Does the consumption-income ratio predict returns on wealth? Long-run evidence for industrial economies*

Lorenzo Pozzi†[1]

[1] *Erasmus University Rotterdam & Tinbergen Institute*

November 2020

Abstract

Intertemporal budget constraint reasoning suggests that the current aggregate consumption to income ratio may contain information about future returns on wealth. This paper investigates whether this ratio has predictive power for real asset returns using historical data over the period $1870-2015$ for four major industrial economies (France, Germany, the UK and the US). The sign and strength of the predictive relationship are investigated for the excess and raw returns obtained from holding equity and long-term government bonds. The short-run and long-run impact estimates of the log consumption-income ratio on these asset returns are calculated using a Bayesian vector autoregression (VAR) approach with time-varying parameters and stochastic volatilities. We find that the consumption-income ratio has substantial predictive power for, in particular, the real (excess) government bond returns of all countries considered. This supports the notion that bond returns may be good proxy's for the returns on total wealth, as is suggested in the literature. The predictive ability of the consumption-income ratio does not appear to be driven by business cycle fluctuations.

JEL Classification: E21, C32, C11

Keywords: consumption, income, asset returns, return predictability, intertemporal budget constraint, historical returns data, historical macro data

---

*For constructive comments received, the author gratefully acknowledges Michel van der Wel and participants to different conferences and seminars (2019 Tinbergen Institute macro research day, 2020 Virtual EEA conference, 2020 Virtual SNDE conference, 2020 seminar at the Dutch Central Planning Bureau CPB).

†Department of Economics, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands. Email: pozzi@ese.eur.nl. Website: http://people.few.eur.nl/pozzi.

1 Introduction

Whether and to what extent returns on wealth are predictable is a question of great relevance for both practitioners - who aim to enhance asset allocation and investment performance - and academics - who aim to construct models that explain the data convincingly. A large literature has argued that wealth returns, in particular US stock returns during the postwar period, can be predicted using economic variables. We refer to Rapach and Zhou (2013) for a comprehensive overview of the literature. Ratios consisting of financial and/or macroeconomic variables are among the predictors that have received considerable attention. Prominent and widely used ratios are, for instance, the dividend-price ratio (see e.g., Campbell and Shiller, 1988; Fama and French, 1989, for early evidence for respectively stock and bond returns) and the consumption-wealth ratio or 'cay' (see e.g., Lettau and Ludvigson, 2001).

This paper contributes to the literature by proposing an alternative ratio as a potential predictor for returns on wealth, namely the aggregate private consumption to aggregate total disposable income ratio (where total income refers to the sum of labor and capital income). This ratio, which we theoretically motivate from intertemporal budget constraint logic, is sufficiently different from more conventional ratios to warrant an investigation into its own distinctive predictability characteristics. For instance, unlike the dividend-price ratio, it is consumption-based. Also, contrary to both the dividend-price and consumption-wealth ('cay') ratios, whose variation to a large extent reflects fluctuations in equity markets, the less volatile and more persistent consumption-income ratio is less driven by financial markets.

We first derive the predictive relationship between the log consumption-income ratio and future real returns on wealth from the intertemporal budget constraint of a representative consumer, i.e., the intertemporal budget constraint can be satisfied if a high current consumption to total disposable income ratio is followed by high subsequent returns on wealth. We then investigate whether this ratio has predictive power for both excess and raw returns on equity and long-term government bonds using historical data over the period $1870-2015$ for four major industrial economies (France, Germany, the UK and the US). Real equity returns are conventionally used to proxy returns on total wealth but recent research argues that real long-term bond returns may be better proxies for returns on total wealth (see Lustig et al., 2013). Our use of historical data is motivated by the theoretical underpinnings of our predictive relationship which is the intertemporal budget constraint, a long-run concept. Moreover, more so than the conventional ratios, the consumption-income ratio is a slow moving and highly persistent variable that we use to predict highly volatile returns with a sizeable unpredictable component, i.e., the estimations are characterized by a low signal-to-noise ratio. By using a long historical sample period and then combining the results across countries, we obtain a more powerful test for the detection of a predictive relationship between these variables.

Methodologically, our main estimations are conducted using a Bayesian vector autoregression (VAR) approach with time-varying parameters and stochastic volatilities (see e.g., Primiceri, 2005). We explicitly test for time variation in parameters and volatilities following the approach of Frühwirth-Schnatter and Wagner (2010). Allowing and testing for time variation in parameters and volatilities is useful given the long time span $1870 - 2015$ of the data that we use. Furthermore, as we discuss in the paper, the VAR set-up allows to deal with a number of statistical complications inherent to return prediction (see e.g., Stambaugh, 1999). From the VAR, we estimate both the predictive impact of the log consumption-income ratio for (excess) returns in the short run (for a one-year horizon) and in the long-run (for 10-year and 20-year horizons). We further average our per country results across the four considered economies using a 'mean-group' approach. Finally, in line with previous literature on return prediction, we investigate to what extent the predictive power of the consumption-income ratio for (excess) returns is driven by business cycle fluctuations (see e.g., Fama and French, 1989; Dangl and Halling, 2012; Golez and Koudijs, 2018).

Our paper is related to four, not mutually exclusive, strands of the literature. First, it is related to the literature on return predictability that uses intertemporal budget constraint logic (see e.g., Lettau and Ludvigson, 2001, 2004, 2005; Whelan, 2008; Bianchi et al., 2017). Second, it is related to the empirical literature that investigates return predictability using predictive regressions with time-varying parameters (see e.g., Dangl and Halling, 2012). Third, it is related to the literature that focusses on return predictability in an international context (see e.g., Ang and Bekaert, 2007). Finally, the paper is in the vein of studies that use historical data to investigate whether asset returns are predictable (see e.g., Chen, 2009; Della Corte et al., 2010; Golez and Koudijs, 2018).

Our results suggest that the consumption-income ratio has substantial predictive power for the (excess) government bond returns of all four countries considered. On average over time and countries, if the consumption-income ratio increases with 1%, gross excess bond returns increase with about 0.14% the following year and with 1.7% to 2% the following 20 years. The evidence with respect to (excess) equity returns is generally inconclusive however. These results imply that bond returns may be good proxy's for returns on total wealth as suggested by Lustig et al. (2013). We further find that the predictive ability of the consumption-income ratio does not appear to be driven by business cycle fluctuations but may reflect, in the words of Fama and French (1989), long-term structural 'business conditions'. Importantly, when investigating the predictive impact of the log consumption-income ratio for (excess) equity and bond returns using a more conventional dataset (i.e., quarterly postwar data for the US), we also find a positive predictive impact of this ratio on both equity and bond (excess) returns but the results are not fully conclusive. This provides further support for the use of a historical dataset to explore the question at hand.

Our findings are quite different from - and, therefore, complementary to - the findings reported in the literature for conventional ratios used as predictors for returns. They reflect the different nature of these ratios. The results reported in the literature for ratios like the dividend-price and consumption-wealth ratios can generally be obtained using postwar data. These ratios have predictive ability for, in particular, equity returns which reflects the fact that both these ratios are strongly driven by the stock market.[1] And these predictability results stem, at least to a certain extent, from cyclical fluctuations. The predictability results that we report in this paper for the consumption to disposable income ratio necessitate the use of long-run historical data to obtain conclusive results at the country level. This ratio is shown to have predictive ability for bond returns more than for equity returns. And the reported predictability results seem to be structural in nature, rather than cyclical.[2]

The paper proceeds as follows. In Section 2, we present a theoretical framework that links the current consumption to disposable income ratio to expected future returns on wealth. Section 3 motivates and details the dataset used and provides some preliminary evidence. Section 4 presents and discusses the VAR approach, in particular the empirical specification and the estimation method. Section 5 reports the basic predictability results. Section 6 investigates to what extent the predictability results are driven by business cycle fluctuations. Section 7 concludes.

2 Theory

This section presents a simple framework that links the current consumption to (disposable) income ratio to expected future returns on wealth. In line with the dataset used which is discussed in Section 3 below and which consists of consumption, income and returns data at the country level, we assume a representative agent economy. Total wealth - i.e., the sum of asset and human wealth - is assumed to be tradeable (see e.g., Campbell and Mankiw, 1989; Lettau and Ludvigson, 2005). The gross real rate of return R_t on total tradeable wealth can be written as $R_{t+1} = \frac{P_{t+1}+Y_{t+1}}{P_t}$ where P_t is the ex-dividend price of a share of total wealth and Y_t is the real dividend or income obtained from total wealth which consists of labor and capital income (as obtained from human wealth, respectively asset wealth). If the

[1] See, however, Afonso and Sousa (2011), for the predictive power of the consumption-wealth ratio for bond returns.

[2] As such, our paper adds to the literature on bond return predictability by identifying the consumption-income ratio as a variable that has predictive power for real bond returns in the long-run, i.e., predictability is detected by using low frequency historical data. This contrasts with the high frequency predictors typically considered in the literature (e.g., forward rates), whose predictive ability is typically related to business cycle fluctuations. See e.g., Gargano et al. (2019) for a recent paper on postwar US bond return predictability at the monthly frequency.

agent's intertemporal budget constraint holds, we can approximate the aggregate log consumption to income ratio $c_t - y_t$ in period t by,

$$c_t - y_t = E_t \sum_{j=1}^{\infty} \left[\kappa^j \left(\Delta y_{t+j} - r_{t+j}\right) - \rho^j \left(\Delta c_{t+j} - r_{t+j}\right)\right] \tag{1}$$

where E_t is the expectations operator conditional on period t information, r_t is the log of the gross real rate of return on total wealth R_t, c_t is the log of real consumption C_t, y_t is the log of real income Y_t, and where κ and ρ are discount rates which are close to one. We refer to Appendix A for the derivation of equation (1). The intuition behind equation (1) is straightforward. If the budget constraint holds intertemporally, a high consumption-income ratio in period t implies higher subsequent expected discounted income growth rates or lower subsequent expected discounted consumption growth rates.

The direct impact of the log consumption-income ratio $c_t - y_t$ on future returns r_{t+j} in equation (1) is ambiguous and most likely not substantial as the discount rates κ and ρ may not differ much. The total, indirect as well as direct, impact of $c_t - y_t$ on r_{t+j} is obtained by noting that the variables Δy_{t+1} and Δc_{t+1} can be written as functions of returns r_{t+1}.

First, we use the relationship between the gross return on, the income from and the price of wealth, i.e., $R_{t+1} = \frac{P_{t+1}+Y_{t+1}}{P_t}$, to write real income (dividend) growth as,

$$\Delta y_{t+1} = r_{t+1} + \omega^y_{t+1} \tag{2}$$

where $\omega^y_{t+1} \equiv -\left(\ln\left(1 + \frac{P_{t+1}}{Y_{t+1}}\right) - \ln\left(\frac{P_t}{Y_t}\right)\right)$.[3] The term ω^y_{t+1} is assumed to be uncorrelated with time t information, i.e., we have $E_t(\omega^y_{t+1}) = 0$.[4]

Second, with time-varying returns on wealth and isoelastic utility, the log-linear version of the first-order condition or Euler equation of a representative utility maximizing consumer is given by,

$$\Delta c_{t+1} = \sigma r_{t+1} + \omega^c_{t+1} \tag{3}$$

where σ is the elasticity of intertemporal substitution and ω^c_{t+1} is an expectation error that is uncorrelated with time t information, i.e., we have $E_t(\omega^c_{t+1}) = 0$. Theoretically, the elasticity of intertemporal substitution σ is positive while the empirical literature strongly suggests that this elasticity is smaller than one.[5] As such, we can write $0 < \sigma < 1$.

[3] To see this, multiply both sides of $R_{t+1} = \frac{P_{t+1}+Y_{t+1}}{P_t}$ by $\frac{P_t}{Y_t}$ to obtain $\frac{R_{t+1}P_t}{Y_t} = \frac{P_{t+1}+Y_{t+1}}{Y_t}$ which can be written as $\frac{R_{t+1}P_t}{Y_t} = \frac{Y_{t+1}}{Y_t}\left(1 + \frac{P_{t+1}}{Y_{t+1}}\right)$ (see e.g., Cochrane, 2005, page 398). After taking logs, this gives $r_{t+1} = \Delta y_{t+1} + \ln\left(1 + \frac{P_{t+1}}{Y_{t+1}}\right) - \ln\left(\frac{P_t}{Y_t}\right)$.

[4] This can be justified by noting that P_t is typically large compared to Y_t so that $\ln\left(1 + \frac{P_{t+1}}{Y_{t+1}}\right) \approx \ln\left(\frac{P_{t+1}}{Y_{t+1}}\right)$ while $\ln\left(\frac{P_{t+1}}{Y_{t+1}}\right)$ is a very persistent variable that approximates a random walk (see Cochrane, 2005, for evidence for equity).

[5] Using a meta-analysis of 169 studies - both at the micro and the macro level - that cover 104 countries, Havranek et al. (2015) find a mean elasticity of intertemporal substitution equal to 0.5 with the largest value found for Japan at 0.9.

Substituting equations (2) and (3) into equation (1), we obtain,

$$c_t - y_t = E_t \sum_{j=1}^{\infty} \rho^j \left[(1-\sigma) r_{t+j}\right] \tag{4}$$

Given that we typically observe time variation in $c_t - y_t$, equation (4) implies that $c_t - y_t$ has predictive power for future returns on wealth.[6] Moreover, since $\sigma < 1$, the predictive relationship between the log consumption-income ratio and future returns is expected to be positive, i.e., a high consumption-income ratio in period t is likely to be followed by high subsequent rates of return on wealth if the budget constraint is to hold intertemporally.

We note, finally, that the predictive relationship derived between the consumption-income ratio and future returns holds also under more general conditions. For instance, if the economy consists of both permanent income consumers who consume according to equation (3) and rule-of-thumb consumers who, in each period, consume a fraction λ (with $0 < \lambda < 1$) of income growth, then we can replace $(1-\sigma)$ in equation (4) by $(1-\lambda)(1-\sigma)$, i.e., the predictive ability of the log consumption-income ratio for future returns, while reduced, is still positive.[7] Ultimately, the magnitude of the predictive impact of the consumption-income ratio for returns is an empirical issue. To this issue, we now turn.

3 Data and preliminary evidence

We investigate whether the consumption-income ratio has predictive power for real asset returns using historical data over the period $1870-2015$ for four major industrial economies (France, Germany, the UK and the US). Data availability determines the countries included in the dataset and the periods considered per country. The use of a historical dataset to investigate the predictive ability of the consumption-income ratio for the returns on wealth is motivated both by the underlying theory (i.e., the intertemporal budget constraint, a long-run concept) and by the data characteristics (i.e., a highly persistent consumption-income ratio and highly volatile returns with a sizeable unpredictable component). It is also motivated by the lack of conclusive evidence obtained when estimating this relationship for individual countries using a more conventional dataset. To show this, Appendix C presents predictability results obtained for the consumption-income ratio using quarterly postwar data for the US. In the remainder of this section, we discuss the data used for the log consumption-income ratio $c_t - y_t$ and for the returns on wealth r_t. We also present some preliminary evidence on the predictive ability of the consumption-income ratio for returns on wealth.

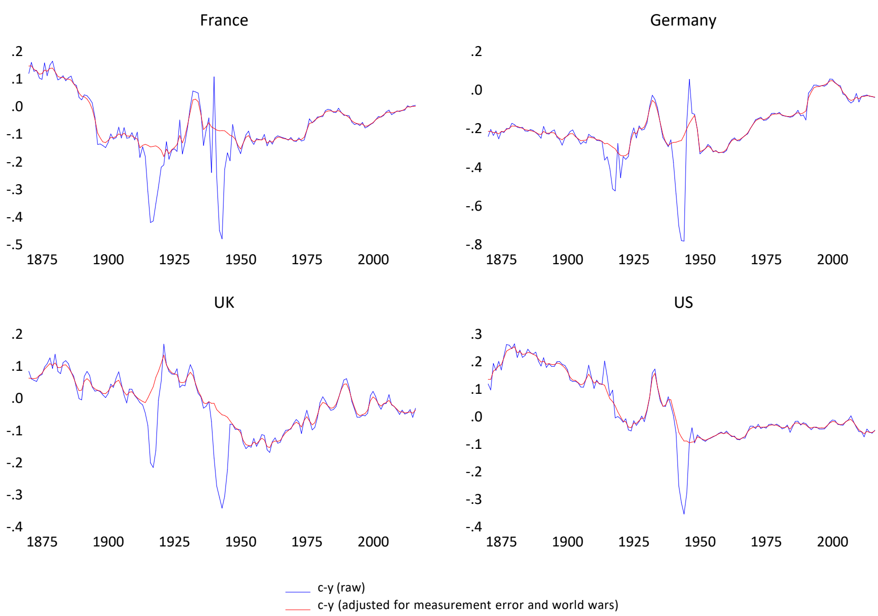
[6] To see this, note that $var\,(c_t - y_t) = cov\left(c_t - y_t, \sum_{j=1}^{\infty} \rho^j (1-\sigma) r_{t+j}\right)$.

