd_t can be thought of as a high frequency component of price/dividends.

To take Equation (9) to the data, we use the same smoothed historical paths as in our decomposition of the level of pd_t (Figure 2). Taking first differences leads to Figure 5.

Figure 5 shows that, in stark contrast to the price-dividend level decomposition, the residual (yellow) plays a minor role in first differences. Instead, long-run growth (green) dominates the chart, accounting for the majority of changes in the price-dividend ratio in nearly all years outside of the late 1990's tech boom.

The two decompositions do agree about the minor roles of long-run volatility and surplus consumption, however. These states show up only during the Great Depression and Great Recession, and even during these crisis events they are not dominant.

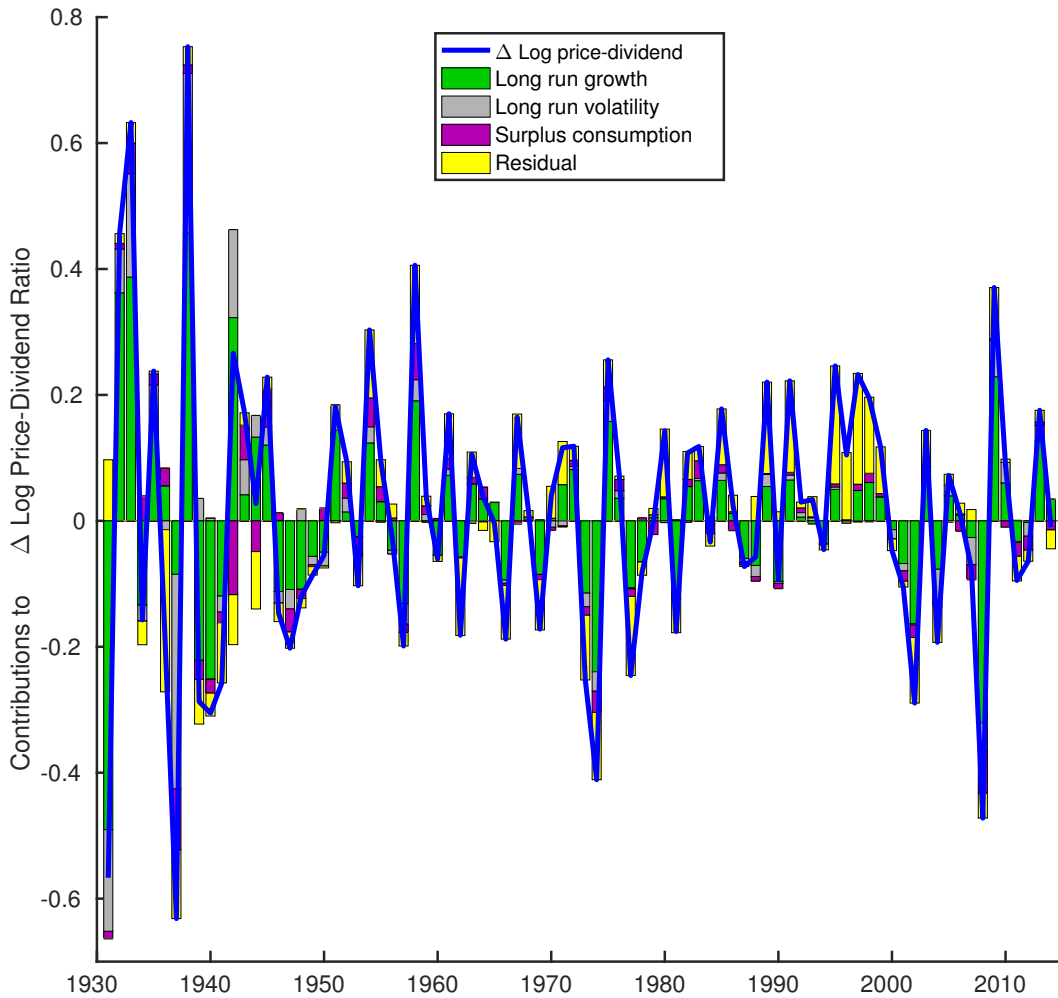
The tight link between first differences in pd_t and expected consumption growth is consistent with the well-known fact that the stock market is a leading indicator (Mitchell and Burns 1938). Indeed, regressions of future real growth on today's stock return produce R^2 's of up to 50% (Fama 1981; Schwert 1990; Cochrane 2011). In contrast, stock *valuations* are unconnected with future growth (Shiller 1981), consistent with our baseline price-dividend decomposition.

The stark change in the decomposition can also be understood from the

Figure 5: Decomposition of First Differences of Price/Dividends. The figure plots the decomposition implied by

$$\Delta pd_t = A_x \Delta x_t + A_V \Delta \tilde{\sigma}_t^2 + A_s \Delta \tilde{s}_t + A_e \Delta e_t$$

where pd_t is log price/dividends, x_t is long-run growth, $\tilde{\sigma}_t$ is long-run volatility, \tilde{s}_t is surplus consumption, and e_t is the residual. States are the mean state found by particle smoother using mean posterior parameters (Table 1). Coefficients are the mean posterior coefficients in Table 1. While expected growth is unimportant for explaining the level of asset valuations (Figure 2), it is important for first differences.



smooth and persistent nature of the residual. The residual is much like the price-dividend ratio with short-term fluctuations stripped away (Figure 1). Thus, it is nearly a random walk with small increments, and first differencing reduces it to a series of small i.i.d. shocks.

Using pd_t to identify expected consumption growth and expected volatility is essential for this result. This identification runs through the price-dividend equation (2) and long-run risks equations (3)-(5), which say that a rise in pd_t suggests an increase in expected growth and a decrease in expected volatility. Estimating the path of long-run risks without using pd_t and then estimating the price-dividend equation in a second stage leads to no role at all for long-run risks (Section 4.3).

As in the baseline estimation, one may be concerned about the statistical power in our relatively small sample of 85 annual observations. Figure 6 should alleviate such concerns.

The figure is based on the following variance decomposition of first differences in the price-dividend ratio:

$$\begin{aligned} \text{Var}(\Delta pd_t) = & \text{Cov}(A_x \Delta x_t, \Delta pd_t) + \text{Cov}(A_V \Delta \tilde{\sigma}_t^2, \Delta pd_t) \\ & + \text{Cov}(A_s \Delta \tilde{s}_t, \Delta pd_t) + \text{Cov}(A_e \Delta e_t, \Delta pd_t). \end{aligned} \quad (10)$$

Analogous to our baseline variance decomposition (Figure 4), this decomposition comes from taking covariance of both sides of Equation (9) with Δpd_t .

Figure 6 plots the posterior distribution of variance shares implied by Equation (9). The figure shows that we can be confident that the residual's share of variance is sharply reduced, despite the short sample. Almost the entire distribution of the residual share is less than 50%, in contrast to the price-dividend variance decomposition (Figure 4), where the vast majority of the distribution lies above 60%.

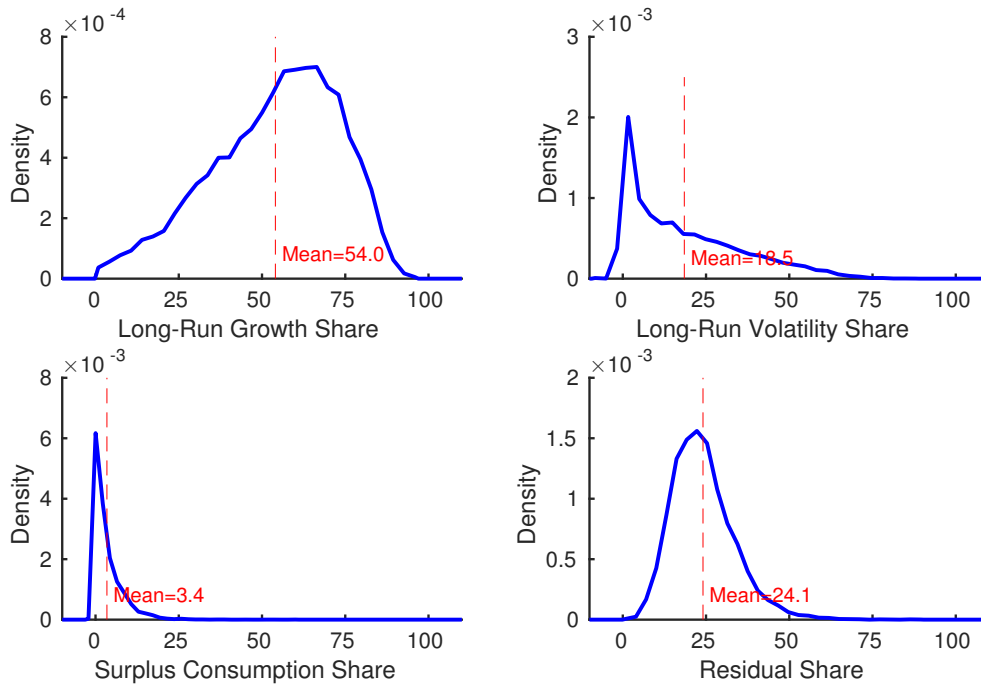
Long-run growth's share is less precisely estimated, but the vast majority of the distribution lies above 25%. This result implies that expected economic growth plays at least a noticeable role in first differences of pd_t . Indeed, on average the posterior implies a large share of 54%, in line with the 50% R^2 's that obtain from forecasting annual GDP growth with stock returns.

Long-run volatility has a smaller but non-negligible 19% share. The distribution has a long right tail, but we can still be quite confident that it accounts

Figure 6: Δpd_t Variance Decomposition and Estimation Uncertainty. Shares are in percent and implied by the decomposition

$$\begin{aligned} \text{Var}(\Delta pd_t) = & \text{Cov}(A_x \Delta x_t, \Delta pd_t) + \text{Cov}(A_V \Delta \tilde{\sigma}_t^2, \Delta pd_t) \\ & + \text{Cov}(A_s \Delta \tilde{s}_t, \Delta pd_t) + \text{Cov}(A_e \Delta e_t, \Delta pd_t) \end{aligned}$$

where pd_t is log price/dividends, x_t is long-run growth, $\tilde{\sigma}_t$ is long-run volatility, \tilde{s}_t is surplus consumption, e_t is the residual, and A 's are the coefficients from equation (2). The densities are computed by drawing parameters from the posterior (Table 1), using the draw to find smoothed mean states, and calculating variance contributions according to the above equation. The residual's relatively minor role in Δpd is robust to estimation uncertainty.



for less than 50% of the variance of Δpd_t .

As in the price-dividend variance decomposition, surplus consumption is negligible. The bulk of its distribution is below 10%. The negligible role of surplus consumption is consistent with the fact that stock returns are essentially uncorrelated with consumption growth.

4.3. Comparison with Two-Stage Estimations

Our baseline results use pd_t to identify the latent states. This identification comes from the fact that the pd_t equation (2) is estimated simultaneously with the consumption and dividend growth equations (3). Thus, our estimation does more than just regress pd_t on long-run risks and habit.

