

ON COMPLEX ECONOMIC SCENARIO GENERATORS: IS LESS MORE?

BY

JEAN-FRANÇOIS BÉGIN

ABSTRACT

This article proposes a complex economic scenario generator that nests versions of well-known actuarial frameworks. The generator estimation relies on the Bayesian paradigm and accounts for both model and parameter uncertainty via Markov chain Monte Carlo methods. So, to the question *is less more?*, we answer maybe, but it depends on your criteria. From an in-sample fit perspective, on the one hand, a complex economic scenario generator seems better. From the conservatism, forecasting and coverage perspectives, on the other hand, the situation is less clear: having more complex models for the short rate, term structure and stock index returns is clearly beneficial. However, that is not the case for inflation and the dividend yield.

KEYWORDS

Economic scenario generators, long-term forecast, risk management, scenario simulation.

1. INTRODUCTION

Economic scenario generators (ESGs) or investment models are comprehensive frameworks that produce scenarios of the joint behaviour of financial market values and economic variables. They are parametric models used by actuaries and risk managers in pension, life insurance and banking for numerous purposes: asset and liability management, financial planning, regulatory compliance, investment strategy and risk attribution, amongst others. These generators are different from other economic and financial models because their primary goal is to generate a set of future scenarios that span the range of realistic outcomes over the long run.

Astin Bulletin 51(3), 779-812. doi:[10.1017/asb.2021.21](https://doi.org/10.1017/asb.2021.21) © The Author(s), 2021. Published by Cambridge University Press on behalf of The International Actuarial Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

ESGs are helpful but—like most models—they are far from perfect. Indeed, “a model is an imperfect representation of a more complex reality” (Collins *et al.*, 2015). As we do not know the true data generating process, one of the major concerns is model risk that results from using insufficiently accurate models.

One could solve this issue by constructing a complicated, general framework that accounts for as many stylized facts as possible. We indeed have the knowledge and the technology to do so nowadays. However, this process would lead to a highly parametrized model, which goes against the idea of parsimony in statistics—the desire to explain phenomena using fewer parameters. This paper investigates this tradeoff: would a more complex ESG perform better, or would a simple model accomplish the same performance?

To answer this question, we develop a new, state-of-the-art ESG that focuses on four different variables: the inflation rate, the risk-free interest rate term structure, the dividend yield and stock index returns.¹ The ESG is based on three main principles. First, we consider a cascade structure to capture some of the dependency amongst the model variables. Second, we rely on regime-switching dynamics modelled by a Markov chain; in this article, they are observable—unlike most studies—and rely on monetary policy. Working with an observable regime is not only beneficial because it allows for a better interpretation of our model variables, but also because it leads to more straightforward parameter estimation procedures. Third, we consider the changing nature of the distribution over time by allowing for dynamic variances via generalized autoregressive conditional heteroskedasticity (GARCH) models. Again, the variances are observable in this case because the one-step-ahead variance is known in GARCH models. This ESG brings the appropriate level of complexity: it captures multiple stylized facts without being overly complicated, and it relies on well-known models and observed variables.

This framework is not the first ESG proposed in the actuarial literature, however (for an in-depth review of ESGs, see Pedersen *et al.*, 2016). The literature on this class of models dates to the mid-1980s with the publication of an influential study by Wilkie (1986). Based on the Box–Jenkins approach, his framework is composed of four connected models: an inflation model, an interest rate model, a dividend yield model and a stock index return model. It has been the object of multiple studies over the years (e.g., Geoghegan *et al.*, 1992; Wilkie, 1995; Huber, 1997; Zhang *et al.*, 2018; Bégin, 2019).

Advances in computing power and technology paved the way for more complex models based on time series methods and applications of stochastic processes in finance and financial econometrics. For instance, Chan (1998) proposed a continuous-time ESG based on Lévy processes. Threshold autoregressive models were employed by Whitten and Thomas (1999) and Chan *et al.* (2004). Chan (2002) generalized Wilkie’s framework by allowing for multiple autoregressive and moving average effects. Ahlgrim *et al.* (2005) proposed a model similar to Wilkie’s, but allowed for regime-switching dynamics

for stock index returns and a second interest rate factor to model long rates. Dynamic volatility—similar to that used in this study—was proposed by Şahin *et al.* (2008), but those authors only considered autoregressive conditional heteroskedasticity (ARCH) dynamics. More recently, Sneddon *et al.* (2016) proposed the Simulation of Uncertainty for Pension Analysis (SUPA) model.

The literature above focused mainly on the modelling issue, paying less attention to the estimation issue. Most studies used classical (frequentist) approaches like the method of least square or maximum likelihood estimation, without substantial emphasis or discussion on the statistical procedures employed. These methods assume that the parameter values are known once the estimation procedure is carried out. It contrasts with another popular school of thought in statistics—the Bayesian paradigm—which allows for a much more general definition of uncertainty: besides the apparent process uncertainty related to the stochastic nature of the random process under study, this paradigm also considers parameter and model uncertainty (see Chapter 6 of Bernardo and Smith, 2001, for details).

The notion of parameter uncertainty, on the one hand, is not a new concern for actuaries and has been studied on multiple occasions; for instance, see Hartman *et al.* (2017) and references therein for a general review, and Bégin (2019) for an application to ESGs. Model uncertainty, on the other hand, has been less of a consideration. One notable exception is Cairns (2000) who discussed a full Bayesian approach for both parameter and model uncertainty in the context of stochastic interest models and ruin theory. Our research question is related to the notion of model uncertainty because it investigates and compares different model features; this makes the use of Bayesian statistics a very natural choice to answer our question.

This paper contributes to the literature in the following ways. First, we introduce a new ESG for the actuarial profession that nests versions of well-known actuarial frameworks (i.e., a Wilkie-like model, an extended Wilkie-like model based on his 1995 extension and an Ahlgrim, D’Arcy, and Gorvett-like model). The model considers inflation, the term structure of risk-free interest rates, the dividend yield and stock index returns. Second, we propose a Bayesian-based Markov chain Monte Carlo (MCMC) estimation procedure that accounts for both parameter and model uncertainty. The method uses an adaptive Metropolis algorithm in the spirit of Haario *et al.* (2001), along with the reversible jump MCMC found in Green (1995). Third, we empirically investigate the performance of the new model and its nested variants on US data. We find that the most complex ESG works better from an in-sample fit perspective. From the conservatism, forecasting and coverage perspectives, the situation is less clear. Models for interest rates and stock index returns benefit from more sophisticated dynamics. It is the opposite for the inflation and the dividend yield, however.