[7] To see this, in line with Campbell and Mankiw (1989), use $\Delta c_{t+1} = (1-\lambda)\sigma r_{t+1} + \lambda \Delta y_{t+1} + (1-\lambda)\omega^c_{t+1}$ instead of equation (3) when deriving equation (4).

3.1 Consumption and income

The log of real per capita consumption c_t is calculated from historical data on total consumer expenditures as reported by (see Jordà et al., 2016) while the log of real per capita disposable income y_t is calculated from historical data on disposable income - i.e., national income after taxes - that accompanies Piketty and Zucman (2014). The latter series are only available for the four countries considered. We have conducted estimations also for other countries using the more widely available GDP series to proxy y_t which captures pre-tax income rather than disposable income, but the obtained results were generally inconclusive. We refer to Appendix B for the construction of the data and for detailed information on the data sources. In this appendix, we also outline how we calculate cleaned series for consumption and income by taking out transitory variation due to measurement error and due the occurrence of both world wars. Since the difference between log consumption and log income is the explanatory variable in our predictive regressions, this transitory variation could potentially obscure a long-run predictive relationship between the consumption-income ratio and asset returns. In Figure 1, we present graphs depicting both the measured (raw) log consumption-income ratio and the adjusted (cleaned) log consumption-income ratio for all four countries of our sample.

Figure 1: The log consumption-income ratio: raw and cleaned for measurement error and world wars

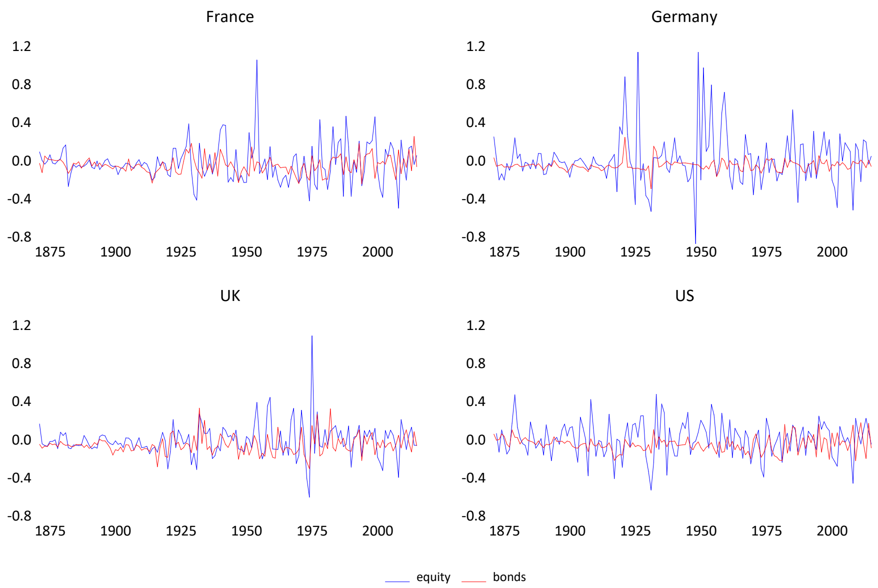


3.2 Returns

For the variable r_t, we consider two asset categories. The conventional variable used as a proxy for the return on total wealth is the real return on equity (see e.g., Lettau and Ludvigson, 2001; Cochrane, 2005).

Recent work by Lustig et al. (2013) argues, however, that human wealth in the US is much larger in size than financial wealth and that total wealth (i.e., the sum of asset and human wealth) has return dynamics that resemble that of long-term bonds rather than that of equity.[8] Hence, we are also interested in the predictive power of the consumption-income ratio for real long-term government bond returns. Following most of the literature, we focus on excess returns, i.e., the difference between returns and a risk-free rate. As such, we investigate the predictive power of the consumption-income ratio for the time-varying reward for risk. Nonetheless, we also report results obtained from estimating our predictive relationship using raw equity and bond returns. All returns used are expressed in real terms using the inflation rate. We refer to Appendix B for further details on the construction of the returns data and on its sources. In Figure 2 we present graphs of both equity excess returns and bond excess returns for all four countries in our sample.

Figure 2: Excess returns on equity and long-term government bonds



3.3 Preliminary evidence

Before turning to the more comprehensive VAR analysis, we take a look at the results of estimating simple predictive regressions of the following form,

$$r_{t+1} = a + b(c_t - y_t) + \epsilon_{t+1} \qquad \epsilon_{t+1} \sim iid\mathcal{N}\left(0, \sigma_\epsilon^2\right) \tag{5}$$

[8]They estimate a correlation for the US economy between their calculated total wealth returns and asset returns equal to 27% for stocks and equal to 94% for 5-year government bonds.

with r_t the period t log gross real return and $c_t - y_t$ the log consumption to income ratio with the parameter of interest being b. We estimate equation (5) using Bayesian OLS with uninformative priors.[9] We use excess and raw equity and bond returns for r_t and raw and cleaned data for y_t and c_t with the data as described in the previous subsections.

Table 1 presents the posterior means and the 90% highest posterior density (HPD) intervals for the parameter b. From the table, we draw two preliminary conclusions. First, the positive impact of the consumption-income ratio on returns predicted by theory seems to hold for bond returns more than for equity returns. Second, while the magnitude of the estimates differs, predictability for bond returns is found *both* with cleaned and with raw income and consumption data. Similarly, we note that the predictability results reported in the remainder of the paper, while based on cleaned income and consumption data, can also be obtained using raw data.

Table 1: The predictive impact of $c_t - y_t$ on r_{t+1}: preliminary evidence

	c_t,y_t raw				c_t,y_t cleaned			
	France	Germany	UK	US	France	Germany	UK	US
Excess equity returns for r_{t+1}								
	0.16	-0.10	-0.15	-0.02	0.11	-0.27	-0.24	0.01
	[-0.08,0.40]	[-0.36,0.17]	[-0.39,0.08]	[-0.23,0.19]	[-0.22,0.46]	[-0.63,0.11]	[-0.55,0.07]	[-0.23,0.25]
Excess bond returns for r_{t+1}								
	0.12	0.01	0.10	0.09	0.23	0.05	0.17	0.13
	[0.01,0.24]	[-0.05,0.07]	[-0.05,0.25]	[0.00,0.19]	[0.07,0.39]	[-0.03,0.13]	[-0.03,0.36]	[0.02,0.24]
Raw equity returns for r_{t+1}								
	0.49	0.12	0.12	-0.03	0.40	0.09	0.00	-0.02
	[0.25,0.74]	[-0.15,0.41]	[-0.13,0.38]	[-0.24,0.18]	[0.05,0.77]	[-0.29,0.49]	[-0.34,0.33]	[-0.26,0.22]
Raw bond returns for r_{t+1}								
	0.46	0.24	0.38	0.08	0.53	0.43	0.41	0.09
	[0.33,0.60]	[0.13,0.34]	[0.22,0.54]	[-0.03,0.19]	[0.33,0.73]	[0.28,0.58]	[0.19,0.62]	[-0.03,0.21]

Notes: Reported are the posterior mean and the 90% highest posterior density interval (in square brackets) of the parameter b in equation (5). Parameter estimates for a and σ_ϵ^2 are unreported but available upon request. The estimation method is Bayesian OLS.

4 VAR estimation

In this section, we detail the VAR approach followed to investigate the predictive ability of the log consumption-income ratio for returns. We first present and discuss the empirical specification. Next, we elaborate on the Bayesian estimation approach.

[9] The Gaussian prior distributions used for a and b have mean zero and unit variance.

4.1 Empirical specification

4.1.1 A time-varying parameter vector autoregression

Based on some of the previous literature on return predictability (see e.g., Stambaugh, 1999; Cochrane, 2008; Rapach and Zhou, 2013), we investigate the predictive ability of the log consumption-income ratio $c_t - y_t$ for asset returns by estimating the following restricted vector autoregression (VAR) model,

$$r_{t+1} = \alpha_{t+1} + \beta_{t+1}(c_t - y_t) + \gamma_{t+1} r_t + \psi_{t+1} \tag{6}$$

$$c_{t+1} - y_{t+1} = \pi_{0,t+1} + \pi_{1,t+1}(c_t - y_t) + \pi_{2,t+1}(c_{t-1} - y_{t-1}) + \eta_{t+1} \tag{7}$$

where r_t is the period t log gross return on wealth (in real terms) and where $\begin{pmatrix} \eta_{t+1} & \psi_{t+1} \end{pmatrix}' \sim \mathcal{N}(0, \Omega_{t+1})$. The number of lags included in equations (6)-(7) is deemed sufficient based on estimated autocorrelation and partial autocorrelation functions of the residuals.[10] The intercepts α_t and $\pi_{0,t}$ and the slope coefficients β_t, γ_t, $\pi_{1,t}$ and $\pi_{2,t}$ as well as the elements in the variance-covariance matrix Ω_t are assumed to be potentially time-varying, i.e., we estimate a time-varying parameter VAR (TVP-VAR) with stochastic volatilities. Following Primiceri (2005), the variance-covariance matrix Ω_{t+1} is decomposed as,

$$\Delta_{t+1} \Omega_{t+1} \Delta'_{t+1} = \Sigma_{t+1} \Sigma'_{t+1} \tag{8}$$

where Δ_{t+1} is the lower triangular matrix $\begin{pmatrix} 1 & 0 \\ -\delta_{t+1} & 1 \end{pmatrix}$ with δ_{t+1} a transformation of the potentially time-varying covariance between the error terms η_{t+1} and ψ_{t+1} and where Σ_{t+1} is a diagonal matrix containing the volatilities of the structural shocks of the VAR.[11] Applying this decomposition to equations (6)-(7), we can write,

$$\begin{pmatrix} \eta_{t+1} \\ \psi_{t+1} \end{pmatrix} = \Delta_{t+1}^{-1} \Sigma_{t+1} \begin{pmatrix} \eta^*_{t+1} \\ \varepsilon^*_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \delta_{t+1} & 1 \end{pmatrix} \begin{pmatrix} \sigma_{\eta,t+1} & 0 \\ 0 & \sigma_{\varepsilon,t+1} \end{pmatrix} \begin{pmatrix} \eta^*_{t+1} \\ \varepsilon^*_{t+1} \end{pmatrix} \tag{9}$$

where $\begin{pmatrix} \eta^*_{t+1} & \varepsilon^*_{t+1} \end{pmatrix}' \sim iid\mathcal{N}(0, I_2)$ and where $\sigma_{\eta,t+1}$ and $\sigma_{\varepsilon,t+1}$ are the time-varying standard deviations of the structural shocks $\eta_{t+1} = \sigma_{\eta,t+1}\eta^*_{t+1}$ and $\varepsilon_{t+1} = \sigma_{\varepsilon,t+1}\varepsilon^*_{t+1}$. Using this result, we rewrite equations (6)-(7) as,

$$r_{t+1} = \alpha_{t+1} + \beta_{t+1}(c_t - y_t) + \gamma_{t+1} r_t + \delta_{t+1}\eta_{t+1} + e^{h_{\varepsilon,t+1}} \varepsilon^*_{t+1} \qquad \varepsilon^*_{t+1} \sim iid\mathcal{N}(0,1) \tag{10}$$

$$c_{t+1} - y_{t+1} = \pi_{0,t+1} + \pi_{1,t+1}(c_t - y_t) + \pi_{2,t+1}(c_{t-1} - y_{t-1}) + e^{h_{\eta,t+1}} \eta^*_{t+1} \qquad \eta^*_{t+1} \sim iid\mathcal{N}(0,1) \tag{11}$$

where $h_{\varepsilon,t+1} = \ln \sigma_{\varepsilon,t+1}$ and $h_{\eta,t+1} = \ln \sigma_{\eta,t+1}$ are the potentially time-varying log volatilities of the structural shocks. The error terms ε^*_{t+1} and η^*_{t+1} are independent while the covariance between the error

[10] More specifically, the autocorrelation and partial autocorrelation functions are calculated from the structural residuals ε^*_{t+1} and η^*_{t+1} obtained from estimating equations (10)-(11) below. Results are unreported but available upon request.

[11] It can be shown that $\delta_{t+1} = \frac{cov(\psi_{t+1}, \eta_{t+1})}{\sigma^2_{\eta,t+1}}$ where $\sigma^2_{\eta,t+1}$ is the potentially time-varying variance of η_{t+1}.

terms ψ_{t+1} and η_{t+1} of equations (6)-(7) is now captured through the term $\delta_{t+1}\eta_{t+1}$ in equation (10). This specification facilitates estimation as the equations can now be estimated one at a time, i.e., first equation (11) and then, conditional on this, equation (10). Full details are provided in Appendix D. Equations (10)-(11) constitute the observation equations from a state space model. The time-varying parameters and time-varying log volatilities are all assumed to follow random walks with innovations that are mutually independent as well as independent from the shocks ε^*_{t+1} and η^*_{t+1} in equations (10)-(11). These random walks constitute the state equations of the state space model. More details on the specification of the state space system are provided in Appendix D.

4.1.2 Discussion

The considered VAR has a number of features that allow for a sound investigation of the considered predictive relationship. First, while the theory discussed above implies a stable predictive relationship between the consumption-income ratio and returns, allowing and testing for structural shifts in parameters and log volatilities through the modelling of these quantities as time-varying processes is nonetheless useful from an empirical viewpoint given the long time span $1870-2015$ of the data that we use. Indeed, our sample includes fundamentally different episodes like the Great Depression, Bretton Woods, etc. Even over more recent periods like the Great Moderation or the Great Recession, time variation in the volatility of asset returns is well documented (see e.g., Tsay, 2005, and references therein) while many studies also deal with potential changes in the volatility of macroeconomic variables like GDP growth (see e.g., Hamilton, 2008; Nakamura et al., 2017). Second, the estimation of an equation for the predictor variable $c_{t+1}-y_{t+1}$ jointly with the equation for the predicted variable r_{t+1} allows for the estimation of a non-zero covariance between the shocks to the variables r_{t+1} and $c_{t+1}-y_{t+1}$. With a very persistent predictor variable like c_t-y_t, if we do not control for the covariance between r_{t+1} and $c_{t+1}-y_{t+1}$, the detection of a predictive impact of c_t-y_t on r_{t+1} could be due to the impact of c_t-y_t on $c_{t+1}-y_{t+1}$ rather than stemming from a true relationship between c_t-y_t and r_{t+1}. This is related to the occurrence in a frequentist context of biases of the type discussed by Stambaugh (1999). Third, our specification also contains a lag of the return in the equation for r_{t+1}. While this term is frequently omitted in empirical studies on return predictability, controlling for r_t in the regression of r_{t+1} on c_t-y_t serves to avoid detecting a relationship between these variables that is driven by the potential covariance between r_t and c_t-y_t. Finally, our VAR approach allows for the calculation of the short-run impact of the log consumption-income ratio c_t-y_t on the asset return in the next period, i.e., the parameter β_{t+1}, but also, indirectly, for the calculation of long-run impacts of c_t-y_t on asset returns, i.e., over horizons of numerous years. Hence, when presenting the estimation results in Section 5 below, a subsection details the calculation of the long-run impacts of the consumption-income ratio on asset returns from our VAR estimation and then presents these long-horizon results.