To illustrate the importance of simultaneous equation estimation, this section presents variance decompositions from two-stage estimations. These two-stage estimations first extract latent states using only consumption and dividend data (with the same priors as in our baseline estimation). In the second stage, we regress pd_t on the smoothed paths for long-run risks and habit from the first stage. Unlike our simultaneous estimation, these two-stage estimations find no role for long-growth, even at business cycle frequencies.

Table 4 shows variance decompositions for pd_t and Δpd_t using two-stage estimation. The table shows two methods for performing the second stage regression: (1) simple OLS, and (2) maximum likelihood regression that allows for an AR1 residual.

Panel A shows that two-stage estimations lead to an even larger role for the residual in $\text{Var}(pd_t)$. The OLS regressions imply an 85% share for the residual, compared with 82% in the baseline. Allowing for an AR1 term in the regression leads to a 101% share for the residual.

The importance of using pd_t to identify long-run risks is seen even more clearly in Panel B, which shows $\text{Var}(\Delta pd_t)$ decompositions. Here we see that using the two-stage approach leads to a nearly 100% residual using either OLS or an AR1 regression, compared to our baseline residual share of 19%.

These results highlight the importance of using asset prices to identify long-run risks. Without using asset prices to identify economic growth, one would arrive at the conclusion that market volatility and growth are completely unrelated. Long-run risks that are identified by asset prices, however, still lead to

$\text{Var}(pd_t)$ being dominated by a residual. Thus, there is a large long-run component of the price-dividend ratio that requires a different theory.

Table 4: Comparison with Two-Stage Regression Estimates

‘Baseline’ applies a particle smoother to the data using mean posterior parameters from the main estimation (Table 1). ‘OLS’ uses the baseline estimation but omits pd_t to extract paths for long-run risks and habit, and then runs OLS to estimate the pd_t equation (2). ‘AR1 Reg’ is the same as ‘OLS,’ but estimates the pd_t equation with an AR(1) error term in the second stage. The ‘OLS’ and ‘AR1 Reg’ estimates lead to huge residual shares in $\text{Var}(\Delta pd_t)$, as they do not use pd_t to identify long-run risks.

Panel A: $\text{Var}(pd_t)$ Shares			
	Baseline	OLS	AR1 Reg
Long-Run Growth	7.36	-0.99	1.10
Long-Run Volatility	10.14	16.21	-2.62
Surplus Consumption	0.86	0.00	0.07
Residual	81.64	84.78	101.44

Panel B: $\text{Var}(\Delta pd_t)$ Shares			
	Baseline	OLS	AR1 Reg
Long-Run Growth	54.16	-2.59	0.62
Long-Run Volatility	20.03	-7.74	1.71
Surplus Consumption	5.85	3.66	0.92
Residual	18.64	106.68	96.75

5. Alternative Full-Information Decompositions

As our findings come from basic features of the historical path of consumption and asset prices, our results are quite robust. All that is required is that we (1) allow for a residual in the price-dividend equation and (2) estimate the historical latent states. These two features of our decomposition are rarely found in the literature, which never allows for omitted influences on asset prices and almost always studies only simulated latent states.

To illustrate this robustness, this section examines four alternative full information decompositions that differ significantly from our baseline estimate.

Section 5.1 estimates a model with structural restrictions implied by an Epstein-Zin-Habit model, Section 5.2 uses quarterly post-war data, Section 5.3 uses diffuse priors, and Section 5.4 uses priors which assume that the prior variance shares are equal.

Table 5 summarizes the alternative decompositions. In this table, long-run growth x_t and long-run volatility $\tilde{\sigma}_t^2$ are renamed expected growth and expected volatility, as the diffuse prior no longer assumes that they are persistent.

Panel A shows price-dividend variance shares across the five specifications listed above, as well as the baseline. Under all six specifications, the residual accounts for the dominant majority of market volatility, with a minimum share of 75%. Moreover, the estimation uncertainty is similar under all specifications, with all priors implying confidence in the importance of the residual.

The same robustness is seen in the variance shares of changes in the price-dividend ratio. Panel B shows the shares of Δpd_t attributed to each latent state. Expected growth and expected volatility dominate the variance of Δpd_t , regardless of the specification.

The remainder of this section describes the alternative estimations and provides additional results.

5.1. Structural Restrictions and Estimation of Preferences

Our first alternative estimation imposes structural restrictions on the price-dividend coefficients. Using an endowment economy with long-run risks and Epstein-Zin-Habit preferences (following Yang 2016), we derive the price-dividend coefficients as a function of preference and other deep parameters. To allow for the possibility that forces outside of this endowment economy drive asset prices, we continue to allow for a residual in the price-dividend equation. We place priors on these preferences and estimate them jointly with our price-dividend decompositions.

The representative investor has Epstein-Zin-Habit utility::

$$U_t = \left((1 - \delta)(C_t - H_t)^{1 - \frac{1}{\psi}} + \delta \mathbb{E}_t [U_{t+1}^{1-\gamma}]^{\frac{1}{1-\gamma} \left(1 - \frac{1}{\psi}\right)} \right)^{\frac{1}{1-\frac{1}{\psi}}}. \quad (11)$$

where δ is time preference, γ is risk aversion, ψ is the elasticity of intertemporal

Table 5: Robustness: Variance Shares Under Alternative Model Specifications

The table shows the mean and standard deviation (in parentheses) of the posterior distribution of variance shares. “Baseline” uses priors that believe in long-run risks and habit but places flat priors on the price-dividend coefficients (Table 1). “Struct restrict” imposes restrictions from an endowment economy ((11)-(13)) and estimates preferences simultaneously (Table 6). “Quarterly” uses quarterly post-war data. “Diffuse” uses uniform priors with wide bounds on all parameters (Table 7). “Eq share” imposes identical and normal price-dividend variance shares in the prior. “Expected growth” x_t and “expected volatility” $\tilde{\sigma}_t^2$ are not necessarily persistent under the “diffuse” specification. Shares are computed using mean smoothed states evaluated at 5,000 draws from the posterior parameter distribution. The variance decompositions are robust.

Panel A: Var($p d_t$) Shares					
	(1)	(2)	(3)	(4)	(5)
	Baseline	Struct Restrict	Quarterly	Diffuse	Eq Prior Share
Expected Growth	13.25 (7.04)	14.61 (7.16)	-0.95 (1.36)	10.21 (3.41)	11.52 (5.79)
Expected Volatility	10.76 (8.12)	6.17 (9.32)	3.57 (1.54)	11.71 (7.50)	14.55 (8.49)
Surplus Consumption	-1.58 (4.05)	-0.02 (3.06)	0.96 (2.55)	-1.99 (4.17)	-1.48 (5.08)
Residual	77.57 (7.36)	79.24 (7.16)	96.42 (3.06)	80.06 (7.78)	75.42 (8.48)

Panel B: Var($\Delta p d_t$) Shares					
	(1)	(2)	(3)	(4)	(5)
	Baseline	Struct Restrict	Quarterly	Diffuse	Eq Prior Share
Expected Growth	54.01 (7.04)	54.61 (19.46)	21.45 (15.94)	53.48 (3.41)	47.18 (5.79)
Expected Volatility	18.48 (8.12)	17.38 (18.00)	45.27 (17.71)	16.96 (7.50)	24.78 (8.49)
Surplus Consumption	3.42 (4.05)	0.8 (1.95)	9.58 (5.30)	8.46 (4.17)	3.95 (5.08)
Residual	24.09 (7.36)	27.22 (8.05)	23.71 (7.07)	21.1 (7.78)	24.09 (8.48)

substitution (EIS), and external habit H_t is defined by

$$(C_t - H_t)/C_t = \bar{S}^{1-\ell} S_t^\ell \quad (12)$$

and the capitalized variables C_t and S_t are in levels rather than logs ($C_t = \log c_t$). The parameter $\ell \in [0, 1]$ measures the strength of habit.

Consumption and dividends as well as the state variables x_t , h_t , s_t , and e_t

follow the same processes as in the semi-structural model (Equations (4) - (7)). Using log-linear approximations (Bansal and Yaron 2004; Yang 2016), one can show that the model-implied log price-dividend ratio z_t is given by

$$\begin{aligned}
z_t = & \left[\frac{\phi_x - \frac{1}{\psi}}{1 - \kappa_1 \rho_x} \right] x_t \\
& + \left[\frac{1}{2} \frac{1}{1 - \kappa_1 \rho_h} \left(\begin{aligned} & \left(\gamma - \frac{1}{\psi} \right) (1 - \gamma) \left((V_s \bar{\lambda} + \ell \bar{\lambda} + 1)^2 + V_x^2 \varphi_x^2 \right) \\ & + (-\bar{\xi}_c + \kappa_1 A_s \bar{\lambda} + \phi_{\eta c})^2 + \varphi_d^2 \\ & + (-\bar{\xi}_x + \kappa_1 A_x \varphi_x)^2 \end{aligned} \right) \right] \sigma_t^2 \\
& + \left[\frac{1 - \rho_s}{1 - \kappa_1 \rho_s} \frac{1}{\psi} \ell \right] s_t
\end{aligned} \tag{13}$$

where $\bar{\lambda} = \lambda(0)$, and V_x , V_s , κ_1 are functions of the structural parameters. The terms in square brackets correspond to A_x , A_V , and A_s in our semi-structural model. Derivations are relegated to Appendix A.1.

We then assume that the observed price-dividend ratio pd_t equals the sum of the model-implied price-dividend ratio and a residual:

$$pd_t = z_t + A_e e_t, \tag{14}$$

where e_t is an AR(1) process with standard normal shocks. The resulting expression is the same as in the baseline (Equation (2)), but with structural restrictions imposed.

The structural model contains a couple additional parameters compared to the semi-structural model. We set these parameters to standard values in the literature, before the estimation, to avoid identification problems. Namely, we set $\delta = 0.987$ to the annualized version of Bansal, Kiku, and Yaron's (2012) monthly $\delta = 0.9989$, and $\kappa_1 = \exp(\overline{pd}_t) / (1 + \exp(\overline{pd}_t)) = 0.9675$ to match the average log price-dividend ratio of 3.36 in our sample.

Table 6 shows the prior and posterior preference parameters from this estimation. Priors are chosen to be flat but encompass the range of values used in the literature. For example, since Mehra and Prescott (1985), values of γ that have been considered ranged from 0 to 10, matching our uniform prior that spans 0 to 10.

The estimated risk aversion and EIS are, overall, consistent with the litera-

Table 6: Structural Restrictions and Estimation of Preferences

We derive the price-dividend coefficients as a function of preferences and other parameters from an endowment economy (Equations (11)-(13)) and allow for a residual that captures risks missing from the endowment economy. We place priors on the preferences and estimate them jointly with other parameters. The posteriors are used in our robustness check for structural restrictions (Table 5).