The remainder of this article is organized as follows. Section 2 introduces a novel, state-of-the-art ESG for the actuarial profession and discusses models that are nested in this general framework. The Bayesian estimation

methodology used to estimate the new ESG is presented in Section 3. Section 4 explains our datasets and displays some of our empirical results. Section 5 goes further and provides a coherent mechanism to assess model uncertainty and selection. Then, in Section 6, the new ESG and other relevant generators are compared in an out-of-sample exercise that focuses on conservatism, forecasting and coverage. Section 7 concludes and discusses possible extensions.

2. A NEW ECONOMIC SCENARIO GENERATOR FOR THE ACTUARIAL PROFESSION

To assess the usefulness of more complex ESGs, we introduce a general framework that embeds models with features similar to those of Wilkie (1986, 1995) and Ahlgrim *et al.* (2005). The new ESG is based on three main principles:

1. **Cascade structure.** Like Wilkie (1986, 1995) and Ahlgrim *et al.* (2005), we capture some of the dependency amongst the model variables via a cascade structure—each subsequent variable in the structure depends only on prior variable values. For the proposed model, the monetary policy is the primary driver and impacts most variables in our framework.
2. **Observable regime-switching dynamics.** Unlike most ESGs proposed in the existing actuarial literature, we rely on a regime variable—monetary policy—to give our model more structure. Indeed, monetary policy impacts multiple facets of the economy such as the inflation, the short rate and stock returns (see, e.g., Engle *et al.*, 2017, for the impact on the short rate dynamics and Ioannidis and Kontonikas, 2008, for the effects on stock returns). We rely on observable instead of latent regimes because they allow for a better interpretation of our model variables and are easier to deal with from an estimation perspective.
3. **Observable dynamic variances.** This study relies on GARCH models to capture the varying nature of volatility over time. Specifically, we use non-linear asymmetric GARCH dynamics in the spirit of Engle and Ng (1993). In addition, as with regime-switching dynamics, we aim for a model that only allows for observable variables. GARCH dynamics are indeed very convenient to work with because the one-step-ahead conditional variance depends only on past observations, unlike stochastic volatility-type models.

Based on these three principles, we build a general framework. As shown in Figure 1, the framework relies on six models: monetary policy, inflation, the short rate, the term structure of interest rates, the dividend yield and stock index returns. We start by introducing each component. Then, we describe three nested economic scenario generators: (1) a simplified version of Wilkie's (1986) model, (2) a model based on his 1995 extension and (3) a variant of the framework proposed by Ahlgrim *et al.* (2005).

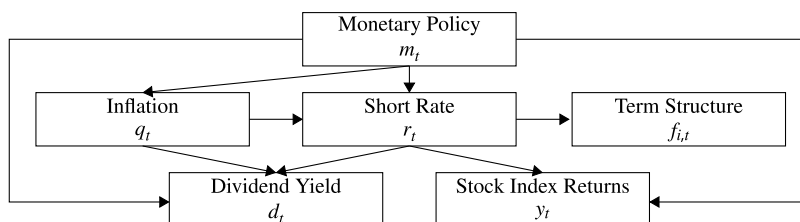


FIGURE 1: Cascade structure of the new economic scenario generator.

2.1. Monetary policy regimes

The first level of our proposed ESG models the monetary policy. We suggest a regime-switching dynamics to capture the changing nature of the policy in the spirit of Renne (2017).

Let m_t be a discrete-time observable Markov chain with three states: tightening or *upward* (u), *status quo* (s) and accommodating or *downward* (d). The transition probabilities of this Markov chain are given as follows:

$$\mathbf{\Pi} = \begin{bmatrix} p_{uu} & p_{us} & 0 \\ p_{su} & p_{ss} & p_{sd} \\ 0 & p_{ds} & p_{dd} \end{bmatrix} = \begin{bmatrix} p_{uu} & 1 - p_{uu} & 0 \\ p_{su} & 1 - p_{su} - p_{sd} & p_{sd} \\ 0 & 1 - p_{dd} & p_{dd} \end{bmatrix},$$

because $p_{uu} + p_{us} = 1$, $p_{su} + p_{ss} + p_{sd} = 1$, and $p_{ds} + p_{dd} = 1$.

We infer the states of this Markov chain from a reference rate R_t —fixed by the central bank—following Engle *et al.* (2017):²

$$m_t = \begin{cases} \text{u} & \text{if } \exists t' \in [t-3, t] \text{ and } t'' \in [t, t+3] \text{ such that } R_t - R_{t'} > 0 \text{ and } R_{t''} - R_t > 0 \\ \text{d} & \text{if } \exists t' \in [t-3, t] \text{ and } t'' \in [t, t+3] \text{ such that } R_t - R_{t'} < 0 \text{ and } R_{t''} - R_t < 0 \\ \text{s} & \text{otherwise.} \end{cases} \quad (2.1)$$

In lay terms, we consider an upward regime at time t whenever an increase in the reference rate happened less than 3 months earlier and the subsequent increase happens during the next 3 months. Similarly, we consider a downward regime at time t whenever a decrease in the reference rate happened less than 3 months earlier and the subsequent decrease happens during the next 3 months.

Note that, as opposed to most studies using regime-switching environments in the ESG literature, our model uses observable—economically meaningful—states instead of latent ones.