4.2 MCMC algorithm and stochastic model specification search

We estimate our TVP-VAR model in a state space framework using Bayesian methods. In particular, we use a Gibbs sampling approach which is a Markov Chain Monte Carlo (MCMC) method that allows to draw from the intractable joint posterior distribution of fixed parameters and latent states using tractable conditional distributions (see e.g., Carter and Kohn, 1994; Kim and Nelson, 1999). Our Bayesian approach is motivated from the presence of stochastic volatilities in the model that make the state space system non-linear, the standard Kalman filter inapplicable and the exact likelihood hard to evaluate (see Kim et al., 1998). It also provides a straightforward manner to test whether the time variation in the parameters α_t, β_t, γ_t, δ_t, $\pi_{0,t}$, $\pi_{1,t}$, $\pi_{2,t}$ and in the log volatilities $h_{\varepsilon,t}$ and $h_{\eta,t}$ - all of which are modelled as random walks - is relevant or not. Our testing procedure follows the stochastic model specification search of Frühwirth-Schnatter and Wagner (2010). For each of the parameters or volatilities that are potentially time-varying, this procedure provides probabilities that time variation is a relevant model attribute. Through this type of model specification, a more parsimonious VAR model can be obtained that helps to avoid potential over-parameterization problems mentioned in the literature (see e.g., Giannone et al., 2015). Full details of our approach are provided in Appendix D, where we also discuss our prior choices and provide details on the steps of the Gibbs sampler and on its convergence.

5 Main predictability results

In this section, we present our main results. First, results on the prevalence of time variation in the parameters and log volatilities of the estimated VAR are discussed. Second, we present the results from estimating the VAR under the restrictions imposed on the time variation in parameters and log volatilities. We note that, to save space, in these sections we only report the results obtained for excess equity returns and excess bond returns and not for raw returns.[12] Next, we discuss the results concerning the predictive power of the log consumption-income ratio for excess equity and bond returns as well as raw equity and bond returns for the four countries in our sample. We also report cross-country averages of this predictive impact, i.e., mean group results. Finally, we present and discuss the predictability results obtained over longer horizons, i.e., over periods of 10 and 20 years.

[12]The results for raw returns are available upon request.

5.1 Testing for time variation

We start by testing whether the time variation that we allow for in the VAR of equations (10)-(11) is actually relevant using the stochastic model specification search of Frühwirth-Schnatter and Wagner (2010). Table 2 reports posterior probabilities of time variation in the parameters α_{t+1}, β_{t+1}, γ_{t+1}, δ_{t+1} and $\pi_{l,t+1}$ (with $l = 0, 1, 2$) and in the log volatilities $h_{\varepsilon,t+1}$ and $h_{\eta,t+1}$ for the four countries that we consider. The prior probability of time variation is set to 50% in all cases. From the table, we note that there is evidence of substantial time variation only for the log volatility of excess equity and bond returns. Except for the US for which the annual time series for excess equity returns reveals rather stable volatility as can be seen in Figure 2, the posterior probabilities of time variation in the log volatility $h_{\varepsilon,t+1}$ are very close to or exactly equal to one in all cases. The absence of time variation in the parameter β_{t+1}, i.e., the predictive impact of the log consumption-income ratio on returns, suggests that, despite the long historical sample period considered, predictability - if present - is stable.[13] Given the absence of time variation in the AR parameters γ_{t+1}, $\pi_{1,t+1}$ and $\pi_{2,t+1}$, this also implies stable potential long-horizon effects of $c_t - y_t$ on returns.

Table 2: Posterior probabilities of time variation in the parameters and log volatilities of equations (10)-(11)

		Country			
	Parameter/log volatility	France	Germany	UK	US
Equation (10) with excess equity returns for r					
	α	0.01	0.01	0.02	0.01
	β	0.06	0.05	0.09	0.11
	γ	0.03	0.03	0.05	0.03
	δ	0.21	0.33	0.44	0.51
	h_ε	1.00	1.00	1.00	0.07
Equation (10) with excess bond returns for r					
	α	0.01	0.00	0.03	0.04
	β	0.06	0.01	0.06	0.07
	γ	0.06	0.04	0.03	0.09
	δ	0.17	0.10	0.17	0.15
	h_ε	1.00	0.98	1.00	1.00
Equation (11) for log-consumption ratio $c - y$					
	π_0	0.00	0.00	0.00	0.00
	π_1	0.01	0.00	0.01	0.01
	π_2	0.01	0.00	0.02	0.01
	h_η	0.02	0.03	0.02	0.02

Notes: The prior probabilities of time variation are set to 50% in all cases. For details on the calculation of these probabilities, we refer to Appendix D.

[13] Golez and Koudijs (2018) also report stable predictive ability of the dividend-price ratio for equity returns using historical data. Della Corte et al. (2010), on the other hand, document that there is no predictive ability of the consumption-wealth ratio ('cay') for equity returns prior to WWII.

5.2 VAR estimation results

Based on the results reported in the previous section, we estimate the VAR presented in equations (10)-(11) with restrictions imposed on the time variation of the parameters and log volatilities, i.e., time variation is withheld only for the log volatility $h_{\varepsilon,t+1}$ of the innovation of the return equation. Hence, we estimate,

$$r_{t+1} = \alpha + \beta(c_t - y_t) + \gamma r_t + \delta\eta_{t+1} + e^{h_{\varepsilon,t+1}}\varepsilon^*_{t+1} \qquad \varepsilon^*_{t+1} \sim iid\mathcal{N}(0,1) \tag{12}$$

$$c_{t+1} - y_{t+1} = \pi_0 + \pi_1(c_t - y_t) + \pi_2(c_{t-1} - y_{t-1}) + e^{h_\eta}\eta^*_{t+1} \qquad \eta^*_{t+1} \sim iid\mathcal{N}(0,1) \tag{13}$$

As noted in Section 4.1.1, equations (12)-(13) of the VAR are estimated equation-by-equation with equation (12) estimated conditional on equation (13), i.e., we estimate equation (12) with the innovation η_{t+1} to the log consumption-income ratio added as a regressor and with δ reflecting the covariance between the shocks to r_{t+1} and $c_{t+1} - y_{t+1}$. Hence, we estimate equation (13) only once and then estimate equation (12) conditional on equation (13) for every return considered. The estimation results for equations (12) and (13) are therefore presented in separate tables.

The posterior means and 90% highest posterior density (HPD) intervals of the parameters π_0, π_1 and π_2 and the log volatility h_η of the AR(2) process estimated in equation (13) are presented in Table 3. From the table, we note that the log consumption-income ratio is very persistent in all countries - as can also be observed in Figure 1 - with values for π_1 and π_2 close to the unit root case.[14] This high persistence is even more outspoken in our potential predictor of returns $c_t - y_t$, when compared to other ratios commonly considered in the literature, such as the price to dividend ratio or the consumption to wealth ratio.

Table 3: Estimation of the AR(2) process of equation (13) for the log consumption-income ratio

		Country			
	Parameter/log volatility	France	Germany	UK	US
Dependent variable is $c_{t+1} - y_{t+1}$					
	π_0	-.002	-.004	-.000	.000
		[-.007,.003]	[-0.01,.005]	[-.005,.004]	[-.005,.005]
	π_1	1.32	1.49	1.46	1.48
		[1.00,1.63]	[1.26,1.72]	[1.12,1.79]	[1.17,1.79]
	π_2	-0.36	-0.52	-0.49	-0.50
		[-0.67,-0.04]	[-0.76,-0.29]	[-0.82,-0.14]	[-0.81,-0.19]
	h_η	-3.36	-3.31	-3.39	-3.38
		[-3.46,-3.26]	[-3.40,-3.21]	[-3.48,-3.29]	[-3.47,-3.28]

Notes: Reported are the posterior mean of the parameters and constant log volatility of equation (13) with the 90% highest posterior density interval in square brackets. Estimation details can be found in Appendix D.

[14] A unit root is present in $c_t - y_y$ if $\pi_1 + \pi_2 = 1$, while it is stationarity if $\pi_1 + \pi_2 < 1$.

In Table 4, we report the posterior means and 90% HPD intervals of the parameters α, β, γ and δ of the predictive equation (12) for both excess equity and excess bond returns. The table further reports the standard full-sample R-square and the out-of-sample R-square.[15] Figures for the time-varying log volatility $h_{\varepsilon,t+1}$ of the return innovations are relegated to Appendix D.

From the table, we note the following. First, the predictive impact β of the log consumption-income ratio for excess equity returns is positive and of substantial magnitude for France and Germany but close to zero for the US and even negative for the UK. Values for β obtained using excess bond returns however are positive and of substantial magnitude for all four countries and in particular for France and the US. In the next section, we provide more details on this result. Second, the impact of lagged returns captured by the AR parameter γ in general is positive but in all cases has rather wide HPD intervals that contain the value of zero suggesting that lagged returns have little predictive ability for current returns. Third, the parameter δ, which reflects the contemporaneous covariance between innovations to $c_t - y_t$ and r_t, has the expected negative sign for equity returns suggesting that controlling for this covariance through a VAR set-up is useful. This seems less urgent when dealing with excess bond returns however. Finally, we note that the values for the R^2's are low but are in line with values typically encountered in the literature on return prediction using highly persistent predictors (see e.g., Lettau and Ludvigson, 2001; Whelan, 2008; Golez and Koudijs, 2018). Consistent with our findings for the parameter β, the fit of the equation for excess government bond returns is generally much better than the fit of the equation for excess equity returns. With respect to out-of-sample predictability, we note that values found for R^2_{OOS} tend to be in accordance with the values found for R^2. When the in-sample fit is good, the out-of-sample predictability is also reasonably good. When the in-sample fit is not so good, the out-of-sample predictability is not very good either resulting in negative values found for R^2_{OOS}. As noted by Cochrane (2008), the finding of a negative value for R^2_{OOS} is not surprising when estimating predictive regressions with a highly persistent predictor. Nor does it necessarily imply that (excess) returns are unpredictable. Rather, it means that the predictor used - in this case, the highly persistent log consumption-income ratio - is not useful for real time predictions. When looking at the predictive ability of $c_t - y_t$ for excess bond returns in the UK, for example, the results for β and the in-sample R^2 reported in Table 4 imply that the log consumption-income ratio does have some predictive power for bond returns. Nonetheless, the R^2_{OOS} is negative which suggests that it is not useful for out-of-sample prediction in this case.

[15]The out-of-sample R-square is calculated as $R^2_{OOS} = 1 - \frac{\sum_{\tau=1}^{T}(r_\tau - \hat{r}_\tau)^2}{\sum_{\tau=1}^{T}(r_\tau - \bar{r}_\tau)^2}$ where r_τ is the actual return, $\hat{r}_\tau$ is the return predicted from the VAR system of eqs.(12)-(13) estimated on the sample until period $\tau - 1$, and $\bar{r}_\tau$ is the mean return up to period $\tau - 1$ (see e.g., Golez and Koudijs, 2018). The first prediction $\hat{r}_{\tau=1}$ is based on estimating the VAR on the sample period until 1950, while the last prediction $\hat{r}_{\tau=T}$ is based on estimating the VAR on the sample period until 2014.

Table 4: Estimation of the predictive regression equation (12) for excess returns on equity and government bonds

	Parameter/R^2	Country			
		France	Germany	UK	US
Excess equity returns for r_{t+1}					
	α	0.02	0.07	0.04	0.07
		[.005,0.04]	[0.01,0.13]	[0.02,0.06]	[0.04,0.10]
	β	0.12	0.19	-0.10	0.01
		[-0.05,0.29]	[-0.12,0.49]	[-0.38,0.19]	[-0.23,0.25]
	γ	0.13	0.05	0.07	0.02
		[-0.02,0.28]	[-0.09,0.20]	[-0.08,0.23]	[-0.12,0.17]
	δ	-0.32	-1.22	-1.44	-0.42
		[-1.32,0.67]	[-2.42,-.005]	[-2.52,-0.40]	[-1.80,0.95]
	R^2	0.02	.001	.003	.003
	R^2_{OOS}	0.02	-0.02	-0.02	-.008
Excess bond returns for r_{t+1}					
	α	0.02	0.02	.003	.004
		[0.01,0.03]	[0.01,0.04]	[-0.01,0.02]	[-0.01,0.02]
	β	0.21	0.06	0.11	0.15
		[0.09,0.34]	[-0.02,0.15]	[-0.06,0.28]	[0.06,0.25]
	γ	0.06	0.08	-.001	0.04
		[-0.10,0.22]	[-0.10,0.26]	[-0.16,0.16]	[-0.12,0.21]
	δ	0.44	-0.09	-0.77	0.25
		[-0.32,1.18]	[-0.57,0.37]	[-1.65,0.09]	[-0.46,0.95]
	R^2	0.10	0.02	0.01	0.10
	R^2_{OOS}	0.04	0.01	-0.01	0.03

Notes: Reported are the posterior mean of the fixed parameters of equation (12) with the 90% highest posterior density interval in square brackets. R^2 denotes the standard full-sample R-square while R^2_{OOS} denotes the out-of-sample R-square. Estimation details can be found in Appendix D.

The generally better fit of the predictive regression estimated for excess bond returns as compared to excess equity returns can be observed graphically in Figures 3 and 4 which shows actual and fitted returns for the cases reported in Table 4.

Figure 3: Fit predictive regression for excess equity returns

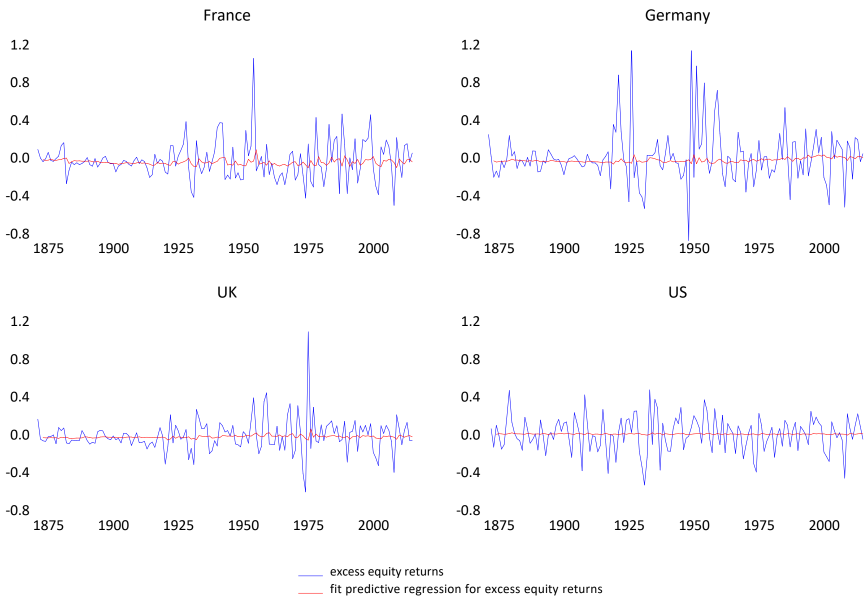
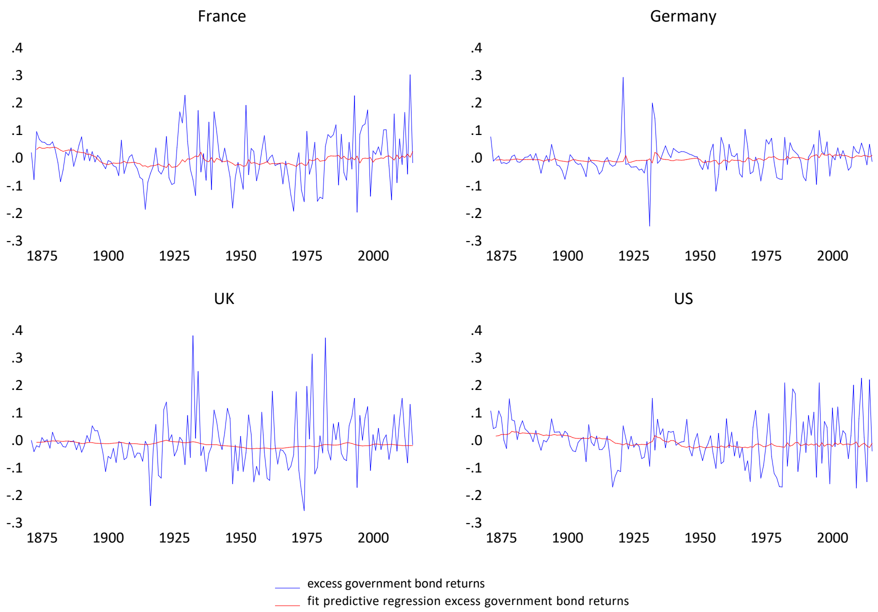


Figure 4: Fit predictive regression for excess government bond returns



5.3 Predictive impact

We now take a closer look at the predictive impact of the log consumption-income ratio for returns as captured by the parameter β. Table 5 reports, as before Table 4, the posterior mean and 90% HPD interval of the parameter β estimated from equations (10)-(11) for equity excess returns and government bond excess returns. Additionally, we now also report results for raw equity and bond returns. Furthermore, the table also reports the posterior probability that the predictive impact of $c_t - y_t$ on r_{t+1} is larger than zero, i.e., *prob*($\beta > 0$). As discussed in Section 2, we theoretically expect a positive impact of the log consumption-income ratio on future returns. Given the Gaussian prior centered at zero for β (see Appendix D for details), the prior probability that $\beta > 0$ equals 0.5. From the table we note that, for excess bond returns, the values for *prob*($\beta > 0$) are generally well above 0.5 and even equal to one for France and the US. For excess equity returns, these probabilities are high for France and Germany but close to and even lower than 0.5 for the US, respectively the UK. When comparing excess and raw returns, we note that the results for raw bond returns are rather similar to those for excess bond returns, while the evidence in favor of predictability is a bit more convincing for raw equity returns in all countries but the US.