Parameter		Prior	Posterior		
			Mean	5%	95%
Preferences					
Risk aversion	γ	U(0,10)	3.79	0.748	8.45
EIS	ψ	U(0,6)	3.39	1.02	5.73
Habit strength	ℓ	U(0,1)	0.0966	0.00737	0.289
Simple Consumption and Dividends					
Vol of Cons Shock	$\bar{\sigma}$	$N_T(0.025, 0.010, 0, \infty)$	0.0178	0.0106	0.0279
Div Loading on Cons	$\phi_{\eta c}$	$N(2, 2)$	0.19	-1.32	1.86
Rel Vol of Dividends	φ_d	$N(5, 2)$	5.67	4.35	7.05
Long Run Risks					
Persist of LR Growth	ρ_x	$N_T(0.85, 0.05, 0, 1)$	0.805	0.729	0.884
Rel Vol of LR Growth	φ_x	$N_T(0.35, 0.10, 0, \infty)$	0.479	0.338	0.626
Div Load on LR Growth	ϕ_x	$N(3, 2)$	3.9	1.98	5.84
Persist of log LR Vol	ρ_h	$N_T(0.92, 0.05, 0, 1)$	0.925	0.867	0.971
Rel Vol of LR Vol	σ_h	$N_T(0.10, 0.10, 0, \infty)$	0.232	0.153	0.319
Habit and Residual					
Persistence of Habit	ρ_s	$N_T(0.90, 0.10, 0, 1)$	0.856	0.721	0.968
Persistence of Residual	ρ_e	U(-1,1)	0.951	0.895	0.991
Residual Coefficient	A_e	U(0,2)	0.14	0.111	0.171

ture. Risk aversion γ is estimated to be 3.39, in line with standard values from the long-run risks literature. The EIS parameter ψ is estimated to be 3.39, noticeably larger than Bansal and Yaron's (2004) preferred value of 1.5, though Bansal and Yaron's preferred value is within the 90% credible interval. Moreover, more recent models calibrate a larger value of 2.0 (Barro 2009, Colacito and Croce 2011, Favilukis and Lin 2015), and microeconomic studies often estimate even larger values (Vissing-Jørgensen and Attanasio 2003, for example).

The strength of habit ℓ is estimated to be quite small, at 0.097, for from the 1.0 implied by Campbell and Cochrane (1999) and related models imply a value of 1. This small ℓ is consistent with our baseline finding that Campbell-Cochrane habit does a poor job explaining the price-dividend ratio (see Figure 3).

The remaining parameters are largely similar to the baseline estimates (Table 1). The only exception is the dividend loading on long-run growth ϕ_x , which we examine below.

The structural restrictions have little effect on our price-dividend decompositions. This is seen in column (2) of Table 5. Both the $\text{Var}(pd_t)$ decomposition (Panel A) and $\text{Var}(\Delta pd_t)$ decomposition (Panel B) are little changed from the baseline.

Imposing cross-equation restrictions has little effect because the structural model is still very flexible. Each price-dividend coefficient in Display (13) contains at least three terms. As a result, the price-dividend coefficient implied by the structural model are close to the baseline semi-structural estimate.

For example, plugging the preference estimates into the coefficient on long-run growth A_x in the structural model leads to

$$A_x = \frac{\phi_x - \frac{1}{\psi}}{1 - \kappa_1 \rho_x} = \frac{3.90 - \frac{1}{3.39}}{1 - 0.9675 \times 0.805} = 17.8, \quad (15)$$

not far from the semi-structural estimate of 19.1. Equation (15) also helps explain why ϕ_x is estimated to be larger than in the baseline. The link between expected growth and first differences in the price-dividend ratio is empirically quite strong, implying a large A_x . The structural expression restricts the magnitude of A_x , however, as being proportional to ϕ_x . ϕ_x , in turn, is identified by the relationship between long-run growth in consumption and long-run growth in dividends. This relationship is difficult to identify, leading to large credible in-

tervals in the semi-structural estimate. But imposing Epstein-Zin-Habit implies that ϕ_x is instead identified with the cross-equation restriction (15), and thus a large estimate for ϕ_x .

5.2. Post-War Quarterly Data

Our second alternative estimation uses post-war quarterly data rather than annual data starting in 1929. This alternative estimation should alleviate concerns about structural differences between the pre-war and post-war data. Notably, volatility in consumption growth is much higher in the pre-war sample, due to both the Great Depression and the data construction. This structural change may lead to spurious persistence in volatility models (Messow and Krämer 2013). Using quarterly data also alleviates concerns about the information that is ignored when using the coarser annual data.

When using quarterly data, one must deal with seasonality in consumption and dividends. We choose the simplest method: we seasonally adjust the data before estimation. Our consumption data is the standard NIPA quarterly non-durables and services consumption, which the BEA adjusts with the Census's X-13 program. We apply the same X-13 program to adjust quarterly dividends for seasonality. The primary alternative approach is to introduce seasonal adjustment within the model (Ferson and Harvey 1992). Seasonal adjustment within the model is tricky as there is no standard method of modeling seasonality from an economic perspective.

The resulting parameter estimates are largely similar to the baseline, adjusted for the model frequency. The primary difference is in the long run volatility coefficient A_V , which is significantly larger in the quarterly estimation. Further details can be found in Table A.1.

The quarterly estimation leads to the main results as in the baseline: the residual is the dominant factor in the level of pd_t , but growth and volatility are dominant for first differences. The quarterly estimation implies a larger role for volatility in first differences, but the Bayesian standard errors imply that this result is not statistically powerful.

Figure 7 provides a closer look at this result. The figure shows the decomposition of the historical price-dividend ratio implied by the pd_t equation (2)). As in the baseline, long-run growth and long-run volatility exhibit business-cycle

frequency fluctuations, and are disconnected from the large, low-frequency fluctuations explained by the residual.

5.3. Diffuse Priors

Our third alternative estimation takes on a more agnostic view than the baseline. Here we estimate the model under uniform priors with wide boundaries. This differs from the baseline which imposes the long-run risks are real. These priors and resulting parameter estimates are shown in Table 7.

Under diffuse priors, the name “long-run growth” is no longer appropriate for the x_t process (Equation (4)). Thus, Tables 5 and 7 describes this variable more generically as “expected growth.” Expected growth is still persistent, but its annual autocorrelation of 0.58 is far below long-run risks values of around 0.85.

Moreover, the expected growth process is surprisingly volatile. The relative volatility parameter is 2.48, implying that expected growth accounts for 64% of the variance (81% of the volatility) of consumption growth. This large variance share is consistent with the fact that consumption growth is predicted by stock returns, lagged consumption growth, and lagged dividend growth. Indeed, these three variables together explain 57% of the variance of consumption growth using OLS.

Expected volatility $\tilde{\sigma}_t^2$ (formerly, “long-run volatility”) is still highly persistent with an annual persistence parameter of 0.91. This high persistence is consistent with the fact that consumption volatility has slowly trended downward over most of the sample period.

The persistence of surplus consumption continues to be similar to the prior, but the diffuse prior implies that its posterior distribution is centered around 0.50. Thus, the resulting surplus consumption process no longer captures the important notion of “deep habits” in the literature following Campbell and Cochrane (1999).

Despite the significant changes in estimated parameters, the variance shares are very similar to the baseline. Table 5 shows these shares in column (4) (“diffuse”).

The price-dividend variance shares of all latent states are nearly identical

Figure 7: Decomposition of the Price-Dividend Ratio: Quarterly Model and Data. The figure plots the decomposition implied by equation (2)

$$pd_t = \mu_{pd} + A_x x_t + A_V \tilde{\sigma}_t^2 + A_s \tilde{s}_t + A_e e_t$$

where pd_t is log price/dividends, x_t is long-run growth, $\tilde{\sigma}_t$ is long-run volatility, \tilde{s}_t is surplus consumption, and e_t is the residual. The states are the mean state found by particle smoother using mean posterior parameters from a Bayesian estimation (Table A.1). Coefficients are the mean posterior coefficients. All states are adjusted for their in-sample means. This decomposition uses post-war quarterly data and a quarterly model. Dividends are seasonally adjusted using the Census's X-13 filter, as is consumption. Our baseline decomposition (Figure 2) is robust to post-war quarterly data.

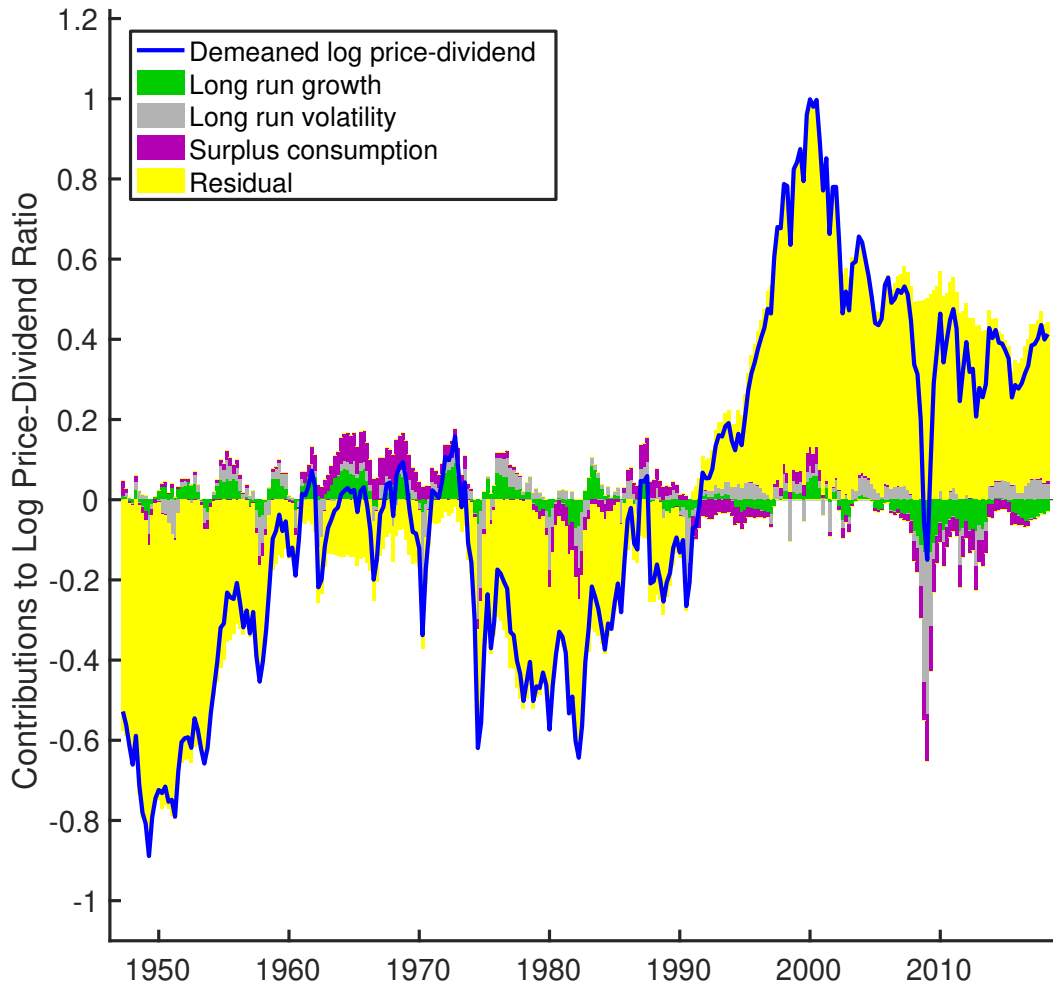


Table 7: Parameter Estimates: Diffuse Priors

The table shows prior and posterior parameter estimates under diffuse priors (Section 5). Priors are uniform U (lower bound, upper bound). Posteriors are computed using annual consumption, dividend, and stock prices from 1929-2014, particle filter, and Metropolis Hastings. $\exp \bar{s} = 0.06$ is chosen outside of the estimation, as are μ_{pd} , μ_c , and μ_d , which are chosen to be their sample means. The posteriors are used in our robustness check for diffuse priors (Table 5).