2.2. Inflation model

Let CPI_t be the level of the consumer price index (CPI) at time t . For most—if not all—economies, the CPI time series contain a unit root. Hence, instead

of modelling the level itself, it is common to model the rate of change of the CPI. In our setting, it is captured by an autoregressive model of order one, AR(1), with a regime-dependent long-run mean level and conditional heteroscedasticity:

$$q_t = \log \left(\frac{\text{CPI}_t}{\text{CPI}_{t-1}} \right) = \mu_{q,m_t} + a_q (q_{t-1} - \mu_{q,m_t}) + \sigma_{q,t} \varepsilon_{q,t}, \quad \varepsilon_{q,t} \sim \mathcal{N}(0, 1), \quad (2.2)$$

$$\sigma_{q,t+1}^2 = \sigma_q^2 + \alpha_q \left((\sigma_{q,t} \varepsilon_{q,t} - \sigma_{q,t} \gamma_q)^2 - \sigma_q^2 (1 + \gamma_q^2) \right) + \beta_q (\sigma_{q,t}^2 - \sigma_q^2), \quad (2.3)$$

where $\mu_{q,u}$, $\mu_{q,s}$ and $\mu_{q,d}$ are the regime-dependent long-run levels of the inflation rate and a_q is the autoregressive parameter. The inflation rate update of Equation (2.2) is reminiscent of the AR model used by Wilkie (1986) and Ahlgrim *et al.* (2005), amongst others. Nonetheless, it also includes the potential for heteroscedasticity (i.e., Equation (2.3)) and $\sigma_{q,t}^2$ is the time- t conditional variance of the inflation. The variance is updated via a GARCH model similar to those used by Kilian and Manganelli (2007) and Bégin (2016) in the context of inflation modelling. The long-run variance level of the inflation rate is given by σ_q^2 , and α_q , β_q and γ_q are the reaction, persistence and asymmetry parameters, respectively.

This model could be generalized by including regime dependence in our volatility process and its parameters. However, our initial analysis of the US data shows that average inflation tends to be impacted by the regimes, whereas the standard deviation is not. To keep the economic scenario generator as parsimonious as possible, we do not include such regime-dependent variance levels, but one could do so if the data warrant their use.

2.3. Risk-free interest rates

Interest rates are one of the most critical components in many economies. They influence the cost of borrowing, the return on savings and the total return of many investments. Consequently, the literature associated with interest rate models is voluminous (the interested reader may refer to Brigo and Mercurio, 2007, for an exhaustive review of interest rate modelling). In the context of ESGs, the risk-free yield curve is also considered as one of the core components because of its importance in discounting liability cash flows and in determining returns on fixed-income instruments.

Our framework's risk-free interest rate is comprised of two components: a short rate model and a term structure component constructed on top of it. We introduce each component separately.

2.3.1. Short rate

Over the years, a debate has raged over the stylized facts required for short rate modelling. Most people agree that shorter rates tend to be more volatile than

longer rates (Ahlgrim *et al.*, 2005) and that the short rate appears to revert to some average level—a feature modelled in very early interest rate studies, such as Vasicek (1977). Nonetheless, the question of the possibility of negative rates has not yet been settled. Some authors proposed dynamics that could become negative (e.g., Vasicek, 1977), while others prevented the short rate from being negative (e.g., Rendleman and Bartter, 1980; Cox *et al.*, 1985). Another avenue is to apply a logarithmic transform to the rates directly before modelling them so that they remain positive.

The literature on ESG also shows divisions on this issue: Wilkie (1986, 1995) uses a logarithm transformation to ensure that real interest rates are positive, whereas Ahlgrim *et al.* (2005) allow for negative rates. Our model embeds both approaches. Specifically, we rely on a transformation reminiscent of Engle *et al.* (2017) by incorporating a piecewise function that deals differently with high and low rates: we use a linear transform for higher rates and a logarithmic transform for lower rates. The transformed (continuously compounded) short rate is given by

$$\tilde{r}_t \equiv T_r(r_t) = \begin{cases} r_t & \text{if } r_t > \bar{r} \\ c_{r,0} + c_{r,1} \log(r_t - c_r) & \text{if } r_t \leq \bar{r} \end{cases}$$

where $c_{r,0} = \bar{r} - (\bar{r} - c_r) \log(\bar{r} - c_r)$ and $c_{r,1} = \bar{r} - c_r$, which makes function T_r continuously differentiable and strictly monotonic for all $r_t > c_r$. Our approach is slightly more general, as negative rates are permitted if $c_r < 0$; for instance, an actuary may determine that the natural limit for the short rate could be lower than zero. Our approach allows for some flexibility in this respect.³

Similar to the inflation model introduced above, the transformed short rate is modelled by an AR(1) model with a regime-dependent mean level and conditional heteroscedasticity:

$$\begin{aligned} \tilde{r}_t &= \mu_{r,m_t} + a_r (\tilde{r}_{t-1} - \mu_{r,m_t}) + \sigma_{r,t} \varepsilon_{r,t}, & \varepsilon_{r,t} &\sim \mathcal{N}(0, 1), \\ \sigma_{r,t+1}^2 &= \sigma_r^2 + \alpha_r \left((\sigma_{r,t} \varepsilon_{r,t} - \sigma_{r,t} \gamma_r)^2 - \sigma_r^2 (1 + \gamma_r^2) \right) + \beta_r (\sigma_{r,t}^2 - \sigma_r^2), \end{aligned}$$

where $\mu_{r,u}$, $\mu_{r,s}$ and $\mu_{r,d}$ are the regime-dependent long-run levels of the short rate. The other parameters are analogous to those introduced in Section 2.2.

Finally, to capture the potential relationship between inflation and the short rate innovations, we assume a non-nil correlation between $\varepsilon_{q,t}$ and $\varepsilon_{r,t}$; that is, $\text{Corr}(\varepsilon_{q,t}, \varepsilon_{r,t}) = \rho_{q,r}$. Note that the correlation captures similar dynamics to Wilkie's explicit loading on q_t in his interest rate model.⁴

2.3.2. Term structure

As a basis for the rest of our term structure, we use (continuously compounded) forward rates constructed from yields. We denote $f_{i,t}$ as the forward rate observed at time t for a contract starting at $t + \tau_{i-1}$ and ending at the next

available maturity, $t + \tau_i$. The latter is computed in the following way:

$$f_{i,t} = \frac{1}{\tau_i - \tau_{i-1}} (\tau_i r_{i,t} - \tau_{i-1} r_{i-1,t}), \quad i \in \{1, 2, \dots, n\},$$

where $r_{i,t}$ is the zero-coupon bond yield for tenor τ_i and the shortest maturity available is assumed to be our short rate, that is, $r_{0,t} = r_t$.