Table 5: The predictive impact of $c_t - y_t$ on r_{t+1}

		Country			
		France	Germany	UK	US
Excess equity returns for r_{t+1}					
	β	0.12	0.19	-0.10	0.01
		[-0.05,0.29]	[-0.12,0.49]	[-0.38,0.19]	[-0.23,0.25]
	prob($\beta > 0$)	0.87	0.85	0.28	0.52
Excess bond returns for r_{t+1}					
	β	0.21	0.06	0.11	0.15
		[0.09,0.34]	[-0.02,0.15]	[-0.06,0.28]	[0.06,0.25]
	prob($\beta > 0$)	1.00	0.90	0.85	1.00
Raw equity returns for r_{t+1}					
	β	0.23	0.30	0.17	-0.03
		[0.04,0.42]	[-.004,0.59]	[-0.12,0.47]	[-0.27,0.21]
	prob($\beta > 0$)	0.97	0.95	0.83	0.42
Raw bond returns for r_{t+1}					
	β	0.28	0.04	0.29	0.11
		[0.14,0.42]	[-0.05,0.13]	[0.11,0.47]	[0.01,0.21]
	prob($\beta > 0$)	1.00	0.75	1.00	0.96

Notes: From the posterior distribution of the parameter β in equation (12), we report the posterior mean, the 90% highest posterior density interval (between square brackets) and the posterior probability that $\beta > 0$. Estimation details can be found in Appendix D.

Table 6 then reports averages of the predictive impact of $c_t - y_t$ on r_{t+1} over the four considered economies, i.e., the mean-group (MG) results. The parameter β^{MG} is calculated as the weighted average of β over the four countries in our sample where the weights are either identical or determined by the size of a country's GDP (respectively, population) vis-a-vis total GDP (respectively, population) of all four countries in the sample. From the table, we note again that the results for bond returns are more conclusive than those for equity returns with probabilities that $\beta^{MG} > 0$ (almost) equal to one. The value for β^{MG} for excess bond returns lies around 0.14 depending on the weights used in its construction.

As the β's can be interpreted as elasticities - i.e., $\beta = \frac{dr}{d(c-y)} = \frac{d \ln R}{d \ln(\frac{C}{Y})}$ - this value implies that if the consumption-income ratio $\frac{C}{Y}$ increases with 1%, gross excess bond returns R increase with 0.14% the following year (on average over time and countries). While an elasticity of 0.14 is not very large, the numbers become more impressive once we consider longer horizons which is the topic of the next section.

Table 6: The predictive impact of $c_t - y_t$ on r_{t+1}: cross-country (weighted) averages

		Weights		
		Equal	GDP	Population
Excess equity returns for r_{t+1}				
	β^{MG}	0.05	0.03	0.04
		[-0.09,0.20]	[-0.13,0.19]	[-0.10,0.19]
	$prob(\beta^{MG} > 0)$	0.74	0.63	0.69
Excess bond returns for r_{t+1}				
	β^{MG}	0.13	0.14	0.14
		[0.04,0.23]	[0.05,0.23]	[0.05,0.22]
	$prob(\beta^{MG} > 0)$	0.99	1.00	1.00
Raw equity returns for r_{t+1}				
	β^{MG}	0.17	0.07	0.10
		[0.02,0.32]	[-0.10,0.23]	[-0.05,0.25]
	$prob(\beta^{MG} > 0)$	0.97	0.75	0.87
Raw bond returns for r_{t+1}				
	β^{MG}	0.18	0.14	0.15
		[0.08,0.28]	[0.04,0.24]	[0.05,0.24]
	$prob(\beta^{MG} > 0)$	1.00	0.99	0.99

Notes: The mean-group parameter β^{MG} is calculated as $\beta^{MG} = \sum_{i=1}^{4} weight^{(i)} \beta^{(i)}$ where $\beta^{(i)}$ is β for country i and $weight^{(i)}$ is the weight given to country i. For equal weights, we have $weight^{(i)} = 1/4$. For GDP weights, $weight^{(i)}$ is equal to country i's GDP (average over the sample period) divided by the total GDP of all four countries in the sample (average over the sample period). For population weights, $weight^{(i)}$ is equal to country i's population (average over the sample period) divided by the total population of all four countries in the sample (average over the sample period). From the posterior distribution of the parameter β^{MG}, we report the posterior mean, the 90% highest posterior density interval (between square brackets) and the posterior probability that $\beta^{MG} > 0$. Estimation details can be found in Appendix D.

5.4 Long-horizon predictability

This section discusses the long-run impact of the log consumption-income ratio on (excess) returns. From the estimates obtained when estimating the VAR system of equations (12)-(13), we calculate the total (cumulative) impact of the log consumption-income ratio $c_t - y_t$ on returns over a horizon of k years through the following formula,

$$\beta^{(k)} = \sum_{j=1}^{k} \beta_j \tag{14}$$

where β_j denotes the impact of $c_t - y_t$ on period $t + j$ returns r_{t+j} which can be calculated recursively from,

$$\beta_j = \gamma \beta_{j-1} + \beta \varrho_{j-1} \tag{15}$$

with initializations $\beta_0 = 0$ and $\varrho_0 = 1$ and where β is the coefficient on the lagged log consumption-income ratio in equation (12), where γ is the coefficient on the lagged return r_t in equation (12) and where ϱ_j is the j-th order sample autocorrelation coefficient of $c_t - y_t$. We calculate the posterior distributions of the long-run coefficients $\beta^{(k)}$ for horizons of ten and twenty years by calculating $\beta^{(k)}$ for $k = 10, 20$ in every iteration of the Gibbs sampler.

Table 7: Long-term k-year horizon impact $\beta^{(k)}$ of $c_t - y_t$ on returns

		Country				Weighted average		
		France	Germany	UK	US	Equal	GDP	Population
Excess equity returns for r_{t+1}								
	$\beta^{(1)}$	0.12	0.19	-0.10	0.01	0.05	0.03	0.04
		[-0.05,0.29]	[-0.12,0.49]	[-0.38,0.19]	[-0.23,0.25]	[-0.09,0.20]	[-0.13,0.19]	[-0.10,0.19]
	$\beta^{(10)}$	0.99	1.56	-0.86	0.08	0.44	0.26	0.36
		[-0.48,2.46]	[-0.94,4.04]	[-3.30,1.68]	[-2.07,2.17]	[-0.78,1.68]	[-1.20,1.68]	[-0.95,1.63]
	$\beta^{(20)}$	1.27	2.44	-1.34	0.14	0.63	0.38	0.53
		[-0.61,3.15]	[-1.47,6.31]	[-5.13,2.60]	[-3.49,3.67]	[-1.24,2.50]	[-2.04,2.76]	[-1.59,2.59]
Excess bond returns for r_{t+1}								
	$\beta^{(1)}$	0.21	0.06	0.11	0.15	0.13	0.14	0.14
		[0.09,0.34]	[-0.02,0.15]	[-0.06,0.28]	[0.06,0.25]	[0.04,0.23]	[0.05,0.23]	[0.05,0.22]
	$\beta^{(10)}$	1.69	0.53	0.91	1.38	1.13	1.25	1.19
		[0.72,2.65]	[-0.16,1.23]	[-0.52,2.36]	[0.56,2.19]	[0.37,1.89]	[0.49,1.99]	[0.46,1.91]
	$\beta^{(20)}$	2.16	0.83	1.41	2.34	1.68	2.01	1.87
		[0.92,3.40]	[-0.25,1.92]	[-0.81,3.66]	[0.95,3.70]	[0.53,2.83]	[0.78,3.20]	[0.71,3.02]
Raw equity returns for r_{t+1}								
	$\beta^{(1)}$	0.23	0.30	0.17	-0.03	0.17	0.07	0.10
		[0.04,0.42]	[-.004,0.59]	[-0.12,0.47]	[-0.27,0.21]	[0.02,0.32]	[-0.10,0.23]	[-0.05,0.25]
	$\beta^{(10)}$	2.00	2.42	1.64	-0.25	1.45	0.58	0.90
		[0.34,3.66]	[-0.03,4.85]	[-1.08,4.41]	[-2.39,1.86]	[0.12,2.78]	[-0.88,2.05]	[-0.44,2.21]
	$\beta^{(20)}$	2.57	3.78	2.56	-0.41	2.12	0.83	1.31
		[0.43,4.71]	[-0.05,7.58]	[-1.67,6.88]	[-4.03,3.12]	[0.10,4.13]	[-1.59,3.26]	[-0.85,3.43]
Raw bond returns for r_{t+1}								
	$\beta^{(1)}$	0.28	0.04	0.29	0.11	0.18	0.14	0.15
		[0.14,0.42]	[-0.05,0.13]	[0.11,0.47]	[0.01,0.21]	[0.08,0.28]	[0.04,0.24]	[0.05,0.24]
	$\beta^{(10)}$	2.67	0.38	2.72	1.07	1.71	1.36	1.43
		[1.38,3.96]	[-0.55,1.33]	[1.04,4.43]	[0.08,2.05]	[0.75,2.67]	[0.44,2.25]	[0.53,2.32]
	$\beta^{(20)}$	3.46	0.60	4.24	1.82	2.53	2.13	2.20
		[1.78,5.11]	[-0.87,2.10]	[1.62,6.92]	[0.14,3.48]	[1.08,3.97]	[0.65,3.59]	[0.78,3.60]

Notes: The calculation of $\beta^{(k)}$ for $k = 1, 10, 20$ is given by equations (14)-(15). We note that the one-year horizon parameter $\beta^{(1)}$ equals the coefficient β on the lagged log consumption-income ratio in the predictive equation (12). From the posterior distribution of the parameter $\beta^{(k)}$, we report the posterior mean and the 90% highest posterior density interval between square brackets. Estimation details can be found in Appendix D.

The results are reported in Table 7. We report long-run impact estimates of $c_t - y_t$ on returns over horizons of ten and twenty years, i.e., the parameters $\beta^{(10)}$ and $\beta^{(20)}$. To facilitate the comparison with the short-run estimates discussed in the previous section, we also add the previously estimated β's in the table, denoted here by $\beta^{(1)}$. Per country estimates are reported as well as cross-country averages using either equal weights, GDP-based weights or weights based on population size. The returns considered are excess and raw equity and government bond returns. We note that, in particular for the excess and raw bond returns, the long-run impact estimates are considerably larger than their short-run counterparts. The mean-group 20-year horizon elasticities for excess bond returns, for example, lie between 1.7 an 2 implying that if the consumption-income ratio $\frac{C}{Y}$ increases with 1%, the increase in gross excess bond returns R lies between 1.7% and 2% over the following 20 years. The long-run elasticities are even higher for raw bond returns. For excess equity returns, the long-run impact estimates $\beta^{(10)}$ and $\beta^{(20)}$ are also of larger magnitude (and more negative for the UK) but the per country results and the mean group results are, in accordance with the short-run estimates $\beta^{(1)}$, all characterized by rather wide HPD intervals that include the value of zero. As noted above, the results for raw equity returns are somewhat more convincing than those for excess equity returns and this can also be observed in the long-horizon results. The predictive power of the log consumption-income ratio for raw equity returns is substantial for France and Germany and, to a lesser extent, for the UK. For the mean-group results based on equal weights (which gives relatively less weight to the US) this translates into a long-run elasticity equal to 2.1 for the 20-year horizon.

6 Business cycle fluctuations

In this section, we investigate whether the predictability results presented in the previous section can be (partially) attributed to business cycle fluctuations or whether they are more structural in nature. Expected (excess) returns are generally considered to be countercyclical (see Fama and French, 1989, for early evidence, and Golez and Koudijs, 2018, for recent evidence using historical data) while the evidence that we present below suggests that the log consumption-income ratio in our sample tends to be countercyclical as well. As such, the positive impact of the log consumption-income ratio on expected returns reported for the cases discussed above could be (partially) driven by the business cycle, i.e., in recessions the consumption-income ratio and expected (excess) returns both increase. First, we present evidence on the cyclicality of the log consumption-income ratio in our sample. Next, we estimate an extended VAR specification that also includes a variable that captures the business cycle and we check whether our predictability results are robust to the inclusion of this variable in the predictive equation for (excess) returns.

6.1 Cyclicality of the consumption-income ratio: some evidence

We start by investigating the cyclicality of the consumption-income ratio in our historical dataset by estimating regressions of the following form,

$$f\left(c_{t+1} - y_{t+1}\right) = \vartheta_0 + \vartheta_1 bc_{t+1} + \zeta_{t+1} \qquad \zeta_{t+1} \sim iid\mathcal{N}\left(0, \sigma_\zeta^2\right) \tag{16}$$

where $c_t - y_t$ is the log consumption-income ratio used previously and where $f(c-y)$ is a transformation of $c-y$, i.e., either the first difference $f(c-y) = \Delta(c-y)$ or the deviation of $c-y$ from its (stochastic) trend $f(c-y) = (c-y) - (\overline{c-y})$ where $\overline{c-y}$ is approximated via a five-year moving average filter applied to $c-y$ as $\overline{c-y} = \frac{1}{5}\sum_{j=0}^{4}(c_{-j} - y_{-j})$. For informative purposes, we also report results obtained from estimating equation (16) with the identity function $f(c-y) = c-y$ even though this violates the *iid* assumption on the error term ζ_{t+1} as we know from the results in Section 5 that $c_{t+1} - y_{t+1}$ is very persistent. The variable bc_t is the business cycle indicator for which we use real per capita GDP growth.[16] The estimation method is Bayesian OLS.[17]

Table 8: The cyclicality of the log consumption-income ratio $c_{t+1} - y_{t+1}$

		Country			
$f(c-y)$	Coefficient on bc	France	Germany	UK	US
$\Delta(c-y)$	ϑ_1	-0.03	-0.05	-0.14	-0.16
		[-0.09,0.03]	[-0.09,-0.01]	[-0.27,-0.02]	[-0.23,-0.09]
$(c-y) - (\overline{c-y})$	ϑ_1	-0.03	-0.07	-0.21	-0.19
		[-0.11,0.03]	[-0.11,-0.02]	[-0.35,-0.07]	[-0.28,-0.11]
$c-y$	ϑ_1	-0.24	-0.05	-0.55	-0.14
		[-0.41,-0.06]	[-0.18,0.08]	[-0.90,-0.20]	[-0.45,0.17]

Notes: Reported are the posterior mean and the 90% highest posterior density interval (in square brackets) of the parameter ϑ_1 in equation (16). Parameter estimates for ϑ_0 and σ_ζ^2 are unreported but available upon request. The estimation method is Bayesian OLS.

Table 8 reports the mean and 90% HPD interval of the posterior distribution of the parameter ϑ_1, i.e., the impact of the business cycle indicator on both transformations of the log consumption-income ratio and on the untransformed log consumption-income ratio. The reported results show that, over the considered historical period, the estimates of ϑ_1 are negative in all countries irrespective of the applied transformation or lack thereof. While the evidence is weaker for France as far as both transformed log

[16] Historical data for this variable are calculated from the Jordà-Schularick-Taylor macro-history database (see Jordà et al., 2016) which report real per capita GDP data for the countries we consider from 1870 onward, uninterrupted. The website is http://www.macrohistory.net/data. The data has code 'rgdppc'.