Parameter	Prior	Posterior			
		Mean	5%	95%	
Simple Consumption and Dividends					
Vol of Cons Shock	$\bar{\sigma}$ $U(0.001, 0.040)$	0.00714	0.00368	0.0114	
Div Loading on Cons	$\phi_{\eta c}$ $U(0, 10)$	0.814	0.0426	2.43	
Rel Vol of Dividends	φ_d $U(0, 40)$	15	9.16	21.2	
Expected Growth and Volatility					
Persist of Growth	ρ_x $U(0.00, 1.00)$	0.577	0.404	0.712	
Rel Vol of Growth	φ_x $U(0.00, 10.00)$	2.48	1.42	3.82	
Div Load on Growth	ϕ_x $U(0, 10)$	1.94	0.771	3.22	
Persist of log Vol	ρ_h $U(0, 1)$	0.912	0.81	0.984	
Rel Vol of Vol	σ_h $U(0.00, 1.50)$	0.248	0.129	0.397	
Habit and Residual					
Persistence of Habit	ρ_s $U(0.00, 1.00)$	0.453	0.0934	0.945	
Persistence of Residual	ρ_e $U(0, 1)$	0.948	0.897	0.99	
Price-Dividend Coefficients					
LR Growth Coefficient	A_x $U(0, 2000)$	7.48	5.15	10.1	
LR Vol Coefficient	A_V $U(-5000, 0)$	-1.56e+03	-3.5e+03	-110	
Habit Coefficient	A_s $U(0, 1)$	0.504	0.204	0.925	
Residual Coefficient	A_e $U(0, 1)$	0.14	0.11	0.175	

to the baseline values. The variance shares are 10%, 12%, -2%, and 80% for expected growth, expected volatility, surplus consumption, and the residual, respectively, compared to 13%, 11%, -2%, and 78% in the baseline.

The same robustness is seen in the Δpd_t variance shares. All four variance shares are nearly identical to the baseline.

5.4. Priors with Identical Price-Dividend Variance Shares

In our final alternative estimation, we choose priors which guarantee that each latent state has an a priori equal explanatory power for the variance of pd_t . Specifically, these priors impose that the theoretical variance shares are identically and log-normally distributed.

To create these priors, we construct four independent random variables T_x, T_V, T_s, T_e that are log-normally distributed with $T_i \sim \log \mathcal{N}(\mu_T, \sigma_T^2)$, $i = x, V, s, e$. We then construct the A_x coefficients conditional on the values of the remaining model parameters θ as follows:

$$\begin{aligned} A_x &= \sqrt{\frac{T_x}{\mathbb{V}[x_t | \theta]}}, & A_V &= -\sqrt{\frac{T_V}{\mathbb{V}[\tilde{\sigma}_t | \theta]}}, \\ A_s &= \sqrt{\frac{T_s}{\mathbb{V}[\tilde{s}_t | \theta]}}, & A_e &= \sqrt{\frac{T_e}{\mathbb{V}[e_t | \theta]}}. \end{aligned} \quad (16)$$

Here, $\mathbb{V}[x_t | \theta]$ etc. are the theoretical variances of the state variables conditional on the other model parameters. Note that we restrict the signs of the coefficients to conform to economic intuition. That is, we restrict the coefficients on long run growth and surplus consumption to be positive, and that on long run volatility to be negative. The result of this prior choice is that the prior distribution of the variances of the factors conditional on any θ are given simply by the T_i 's, and in particular i.i.d. among each other. We set $\sigma_T^2 = 2$ and μ_T such that the unconditional prior variance of the price-dividend ratio equals the observed variance in the data. Other, similarly diffuse distributions of the T_i 's produce very similar results.

Table 5's column (5) shows that priors with equal variance shares has little effect on the posterior variance shares. All four variance shares are very similar to the baseline, for both the log price-dividend ratio and first differences.

6. Conclusion

The consumption-based asset pricing literature has found many models that provide a good fit to the data. This success, however, comes almost entirely from moment-matching methods.

In this paper, we find that full-information methods lead to a less positive picture. Our Bayesian estimation finds that habit has negligible explanatory power for the variance of the price-dividend ratio, and long-run risks explain at most 25%. Moreover, the extracted paths of long-run risks and habit bear little resemblance with the visible features of stock market history. These prominent asset pricing models do not explain the large valuation booms of the 1960's and 1990's, nor the subsequent busts.

Our results suggest that there is plenty of room for new models for understanding market volatility. While consumption-based explanations play a large role in short-term movements in the stock market, there is a large, highly persistent component of the price-dividend ratio that must be attributed to something else.

There are many theories that are consistent with our results. These theories include hard-to-observe changes in the higher moments of consumption growth (Gabaix 2012; Wachter 2013; Marfè and Penasse 2016; Schorfheide, Song, and Yaron Forthcoming), dynamic financial frictions (Vissing-Jørgensen 2002, Lustig and Van Nieuwerburgh 2010, He and Krishnamurthy 2013), time-varying preferences (Albuquerque et al. 2016), time-varying ambiguity (Sbuelz and Trojani 2008), imperfect learning (Adam, Marcet, and Nicolini 2016), and hard-to-observe changes in long-term interest rates (Bianchi, Lettau, and Ludvigson 2016). Measurement issues related to slow changes in payout policy (Fama and French 2001, Chen, Da, and Priestley 2012) may also explain our results.

Overall, we find that more empirical comparisons of macro-finance asset pricing models are needed. Our Bayesian framework should be helpful in these comparisons, as it avoids the contentious choice of moment conditions, and provides a detailed understanding of identification.

A. Appendix

A.1. One Structural Interpretation: Derivations

A.1.1. Volatility and disaster probability process approximations

We first establish the following approximation to the demeaned volatility process $\tilde{\sigma}_t$:

$$\begin{aligned}
 \tilde{\sigma}_{t+1}^2 &= \sigma_{t+1}^2 - \mathbb{E}\sigma_t^2 = \bar{\sigma}^2 (\exp(2h_{t+1}) - \mathbb{E}\exp(2h_t)) \\
 &\approx \bar{\sigma}^2 (2h_{t+1} - 2\mathbb{E}h_t) \\
 &= \bar{\sigma}^2 (\rho_h (2h_t - 2\mathbb{E}h_t) + 2\sigma_h \eta_{ht+1}) \\
 &\approx \rho_h \bar{\sigma}_t^2 + 2\bar{\sigma}^2 \sigma_h \eta_{ht+1}.
 \end{aligned} \tag{17}$$

A.1.2. Utility-surplus consumption ratio

We now establish an equation for the utility-surplus consumption ratio. We have

$$\log \frac{C_{t+1} - H_{t+1}}{C_t - H_t} = \log \frac{C_{t+1} \bar{S}^{1-\ell} S_{t+1}^\ell}{C_t \bar{S}^{1-\ell} S_t^\ell} = \Delta c_{t+1} + \ell \Delta s_{t+1}. \tag{18}$$

If $\gamma \neq 1$ and $\psi \neq 1$, the log utility-surplus consumption ratio y_t can then be rewritten as

$$y_t = \log \frac{U_t}{C_t - H_t} = \frac{1}{1 - \frac{1}{\psi}} \log \left(1 - \delta + \delta \exp \left(\left(1 - \frac{1}{\psi} \right) q_t \right) \right), \tag{19}$$

where

$$q_t = \frac{1}{1 - \gamma} \log \mathbb{E}_t \exp \left((1 - \gamma) (y_{t+1} + \Delta c_{t+1} + \ell \Delta s_{t+1}) \right). \tag{20}$$

We linearize y_t as a function of q_t to obtain

$$y_t \approx \kappa_{0y} + \kappa_{1y} q_t \tag{21}$$

where $\kappa_{1y} = \delta \exp((1 - 1/\psi)\mu_c)$. Next, we guess and verify a linear approximation to y_t :

$$y_t \approx V_0 + V_s \tilde{s}_t + V_x x_t + V_V \tilde{\sigma}_t^2. \tag{22}$$

We then obtain

$$\begin{aligned}
(1-\gamma)q_t &= \\
&= \mathbb{E}_t[(1-\gamma)(y_{t+1} + \Delta c_{t+1} + \ell \Delta s_{t+1})] \\
&+ \frac{1}{2} \mathbb{V}_t[(1-\gamma)(y_{t+1} + \Delta c_{t+1} + \ell \Delta s_{t+1})] \\
&= (1-\gamma)(V_0 + V_s \rho_s \tilde{s}_t + V_x \rho_x x_t + V_V \rho_h \tilde{\sigma}_t^2 + \mu_c + x_t + \ell(\rho_s - 1) \tilde{s}_t) \\
&+ \frac{(1-\gamma)^2}{2} \left((V_s \lambda(\tilde{s}_t) + \ell \lambda(\tilde{s}_t) + 1)^2 \sigma_t^2 + V_x^2 \varphi_x^2 \sigma_t^2 \right. \\
&\quad \left. + 4V_V^2 \tilde{\sigma}^4 \sigma_h^2 \right). \tag{23}
\end{aligned}$$

We can now solve for the coefficients for y_t by combining equations (21)–(23) and approximating $\lambda(\tilde{s}_t)$ by its steady-state value $\bar{\lambda}$. We find:

$$V_s = \frac{\kappa_{1y}(\rho_s - 1)}{1 - \kappa_{1y}\rho_s} \ell \tag{24}$$

$$V_x = \frac{\kappa_{1y}}{1 - \kappa_{1y}\rho_x} \tag{25}$$

$$V_V = \kappa_{1y} \frac{1-\gamma}{2} \frac{(V_s \bar{\lambda} + \ell \bar{\lambda} + 1)^2 + V_x^2 \varphi_x^2}{1 - \kappa_{1y}\rho_h}. \tag{26}$$

A.1.3. Discount factor and market prices of risk

Given wealth, the agent chooses consumption and invests in a portfolio of a risk-free and risky assets. The first order conditions imply the following log pricing kernel, as derived in Yang 2016:

$$m_{t+1} = \log \delta + \left(\gamma - \frac{1}{\psi} \right) q_t - \gamma (y_{t+1} + \Delta c_{t+1} + \ell \Delta s_{t+1}) + \frac{1}{\psi} y_{t+1}. \tag{27}$$

We write down the innovation to the log pricing kernel. This also lets us define “market prices of risk” ξ_{ct} , ξ_x , and ξ_h :

$$\begin{aligned}
m_{t+1} - \mathbb{E}_t[m_{t+1}] &= \left(\left(\frac{1}{\psi} - \gamma \right) V_s \lambda(\tilde{s}_t) - \gamma (1 + \ell \lambda(\tilde{s}_t)) \right) \sigma_t \eta_{ct+1} \\
&+ \left(\frac{1}{\psi} - \gamma \right) V_x \varphi_x \sigma_t \eta_{xt+1} \\
&+ 2 \left(\frac{1}{\psi} - \gamma \right) V_V \tilde{\sigma}^2 \sigma_h \eta_{ht+1} \\
&= -\xi_{ct} \sigma_t \eta_{ct+1} - \xi_x \sigma_t \eta_{xt+1} - \xi_h \sigma_h \eta_{ht+1}. \tag{28}
\end{aligned}$$