To prevent the chance of any forward rates of becoming negative, we use a similar transformation to that proposed in Section 2.3.1:

$$\tilde{f}_{i,t} \equiv T_f(f_{i,t}) = \begin{cases} f_{i,t} & \text{if } f_{i,t} > \bar{f} \\ c_{f,0} + c_{f,1} \log(f_{i,t} - c_f) & \text{if } f_{i,t} \leq \bar{f} \end{cases},$$

where $c_{f,0} = \bar{f} - (\bar{f} - c_f) \log(\bar{f} - c_f)$ and $c_{f,1} = \bar{f} - c_f$.⁵

While interest rate—and forward rates for that matter—movements are complex, Litterman and Scheinkman (1991) show that 99% of the yield curve's total variation can be explained by three fundamental shifts: a level component, a slope component and a curvature component. We apply a similar idea to our transformed forward rates using the slope and the curvature as observable factors to explain the term structure; we do not consider the level as part of our forward rate model because it already accounted for by the short rate model.⁶ This strategy also allows us to cope with the dimensionality associated with modelling each forward rate individually.

Specifically, we model the term structure by assuming that the spreads—defined as the difference between the transformed forward rates and the transformed short rate—are generated by observable factors:

$$\tilde{f}_t - \mathbf{1}_n \tilde{r}_t = \boldsymbol{\mu}_f + \mathbf{A}_f \mathbf{F}_t + \boldsymbol{\Sigma}_f \boldsymbol{\varepsilon}_{f,t}, \quad \boldsymbol{\varepsilon}_{f,t} \sim \mathcal{N}_n(\mathbf{0}_n, \mathbf{I}_n),$$

where $\tilde{f}_t = [\tilde{f}_{1,t} \ \tilde{f}_{2,t} \ \dots \ \tilde{f}_{n,t}]^\top$, $\mathbf{1}_n$ is an n -dimensional vector of ones, $\boldsymbol{\mu}_f$ is an n -dimensional vector containing average spread levels, \mathbf{A}_f is an $n \times 2$ matrix defined below, $\boldsymbol{\Sigma}_f$ is an $n \times n$ diagonal matrix that contains the standard deviations of the measurement errors and $\mathcal{N}_n(\mathbf{0}_n, \mathbf{I}_n)$ is an n -dimensional multivariate normal distribution with mean $\mathbf{0}_n$ and variance \mathbf{I}_n denoting the $n \times n$ identity matrix. Each row of matrix \mathbf{A}_f is given by

$$\mathbf{A}_{f_i} = [a_{f_i,1} \ a_{f_i,2}].$$

As commonly done in the interest rate literature, the factors are calculated in the following way:

$$\begin{aligned} \text{Slope:} \quad F_{1,t} &\equiv \tilde{f}_{n,t} - \tilde{f}_{1,t}, \\ \text{Curvature:} \quad F_{2,t} &\equiv \tilde{f}_{1,t} + \tilde{f}_{n,t} - 2\tilde{f}_{j,t}, \end{aligned}$$

where $1 < j < n$. We then assume that the observable factors summarizing the spreads follow a two-dimensional autoregressive model given by the following

dynamics:

$$\mathbf{F}_t = \boldsymbol{\mu}_F + \mathbf{A}_F (\mathbf{F}_{t-1} - \boldsymbol{\mu}_F) + \boldsymbol{\Sigma}_F \boldsymbol{\varepsilon}_{F,t}, \quad \boldsymbol{\varepsilon}_{F,t} \sim \mathcal{N}_2(\mathbf{0}_2, \mathbf{I}_2),$$

where $\mathbf{F}_t = [F_{1,t} \ F_{2,t}]^\top$, $\boldsymbol{\mu}_F$ is a two-dimensional vector containing the long-run mean parameters and \mathbf{A}_F and $\boldsymbol{\Sigma}_F$ are 2×2 diagonal matrices containing the autoregressive parameters and variance parameters, respectively.⁷

2.4. Dividend yield

Most ESGs in the actuarial literature rely on AR(1) dynamics to model the logarithm of the dividend yield (Wilkie, 1986, 1995; Hibbert *et al.*, 2001; Ahlgrim *et al.*, 2005; Sneddon *et al.*, 2016). We generalize these dynamics to account for a regime-dependent long-run mean and conditional heteroscedasticity:

$$\log(d_t) = \log(\mu_{d,m_t}) + a_d (\log(d_{t-1}) - \log(\mu_{d,m_t})) + \sigma_{d,t} \varepsilon_{d,t}, \quad \varepsilon_{d,t} \sim \mathcal{N}(0, 1),$$

$$\sigma_{d,t+1}^2 = \sigma_d^2 + \alpha_d \left((\sigma_{d,t} \varepsilon_{d,t} - \sigma_{d,t} \gamma_d)^2 - \sigma_d^2 (1 + \gamma_d^2) \right) + \beta_d (\sigma_{d,t}^2 - \sigma_d^2),$$

where $\mu_{d,u}$, $\mu_{d,s}$ and $\mu_{d,d}$ are the regime-dependent long-run levels of the dividend yield. The interpretation of parameters a_d , σ_d^2 , α_d , β_d and γ_d is similar to their inflation and short rate counterparts.

Also, identically to the relationship between inflation and the short rate, we assume a non-nil correlation between $\varepsilon_{q,t}$, $\varepsilon_{r,t}$ and $\varepsilon_{d,t}$; that is, $\text{Corr}(\varepsilon_{q,t}, \varepsilon_{d,t}) = \rho_{q,d}$ and $\text{Corr}(\varepsilon_{r,t}, \varepsilon_{d,t}) = \rho_{r,d}$.

2.5. Stock index returns

Like interest rates, many studies have examined the behaviour of stock index and equity returns. Frequently, stock index returns are assumed to follow a normal distribution, similar to the assumption used in Black and Scholes (1973). Unfortunately, this very simplistic model does not fit historical observations of stock index returns; they reveal fatter tails than those implied by the normality assumption (Campbell *et al.*, 1997).

Several alternative approaches have been proposed to cope with this issue: GARCH models (e.g., Bollerslev, 1986), regime-switching frameworks (e.g., Hamilton, 1989) and jump-diffusion models (e.g., Heston, 1993; Duffie *et al.*, 2000). The two former models have been popularized in actuarial science by Hardy (2001) and the Canadian Institute of Actuaries (2002). The latter class of models received increasing attention from actuaries and has recently been used in academic studies (e.g., Cui *et al.*, 2017; Bégin, 2020).