[17] Gausian prior distributions are used for ϑ_0 and ϑ_1 with mean zero and variance equal to 10. The prior distribution for σ_ζ^2 is inverse Gamma with belief equal to zero and strength equal to 0.01. See Bauwens et al. (2000) for details. The posterior distributions of the parameters are calculated using Gibbs sampling based on 12.000 iterations of which 2000 serve as burn-in.

consumption-income ratios are concerned, the overall results suggest that the log consumption-income ratio is (moderately) countercyclical.

6.2 Impact of the business cycle on predictability

6.2.1 Specification

To investigate whether our predictability results are (partially) driven by business cycle fluctuations, we now estimate the following extended three-equation VAR,

$$r_{t+1} = \alpha_{t+1} + \beta_{t+1}(c_t - y_t) + \gamma_{t+1} r_t + \mu_{t+1} bc_t + \psi_{t+1} \tag{17}$$

$$c_{t+1} - y_{t+1} = \pi_{0,t+1} + \pi_{1,t+1}(c_t - y_t) + \pi_{2,t+1}(c_{t-1} - y_{t-1}) + \eta_{t+1} \tag{18}$$

$$bc_{t+1} = \lambda_{0,t+1} + \lambda_{1,t+1} bc_t + \chi_{t+1} \tag{19}$$

where $\begin{pmatrix} \chi_{t+1} & \eta_{t+1} & \psi_{t+1} \end{pmatrix}' \sim \mathcal{N}(0, \Omega_{t+1})$. The extended VAR contains an additional equation - i.e., equation (19) - for the business cycle variable bc_{t+1} which, as detailed in Section 6.1, is proxied by real per capita GDP growth. We assume that bc_{t+1} follows an $AR(1)$ process. Moreover, the variable bc_t is added as an additional predictor variable in the equation for the returns r_{t+1}. Applying the decomposition of equation (8) to the variance matrix Ω_{t+1} allows us to rewrite the VAR with independent (structural) error terms so that it can be estimated one equation at a time. Hence, the Gibbs sampling procedure discussed in Appendix D can again be applied. After testing for time variation in the parameters and log volatilities, we withhold time variation only for the log volatility of the innovation in the return equation and for the log volatility of the innovation in the business cycle equation. All other parameters, variances and covariances are restricted to be constant.

6.2.2 Results

We investigate whether the inclusion of bc_t in equation (17) has an impact on the predictive power of $c_t - y_t$ for r_{t+1}. As before, this potential impact is captured by the parameter β. If this parameter is smaller after controlling for the business cycle, this suggests that the business cycle is a channel through which the consumption-income ratio predicts returns. The impact of bc_t on expected (excess) returns is captured by the parameter μ. If expected (excess) returns are countercyclical, then we expect $\mu < 0$.

Table 9: The predictive impact of $c_t - y_t$ on r_{t+1}, after controlling for the business cycle

		Country			
		France	Germany	UK	US
Excess equity returns for r_{t+1}					
	β	0.13	0.18	-0.06	.000
		[-0.04,0.31]	[-0.13,0.48]	[-0.34,0.22]	[-0.23,0.23]
	$prob(\beta > 0)$	0.90	0.83	0.36	0.51
	μ	-0.01	0.29	0.30	-0.23
		[-0.32,0.28]	[-0.07,0.69]	[-0.25,0.85]	[-0.81,0.34]
Excess bond returns for r_{t+1}					
	β	0.19	0.06	0.08	0.15
		[0.06,0.32]	[-0.02,0.15]	[-0.10,0.25]	[0.06,0.25]
	$prob(\beta > 0)$	0.99	0.90	0.77	0.99
	μ	-0.14	-0.01	-0.50	-0.03
		[-0.34,0.05]	[-0.09,0.08]	[-0.92,-0.08]	[-0.23,0.16]
Raw equity returns for r_{t+1}					
	β	0.25	0.32	0.20	-0.04
		[0.05,0.45]	[0.02,0.61]	[-0.09,0.49]	[-0.28,0.19]
	$prob(\beta > 0)$	0.98	0.96	0.87	0.39
	μ	0.07	0.41	0.17	-0.35
		[-0.30,0.44]	[0.03,0.83]	[-0.41,0.75]	[-0.92,0.23]
Raw bond returns for r_{t+1}					
	β	0.25	0.04	0.25	0.11
		[0.11,0.41]	[-0.04,0.14]	[0.07,0.44]	[.003,0.21]
	$prob(\beta > 0)$	1.00	0.78	0.98	0.95
	μ	-0.03	0.04	-0.53	-0.16
		[-0.31,0.23]	[-0.05,0.13]	[-0.99,-0.08]	[-0.40,0.07]

Notes: The posterior distribution of the parameters β and μ are estimated from the VAR in equations (17)-(19) after decomposing the variance matrix of the innovations according to equation (8) and after imposing restrictions on the time variation in the parameters and log volatilities. We report the posterior mean and the corresponding 90% highest posterior density interval (between square brackets). For β, we also report the posterior probability that $\beta > 0$. Estimation details can be found in Appendix D.

In Table 9, we report statistics of the posterior distributions of the parameters β and μ. We do not report the results obtained for the other parameters estimated from the extended VAR but they are available upon request.[18] From a comparison of this table with Table 5 above, we only find evidence that supports a role for the business cycle when looking at the excess and raw bond returns of the UK. In these cases, we find a reduction in the posterior mean of β and in the probability $prob(\beta > 0)$ after controlling for the business cycle *combined with* a negative impact on r_{t+1} of the variable bc_t as captured by the parameter μ. In both these instances, the magnitude of μ is quite substantial even though the reduction

[18]Of interest are, in particular, the covariances between the innovation to the log-consumption ratio in equation (18) and the innovation to the business cycle variable in equation (19) which - in line with the countercyclicality results for $c_t - y_t$ reported in Section 6.1 - are found to be negative. The means with corresponding 90% highest posterior density intervals for this covariance are respectively -0.03 with $[-0.10, 0.05]$ for France, -0.02 with $[-0.07, 0.03]$ for Germany, -0.14 with $[-0.31, 0.04]$ for the UK and -0.12 with $[-0.22, -0.02]$ for the US.

in the posterior mean of β is quite small. Hence, there is only very limited evidence that business cycle fluctuations affect the predictive relationship between $c_t - y_t$ and r_{t+1} documented in this paper. The predictive ability of the log consumption-income ratio for in particular (excess) bond returns seems to be more structural in nature or, in the words of Fama and French (1989), related to long-term 'business conditions'.

7 Conclusions

We have investigated whether the log aggregate consumption to disposable income ratio has predictive power for excess and raw returns on equity and long-term government bonds using historical data over the period $1870 - 2015$ for four major industrial economies, i.e., France, Germany, the UK and the US. The predictive ability of the consumption-income ratio for the returns on wealth is implied by intertemporal budget constraint reasoning. The short-run and long-run impact estimates of the current log consumption-income ratio on next period's returns are calculated using a Bayesian vector autoregression (VAR) model with time-varying parameters and stochastic volatilities. We further average our per country results across the four considered economies using a 'mean-group' approach. Finally, we investigate to what extent the predictive power of the consumption-income ratio for (excess) returns is driven by business cycle fluctuations.

Our results suggest that the consumption-income ratio has substantial predictive power for the (excess) government bond returns of all four countries considered. On average over time and countries, if the consumption-income ratio increases with 1%, gross excess bond returns increase with about 0.14% the following year and with 1.7% to 2% the following 20 years. The evidence with respect to (excess) equity returns is generally inconclusive however. These findings support the notion that bond returns may be good proxy's for the returns on total wealth, as is suggested in the literature. We find that the predictive ability of the consumption-income ratio does not appear to be driven by business cycle fluctuations, but rather may reflect long-term 'business conditions'. We consider our results as complementary to the predictability results reported in the literature that use more conventional ratios, i.e., our results are obtained from long-run lower frequency historical data rather than from postwar quarterly data, they apply to bonds rather than equity, and they are structural in nature rather than cyclical.

References

Afonso, A. and Sousa, R. (2011). Consumption, wealth, stock and government bond returns: international evidence. *The Manchester School*, 79(6):1294–1322.

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

Barro, R. and Ursúa, J. (2008). Macroeconomic crises since 1870. *Brookings Papers on Economic Activity*, 38(1):255–335.

Bauwens, L., Lubrano, M., and Richard, J.-F. (2000). *Bayesian inference in dynamic econometric models.* Oxford University Press.

Bianchi, F., Lettau, M., and Ludvigson, S. (2017). Monetary policy and asset valuation. *Working Paper.*

Campbell, J. and Mankiw, N. (1989). Consumption, income, and interest rates: reinterpreting the time series evidence. *NBER Macroeconomics Annual*, 4:185–216.

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

Carter, C. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81:541–53.

Chen, L. (2009). On the reversal of return and dividend growth predictability: a tale of two periods. *Journal of Financial Economics*, 92:128–151.

Cochrane, J. (2005). *Asset Pricing.* Princeton University Press.

Cochrane, J. (2008). The dog that did not bark: a defense of return predictability. *Review of Financial Studies*, 21(4):1533–1575.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106:157–181.

Del Negro, M. and Primiceri, G. (2015). Time varying structural vector autoregressions and monetary policy: a corrigendum. *Review of Economic Studies*, 82:1342–1345.

Della Corte, P., Sarno, L., and Valente, G. (2010). A century of equity premium predictability and the consumption-wealth ratio: an international perspective. *Journal of Empirical Finance*, 17:313–331.

Everaert, G. and Pozzi, L. (2019). Measuring international consumption risk sharing and the impact of trade and financial globalization. *Working Paper.*

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

Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*, 154(1):85–100.

Gargano, A., Pettenuzzo, D., and Timmermann, A. (2019). Bond return predictability: economic value and link to the macroeconomy. *Management Science*, 65(2):508–540.

George, E. and McCulloch, R. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88:881–889.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Berger, J., Bernardo, J., Dawid, A., and Smith, A., editors, *Bayesian statistics*. Oxford University Press.

Giannone, D., Lenza, M., and Primiceri, G. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, 97(2):436–451.

Golez, B. and Koudijs, P. (2018). Four centuries of return predictability. *Journal of Financial Economics*, 127:248–263.

Groen, J., Paap, R., and Ravazzolo, F. (2013). Real-time inflation forecasting in a changing world. *Journal of Business and Economic Statistics*, 31(1):29–44.

Hamilton, J. (2008). Macroeconomics and ARCH. *NBER Working Papers*, 14151.

Havranek, T., Horvath, R., Irsova, Z., and Rusnak, M. (2015). Cross-country heterogeneity in intertemporal substitution. *Journal of International Economics*, 96:100–118.

Jordà, O., Knoll, K., Kuvshinov, D., Schularick, M., and Taylor, A. (2019). The rate of return on everything, 1870-2015. *Quarterly Journal of Economics*, 134(3):1225–1298.

Jordà, O., Schularick, M., and Taylor, A. (2016). Macrofinancial history and the new business cycle facts. In Eichenbaum, M. and Parker, J., editors, *NBER Macroeconomics Annual 2016*, volume 31. University of Chicago Press, Chicago.

Kim, C.-J. and Nelson, C. (1999). *State-space models with regime switching. Classical and Gibbs-sampling approaches with applications.* MIT Press, Cambridge.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH Models. *Review of Economic Studies*, 65(3):361–93.

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, LVI(3):815–849.

Lettau, M. and Ludvigson, S. (2004). Understanding trend and cycle in asset values: re-evaluating the wealth effect on consumption. *American Economic Review*, 94(1):276–299.

Lettau, M. and Ludvigson, S. (2005). Expected returns and expected dividend growth. *Journal of Financial Economics*, 76(3):583–626.

Lustig, H., Van Nieuwerburgh, S., and Verdelhan, A. (2013). The wealth-consumption ratio. *Review of Asset Pricing Studies*, 3(1):38–94.

Nakamura, E., Sergeyev, D., and Steinsson, J. (2017). Growth-rate and uncertainty shocks in consumption: cross-country evidence. *American Economic Journal: Macroeconomics*, 9(1):1–39.

Nakamura, E., Steinsson, J., Barro, R., and Ursúa, J. (2013). Crises and recoveries in an empirical model of consumption disasters. *American Economic Journal: Macroeconomics*, 5(3):35–74.

Newey, W. and West, K. (1987). A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: fast and efficient likelihood inference. *Journal of Econometrics*, 140:425–449.

Piketty, T. and Zucman, G. (2014). Capital is back: wealth-income ratios in rich countries 1700-2010. *Quarterly Journal of Economics*, 129(3):1255–1310.

Primiceri, G. (2005). Time-varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72(3):821–852.

Rapach, D. and Zhou, G. (2013). Forecasting stock returns. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2A, chapter 6, pages 329–383. Elsevier, Amsterdam.

Sommer, M. (2007). Habit formation and aggregate consumption dynamics. *The B.E. Journal of Macroeconomics - Advances*, 7(1).

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

Tsay, R. (2005). *Analysis of Financial Time Series*. Wiley.

Whelan, K. (2008). Consumption and expected asset returns without assumptions about unobservables. *Journal of Monetary Economics*, 55:1209–1221.

Appendices

Appendix A Derivation of equation (1)

This appendix briefly describes the steps in the derivation of equation (1) in the main text. We refer to Campbell and Mankiw (1989) and Lettau and Ludvigson (2005) for more details. When total wealth is tradeable, the period-by-period budget constraint of a consumer can be written as,

$$W_{t+1} = R_{t+1}(W_t - C_t) \tag{A-1}$$

where W_t is real total wealth. Dividing both sides by W_t, we can write $\frac{W_{t+1}}{W_t} = R_{t+1}\left(1 - \frac{C_t}{W_t}\right)$. After taking logs, this gives $\Delta w_{t+1} = r_{t+1} + \ln(1 - \exp(c_t - w_t))$ with $w_t = \ln W_t$, $r_t = \ln R_t$ and $c_t = \ln C_t$. We linearize this equation by taking a first-order Taylor approximation which gives,

$$\Delta w_{t+1} = r_{t+1} + \left(1 - \frac{1}{\rho}\right)(c_t - w_t) \tag{A-2}$$

where we ignore the unimportant linearization constant and where $\rho = \frac{W-C}{W}$ with W and C the average or steady state values of W_t and C_t and with ρ slightly below one.[1] Note that we can write Δw_{t+1} as $\Delta w_{t+1} = \Delta c_{t+1} + (c_t - w_t) - (c_{t+1} - w_{t+1})$. Upon combining this result with equation (A-2) and rearranging terms, we obtain,

$$c_t - w_t = \rho(r_{t+1} - \Delta c_{t+1}) + \rho(c_{t+1} - w_{t+1}) \tag{A-3}$$

Solving equation (A-3) forward ad infinitum, imposing the transversality condition $\rho^{\infty}(c_{t+\infty} - w_{t+\infty}) = 0$ and taking expectations at period t then gives,

$$c_t - w_t = E_t \sum_{j=1}^{\infty} \rho^j (r_{t+j} - \Delta c_{t+j}) \tag{A-4}$$

Since total wealth is tradeable, we assume it consists of N_t shares with cum-dividend share price given by $P_t + Y_t$, i.e., we have $W_t = N_t(P_t + Y_t)$. From the main text, the gross real return on wealth is given by $R_{t+1} = \frac{P_{t+1}+Y_{t+1}}{P_t}$. By combining these results and rearranging terms, we obtain,

$$\frac{W_{t+1}}{N_{t+1}} = R_{t+1}\left(\frac{W_t}{N_t} - Y_t\right) \tag{A-5}$$

Equation (A-5) is in the same form as equation (A-1) so the same steps (linearization, forward solving) can be applied to obtain,

$$y_t - w_t = -n_t + E_t \sum_{j=1}^{\infty} \kappa^j (r_{t+j} - \Delta y_{t+j}) \tag{A-6}$$

[1] Note that the linearization occurs around the point $c_t - w_t = c - w$ with $c - w = \ln\left(\frac{C}{W}\right)$.

where $n_t = \ln N_t$, $y_t = \ln Y_t$ and where $\kappa = \frac{W-Y}{W}$ with W and Y the average or steady state values of W_t and Y_t and with κ slightly below one.

By combining equations (A-4) and (A-6) while normalizing N_t to be equal to one so that $n_t = 0$, we obtain equation (1) in the main text.

Appendix B Data construction and data sources

This appendix provides details on the sources and on the construction of the data used in the paper. First, the data for consumption and income are discussed. Then, we elaborate on the returns data.