The expected discount factor is

$$\begin{aligned} \mathbb{E}_t[m_{t+1}] &\approx \log \delta - \frac{1}{\psi} (\mu_c + x_t + \ell (\rho_s - 1) \tilde{s}_t) \\ &\quad + \left(\gamma - \frac{1}{\psi} \right) \frac{1-\gamma}{2} \left((V_s \lambda(\tilde{s}_t) + \ell \lambda(\tilde{s}_t) + 1)^2 \sigma_t^2 + V_x^2 \varphi_x^2 \sigma_t^2 \right. \\ &\quad \left. + 4V_V^2 \bar{\sigma}^4 \sigma_h^2 \right). \end{aligned} \quad (29)$$

A.1.4. Risk-free rate

We can approximate the risk-free rate as

$$\begin{aligned} r_{ft} &= -\log \mathbb{E}_t \exp m_{t+1} \\ &\approx -\mathbb{E}_t \exp[m_{t+1}] - \frac{1}{2} \mathbb{V}_t \exp[m_{t+1}] \\ &= -\log \delta + \frac{1}{\psi} \mu_c - \left(\gamma - \frac{1}{\psi} \right) \frac{1-\gamma}{2} (4V_V^2 \bar{\sigma}^4 \sigma_h^2) - \frac{1}{2} \xi_h^2 \sigma_h^2 \\ &\quad + \frac{1}{\psi} (x_t + \ell (\rho_s - 1) \tilde{s}_t) \\ &\quad - \left(\gamma - \frac{1}{\psi} \right) \frac{1-\gamma}{2} ((V_s \lambda(\tilde{s}_t) + \ell \lambda(\tilde{s}_t) + 1)^2 + V_x^2 \varphi_x^2) \sigma_t^2 - \frac{1}{2} \xi_{ct}^2 \sigma_t^2 - \frac{1}{2} \xi_x^2 \sigma_t^2. \end{aligned} \quad (30)$$

A.1.5. Price-dividend ratio

We guess and verify a linear approximation of the log price-dividend ratio:

$$z_t = \log \frac{P_t}{D_t} \approx \mu_{pd} + A_s \tilde{s}_t + A_x x_t + A_V \bar{\sigma}_t^2. \quad (31)$$

We write the log return using the standard Campbell-Shiller approximation:

$$r_{t+1} \approx \kappa_0 + \kappa_1 z_{t+1} + \Delta d_{t+1} - z_t \quad (32)$$

and set up the pricing equation

$$1 = \mathbb{E}_t \exp(m_{t+1} + r_{t+1}). \quad (33)$$

We take the logarithm of Equation (33) and express it as a linear equation of the state variables:

$$\begin{aligned}
z_t &= \kappa_0 + \log \mathbb{E}_t \exp(m_{t+1} + \kappa_1 z_{t+1} + \Delta d_{t+1}) \\
&= \kappa_0 + \log \mathbb{E}_t [\exp(m_{t+1} + \kappa_1 z_{t+1} + \Delta d_{t+1})] \\
&\approx \kappa_0 + \mathbb{E}_t [m_{t+1} + \kappa_1 z_{t+1} + \Delta d_{t+1}] \\
&\quad + \frac{1}{2} \mathbb{V}_t [m_{t+1} + \kappa_1 z_{t+1} + \Delta d_{t+1}] \\
&= \kappa_1 (A_s \rho_s \tilde{s}_t + A_x \rho_x x_t + A_V \rho_h \tilde{\sigma}_t^2) + \phi_x x_t \\
&\quad - \frac{1}{\psi} (x_t + \ell (\rho_s - 1) \tilde{s}_t) \\
&\quad + \left(\gamma - \frac{1}{\psi} \right) \frac{1 - \gamma}{2} ((V_s \lambda(\tilde{s}_t) + \ell \lambda(\tilde{s}_t) + 1)^2 \sigma_t^2 + V_x^2 \varphi_x^2 \sigma_t^2) \\
&\quad + \frac{1}{2} (-\xi_{ct} + \kappa_1 A_s \lambda(\tilde{s}_t) + \phi_{\eta c})^2 \sigma_t^2 + \frac{1}{2} \varphi_d^2 \sigma_t^2 \\
&\quad + \frac{1}{2} (-\xi_x + \kappa_1 A_x \varphi_x)^2 \sigma_t^2 \\
&\quad + \text{const.} \tag{34}
\end{aligned}$$

Combining equations (31) and (34), and approximating $\lambda(\tilde{s}_t)$ and ξ_{ct} by their steady-state values $\bar{\lambda}$ and $\bar{\xi}_c$ we obtain the structural price-dividend coefficients in (13).

A.2. State Space Formulation

To estimate the model, we write it in a state space formulation following Schorfheide, Song, and Yaron (Forthcoming). In the end, we have transition equations

$$\begin{aligned}
h_t &= \rho_h h_{t-1} + \sigma_h w_t \\
m_t &= \Phi(m_{t-1}) m_{t-1} + \Sigma_s(m_{t-1}) \eta_t.
\end{aligned} \tag{35}$$

and observation equations

$$y_t = \mu_y + Z m_t + Z_v (\exp(2h_t) - \exp(2\sigma_h^2)) \tag{36}$$

where m_t is a vector of mean “states,” y_t is a vector of observables, w_t, η_t are vectors of standard normal independent noise, $\Phi(m_{t-1}), \Sigma_s(m_{t-1})$ are matrices that describe the evolution of the mean “states,” and μ_y, Z, Z_v are vectors and

matrices that map states to observables. We put quotes around “states” because the elements of m_t include terms which are not state variables in the traditional economic sense. These additional “states” help simplify notation.

Equations (35) and (36) are convenient forms for the asset pricing models with time varying volatility and normal shocks. As this class of models is conditionally normal, it helps to express the model as close to a state space form as possible. Moreover, this formulation allows the model to be extended to account for mixed frequency data.

These equations are mapped to the model in Section 2.1 by a careful definition of vectors and matrices. We’ll now define these vectors and matrices.

Observables and States Equations (2) and (3) can be mapped into the observation equation (36) as follows:

$$\underbrace{\begin{bmatrix} \Delta c_t \\ \Delta d_t \\ p d_t \end{bmatrix}}_{y_t} = \underbrace{\begin{bmatrix} \mu_c \\ \mu_d \\ \mu_{pd} \end{bmatrix}}_{\mu_y} + \underbrace{\begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & \phi_x & \phi_{\eta_c} & 1 \\ [A_x, A_s, A_e] & 0 & 0 & 0 \end{bmatrix}}_Z \underbrace{\begin{bmatrix} z_t \\ x_{t-1} \\ \tilde{\eta}_{c,t} \\ \tilde{\eta}_{d,t} \end{bmatrix}}_{m_t} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ A_V \bar{\sigma}^2 \end{bmatrix}}_{Z_v} (\exp(2h_t) - \exp(2\sigma_h^2)).$$

where

$$\begin{aligned} z_t &\equiv [x_t, \tilde{s}_t, e_t]' \\ \tilde{\eta}_{c,t} &= \bar{\sigma} \exp(h_{t-1}) \eta_{c,t} \\ \tilde{\eta}_{d,t} &= \varphi_d \bar{\sigma} \exp(h_{t-1}) \eta_{d,t} \end{aligned}$$

and $'$ indicates a transpose.

State Transition Then state transitions (4)-(7) can be expressed in terms of the augmented state space transitions (35) as follows:

$$\underbrace{\begin{bmatrix} z_t \\ x_{t-1} \\ \tilde{\eta}_{c,t} \\ \tilde{\eta}_{d,t} \end{bmatrix}}_{m_t} = \underbrace{\begin{bmatrix} \left[\begin{array}{ccc|ccc} \rho_x & 0 & 0 & 0 & 0 & 0 \\ \psi_x \lambda(u_{t-1}) & \rho_u & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_e & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & & & 0 & 0 & 0 \\ 0 & & & 0 & 0 & 0 \end{array} \right] \begin{bmatrix} z_{t-1} \\ x_{t-2} \\ \tilde{\eta}_{c,t-1} \\ \tilde{\eta}_{d,t-1} \end{bmatrix}}_{\Phi(m_{t-1})} \underbrace{\begin{bmatrix} z_{t-1} \\ x_{t-2} \\ \tilde{\eta}_{c,t-1} \\ \tilde{\eta}_{d,t-1} \end{bmatrix}}_{m_{t-1}} \\
 + \underbrace{\begin{bmatrix} \left[\begin{array}{ccc|ccc} \varphi_x \bar{\sigma} \exp(h_{t-1}) & & & 0 & & \\ 0 & & \lambda(u_{t-1}) \bar{\sigma} \exp(h_{t-1}) & & & 0 \\ 0 & & 0 & & & 1 \\ 0 & & \bar{\sigma} \exp(h_{t-1}) & & & 0 \\ 0 & & 0 & & \varphi_d \bar{\sigma} \exp(h_{t-1}) & 0 \end{array} \right] \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{\Sigma_s(m_{t-1})} \underbrace{\begin{bmatrix} \eta_{x,t} \\ \eta_{c,t} \\ \eta_{e,t} \\ \eta_{d,t} \end{bmatrix}}_{\eta_t},$$

where ψ_x is an indicator variable which depends on the habit specification ($\psi_x = 0 \Rightarrow$ habit responds to innovations in consumption growth).

A.3. Particle Filter Details

With the state space formulation in hand, we can now write down the particle filter algorithm in a compact form. We first describe the big picture of the algorithm. We then go on to give the details of how each distribution is defined.

For each $t = 1, \dots, T$, do the following for particles $i = 1, \dots, M$.

1. Begin with a set of particles $[m_{t-1}^i, h_{t-1}^i]$ and weights π_{t-1}^i .
2. Draw $h_t^i \sim q(h_t^i | p d_t^i, h_{t-1}^i, m_{t-1}^i)$ for each i , where $q(h_t^i | p d_t^i, h_{t-1}^i, m_{t-1}^i)$ is a proposal distribution which we'll describe shortly.
3. Draw $m_t^i \sim p(m_t | y_t, h_t^i, h_{t-1}^i, m_{t-1}^i)$ for each i , where $p(m_t | y_t, h_t^i, h_{t-1}^i, m_{t-1}^i)$ is the conditional density of m_t . We'll describe how this density is computed shortly. Throughout this Appendix $p(x|y)$ means the conditional density of x given y .

4. Update particle weights using

$$\pi_t^i = \pi_{t-1}^i [\text{update factor}]^i \quad (37)$$

$$[\text{update factor}]^i = p(y_t | h_t^i, h_{t-1}^i, m_{t-1}^i) \left[\frac{p(h_t^i | h_{t-1}^i)}{q(h_t^i | p d_t, h_{t-1}^i, m_{t-1}^i)} \right] \quad (38)$$

In a simple bootstrap particle filter, the update factor is just the likelihood of y_t given m_t^i and h_t^i . We'll explain how to derive the above update factor shortly.