In the spirit of Hardy (2001), Ahlgrim *et al.* (2005) and Collins *et al.* (2015), we propose a regime-switching model for stock returns in this study. Instead of considering latent regimes, however, we use the observable monetary regimes of Section 2.1, which capture the changing nature of the average return. In

addition, we add a GARCH structure to capture the changing nature of volatility over time. Assuming that S_t is the time- t price of the index and y_t is the time- t return, we have the following:

$$\begin{aligned} y_t &= \log \left(\frac{S_t}{S_{t-1}} \right) \\ &= \frac{r_t}{12} + \mu_{y,m_t} + \sigma_{y,t} \varepsilon_{y,t}, \quad \varepsilon_{y,t} \sim \mathcal{N}(0, 1), \\ \sigma_{y,t+1}^2 &= \sigma_y^2 + \alpha_y \left((\sigma_{y,t} \varepsilon_{y,t} - \sigma_y \gamma_y)^2 - \sigma_y^2 (1 + \gamma_y^2) \right) + \beta_y (\sigma_{y,t}^2 - \sigma_y^2), \end{aligned}$$

where $\mu_{y,u}$, $\mu_{y,s}$ and $\mu_{y,d}$ are the regime-dependent long-run average returns in excess of the short rate. Again, the GARCH parameters have a similar interpretation to those already used for the inflation, short rate and dividend models.⁸

2.6. Nested specifications

In addition to the full framework explained above, three nested cases are used in this study. These three nested models were selected because they are comparable to well-known frameworks in the ESG literature. They allow us to make comparisons readily without relying on additional modelling frameworks.

- **\mathcal{M}_1 : A Wilkie-like framework.** The most elementary model considered in this study is a framework reminiscent of Wilkie (1986). It is obtained by assuming only one monetary regime and homoscedastic innovations for the inflation rate, short rate, dividend yield and stock index returns. To force the interest rate model to use only one factor—the one associated with the short rate—we set all the entries of \mathbf{A}_f to zero. More technical details on this model are given in Section SM.A.1 of the Supplementary Material.
- **\mathcal{M}_2 : An extended Wilkie-like framework.** The term structure model consists of one factor (i.e., slope), bringing the total number of interest rate factors to two, similar to Wilkie (1995). For more technical details on this model, refer to Section SM.A.2 of the Supplementary Material.
- **\mathcal{M}_3 : An Ahlgrim, D’Arcy, and Gorrivett-like framework.** The model is slightly more complex than \mathcal{M}_2 , as it allows for regimes. The regime only impacts the mean and the variance of the stock index returns, however. All other dynamics, except for the stock index returns, are homoscedastic and highly similar to those used in \mathcal{M}_2 . The term structure model consists of one factor, similar to \mathcal{M}_2 .⁹ More details on this Ahlgrim, D’Arcy, and Gorrivett-like framework are given in Section SM.A.3 of the Supplementary Material.
- **\mathcal{M}_4 : The new framework.** This is the full ESG described in the current section.

Table 1 summarizes the main features of the three nested models as well as the entire model.

TABLE 1
SUMMARY OF FRAMEWORK PROPERTIES.

	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_4
Inflation				
Autoregressive	✓	✓	✓	✓
Regime-switching long-run mean	✗	✗	✗	✓
Dynamic GARCH variance	✗	✗	✗	✓
Risk-free interest rates				
Autoregressive	✓	✓	✓	✓
Multiple factors	✗	✓	✓	✓
Regime-switching long-run mean	✗	✗	✗	✓
Dynamic GARCH variance	✗	✗	✗	✓
Dividend yield				
Autoregressive	✓	✓	✓	✓
Regime-switching long-run mean	✗	✗	✗	✓
Dynamic GARCH variance	✗	✗	✗	✓
Stock index returns				
Regime-switching long-run mean	✗	✗	✓	✓
Regime-switching long-run variance	✗	✗	✓	✗
Dynamic GARCH variance	✗	✗	✗	✓

\mathcal{M}_1 stands for the Wilkie-like framework, \mathcal{M}_2 for the 1995 extended version of Wilkie's model, \mathcal{M}_3 for Ahlgrim *et al.*'s (2005) framework and \mathcal{M}_4 for the new ESG.

3. ESTIMATION

This section introduces the main building blocks of Bayesian inference: we describe what is needed to estimate our models while accounting for parameter uncertainty. We then briefly present the parameter prior distributions employed in this study. Finally, the MCMC-based methodology used to estimate the models is explained.

3.1. Building blocks of Bayesian inference

The ESG parameters introduced above are unknown and therefore need to be estimated. In this study, we consider a Bayesian inference about Θ , the model parameters. This paradigm has the benefit of allowing for parameter and model uncertainty in a consistent manner (the notion of model uncertainty is investigated in Section 5). As a result of this inference step, we understand the plausible values of Θ and the extent of the uncertainty associated with these estimates.

Parameter estimation under the Bayesian paradigm relies on three building blocks: the likelihood function, the parameter prior and the posterior distribution of the parameters. The latter distribution is obtained by employing the

Bayes theorem and combining the likelihood and the prior distribution:

$$\underbrace{\pi(\Theta | \mathbf{X}, \mathcal{M})}_{\text{Parameter Posterior}} \propto \underbrace{\mathcal{L}(\mathbf{X} | \Theta, \mathcal{M})}_{\text{Likelihood}} \underbrace{\pi(\Theta | \mathcal{M})}_{\text{Parameter Prior}}, \quad (3.1)$$

where $\mathbf{X} = \{q_t, \tilde{r}_t, \mathbf{F}_t, \tilde{\mathbf{f}}_t, d_t, y_t\}_{t=1}^T$ denotes the data. Note that this parameter posterior distribution depends on model \mathcal{M} ; indeed, we will obtain one parameter posterior distribution per model at first.

Under the modelling assumptions of Section 2, the likelihood function is easily derived. Section SM.B.1 of the Supplementary Material explains the construction of this function.

3.2. Parameter prior

Information about the model parameters' value should be incorporated in the estimation somehow; this is achieved by setting a prior distribution in Bayesian inference. When the modeller does not have strong prior beliefs or relevant information about the parameters, however, the prior distribution should have minimal influence on the inference. Typically, in this case, non-informative priors are used. In their most elementary forms, these priors are rectangular (i.e., flat) distributions over the feasible set of parameter values (Upton and Cook, 2014).