B.1 Consumption and income

The log of real per capita consumption c_t is calculated from historical data as reported in the the Jordà-Schularick-Taylor macro-history database (see Jordà et al., 2016) which reports real per capita consumption data for the countries we consider from 1870 onward, uninterrupted.[2] These data correspond to and update the historical consumption data reported by Barro and Ursúa (2008). The log of real per capita income y_t is calculated from historical data that accompanies Piketty and Zucman (2014).[3] In particular, from the reported real per capita national income series and from the reported series for the ratio of national income after taxes to national income, a series is constructed for real per capita disposable income (=national income minus taxes plus transfers). The data used are available uninterruptedly from 1870 onward for the countries in our sample.[4,5] For other countries, these data are insufficiently available.

We calculate cleaned series for consumption and income by taking out two types of shocks of a transitory nature - i.e., measurement error and the occurrence of both world wars - from the log of measured real per capita consumption and the log of measured real per capita disposable income. To this end, we estimate,

$$x_{t+1}^m = x_{t+1} + I_{t+1}\xi_{t+1} + \nu_{t+1} + \theta\nu_t \qquad \nu_{t+1} \sim \mathcal{N}\left(0, \sigma_{\nu,t+1}^2\right) \tag{B-1}$$

where x_t^m is either the log of measured real per capita consumption c_t^m or the log of measured real per capita disposable income y_t^m so that x_t is the corresponding adjusted (i.e., cleaned) variable c_t or y_t. The shock ν_{t+1} captures measurement error which, following the generalized specification suggested by

[2]The website is http://www.macrohistory.net/data. The data has code 'rconpc'.

[3]The website is http://piketty.pse.ens.fr/fr/capitalisback. The data used are in the country excel files, Table 1, columns 9 and 14.

[4]The ratio of disposable income to national income is only available from 1948 onward for the UK, however, so that for this country we extrapolate the 1948 value of this ratio to the period $1870 - 1947$.

[5]We update the calculated historical real per capita disposable income series from 2011 to 2015 using data from OECD Economic Outlook.

Sommer (2007), is modeled as an $MA(1)$ process with MA parameter θ. Importantly, to account for potential shifts in national accounts measurement around 1945, we allow for a structural break in the variance of the measurement error $\sigma^2_{\nu,t}$ in 1945 (see Nakamura et al., 2017, and references therein). The shock $I_{t+1}\xi_{t+1}$ captures transitory variation in consumption and income that is caused by the occurrence of both world wars (WWI, WWII). The variable I_t is a dummy variable that is set to one during the period $1914 - 1920$ (the official period of WWI plus two years) and during the period $1939 - 1947$ (the official period of WWII plus two years). For the world war shock ξ_{t+1}, we have $\xi_{t+1} \sim iid\mathcal{N}(\xi, 1)$, i.e., we fix its variance to one - which is a large value - to ensure that it soaks up all transitory variation in consumption and income during both world wars. The specification of the term $I_{t+1}\xi_{t+1}$ is identical to the specification considered by Nakamura et al. (2017) to capture disaster episodes in consumption but we focus more narrowly on world wars.[6] Our results are not qualitatively affected however when using their disaster dummies instead of our world war dummies. We use a Bayesian state space approach to estimate equation (B-1) for $x^m_t = c^m_t$, respectively $x^m_t = y^m_t$, for every country in the sample to obtain the unobserved adjusted (i.e., cleaned) variable $x_t = c_t$, respectively $x_t = y_t$. Methodological details are provided in Everaert and Pozzi (2019).

B.2 Returns

Historical data for equity returns, for long-term government bond returns and for returns on short-term bills (which are used as a proxy for the risk-free rate) going back to the 1870's are provided by Jordà et al. (2019).[7] For the US and the UK, time series for equity, bonds and bills returns are available uninterruptedly starting in either 1870, 1871 or 1872. For France and Germany the returns data are available from 1870 or 1871 onward but there are some missing observations during the sample period. For France, data on bill returns are missing over the period $1915 - 1921$. For Germany, data are missing for equity in the year 1923, for bonds over the period $1944 - 1948$ and for bills in the year 1923 and over the period $1945 - 1949$. We apply linear interpolation to deal with these missing values. With respect to interpolation, we note that, since the log consumption-income ratio is a very persistent slowly moving regressor, its impact - if present - is expected to be on the underlying low frequency trend movements in (excess) returns. As interpolation tends to preserve the underlying trend in returns, the few instances

[6]The disaster dummies used by Nakamura et al. (2017) and identified in Nakamura et al. (2013) include both world wars for France, Germany and the UK but include WWI and the Great Depression and exclude WWII for the US. Moreover, the timing of their world war disasters is somewhat different from our 'official' timing (see Table 2 in Nakamura et al., 2013).

[7]The data can be found at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GGDQGJ. The nominal equity returns have code 'eq-tr' the nominal bond returns have code 'bond-tr' and the nominal bill returns have code 'bill-rate'. Details on the data sources used for equity, bond and bill returns are discussed for all four countries in our sample in appendices J and L of the online Appendix of Jordà et al. (2019)'s paper.

where interpolation is applied to our returns data are not expected to have an important impact on our results. Next, we deflate all nominal returns (equity, bonds, bills) using the inflation rate as calculated from the Consumer Price Index (CPI). Historical data for this index are reported for all four countries in our sample in the Jordà-Schularick-Taylor macro-history database (Jordà et al., 2016).[8] Equity and bond excess returns are then calculated as the difference between real equity returns, respectively real bond returns, and real bill returns.

Appendix C Results using postwar quarterly data for the US

This appendix presents the results from estimating the predictive relationship between the log consumption-income ratio and asset returns with the use of a conventional dataset, namely quarterly data for the US over the period $1947Q2 - 2015Q4$.

C.1 Estimation

As in the main text, we investigate the predictive relationship by estimating a TVP-VAR with stochastic volatility in the error terms. The VAR system consists of equation (10) and an AR(4) process for $c_{t+1} - y_{t+1}$. Estimation details are provided in Appendix D where - given that an AR(4) process for $c_{t+1} - y_{t+1}$ replaces the AR(2) process given by equation (11) - we replace $l = 0, 1, 2$ by $l = 0, 1, 2, 3, 4$. We test for time variation in all parameters and log volatilities and withhold it only for the log volatility of the return equation.[9] Then, we estimate a VAR system consisting of equation (12) and a constant parameter AR(4) process for $c_{t+1} - y_{t+1}$ that replaces the AR(2) process given by equation (13).

C.2 Data

Data for consumption, personal disposable income, the price index for personal consumption expenditures (all seasonally adjusted) and population are taken from the US National Income and Product Accounts (NIPA) collected via the Bureau of Economic Analysis (BEA). For consumption, we take either personal consumer expenditures (line 29, NIPA table 2.1) or personal outlays (line 28, NIPA table 2.1). The latter series is used in the calculation of the personal saving series reported in the NIPA. For income, we take disposable personal income (line 27, NIPA table 2.1). To obtain real per capita consumption and income series, these data are deflated with the price index for personal consumption expenditures (line 1, NIPA table 2.3.4) and further expressed in per capita terms using population data (line 40, NIPA table 2.1). For the returns, all series are from the Center for Research in Security Prices (CRSP) collected via

[8]The website is http://www.macrohistory.net/data. The data used has code 'cpi'.

[9]Results are unreported but available upon request.

Wharton Research Data Services (WRDS). Equity returns are obtained from the value-weighted equity index, government bond returns are obtained from the 10-year bond index and the risk-free rate used to calculate excess equity and bond returns is obtained from the 90-day bill index. Raw and excess equity and bond returns are deflated using the inflation rate as calculated from the price index for personal consumption expenditures.

C.3 Results

Table C-1 presents the results. We report results only for the posterior distribution of β in equation (12), i.e., the posterior mean, the 90% highest posterior density interval and the posterior probability that $\beta > 0$.[10] Results are reported both for excess and raw equity and bond returns. Additionally, results are reported with c_t being log real per capita personal consumer expenditures and with c_t being log real per capita personal outlays. In the latter case, the variable $1 - e^{c-y}$ equals the personal saving to disposable income ratio. Finally, results are reported that are obtained with raw data used for c_t and y_t as well as with data used for c_t and y_t that are cleansed of measurement error. Cleaned data are obtained through the estimation of a simplified version of equation (B-1) following the methodology detailed in Everaert and Pozzi (2019).[11]

The results in Table C-1 show that, in all cases, there is a positive predictive impact of $c_t - y_t$ on returns. The HPD intervals, however, are rather wide and always contain the value of zero. The posterior probabilities that $\beta > 0$ are always larger - sometimes, considerably so - than the prior probability which equals 50%. They never surpass 90% however. While these results suggest that the log consumption-income ratio has predictive ability for (excess) returns on equity and government bonds, we do not obtain fully conclusive results when using US postwar quarterly data.

[10] The results for the other parameters are unreported but available upon request.

[11] In particular, we estimate $x^m_{t+1} = x_{t+1} + \nu_{t+1} + \theta\nu_t$ with $\nu_{t+1} \sim \mathcal{N}\left(0, \sigma^2_\nu\right)$ where x^m_t is either the log of measured real per capita consumption c^m_t or the log of measured real per capita disposable income y^m_t so that x_t is the corresponding adjusted (i.e., cleaned) variable c_t or y_t. The shock ν_{t+1} captures measurement error and is modeled as an $MA(1)$ process with MA parameter θ and constant error variance σ^2_ν.

Table C-1: The predictive impact of $c_t - y_t$ on r_{t+1}: quarterly US data (1947Q2-2015Q4)

		Predictor $c_t - y_t$			
		Pers. cons. exp. for c_t		Pers. outlays for c_t	
		c_t, y_t raw	c_t, y_t cleaned	c_t, y_t raw	c_t, y_t cleaned
Excess equity returns for r_{t+1}					
	β	0.13	0.15	0.07	0.06
		[-0.21,0.48]	[-0.21,0.52]	[-0.21,0.35]	[-0.23,0.37]
	$prob(\beta > 0)$	0.75	0.77	0.66	0.64
Excess bond returns for r_{t+1}					
	β	0.07	0.08	0.10	0.11
		[-0.12,0.26]	[-0.12,0.30]	[-0.06,0.26]	[-0.06,0.28]
	$prob(\beta > 0)$	0.74	0.76	0.85	0.87
Raw equity returns for r_{t+1}					
	β	0.09	0.11	0.06	0.05
		[-0.25,0.43]	[-0.25,0.47]	[-0.23,0.34]	[-0.25,0.35]
	$prob(\beta > 0)$	0.68	0.70	0.64	0.61
Raw bond returns for r_{t+1}					
	β	0.04	0.05	0.09	0.10
		[-0.16,0.24]	[-0.17,0.26]	[-0.07,0.26]	[-0.07,0.28]
	$prob(\beta > 0)$	0.64	0.65	0.83	0.84

Notes: The estimated system consists of equation (12) and an $AR(4)$ process for the equation for the log consumption-income ratio. The effective sample period is $1948Q2 - 2015Q4$ due to the use of four lags in estimation. Estimation details can be found in Appendix D with $l = 0, 1, 2, 3, 4$ replacing $l = 0, 1, 2$. From the posterior distribution of the parameter β in equation (12), we report the posterior mean, the 90% highest posterior density interval (between square brackets) and the posterior probability that $\beta > 0$. For the cleaned series for c_t and y_t, measurement error has been taken out.

Appendix D Estimation details VAR

This appendix presents the details of the estimation of the VAR presented in Section 4. Details are given on the state space system, on how we test for time variation, and on the use of parameter priors. Further, the details of the Gibbs sampler are provided, followed by a convergence analysis. Finally, we present some estimation results not reported in the main text, i.e., the estimated time-varying volatilities of the return innovations.

D.1 State space system

The observation equations (10)-(11) are complemented by state equations for the time-varying parameters α_{t+1}, β_{t+1}, γ_{t+1}, δ_{t+1} and $\pi_{l,t+1}$ with $l = \{0, 1, 2\}$ and log volatilities $h_{k,t+1}$ with $k = \{\varepsilon, \eta\}$. All are assumed to follow random walks. We follow Frühwirth-Schnatter and Wagner (2010) and write these random walks in non-centered form as,

$$\alpha_{t+1} = \alpha + \iota_\alpha \sigma_\alpha \alpha^*_{t+1} \tag{D-1}$$

$$\beta_{t+1} = \beta + \iota_\beta \sigma_\beta \beta^*_{t+1} \tag{D-2}$$

$$\gamma_{t+1} = \gamma + \iota_\gamma \sigma_\gamma \gamma^*_{t+1} \tag{D-3}$$

$$\delta_{t+1} = \delta + \iota_\delta \sigma_\delta \delta^*_{t+1} \tag{D-4}$$

$$\pi_{l,t+1} = \pi_l + \iota_{\pi_l} \sigma_{\pi_l} \pi^*_{l,t+1} \qquad l = \{0,1,2\} \tag{D-5}$$

$$h_{k,t+1} = h_k + \iota_{h_k} \sigma_{h_k} h^*_{k,t+1} \qquad k = \{\varepsilon, \eta\} \tag{D-6}$$

with standardized random walks given by,

$$\alpha^*_{t+1} = \alpha^*_t + \varpi^\alpha_{t+1} \qquad \alpha^*_0 = 0 \qquad \varpi^\alpha_{t+1} \sim iid\mathcal{N}(0,1)$$

$$\beta^*_{t+1} = \beta^*_t + \varpi^\beta_{t+1} \qquad \beta^*_0 = 0 \qquad \varpi^\beta_{t+1} \sim iid\mathcal{N}(0,1)$$

$$\gamma^*_{t+1} = \gamma^*_t + \varpi^\gamma_{t+1} \qquad \gamma^*_0 = 0 \qquad \varpi^\gamma_{t+1} \sim iid\mathcal{N}(0,1)$$

$$\delta^*_{t+1} = \delta^*_t + \varpi^\delta_{t+1} \qquad \delta^*_0 = 0 \qquad \varpi^\delta_{t+1} \sim iid\mathcal{N}(0,1)$$

$$\pi^*_{l,t+1} = \pi^*_{l,t} + \varpi^\pi_{l,t+1} \qquad \pi^*_{l,0} = 0 \qquad \varpi^\pi_{l,t+1} \sim iid\mathcal{N}(0,1) \qquad l = \{0,1,2\}$$

$$h^*_{k,t+1} = h^*_{k,t} + \varpi^h_{k,t+1} \qquad h^*_{k,0} = 0 \qquad \varpi^h_{k,t+1} \sim iid\mathcal{N}(0,1) \qquad k = \{\varepsilon, \eta\}$$

where α, β, γ, δ, π_l and h_k are constants which correspond to the initial values of the random walks when the random walks are time-varying, where σ_α, σ_β, σ_γ, σ_δ, σ_{π_l} and σ_{h_k} are the square roots of the innovation variances of the random walks and where ι_α, ι_β, ι_γ, ι_δ, ι_{π_l} and ι_{h_k} are binary indicators that are equal to either zero or one.

After applying a triangular reduction to the variance covariance matrix of the innovations of the VAR given by equations (17)-(19) in Section 6.2.1, similar state equations are assumed for the time-varying parameters and log volatilities of that model (which we leave out here due to space considerations).