5. Estimate log-likelihood contribution

$$\log \hat{p}(y_t) = \log \left(\sum_i \pi_{t-1}^i [\text{update factor}]^i \right) \quad (39)$$

6. Resample: if $\frac{1}{M^2 \sum_i (\pi_t^i)^2} < 0.5$ redraw $\{\pi_t^i\}$ using a multinomial distribution with probabilities $\{\pi_t^i\}$.

Since the remainder of this section discusses operations applied to every particle i , we drop the superscript for ease of reading.

A.3.1. Proposal Distribution for $h_t \sim q(h_t | p d_t, h_{t-1}, m_{t-1})$

We draw h_t based off of $p d_t$ and the previous state (h_{t-1}, m_{t-1}) . The basic idea is that we want to draw h_t as close to the true probability $p(h_t | p d_t, h_{t-1}, m_{t-1})$ as possible in order to minimize Monte Carlo noise in the particle filter. Unfortunately, the relationship between $p d_t$ and h_t is nonlinear (equations (2) and (5)). We can, however, use the following Taylor expansion

$$\exp(2h_t) \approx \exp(2\rho_h h_{t-1})(1 - 2\rho_h h_{t-1}) + 2 \exp(2\rho_h h_{t-1})h_t \quad (40)$$

which leads to an approximation (36) that is linear in h_t

$$\begin{aligned} y_t &= \mu_y + Z m_t + Z_v (\exp(2h_t) - \exp(2\sigma_h^2)) \\ &\approx \mu_y + Z m_t + Z_v \exp(2\rho_h h_{t-1})(1 - 2\rho_h h_{t-1}) + Z_v 2 \exp(2\rho_h h_{t-1})h_t. \end{aligned} \quad (41)$$

This approximation, combined with the equation (35) and the definition of y_t lets us write a mini state space system

$$\begin{aligned} pd_t &= pd_0 + \tilde{A}_{V,t} h_t + \sigma_{pd} \eta_{pd,t} \\ h_t &= \rho_h h_{t-1} + \sigma_h w_t \end{aligned} \quad (42)$$

where $\eta_{pd,t} \sim N(0, 1)$ i.i.d. and

$$\begin{aligned} pd_0 &= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} [\mu_y + Z\Phi(m_{t-1})m_{t-1} + Z_v \exp(2\rho_h h_{t-1})(1 - 2\rho_h h_{t-1})] \\ \tilde{A}_V &= A_V \bar{\sigma}^2 2 \exp(2\rho_h h_{t-1}) \\ \tilde{\sigma}_{pd} &= \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} Z \Sigma_s(m_{t-1}). \end{aligned}$$

In this approximation, $h_t | pd_t, h_{t-1}, m_{t-1}$ is normally distributed, and a one-step Kalman filter gives the mean and variance. We use this distribution as $q(h_t | pd_t, h_{t-1}, m_{t-1})$.

In principle, this approximation could be used to generate a proposal distribution for all states $[h_t, s_t]$. But drawing h_t separately lets us nest Schorfheide, Song, and Yaron (Forthcoming)'s specification and helps error checking.

A.3.2. Proposal Distribution for $m_t \sim p(m_t | y_t, h_t, h_{t-1}, m_{t-1})$

We draw m_t in a similar way to h_t . The only difference is we draw m_t given h_t , and thus y_t is linear in the unobserved m_t and so no approximations are needed.

Explicitly, given h_t, h_{t-1} , the state space system

$$\begin{aligned} y_t &= \mu_y + Z m_t + Z_v (\exp(2h_t) - \exp(2\sigma_h^2)) \\ m_t &= \Phi(m_{t-1})m_{t-1} + \Sigma_s(m_{t-1})\eta_t \end{aligned} \quad (43)$$

shows that $p(m_t | y_t, h_t, h_{t-1}, m_{t-1})$ is normally distributed, and a one-step Kalman filter gives the mean and variance of this distribution. We use this distribution to draw m_t in the particle filter.

A.3.3. Simplifying the Update Factor

We simplify the particle filter update step by taking advantage of the conditional Gaussian properties of the model and using Bayes' theorem.

The standard generic particle filter update follows

$$\pi_t = \pi_{t-1} \text{update weight} \quad (44)$$

where the update weight is

$$\text{update weight} \equiv p(y_t | m_t, h_t) \frac{p(m_t, h_t | m_{t-1}, h_{t-1})}{q(m_t, h_t | m_{t-1}, h_{t-1}, y_t)}$$

and q is the proposal distribution in the propagation step (see Herbst and Schorfheide (2014)). In our case, $p(m_t, h_t | m_{t-1}, h_{t-1})$ and $q(m_t, h_t | m_{t-1}, h_{t-1}, y_t)$ can be broken up into mean and volatility components

$$\text{update weight} = p(y_t | m_t, h_t) \frac{p(m_t | m_{t-1}, h_{t-1})}{p(m_t | y_t, h_t, m_{t-1}, h_{t-1})} \frac{p(h_t | h_{t-1})}{q(h_t | y_t, m_{t-1}, h_{t-1})} \quad (45)$$

The above expression can be further simplified. First note that (m_t, h_t) are sufficient to determine the density of y_t . Similarly, h_t adds no information regarding m_t given (m_{t-1}, h_{t-1}) . Thus,

$$\text{update weight} = \left[p(y_t | m_t, h_t, m_{t-1}, h_{t-1}) \frac{p(m_t | h_t, m_{t-1}, h_{t-1})}{p(m_t | y_t, h_t, m_{t-1}, h_{t-1})} \right] \frac{p(h_t | h_{t-1})}{q(h_t | y_t, m_{t-1}, h_{t-1})} \quad (46)$$

The term in the brackets can be simplified using Bayes' rule twice

$$\begin{aligned} p(y_t | m_t, h_t, m_{t-1}, h_{t-1}) \frac{p(m_t | h_t, m_{t-1}, h_{t-1})}{p(m_t | y_t, h_t, m_{t-1}, h_{t-1})} &= p(m_t | y_t, h_t, m_{t-1}, h_{t-1}) \frac{p(y_t, h_t, m_{t-1}, h_{t-1})}{p(m_t, h_t, m_{t-1}, h_{t-1})} \frac{p(m_t | h_t, m_{t-1}, h_{t-1})}{p(m_t | y_t, h_t, m_{t-1}, h_{t-1})} \\ &= \frac{p(y_t, h_t, m_{t-1}, h_{t-1})}{p(m_t, h_t, m_{t-1}, h_{t-1})} p(m_t | h_t, m_{t-1}, h_{t-1}) \\ &= \frac{p(y_t, h_t, m_{t-1}, h_{t-1})}{p(m_t, h_t, m_{t-1}, h_{t-1})} \frac{p(m_t, h_t, m_{t-1}, h_{t-1})}{p(h_t, m_{t-1}, h_{t-1})} \\ &= p(y_t | h_t, m_{t-1}, h_{t-1}). \end{aligned} \quad (47)$$

Finally, combining equations (46) and (47) gives the update factor expression in the particle filter algorithm (38).

A.4. Particle smoother details

We use a variant of the backward-simulation particle smoother of Godsill, Doucet, and West (2004). Our procedure explicitly takes care of the possibility that the transitional likelihood is degenerate whenever the habit innovation $\lambda(u_{t-1}) = 0$ is zero.

We start with a set of filtered particles (m_t^i, h_t^i) with weights π_t^i , $t = 1, \dots, T$, $i = 1, \dots, M$. We then compute a set of smoothed particles $(\tilde{m}_t^i, \tilde{h}_t^i)$ as follows.

- Draw k_{iT} from the distribution π_t . Set $(\tilde{m}_T^i, \tilde{h}_T^i) = (m_T^{k_{iT}}, h_T^{k_{iT}})$.
- For $t = T - 1 \dots 1$:
 1. Check whether $\lambda(u_t^{k_{it+1}}) = 0$:
 - If it is, then the filtered particle drawn in $t + 1$ came from a degenerate distribution, and so

$$\tilde{\pi}_t^k = p(m_t^k, h_t^k \mid \tilde{m}_{t+1}^i, \tilde{h}_{t+1}^i, y_{1:T}^o) = I(i = k)$$

where $I()$ is an indicator function. Set $k_{it} = k_{it+1}$.

- If not, then

$$\tilde{\pi}_t^k = p(m_t^k, h_t^k \mid \tilde{m}_{t+1}^i, \tilde{h}_{t+1}^i, y_{1:T}^o) \sim \pi_t^j p(\tilde{m}_{t+1}^i, \tilde{h}_{t+1}^i \mid m_t^k, h_t^k)$$

and these densities are finite. Draw k_{it} from the distribution $\tilde{\pi}_t$.

2. Set $(\tilde{m}_t^i, \tilde{h}_t^i) = (m_t^{k_{it}}, h_t^{k_{it}})$.

In particular, whenever $\lambda(u_t^{k_{it+1}}) > 0$ we have that $p(\tilde{m}_{t+1}^i, \tilde{h}_{t+1}^i \mid m_t^k, h_t^k) = 0$ for all k with $\lambda(u_t^k) = 0$.

A.5. Bayesian MCMC Method

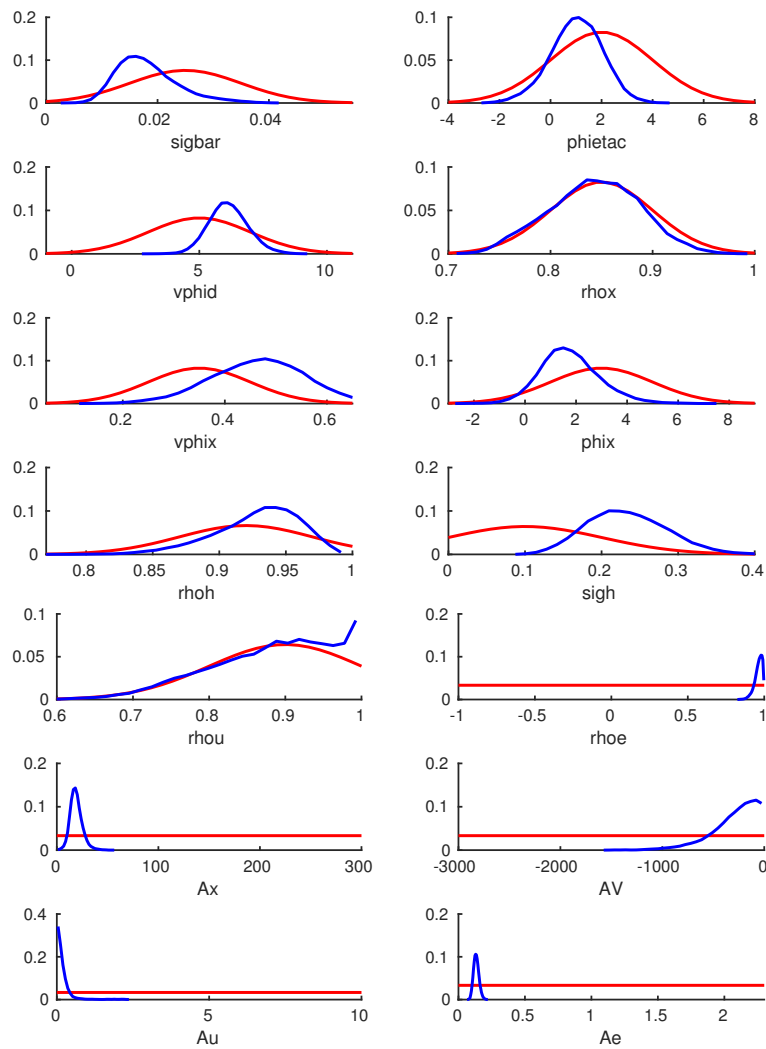
We wrap the filter in a standard Random Walk Metropolis-Hastings algorithm in order to derive parameter estimates (Herbst and Schorfheide 2014). We run standard initial tuning runs of the algorithm in order to choose a good proposal distribution. That is, we begin by finding a local maximum of the likelihood function using numerical optimization. We then run a chain of length 5,000 with a symmetric step direction and use the variance of the posterior as

the step direction in the next steps. We also choose the step size such that the acceptance rate is a little larger than 0.3. The final MCMC chain has length 500,000.