A flat prior is improper if the feasible set is infinite (i.e., when the model parameters are unbounded). This may lead to an improper posterior, meaning that no Bayesian inference can be reliably achieved. One way to cope with this issue is by using diffuse priors—distributions with a considerable dispersion.

In this study, we rely on such non-informative priors. Specifically, all the mean and long-run level parameters have normal priors with unit variance. The autoregressive parameters rely on uniform priors over the interval $(-1, 1)$ to ensure that the dynamics are stationary. The priors associated with variance parameters are exponentially distributed with a rate parameter of one. The GARCH reaction and persistence parameters have prior distributions that are uniform over the unit interval, and we consider only parameters that lead to a stationary variance process. The GARCH asymmetry parameters, on the other hand, have normally distributed priors with a large variance of 100. Finally, correlation parameters use a prior based on the Lewandowski *et al.* (2009) distribution, thus assuring that all correlation matrices are positive semidefinite.

3.3. Estimation methodology

Equation (3.1) lays out a theoretical foundation to understand parameter uncertainty for a given model. Computing this posterior distribution, however, can be quite cumbersome in practice, as it involves multiple high-dimensional

integrals. Many methods have been proposed to recover such *a posteriori* distributions numerically.

In this study, we use an MCMC scheme to generate parameter samples from the posterior distribution. Specifically, we rely on a version of the adaptive Metropolis algorithm of Haario *et al.* (2001). This procedure employs multidimensional random walk Metropolis steps along with a proposal based on a mixture of two normal distributions centred at the current parameter value in the Markov chain. The variance of the first mixture component is based on the target distribution's covariance structure based on the run so far. The variance of the second mixture component, on the other hand, relies on the identity matrix multiplied by some scaling parameters. This second mixture component assures that the proposal does not get stuck.

In our implementation, we use a blocking strategy—separation of sampling between certain sets of parameters—divided according to the different models introduced in Section 2. More details on the estimation methodology are given in Section SM.B.2 of the Supplementary Material.

4. EMPIRICAL RESULTS

4.1. Data

We begin this section by describing the data. This study focuses exclusively on post-1982 monthly data for the United States as some of our series starts in 1982. Specifically, our data begin in September 1982 and finish in January 2020.

We select the intended US Federal Reserve (Fed) funds rate as the reference interest rate for the US economy; it is the intended rate at which commercial banks in the United States lend to each other overnight.¹⁰ The data—the target rate before 2009 and the target range after 2009—are extracted from the Federal Reserve Economic Data (FRED) website maintained by the Federal Bank of St. Louis and converted into monthly time series by taking the beginning-of-month rate.¹¹ To make the range comparable with pre-2009 data, we take the midpoint as the intended Fed funds rate. The black line of Figure 2 shows the intended Fed funds rate used. Its average is almost 4% over the past 37 years. Over time, the intended Fed funds rate decreased: it was about 10% in 1982 and dropped to almost 0% between 2009 and 2015.

Second, the monthly inflation rate is constructed from the non-seasonally adjusted US Consumer Price Index for All Urban Consumers (CPURNSA) series. The CPI level data are also extracted from FRED and converted into monthly rates by taking the (log) return of the index. The annualized standard deviation of the series is 0.86%, and the series exhibits heteroscedasticity; for instance, the series variance is smaller between 1991 and 1999 and larger between 2005 and 2010.

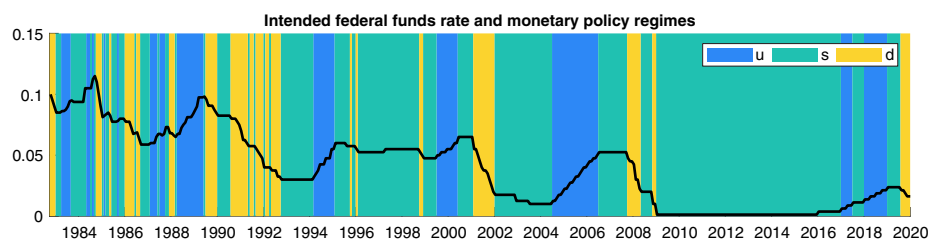


FIGURE 2: Intended US federal reserve funds rate and monetary policy regimes.

This figure shows the intended federal funds rate and the estimated monetary policy regimes. State *u* denotes a tightening or *upward* regime, *s* a *status quo* regime and *d* an accommodating or *downward* regime.

Third, the monthly term structure of risk-free interest rates is based upon the Federal Reserve Board's H.15 release and obtained via the Wharton Research Data Service. This study focuses on 3-month and 1-, 2-, 3-, 5-, 7-, 10-, and 30-year interest rates. The 3-month rate is used as a proxy for the short rate, and the other rates are used to estimate the term structure model. The yield curve is upward sloping during most months, except during turmoil such as the bursting of the dot-com bubble and the Great Recession. Like the intended Fed funds rate, the risk-free interest rates decreased over the last 37 years.

Fourth, the stock index dividend yields in the US economy are proxied by the dividends paid out on the stocks of the S&P 500. They are extracted from the Bloomberg terminal and constructed by taking the sum of the gross dividend payments over the previous 12 months and dividing it by the value of the index at the end of this period.

Finally, the stock index returns are constructed from the S&P 500 level and also obtained from the Bloomberg terminal. Similar to our process for inflation, we build our returns out of monthly index values. Interestingly, and consistently with previous studies (e.g., Bégin and Boudreault, [forthcoming](#)), we find heteroscedasticity, negative skewness (-1.02) and large positive excess kurtosis (3.63) for the monthly S&P 500 returns.

4.2. Monetary policy regimes

Before estimating the four models with our MCMC method, we first need to extract the monetary policy regimes. They are constructed from the intended Fed funds rate using Equation (2.1).

Figure 2 reports the time series of the intended Fed funds rate and their associated monetary regimes: state *u* (blue) denotes a tightening or *upward* regime, *s* (turquoise) a *status quo* regime and *d* (yellow) an accommodating or *downward* regime. Between 1982 and 2020, the US economy has been in a tight—contractionary—monetary policy about 21% of the time. This figure is about 17% for the accommodative monetary policy. The economy has been in a *status quo* state for the rest, which accounts for 62%.