D.2 Testing for time variation

The non-centered random walks put forward in the previous section facilitate the estimation of and testing for time variation in parameters and log volatilities. First, these specifications improve on the estimation of the posterior distributions of the variances σ^2 of the random walk innovations. As the specifications are in terms of the square roots of the innovation variances σ, rather than the variances σ^2, the standard inverse gamma ($\mathcal{IG}$) prior normally used for σ^2 can be replaced by Gaussian prior for σ centered at zero.[12] This is advantageous because the choice of the scale and shape parameters that define the $\mathcal{IG}$

[12]The square roots of the innovation variances σ enter the model as regression coefficients. The centering of the Gaussian prior distributions of these coefficients at zero makes sense as, for both $\sigma^2 = 0$ and $\sigma^2 > 0$, σ is symmetric around zero.

prior distribution has a strong impact on the posterior distribution when the true value of the variance is close to zero. In particular, as the $\mathcal{IG}$ distribution does not have probability mass at zero, using it as a prior distribution tends to push the posterior away from zero. This is of particular importance when estimating the variances σ^2 of the random walk innovations as these variances determine whether or not there is time variation in the parameters and log volatilities of the model. Frühwirth-Schnatter and Wagner (2010) show that the posterior density of σ is much less sensitive to the parameters of the Gaussian distribution and is not pushed away from zero when $\sigma^2 = 0$. Second, in equations (D-1)-(D-6), the signs of the square roots of the variance σ and of the corresponding standardized random walk can be changed without changing their products. This non-identification provides further useful information on whether $\sigma^2 = 0$ or $\sigma^2 > 0$. When $\sigma^2 > 0$, the posterior distribution of σ is bimodal with modes $-\sigma$ and σ. When $\sigma^2 = 0$, the posterior is unimodal around zero. Third, the non-centered parameterization is useful for model selection because if the variance $\sigma^2 = 0$, then $\sigma = 0$ and the time-varying part of the random walk - i.e., the product of the square root of the innovation variance and the corresponding standardized random walk - drops from the corresponding equation in (D-1)-(D-6). Hence, in the non-centered parameterization, the presence or absence of the time-varying component in the random walk can be expressed as a variable selection problem through the use of the binary selection indicator ι. If $\iota = 0$, the time-varying component of the corresponding parameter or log volatility is excluded from the model and the constant represents the constant level of the parameter or log volatility. In this case, the parameter σ is set to zero. If $\iota = 1$, the time-varying component of the corresponding parameter or log volatility is included in the model and the constant represents the initial (= period zero) value. In this case, the parameter σ is estimated from the data. The binary indicators ι are sampled jointly with the other parameters. From their posterior distributions, we can calculate the posterior probabilities that the parameters and log volatilities are time-varying.

D.3 Parameter priors

Statistics for the prior distributions of the parameters of the VAR of equations (10)-(11) are presented in Table D-1. For the binary selection indicators, we use a Bernoulli distribution with prior probability equal to $p_0 = 0.5$.[13] For the other parameters, we use mean-zero Gaussian prior distributions with variances chosen so that the support contains a sufficiently large range of relevant parameter values. In particular, the initial/constant values of the random walk processes in equations (D-1)-(D-6) have prior variances equal to one with the exception of the constants of the log volatility processes which have a prior variance

[13] The results are robust to the use of alternative prior probabilities such as $p_0 = 0.25$ and $p_0 = 0.75$. These results are unreported but available upon request.

equal to ten. The square roots of the innovation variances of the random walk processes all have prior variances equal to one. These prior variances imply rather uninformative priors so that the reported results are driven mostly by the data and are not sensitive to the chosen priors. We note, finally, that the same prior values are used for the corresponding parameters of the extended three-equation VAR discussed in Section 6.2.1.

Table D-1: Prior distributions fixed parameters VAR equations (10)-(11)

Gaussian priors $\mathcal{N}(m_0, V_0)$			Percentiles	
	mean (m_0)	variance (V_0)	5%	95%
1. $\{\alpha, \beta, \gamma, \delta, \pi_l\}$	0.00	1.00	-1.64	1.64
2. $\{h_k\}$	0.00	10.00	-5.20	5.20
3. $\{\sigma_\alpha, \sigma_\beta, \sigma_\gamma, \sigma_\delta, \sigma_{\pi_l}, \sigma_{h_k}\}$	0.00	1.00	-1.64	1.64
Bernoulli priors $\mathcal{B}(p_0)$				
	mean (p_0)	variance ($p_0(1-p_0)$)		
$\{\iota_\alpha, \iota_\beta, \iota_\gamma, \iota_\delta, \iota_{\pi_l}, \iota_{h_k}\}$	0.50	0.25		

Notes: The constants of the random walks, i.e., α, β, γ, δ, π_l and h_k, and the square roots of the innovation variances of the random walks, i.e., σ_α, σ_β, σ_γ, σ_δ, σ_{π_l} and σ_{h_k} have Gaussian priors. The binary indicators ι_α, ι_β, ι_γ, ι_δ, ι_{π_l} and ι_{h_k} have Bernoulli priors. We note that $l = \{0, 1, 2\}$ and $k = \{\varepsilon, \eta\}$.

D.4 Gibbs sampler

D.4.1 Note

The equations that constitute the VAR can be estimated one equation at the time. The ordering of variables implied by the decomposition of the variance covariance matrix of the VAR innovations determines the ordering in which sampling occurs. For equations (10)-(11), first, the equation for $c - y$ is estimated and then, conditional on this, the equation for r. After applying a triangular reduction to the variance covariance matrix of the innovations of the VAR given by equations (17)-(19) in Section 6.2.1, first, the equation for bc is estimated, then the equation for $c - y$ and, finally, the equation for r.

D.4.2 Model

Each separate equation from the VAR systems considered in the paper can be written as,

$$x_{t+1} = z_{t+1}\psi_{t+1} + e^{h_{t+1}}\tau_{t+1} \qquad \tau_{t+1} \sim iid\mathcal{N}(0, 1) \tag{D-7}$$

where z_{t+1} is a $1 \times N$ vector of regressors - which includes the value of one for the intercept and which can include one or more innovations from other equations - with ψ_{t+1} the corresponding $N \times 1$ vector of time-varying coefficients and where h_{t+1} is log volatility. The time-varying parameters and log volatility follow random walks. Written in non-centered form, these are given by,

$$\psi_{n,t+1} = \psi_n + \iota_n \sigma_n \psi^*_{n,t+1} \tag{D-8}$$

for $n = 1, ..., N$ and,

$$h_{t+1} = h + \iota_h \sigma_h h^*_{t+1} \tag{D-9}$$

with standardized random walks given by,

$$\psi^*_{n,t+1} = \psi^*_{n,t} + \varpi^{\psi}_{n,t+1} \qquad \psi^*_{n,0} = 0 \qquad \varpi^{\psi}_{n,t+1} \sim iid\mathcal{N}(0,1) \tag{D-10}$$

$$h^*_{t+1} = h^*_t + \varpi^h_{t+1} \qquad h^*_0 = 0 \qquad \varpi^h_{t+1} \sim iid\mathcal{N}(0,1) \tag{D-11}$$

where ψ_n (for $n = 1, ..., N$) and h are constants which correspond to the initial values of the random walks when the random walks are time-varying, where σ_n (for $n = 1, ..., N$) and σ_h are the square roots of the innovation variances of the random walks and where ι_n (for $n = 1, ..., N$) and ι_h are binary indicators that are equal to either zero or one.

D.4.3 Offset mixture representation for the stochastic volatility component

The stochastic component $e^{h_{t+1}}\tau_{t+1}$ in equation (D-7) is nonlinear but can be transformed into a linear component by taking the logarithm of its square, i.e., $\ln(e^{h_{t+1}}\tau_{t+1})^2 = 2h_{t+1} + \ln \tau^2_{t+1}$ where $\ln \tau^2_{t+1}$ is log-chi-square distributed with expected value -1.27 and variance 4.93. Following Kim et al. (1998), we approximate this model by an offset mixture time series model as,

$$g_{t+1} = 2h_{t+1} + \tau^*_{t+1} \tag{D-12}$$

where $g_{t+1} \equiv \ln\left((e^{h_{t+1}}\tau_{t+1})^2 + c\right)$ with $c = 0.001$ an offset constant and the distribution of τ^*_{t+1} a mixture of J normal distributions given by,

$$\sum_{j=1}^{J} q^j f_N(\tau^*_{t+1} | m^j - 1.27, v^j) \tag{D-13}$$

with component probabilities q^j, means $m^j - 1.27$ and variances v^j. Equivalently, this mixture density can be written in terms of the mixture indicator variable ℓ_{t+1} as,

$$\tau^*_{t+1} | (\ell_{t+1} = j) \sim \mathcal{N}\left(m^j - 1.27, v^j\right) \tag{D-14}$$

with $prob(\ell_{t+1} = j) = q^j$. We set $J = 10$ to make the approximation of the mixture distribution to the log-chi-square distribution sufficiently good. We refer to Omori et al. (2007) who provide values for $\{q^j, m^j, v^j\}$.

D.4.4 Sampling algorithm

We use the Gibbs sampler to simulate draws from the joint posterior distribution the fixed parameters (i.e., binary indicators, constants, square roots of variances) and unobserved states (i.e., time-varying parameters and log volatilities). Conditional on the data, on the innovations of previous equations in the VAR, on the prior distributions of fixed parameters and on initial draws of the unobserved time-varying states taken from their prior distributions[14], the following steps are implemented where each step is conditional on the previous one:

1. Sample the binary indicators ι_n (for $n = 1, ..., N$).

2. Sample the parameters ψ_n and σ_n (for $n = 1, ..., N$).

3. Sample the unobserved states $\psi^*_{n,t+1}$ (for $n = 1, ..., N$).

4. Sample the mixture indicators ℓ_{t+1} as introduced in Section D.4.3.[15]

5. Sample the binary indicator ι_h.

6. Sample the parameters h and σ_h.

7. Sample the unobserved state h^*_{t+1}.

These steps are iterated 12.000 times and in each iteration values for fixed parameters and states are sampled. Of these draws, we discard the first 2.000 draws as a burn-in sequence. As such, we have 10.000 retained draws. We find that the Markov chain is sufficiently converged using this number of draws so that the retained sequence of draws for fixed parameters and states can be considered a sample from the posterior distribution. Convergence statistics are reported below in Section D.5 for the restricted VAR of equations (12)-(13) presented in Section 5.2. For the other VAR's estimated in the paper, convergence statistics are not reported but they are available upon request.

D.4.5 Framework for sampling the fixed parameters

The fixed parameters can be sampled from a regression model with heteroskedastic Gaussian errors with known variances,

$$u = w^r b^r + \epsilon \qquad\qquad \epsilon \sim \mathcal{N}(0, \Gamma) \tag{D-15}$$

[14] These draws can be obtained for $\psi^*_{n,t+1}$ and h^*_{t+1} from equation (D-10), respectively equation (D-11).

[15] The ordering is in line with Del Negro and Primiceri (2015) who note that the mixture indicators ℓ should be drawn after the quantities that do not condition on ℓ directly but before the quantities that condition on ℓ.

where u is a $T \times 1$ vector containing T observations on the dependent variable, w^r is a $T \times M$ matrix containing T observations of M regressors, b^r is a $M \times 1$ parameter vector, ϵ is a $T \times 1$ vector of error terms with $T \times T$ variance covariance matrix Γ. The binary indicators ι impose restrictions on the regressors. If all binary indicators in the regression are equal to one, then $w^r = w$ and $b^r = b$ where w and b are the unrestricted regressor matrix and the corresponding unrestricted coefficient vector. Otherwise, the restricted parameter vector b^r and the corresponding restricted regressor matrix w^r exclude those elements of w and b for which the corresponding binary indicators ι are equal to zero. The prior distribution of b^r is given by $b^r \sim \mathcal{N}(a_0^r, A_0^r)$ with a_0^r a $M \times 1$ vector and A_0^r a $M \times M$ matrix. The posterior distribution of b^r is then given by $b^r \sim \mathcal{N}(a^r, A^r)$ with,

$$A^r = \left[(w^r)'\Gamma^{-1}w^r + (A_0^r)^{-1}\right]^{-1}, \tag{D-16}$$

$$a^r = A^r \left[(w^r)'\Gamma^{-1}u + (A_0^r)^{-1}a_0^r\right], \tag{D-17}$$

When sampling the binary indicators ι, as in Frühwirth-Schnatter and Wagner (2010), we marginalize over the parameters in b that are subject to model selection. Next, we draw b^r conditional on ι. The posterior distribution of the binary indicators ι is obtained from Bayes' theorem as,

$$p(\iota|u, w) \propto p(u|\iota, w)p(\iota) \tag{D-18}$$

where $p(\iota)$ is the prior distribution of ι and $p(u|w, \iota)$ is the marginal likelihood of regression (D-15). We refer to equation (45) in Frühwirth-Schnatter and Wagner (2010) for the closed form expression of the marginal likelihood for the regression model (D-15).

D.4.6 Framework for sampling the time-varying unobserved states

The time-varying parameters and log volatilities can be sampled using a state space framework. The general form of the state space model is given by,

$$Y_{t+1} = Z_{t+1}S_{t+1} + V_{t+1}, \qquad V_{t+1} \sim \mathcal{N}(0, H_{t+1}) \tag{D-19}$$

$$S_{t+1} = T_{t+1}S_t + K_{t+1}E_{t+1}, \qquad E_{t+1} \sim \mathcal{N}(0, Q_{t+1}) \tag{D-20}$$

$$S_0 \sim \mathcal{N}(s_0, P_0) \tag{D-21}$$

where Y_{t+1} is a $N^o \times 1$ vector of observations and S_{t+1} an unobserved $N^s \times 1$ state vector. The matrices Z_{t+1}, T_{t+1}, K_{t+1}, H_{t+1}, Q_{t+1} and the mean s_0 and variance P_0 of the initial state vector S_0 are assumed to be known (conditioned upon) and the error terms V_{t+1} and E_{t+1} are assumed to be serially uncorrelated and independent of each other at all points in time. Note that E_{t+1} is a $N^{ss} \times 1$ matrix (where $N^{ss} \leq N^s$). As equations (D-19)-(D-21) constitute a linear Gaussian state space model, the unknown state variables in S_t can be filtered using the standard Kalman filter. Sampling $S = [S_1, \ldots, S_T]$ from its conditional distribution is then implemented using the multimove Gibbs sampler of Carter and Kohn (1994) which is discussed extensively in Kim and Nelson (1999).

D.4.7 Sampling details

Sampling ι_n, ψ_n and σ_n (for $n = 1, ..., N$)

Conditional on h_{t+1} and $\psi^*_{n,t+1}$ (for $n = 1, ..., N$), equation (D-7) combined with equation (D-8) can be written in the general notation of equation (D-15) setting $u_{t+1} = x_{t+1}$, $b = \left[\begin{array}{cccccc} \psi_1 & ... & \psi_N & \sigma_1 & ... & \sigma_N \end{array} \right]'$, $w_{t+1} = \left[\begin{array}{cccccc} z_{1,t+1} & ... & z_{N,t+1} & \psi^*_{1,t+1} z_{1,t+1} & ... & \psi^*_{N,t+1} z_{N,t+1} \end{array} \right]$, $\epsilon_{t+1} = e^{h_{t+1}} \tau_{t+1}$ and where we have $\Gamma = diag\left(e^{2h_1}, ..., e^{2h_T}\right)$. If one or more binary indicators $(\iota_1, ..., \iota_N)$ take on the value of zero, this implies the restricted w^r_{t+1} and b^r.

First, the binary indicators ι are sampled from their posterior distribution in equation (D-18). In particular, we follow George and McCulloch (1993) and use a single-move sampler where the N binary indicators ι_n are sampled one-by-one. We calculate the marginal likelihoods $p(u|\iota_n = 1, w, \iota_{-n})$ and $p(u|\iota_n = 0, w, \iota_{-n})$ (see Frühwirth-Schnatter and Wagner, 2010, for the expression). Upon combining the marginal likelihoods with the Bernoulli prior distributions of the binary indicators $p(\iota_n = 1) = p_0 = 0.5$ and $p(\iota_n = 0) = 1 - p_0$, the posterior distributions $p(\iota_n = 1|u, w, \iota_{-n})$ and $p(\iota_n = 0|u, w, \iota_{-n})$ are obtained from which the probability $prob(\iota_n = 1|u, w, \iota_{-n}) = \frac{p(\iota_n=1|u,w,\iota_{-n})}{p(\iota_n=1|u,w,\iota_{-n})+p(\iota_n=0|u,w,\iota_{-n})}$ is calculated which is used to sample ι_n, i.e., draw a random number r from a uniform distribution with support between zero and one and set $\iota_n = 1$ if $r < prob(.)$ and $\iota_n = 0$ if $r > prob(.)$.

Second, conditional on ι, the parameters ψ and σ can be sampled. If all binary indicators in the regression are equal to one, then $w^r = w$ and $b^r = b$. If some binary indicators are equal to zero, the unrestricted w and b are restricted to obtain w^r and b^r by excluding those coefficients σ_n for which the corresponding binary indicators ι_n are zero. These σ_n are not sampled but set to zero. We sample b^r from the distribution $\mathcal{N}\left(a^r, A^r\right)$ with a^r and A^r given by equations (D-16)-(D-17). The means and variances of the prior distributions are a^r_0 and A^r_0. For the estimation of the VAR's of equations (10)-(11) and equations (12)-(13), the numbers used are reported in Table D-1.