A.6. Baseline Posterior Details

Figure A.1 shows the distributions of all posterior parameters in the baseline estimation (Table 1).

Figure A.1: Baseline Posterior Details.



A.7. Additional Results

Figure A.2: Simple Regressions of P/D on Surplus Consumption. Data is quarterly, unlike in the main text. CC 1999 is the surplus consumption ratio using the process and parameter values in Campbell and Cochrane (1999). Cochrane 2018 is the log surplus consumption using the process and parameter values in Cochrane (2017). Consumption is real non-durable and services consumption, as in the main text. Both surplus consumption series are scaled to have a unit standard deviation and zero mean. ‘High pass filtered’ applies the Christiano and Fitzgerald (2003) filter to all series, with a maximum period of 8 years. Dotted lines in the right panels show OLS regressions. Alternative habit processes also show a very weak relationship between P/D and surplus consumption. Applying filters to both P/D and surplus consumption results in a stronger relationship, but still leaves the vast majority of P/D variation unexplained. Alternative filters lead to similar results.

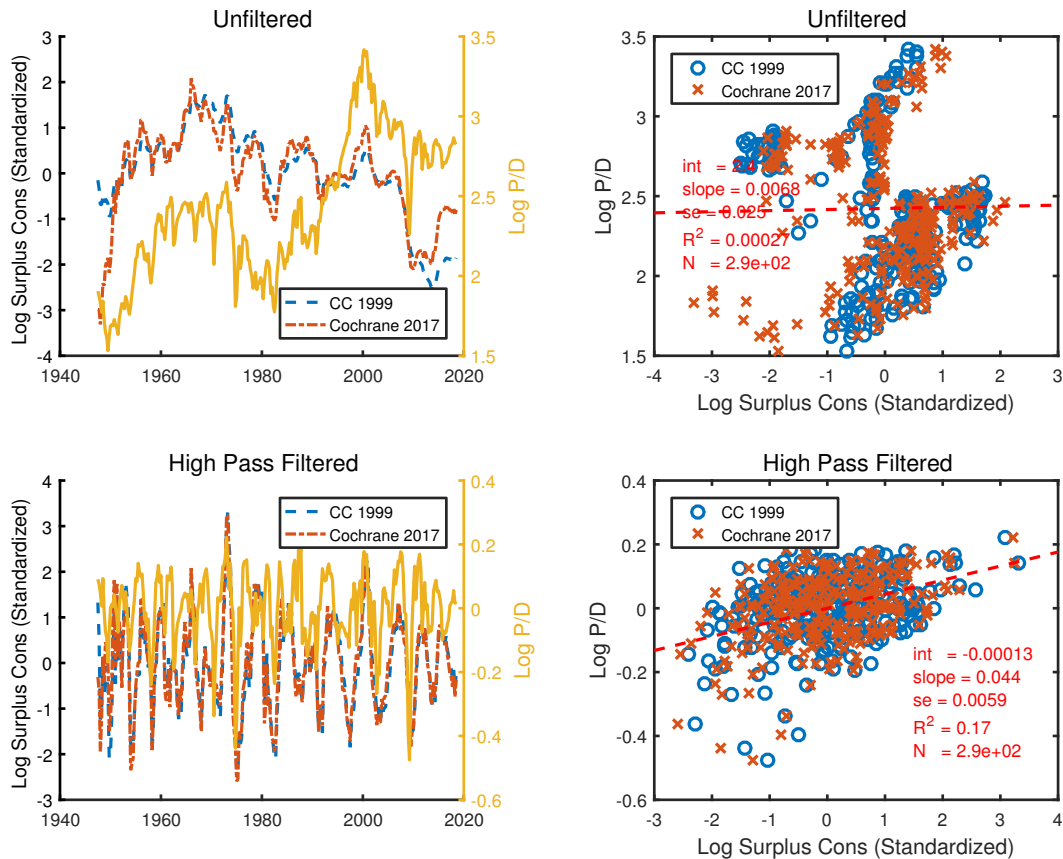


Figure A.3: Decomposition of the Business-Cycle Component of the Price-Dividend Ratio. The plot shows the decomposition implied by applying the Christiano and Fitzgerald (2003) high pass filter with a long period cutoff of 8 years to the components of Equation (2):

$$pd_t = \mu_{pd} + A_x x_t + A_V \tilde{\sigma}_t^2 + A_s \tilde{s}_t + A_e e_t.$$

where pd_t is log price/dividends, x_t is long-run growth, $\tilde{\sigma}_t$ is long-run volatility, \tilde{s}_t is surplus consumption, and e_t is the residual. The states are the mean state found by particle smoother using mean posterior parameters from a Bayesian estimation (Table 1). Coefficients are the mean posterior coefficients. As in the Δpd_t decomposition (Figure 5), long-run growth plays a key role and the role of the residual is reduced significantly relative to the price-dividend decomposition (Figure 2).

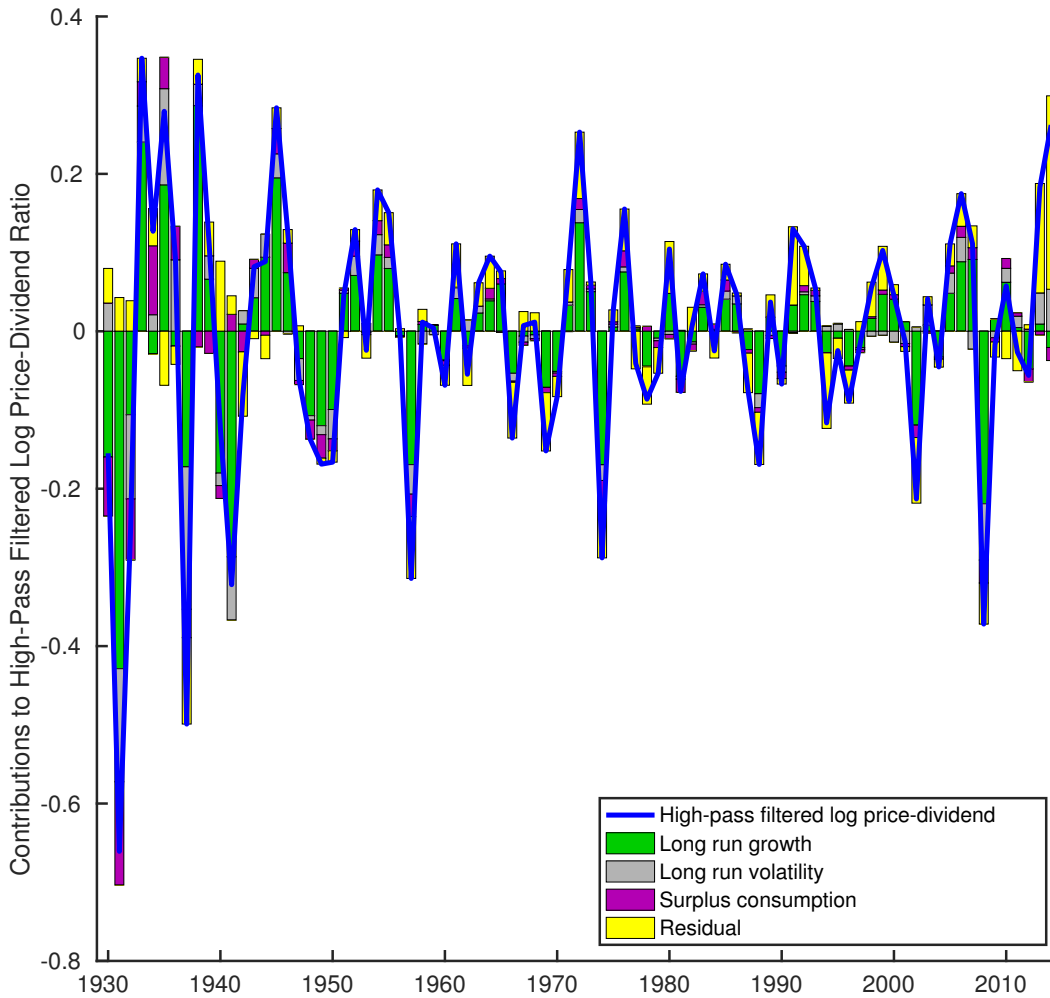


Table A.1: Parameter Estimates: Quarterly

The table shows prior and posterior parameter estimates for the quarterly model on quarterly data. The posteriors are used in our robustness check using quarterly data (Table 5 and Figure 7).