4.3. Results and discussion

We apply the adaptive Metropolis algorithm from Section 3 to the four models of Section 2.6. This process yields four Markov chains that represent the posterior distributions of the model parameters. For each parameter, we get a sample of 100,000 values from which we compute the average and standard deviation for the Wilkie-like framework (\mathcal{M}_1), the extended Wilkie-like framework (\mathcal{M}_2), the Ahlgrim, D'Arcy, and Gorvett-like framework (\mathcal{M}_3) and the new ESG (\mathcal{M}_4). The average parameters and their standard deviations are available in Table SM.1 of the Supplementary Material.

From these estimated parameters, we can also obtain future scenarios by using the following posterior-predictive density:

$$f(\mathbf{Y} | \mathbf{X}, \mathcal{M}) = \int f(\mathbf{Y} | \boldsymbol{\Theta}, \mathcal{M}) \pi(\boldsymbol{\Theta} | \mathbf{X}, \mathcal{M}) d\boldsymbol{\Theta}, \quad (4.1)$$

where $\mathbf{Y} = \{q_t, \tilde{r}_t, \mathbf{F}_t, \tilde{\mathbf{f}}_t, d_t, y_t\}_{t=T+1}^{T+\tau}$ and $\pi(\boldsymbol{\Theta} | \mathbf{X}, \mathcal{M})$ is obtained from Equation (3.1). The latter is used to construct funnels of doubt.¹²

Regarding the monetary regime parameters, all regimes tend to be somewhat persistent, with a probability of staying in the same regime above 75% in all cases. The uncertainty about these values is rather small (i.e., standard deviations below 5%). Based on the average, the stationary state probabilities are consistent with the empirical proportion of months spent in each regime: 21% for the upward regime, 62% for the *status quo* regime and 17% for the downward regime.

The average inflation parameters for \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 are virtually the same, which is to be expected because the three frameworks share the same inflation model. Interestingly, the estimated regime-dependent long-run levels for \mathcal{M}_4 are consistent with theory, on average. On the one hand, the upward regime aims at slowing down overheated economic growth as inflation is rising too fast, consistently with the higher value estimated in this regime. On the other hand, the downward regime does the opposite, and we witness a lower average long-run level in this state. All four models find that the inflation rate is not persistent, with a_q ranging between 0.341 and 0.433, on average. The average variance parameters are similar for all four models, yielding annualized volatility between 0.7 and 1.1%. This figure is indeed consistent with the sample inflation rate standard deviation (i.e., 0.86% when annualized).

Panel A of Figure 3 shows the past inflation rate and the inflation funnels of doubt for the four models over the next 30 years. \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 unsurprisingly yield the same uncertainty. Overall, the new ESG allows for higher inflation rates than the three other models, leading to larger inflation rates and index values, on average. The distributions are nonetheless very similar and the total uncertainty not significantly different.

For the new ESG, the short rate long-run levels are in accordance with theory: $\mu_{r,u}$ is positive, $\mu_{r,s}$ is close to zero and $\mu_{r,d}$ is negative, as expected. All

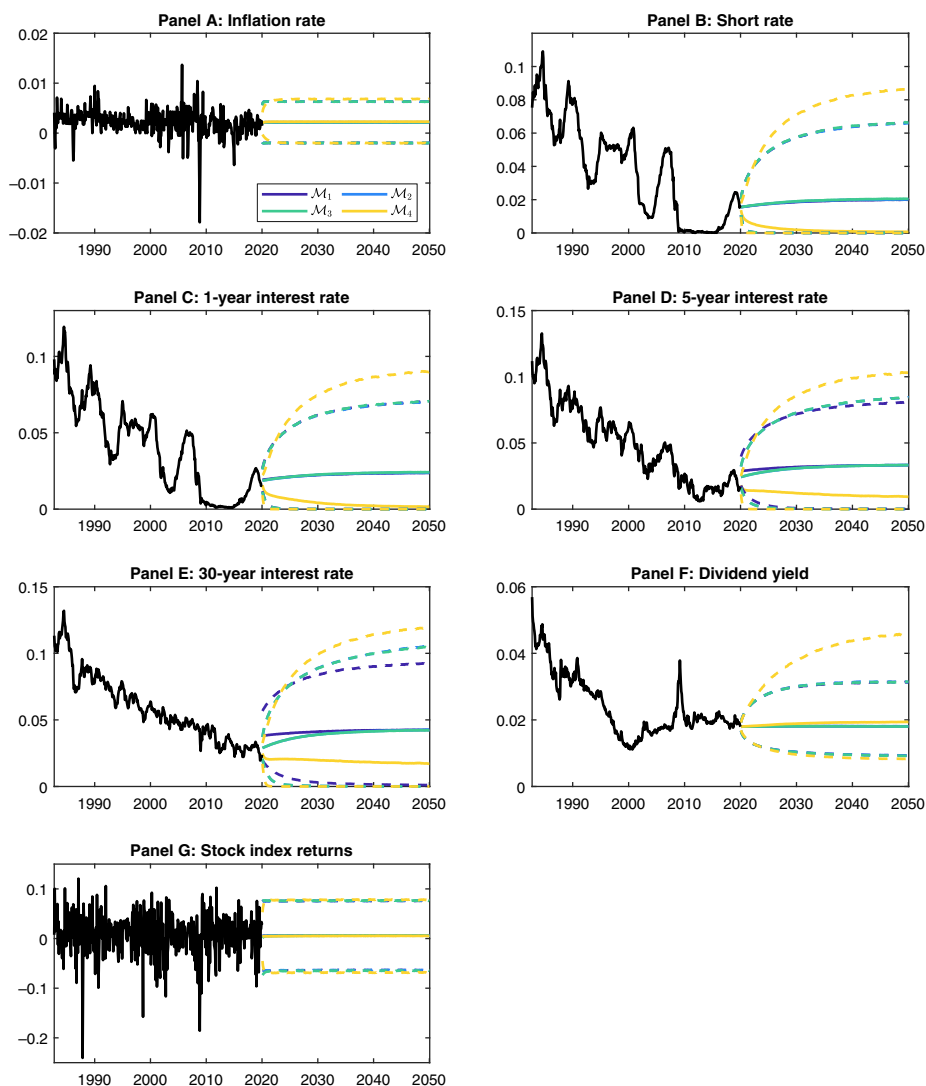


FIGURE 3: Funnel of doubt for relevant economic and financial variables.