Sampling $\psi^*_{n,t+1}$ (for $n = 1, ..., N$)

Conditional on h_{t+1}, ψ_n and σ_n (for $n = 1, ..., N$), equation (D-7) combined with equation (D-8) can be written as the observation equation $x_{t+1} - \sum_{n=1}^{N} \psi_n z_{n,t+1} = \sigma_1 z_{1,t+1} \psi^*_{1,t+1} + ... + \sigma_N z_{N,t+1} \psi^*_{N,t+1} + e^{h_{t+1}} \tau_{t+1}$ of a state space system with state equations given by equation (D-10) (for $n = 1, ..., N$). Using the state space framework of equations (D-19)-(D-21), we then have the matrices $Y_{t+1} = x_{t+1} -$

$\sum_{n=1}^{N} \psi_n z_{n,t+1}$, $Z_{t+1} = \begin{bmatrix} \sigma_1 z_{1,t+1} & \dots & \sigma_N z_{N,t+1} \end{bmatrix}$, $S_{t+1} = \begin{bmatrix} \psi^*_{1,t+1} & \dots & \psi^*_{N,t+1} \end{bmatrix}'$, $V_{t+1} = e^{h_{t+1}}\tau_{t+1}$, $H_{t+1} = e^{2h_{t+1}}$, $T_{t+1} = I_N$, $K_{t+1} = I_N$, $E_{t+1} = \begin{bmatrix} \varpi^{\psi}_{1,t+1} & \dots & \varpi^{\psi}_{N,t+1} \end{bmatrix}'$, $Q_{t+1} = I_N$, $s_0 = \begin{bmatrix} 0 & \dots & 0 \end{bmatrix}'$ and $P_0 = 10^{-5} I_N$.

For binary indicators ι_n equal to zero, the corresponding values of σ_n are zero and the corresponding terms $\sigma_n z_{n,t+1}\psi^*_{n,t+1}$ drop out of the observation equation. The states $\psi^*_{n,t+1}$ for which $\iota_n = 0$ are then excluded from the state vector S_{t+1}. They are sampled from their prior distribution using equation (D-10). Using ψ_n, σ_n and $\psi^*_{n,t+1}$ (for $n = 1, ..., N$), the centered random walks $\psi_{n,t+1}$ can be reconstructed using equation (D-8).

Sampling ℓ_{t+1}

Conditional on g_{t+1} and h_{t+1}, the mixture indicator series ℓ_{t+1} is sampled from its conditional probability mass function which is given by $prob(\ell_{t+1} = j | g_{t+1}, h_{t+1}) \propto q^j f_N(g_{t+1} | 2h_{t+1} + m^j - 1.27, v^j)$ (see Kim et al., 1998). From the indicator series ℓ_{t+1}, we then obtain the mean series m_{t+1} and the variance series v_{t+1} as for every j there is a corresponding value of m^j and v^j.

Sampling ι_h, h and σ_h

Conditional on g_{t+1}, h^*_{t+1} and ℓ_{t+1}, equation (D-12) combined with equation (D-9) can be written in the general notation of equation (D-15) setting $u_{t+1} = g_{t+1} - (m_{t+1} - 1.27)$, $w_{t+1} = \begin{bmatrix} 2 & 2h^*_{t+1} \end{bmatrix}$, $b = \begin{bmatrix} h & \sigma_h \end{bmatrix}'$, $\epsilon_{t+1} = \tau^*_{t+1} - (m_{t+1} - 1.27)$. Given the mixture distribution of τ^*_{t+1} defined in equation (D-14), the centered error term $\tau^*_{t+1} - (m_{t+1} - 1.27)$ has a heteroskedastic variance v_{t+1} such that $\Gamma = diag\,(v_1, ..., v_T)$. If the binary indicator ι_h takes on the value of zero, this implies the restricted w^r_{t+1} and b^r.

First, the binary indicator ι_h is sampled from its posterior distribution in equation (D-18). We calculate the marginal likelihood $p(u|\iota_h = 1, w)$ and $p(u|\iota_h = 0, w)$ (see Frühwirth-Schnatter and Wagner, 2010, for the expression). Upon combining the marginal likelihoods with the Bernoulli prior distributions of the binary indicators $p(\iota_h = 1) = p_0 = 0.5$ and $p(\iota_h = 0) = 1 - p_0$, the posterior distributions $p(\iota_h = 1|u, w)$ and $p(\iota_h = 0|u, w)$ are obtained from which the probability $prob(\iota_h = 1|u, w) = \frac{p(\iota_h=1|u,w)}{p(\iota_h=1|u,w)+p(\iota_h=0|u,w)}$ is calculated which is used to sample ι_h, i.e., draw a random number r from a uniform distribution with support between zero and one and set $\iota_h = 1$ if $r < prob(.)$ and $\iota_h = 0$ if $r > prob(.)$.

Second, conditional on ι_h, the parameters h and σ_h can be sampled. If ι_h is equal to one, then $w^r = w$ and $b^r = b$. If ι_h is equal to zero, the unrestricted b is restricted to obtain b^r by excluding σ_h. In this case, σ_h is not sampled but set to zero. We sample b^r from the distribution $\mathcal{N}(a^r, A^r)$ with a^r and A^r given by equations (D-16)-(D-17). The means and variances of the prior distributions are a^r_0 and A^r_0.

For the estimation of the VAR's of equations (10)-(11) and equations (12)-(13), the numbers used are reported in Table D-1.

Sampling h^*_{t+1}

Conditional on h, σ_h and on the mixture indicator series ℓ_{t+1} from which we have the time-varying means $m_{t+1} - 1.27$ and variances v_{t+1} of the error term τ^*_{t+1}, equation (D-12) combined with equation (D-9) can be written as the observation equation $g_{t+1} - (m_{t+1} - 1.27) - 2h = 2\sigma_h h^*_{t+1} + \tau^*_{t+1} - (m_{t+1} - 1.27)$ of a state space system with state equation given by equation (D-11). Using the state space framework of equations (D-19)-(D-21), we then have the matrices $Y_{t+1} = g_{t+1} - (m_{t+1} - 1.27) - 2h$, $Z_{t+1} = 2\sigma_h$, $S_{t+1} = h^*_{t+1}$, $V_{t+1} = \tau^*_{t+1} - (m_{t+1} - 1.27)$, $H_{t+1} = v_{t+1}$, $T_{t+1} = 1$, $K_{t+1} = 1$, $E_{t+1} = \varpi^h_{t+1}$, $Q_{t+1} = 1$, $s_0 = 0$ and $P_0 = 10^{-5}$.

If the binary indicator ι_h equals zero, the corresponding value of σ_h equals zero and the state h^*_{t+1} is sampled from its prior distribution using equation (D-11). Using h, σ_h and h^*_{t+1}, the centered random walk h_{t+1} can be reconstructed using equation (D-9).

D.5 Convergence analysis Gibbs sampler

We analyse the convergence of the Gibbs sampler for the estimation of the restricted VAR of equations (12)-(13) as presented in Section 5.2. To this end, we use the simulation inefficiency factors as proposed by Kim et al. (1998) and the convergence diagnostic of Geweke (1992) for equality of means across subsamples of draws from the Markov chain (see Groen et al., 2013, for a similar convergence analysis). For the other VAR's estimated in the paper, convergence statistics are not reported but they are available upon request.

For each fixed parameter and for every point-in-time estimate of a state, we calculate the inefficiency factor as $IF = 1 + 2\sum_{l=1}^{m} \kappa(l,m)\widehat{\theta}(l)$ where $\widehat{\theta}(l)$ is the estimated l-th order autocorrelation of the chain of retained draws and $\kappa(l,m)$ is the kernel used to weigh the autocorrelations. We use a Bartlett kernel $\kappa(l,m) = 1 - \frac{l}{m+1}$ with bandwidth m set equal to 4% of the 10.000 retained sampler draws. If we assume that n draws are sufficient to cover the posterior distribution in the ideal case where draws from the Markov chain are fully independent, then $n \times IF$ provides an indication of the minimum number of draws that are necessary to cover the posterior distribution when the draws are not independent. Usually, n is set to 100. Then, for example, an inefficiency factor equal to 20 suggests that we need at least 2.000 draws from the sampler for a reasonably accurate analysis of the parameter of interest. Additionally, we compute the p-values of the Geweke (1992) test for the null hypothesis of equality of the means of the first 20% and last 40% of the retained draws obtained from the sampler for each fixed parameter and for every point-in-time estimate of the states. We compute Newey and West (1987) robust variances using a Bartlett kernel with bandwidth equal to 4% of the respective sample sizes.

Table D-2 presents the convergence analysis corresponding to the results reported in Tables 3 and 4 in the text. To economize on space, we group the results across parameters and periods. We report the median and - if applicable - the minima and maxima of the distributions of the inefficiency factors. We also report rejection rates of the Geweke (1992) test conducted both at the 5% and 10% levels of significance. These rates are equal to the number of rejections of the null hypothesis of the test per parameter group divided by the number of parameters in a parameter group. These rates can only be zero or one for individual (non-grouped) parameters but can lie between zero and one for the grouped parameters.

The calculated inefficiency factors show convergence of the chain of 10.000 draws for all parameters. The highest maximum value of an inefficiency factor equals 86. Hence, our sampler would provide a reasonably accurate analysis even with only 8.600 draws. The results for the Geweke (1992) test for equality of means across subsamples of the draws are also good. The rejection rates are generally low and often close to or equal to zero, suggesting that the means of the first 20% and last 40% of the draws are equal. In a number of instances, somewhat higher rejection rates are reported. This is particularly the case for the log volatility $h_{\varepsilon,t}$. We note, however, that the rejection rates in this case are the result of applying the Geweke (1992) test on each of 142 or 143 point-in-time estimates of that state. Upon rerunning the sampler using other seeds, we find that it is not always for the same h_{ε}'s - i.e., for the same periods - that equality of means across subsamples of draws is rejected. Hence, we conclude that the convergence of the sampler for the retained number of draws is generally satisfactory.

Table D-2: Convergence analysis of the Gibbs sampler (corresponding to the results of Tables 3 and 4)

				Inefficiency factors (Stats distribution)			Geweke (1992) test (Rejection rates)	
Country	Dependent variable	Parameter/state	Number	Median	Min	Max	5%	10%
France	Excess equity returns	$\{\alpha, \beta, \gamma, \delta\}$	4	1.71	1.04	2.33	0.00	0.00
		$h_{\varepsilon,t}$	143	3.27	1.52	67.25	0.26	0.34
	Excess bond returns	$\{\alpha, \beta, \gamma, \delta\}$	4	1.41	1.03	1.82	0.00	0.00
		$h_{\varepsilon,t}$	143	3.29	1.57	42.49	0.15	0.23
	Log consumption-income ratio	$\{\pi_0, \pi_1, \pi_2\}$	3	0.95	0.92	1.06	0.00	0.33
		h_η	1	3.71	-	-	0.00	0.00
Germany	Excess equity returns	$\{\alpha, \beta, \gamma, \delta\}$	4	1.84	1.19	2.07	0.00	0.25
		$h_{\varepsilon,t}$	143	2.96	1.58	86.19	0.24	0.33
	Excess bond returns	$\{\alpha, \beta, \gamma, \delta\}$	4	1.27	1.24	1.37	0.00	0.00
		$h_{\varepsilon,t}$	143	2.86	1.52	41.62	0.13	0.25
	Log consumption-income ratio	$\{\pi_0, \pi_1, \pi_2\}$	3	0.94	0.92	1.05	0.00	0.33
		h_η	1	3.34	-	-	0.00	0.00
UK	Excess equity returns	$\{\alpha, \beta, \gamma, \delta\}$	4	1.57	1.19	2.04	0.00	0.50
		$h_{\varepsilon,t}$	143	3.55	1.89	66.10	0.35	0.41
	Excess bond returns	$\{\alpha, \beta, \gamma, \delta\}$	4	1.36	0.93	1.77	0.00	0.00
		$h_{\varepsilon,t}$	143	3.32	2.03	41.29	0.22	0.36
	Log consumption-income ratio	$\{\pi_0, \pi_1, \pi_2\}$	3	0.95	0.93	1.05	0.00	0.33
		h_η	1	3.33	-	-	0.00	0.00
US	Excess equity returns	$\{\alpha, \beta, \gamma, \delta\}$	4	1.28	1.13	1.43	0.00	0.25
		$h_{\varepsilon,t}$	142	3.27	1.44	15.94	0.01	0.04
	Excess bond returns	$\{\alpha, \beta, \gamma, \delta\}$	4	1.81	1.18	1.87	0.00	0.25
		$h_{\varepsilon,t}$	143	3.22	1.95	45.24	0.37	0.43
	Log consumption-income ratio	$\{\pi_0, \pi_1, \pi_2\}$	3	0.94	0.93	1.06	0.00	0.33
		h_η	1	3.79	-	-	0.00	0.00

Notes: The convergence analysis pertains to the estimation of equation (12) for either excess equity returns or for excess government bond returns and to the estimation of equation (13) for the log consumption-income ratio. This analysis complements the results presented in Tables 3 and 4. We note that the effective sample period (after lagging) is $1873 - 2015$ for all countries and variables so that the effective sample size is $T = 143$ with the exception of the US when excess equity returns are used for which the effective sample period is $1874 - 2015$ with corresponding effective sample size $T = 142$. The statistics of the distribution of the inefficiency factors are presented in columns 5 to 7 for every parameter or group of parameters. The inefficiency factors are calculated per fixed parameter and per point-in-time estimate of a state - in this case, the only estimated state is the log volatility $h_{\varepsilon,t}$ - using a Bartlett kernel with bandwidth equal to 4% of the 10.000 retained draws. The rejection rates of the Geweke (1992) test conducted at the 5% and 10% levels of significance are reported in columns 8 and 9. These rates are equal to the number of rejections of the null hypothesis of the test per parameter group divided by the number of parameters in a parameter group. These rates are either 1 or 0 for parameters that are considered individually. They are based on the p-value of the Geweke test of the hypothesis of equal means across the first 20% and last 40% of the 10.000 retained draws which is calculated for every fixed parameter and for every point-in-time estimate of a state. The variances of the means in the Geweke (1992) test are calculated with the Newey and West (1987) robust variance estimator using a Bartlett kernel with bandwidth equal to 4% of the respective sample sizes.

D.6 The estimated time-varying log volatilities of the return innovations

This section reports the time-varying log volatility $h_{\varepsilon,t+1}$ estimated from the VAR for both excess equity returns and excess government bond returns in, respectively, Figures D-1 and D-2. These estimates correspond to the estimation results reported in Tables 3 and 4 in the main text.

Figure D-1: The log volatility $h_{\varepsilon,t+1}$ for the excess equity return innovations

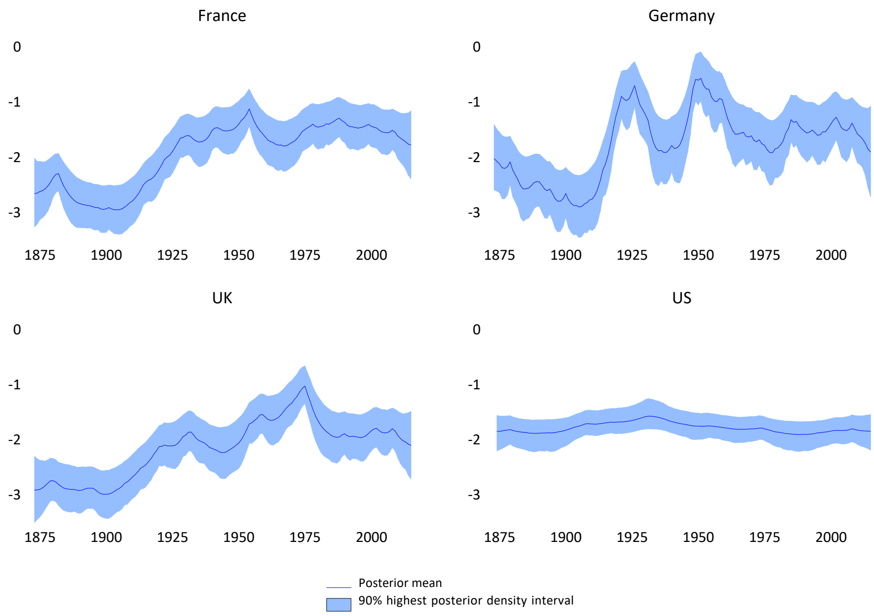


Figure D-2: The log volatility $h_{\varepsilon,t+1}$ for the excess government bond return innovations