Parameter	Prior	Posterior			
		Mean	5%	95%	
Simple Consumption and Dividends					
Vol of Cons Shock	$\bar{\sigma}$	$N_T(0.010, 0.010, 0, \infty)$	0.00357	0.00292	0.00432
Div Loading on Cons	$\phi_{\eta c}$	$N(2, 2)$	0.593	-1.37	2.36
Rel Vol of Dividends	φ_d	$N(19, 5)$	22.8	20.3	25.6
Long Run Risks					
Persist of LR Growth	ρ_x	$N_T(0.90, 0.05, 0, 1)$	0.868	0.812	0.926
Rel Vol of LR Growth	φ_x	$N_T(0.35, 0.10, 0, \infty)$	0.372	0.269	0.476
Div Load on LR Growth	ϕ_x	$N(3, 2)$	1.45	-0.886	3.77
Persist of log LR Vol	ρ_h	$N_T(0.95, 0.05, 0, 1)$	0.858	0.806	0.911
Rel Vol of LR Vol	σ_h	$N_T(0.10, 0.10, 0, \infty)$	0.248	0.195	0.305
Habit and Residual					
Persistence of Habit	ρ_s	$N_T(0.95, 0.05, 0, 1)$	0.914	0.844	0.973
Persistence of Residual	ρ_e	$U(-1, 1)$	0.989	0.976	0.998
Price Dividend Coefficients					
LR Growth Coefficient	A_x	$U(0, 300)$	19.7	6.76	34.8
LR Vol Coefficient	A_V	$U(-6000, 0)$	-3.36e+03	-5.29e+03	-1.37e+03
Habit Coefficient	A_s	$U(0, 10)$	0.303	0.158	0.459
Residual Coefficient	A_e	$U(0, 2.3)$	0.0609	0.0523	0.0691

References

- Adam, Klaus, Albert Marcet, and Johannes Beutel. “Stock price booms and expected capital gains”. *American Economic Review*, *Forthcoming* (2017).
- Adam, Klaus, Albert Marcet, and Juan Pablo Nicolini. “Stock market volatility and learning”. *The Journal of Finance* 71.1 (2016), pp. 33–82.
- Ai, Hengjie et al. “News shocks and the production-based term structure of equity returns”. *The Review of Financial Studies* 31.7 (2018), pp. 2423–2467.
- Albuquerque, Rui et al. “Valuation risk and asset pricing”. *The Journal of Finance* 71.6 (2016), pp. 2861–2904.
- Aldrich, Eric M and A Ronald Gallant. “Habit, long-run risks, prospect? a statistical inquiry”. *Journal of Financial Econometrics* 9.4 (2011), pp. 589–618.
- Backus, David K, Bryan R Routledge, and Stanley E Zin. “The cyclical component of US asset returns” (2010).
- Bansal, Ravi, A Ronald Gallant, and George Tauchen. “Rational pessimism, rational exuberance, and asset pricing models”. *The Review of Economic Studies* 74.4 (2007), pp. 1005–1033.
- Bansal, Ravi, Dana Kiku, and Amir Yaron. “An Empirical Evaluation of the Long-Run Risks Model for Asset Prices”. *Critical Finance Review* 1.1 (2012), pp. 183–221.
- Bansal, Ravi and Amir Yaron. “Risks for the long run: A potential resolution of asset pricing puzzles”. *The Journal of Finance* 59.4 (2004), pp. 1481–1509.
- Barro, Robert J. “Rare disasters and asset markets in the twentieth century”. *The Quarterly Journal of Economics* 121.3 (2006), pp. 823–866.
- “Rare disasters, asset prices, and welfare costs”. *American Economic Review* 99.1 (2009), pp. 243–64.
- Beeler, Jason and John Y. Campbell. “The Long-Run Risks Model and Aggregate Asset Prices: An Empirical Assessment”. *Critical Finance Review* 1.1 (2012), pp. 141–182. ISSN: 2164-5744.
- Bekaert, Geert and Eric Engstrom. “Asset return dynamics under habits and bad environment–good environment fundamentals”. *Journal of Political Economy* 125.3 (2017), pp. 713–760.
- Bianchi, Francesco, Martin Lettau, and Sydney C Ludvigson. *Monetary policy and asset valuation*. Tech. rep. National Bureau of Economic Research, 2016.
- Bidder, Rhys and Ian Dew-Becker. “Long-run risk is the worst-case scenario”. *American Economic Review* 106.9 (2016), pp. 2494–2527.

- Bloom, N. “The Impact of Uncertainty Shocks”. *Econometrica* 77.3 (2009), pp. 623–685.
- Brunnermeier, Markus K and Yuliy Sannikov. “A macroeconomic model with a financial sector”. *The American Economic Review* 104.2 (2014), pp. 379–421.
- Campbell, John Y. “Consumption-based asset pricing”. *Handbook of the Economics of Finance* 1 (2003), pp. 803–887.
- Campbell, John Y. and John H Cochrane. “By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior”. *Journal of Political Economy* 107.2 (1999), pp. 205–251.
- Chen, Andrew Y. “External Habit in a Production Economy: A Model of Asset Prices and Consumption Volatility Risk”. *The Review of Financial Studies* 30.8 (2017), pp. 2890–2932.
- Chen, Long, Zhi Da, and Richard Priestley. “Dividend smoothing and predictability”. *Management science* 58.10 (2012), pp. 1834–1853.
- Christiano, Lawrence J and Terry J Fitzgerald. “The band pass filter”. *international economic review* 44.2 (2003), pp. 435–465.
- Cochrane, J.H. “Presidential address: Discount rates”. *The Journal of Finance* 66.4 (2011), pp. 1047–1108.
- Cochrane, John H. “Macro-finance”. *Review of Finance* 21.3 (2017), pp. 945–985.
- Colacito, Riccardo and Mariano M Croce. “Risks for the long run and the real exchange rate”. *Journal of Political economy* 119.1 (2011), pp. 153–181.
- Colacito, Riccardo et al. “Bkk the ez way”. *American Economic Review* (Forthcoming).
- Croce, Mariano Massimiliano. “Long-run productivity risk: A new hope for production-based asset pricing?” *Journal of Monetary Economics* 66 (2014), pp. 13–31. ISSN: 0304-3932.
- Drechsler, Itamar and Amir Yaron. “What’s vol got to do with it”. *The Review of Financial Studies* 24.1 (2010), pp. 1–45.
- Fama, Eugene F. “Stock returns, real activity, inflation, and money”. *The American economic review* 71.4 (1981), pp. 545–565.
- Fama, Eugene F and Kenneth R French. “Disappearing dividends: changing firm characteristics or lower propensity to pay?” *Journal of Financial economics* 60.1 (2001), pp. 3–43.
- Favilukis, Jack and Xiaoji Lin. “Wage rigidity: A quantitative solution to several asset pricing puzzles”. *Review of Financial Studies* (2015), pp. 148–192.

- Person, Wayne E and Campbell R Harvey. “Seasonality and Consumption-Based Asset Pricing”. *The Journal of Finance* 47.2 (1992), pp. 511–552.
- Gabaix, Xavier. “Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance”. *The Quarterly Journal of Economics* 127.2 (2012), pp. 645–700.
- Gallant, A Ronald, Mohammad R Jahan-Parvar, and Hening Liu. “Does Smooth Ambiguity Matter for Asset Pricing?” (2018).
- Gallant, A Ronald and Robert E McCulloch. “On the determination of general scientific models with application to asset pricing”. *Journal of the American Statistical Association* 104.485 (2009), pp. 117–131.
- Godsill, Simon J, Arnaud Doucet, and Mike West. “Monte Carlo smoothing for nonlinear time series”. *Journal of the american statistical association* 99.465 (2004), pp. 156–168.
- Guvenen, Fatih. “A parsimonious macroeconomic model for asset pricing”. *Econometrica* 77.6 (2009), pp. 1711–1750. ISSN: 1468-0262.
- Hansen, Lars Peter and Thomas J Sargent. “Fragile beliefs and the price of uncertainty”. *Quantitative Economics* 1.1 (2010), pp. 129–162.
- He, Zhiguo and Arvind Krishnamurthy. “Intermediary asset pricing”. *The American Economic Review* 103.2 (2013), pp. 732–770.
- Herbst, Edward and Frank Schorfheide. “Bayesian inference for DSGE models”. *Manuscript. Board of Governors of the Federal Reserve System, Washington, DC* (2014).
- Jahan-Parvar, Mohammad R and Hening Liu. “Ambiguity aversion and asset prices in production economies”. *Review of Financial Studies* (2014), hhu037.
- Jermann, Urban J. “Asset pricing in production economies”. *Journal of Monetary Economics* 41.2 (1998), pp. 257–275.
- Ju, Nengjiu and Jianjun Miao. “Ambiguity, learning, and asset returns”. *Econometrica* 80.2 (2012), pp. 559–591.
- Lettau, Martin and Sydney C Ludvigson. “Shocks and crashes”. *NBER Macroeconomics Annual* 28.1 (2014), pp. 293–354.
- “Understanding trend and cycle in asset values: Reevaluating the wealth effect on consumption”. *american economic review* 94.1 (2004), pp. 276–299.
- Lustig, Hanno and Stijn Van Nieuwerburgh. *Can Housing Collateral Explain Long-Run Swings in Asset Returns?* Tech. rep. National Bureau of Economic Research, 2006.

- Lustig, Hanno and Stijn Van Nieuwerburgh. “How much does household collateral constrain regional risk sharing?” *Review of Economic Dynamics* 13.2 (2010), pp. 265–294.
- Mankiw, N Gregory and Stephen P Zeldes. “The consumption of stockholders and nonstockholders”. *Journal of financial Economics* 29.1 (1991), pp. 97–112.
- Marfè, Roberto and Julien Penasse. “The Time-Varying Risk of Macroeconomic Disasters”. *Available at SSRN 2773874* (2016).
- Mehra, Rajnish and Edward C Prescott. “The equity premium: A puzzle”. *Journal of monetary Economics* 15.2 (1985), pp. 145–161.
- Messow, Philip and Walter Krämer. “Spurious persistence in stochastic volatility”. *Economics Letters* 121.2 (2013), pp. 221–223.
- Mitchell, Wesley Clair and Arthur F Burns. “Statistical indicators of cyclical revivals”. *Statistical indicators of cyclical revivals*. NBER, 1938, pp. 1–12.
- Nakamura, Emi, Dmitriy Sergeyev, and Jon Steinsson. “Growth-rate and uncertainty shocks in consumption: Cross-country evidence”. *American Economic Journal: Macroeconomics* 9.1 (2017), pp. 1–39.
- Papanikolaou, Dimitris. “Investment Shocks and Asset Prices”. *Journal of Political Economy* 119.4 (2011), pp. 639–685.
- Rietz, Thomas A. “The equity risk premium a solution”. *Journal of monetary Economics* 22.1 (1988), pp. 117–131.
- Sbuelz, Alessandro and Fabio Trojani. “Asset prices with locally constrained-entropy recursive multiple-priors utility”. *Journal of Economic Dynamics and Control* 32.11 (2008), pp. 3695–3717.
- Schorfheide, Frank, Dongho Song, and Amir Yaron. “Identifying long-run risks: A bayesian mixed-frequency approach”. *Econometrica* (Forthcoming).
- Schwert, G William. “Stock returns and real activity: A century of evidence”. *The Journal of Finance* 45.4 (1990), pp. 1237–1257.
- Shiller, RJ. “Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?” *American Economic Review* 71.3 (1981), pp. 421–436.
- Verdelhan, Adrien. “A Habit-Based Explanation of the Exchange Rate Risk Premium”. *The Journal of Finance* 65.1 (2010), pp. 123–146.
- Vissing-Jørgensen, Annette. “Limited asset market participation and the elasticity of intertemporal substitution”. *Journal of political Economy* 110.4 (2002), pp. 825–853.

- Vissing-Jørgensen, Annette and Orazio P Attanasio. “Stock-market participation, intertemporal substitution, and risk-aversion”. *American Economic Review* 93.2 (2003), pp. 383–391.
- Wachter, Jessica A. “A consumption-based model of the term structure of interest rates”. *Journal of Financial economics* 79.2 (2006), pp. 365–399.
- “Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?” *The Journal of Finance* 68.3 (2013), pp. 987–1035.
- Wilcox, David W. “The construction of US consumption data: some facts and their implications for empirical work”. *The American economic review* (1992), pp. 922–941.
- Yang, Wei. “Intertemporal Substitution and Equity Premium”. *Review of Finance* 20.1 (2016), pp. 403–445.