This figure shows the past series along with the funnels of doubt for the four models over the next 30 years. The figure reports the inflation rate (Panel A), the short rate (Panel B), the 1-, 5- and 30-year interest rates (Panels C–E, respectively), the dividend yield (Panel F) and the stock index returns (Panel G). Solid lines represent median and dashed lines the 90% confidence interval. Dark blue is used for \mathcal{M}_1 , light blue for \mathcal{M}_2 , turquoise for \mathcal{M}_3 and yellow for \mathcal{M}_4 .

four models exhibit high persistence—above 0.99, on average; this result is consistent with other studies (e.g., Engle *et al.*, 2017). The short rate innovations display a low correlation with the inflation innovations for all four models, on average. Indeed, multiple studies have argued that the dependence between short-run interest rate and inflation is less than that prescribed by Fisher’s

equation (e.g., Mishkin, 1992). GARCH reaction and persistence parameters are high, whereas the asymmetry parameter is close to zero, on average. The GARCH parameters are quite uncertain, however.

The short rate funnels of doubt are shown in Panel B of Figure 3. Overall, \mathcal{M}_4 leads to a more conservative view than the three other models, with wider confidence intervals. The median is also quite different: it trends to zero for \mathcal{M}_4 , whereas it stays relatively constant at about 2% for \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 .

The term structure model builds on the short rate. For \mathcal{M}_1 , the mean and variance parameters increase as a function of the tenor; this captures the average upward slope of the term structure and the additional risk in longer tenors. We obtain a very different behaviour for \mathcal{M}_2 , \mathcal{M}_3 and \mathcal{M}_4 : indeed, the mean and variance parameters are rather constant as a function of the tenor. For \mathcal{M}_2 , \mathcal{M}_3 and \mathcal{M}_4 , the slope factor loading increases as a function of maturity: it is close to zero for 1-year rates and close to one for 30-year rates. The curvature loading—only used in \mathcal{M}_4 —is larger for short and long tenors and smaller for midterm tenors.

Panels C–E of Figure 3 report the funnels of doubt for three tenors: 1-year, 5-year and 30-year rates, respectively. In sum, \mathcal{M}_4 yields higher uncertainty than \mathcal{M}_2 and \mathcal{M}_3 . \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 are similar for short tenors but different for longer ones, with the extended Wilkie- and the Ahlgrim, D’Arcy, and Gorvett-like frameworks being more uncertain than the Wilkie-like framework.

Interestingly, the interest rate funnels of doubt (i.e., Panels B–E of Figure 3) reveal that the first part of the historical observations (roughly 1982–1985) is outside the confidence intervals for all four models. To test whether pre-1986 data impacts our conclusions, we reestimate the models using data between January 1986 and January 2020 instead. The estimated parameters are almost identical for both datasets, except for the long-run level of the short rate for \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 : it is larger for 1982–2020 and smaller for 1986–2020, as expected. The long-run level parameters for \mathcal{M}_4 are not significantly different, however. This impacts the interest rate funnels of doubt, yielding wider ones when using 1982–2020 for \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 . One compelling conjecture explaining the difference between the two sets of results is model misspecification: simpler models are more likely to be misspecified, and parameters tend to be less robust in these cases.

Additional results on the term structure fit are given in Table 2. The latter reports in-sample (average) root-mean-square errors (RMSEs) on the rates used in the estimation. Overall, the most complex model \mathcal{M}_4 yields the smallest errors. The performance of \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 is similar for shorter tenors, but \mathcal{M}_2 and \mathcal{M}_3 are better than \mathcal{M}_1 for longer maturities. This is expected, as the second and third models include the slope of the forward rates in its modelling.

The average values of dividend yield parameters are virtually the same for \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 but very different than those obtained for \mathcal{M}_4 . Indeed, for the new ESG, the long-run average of the dividend yield tends to be lower in the *status quo* monetary regime; it is larger in accommodating and tightening

TABLE 2

IN-SAMPLE AVERAGE ROOT-MEAN-SQUARE ERRORS FOR THE TERM STRUCTURE AS A FUNCTION OF THE TENOR.

	Tenor						
	1-Year (%)	2-Year (%)	3-Year (%)	5-Year (%)	7-Year (%)	10-Year (%)	30-Year (%)
\mathcal{M}_1	0.32	0.55	0.68	0.91	1.06	1.20	1.45
\mathcal{M}_2	0.31	0.51	0.58	0.63	0.63	0.60	0.49
\mathcal{M}_3	0.31	0.51	0.58	0.63	0.63	0.60	0.49
\mathcal{M}_4	0.28	0.35	0.37	0.38	0.39	0.42	0.41

This table reports in-sample average root-mean-square errors (RMSEs) on rates used in the estimation for the four models. Specifically, 1-, 2-, 3-, 5-, 7, 10- and 30-year rates are used. The average of the RMSEs is taken across the 100,000 parameter sets obtained via the MCMC sampler.

regimes. All models show high persistence in the log dividend yield rates, with a parameter close to 0.99, on average. The GARCH dynamics of \mathcal{M}_4 also exhibit high persistence (in addition to a low reaction average parameter): this implies that the conditional heteroskedasticity structure might not be needed for this variable.

The dividend yield funnels of doubt are given in Panel F of Figure 3. Generally speaking, the new ESG yields wider confidence intervals over the next 30 years, especially for the upper bounds. The median, on the other hand, is fairly similar for all four models.

The different monetary regimes allow for different mean parameters for the stock index returns; generally speaking, the *status quo* regime is associated with higher returns and the upward and downward regimes with lower returns. For the Ahlgrim, D’Arcy, and Gorvett-like framework, the return variance is significantly different from one regime to another, with the lowest volatility for the upward regime—when the economy is overheating—and the highest volatility for the downward regime—when the economy is expanding. The GARCH model used in the new ESG has a large asymmetry parameter, on average, implying a negative correlation between stock index returns and their variance. This observation is consistent with Black’s (1976) leverage effect.

Finally, Panel G of Figure 3 reports the stock index return funnels of doubt. The new ESG yields more conservative estimates, as the funnel is lower than the three other frameworks. It is closely followed by that of \mathcal{M}_3 . The Wilkie- and extended Wilkie-like frameworks unsurprisingly generate larger returns.

In sum, the four ESGs yield different results. The new ESG inflation and dividend yield future paths are not too different from those of \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 . However, they are utterly different when it comes to interest rates and stock index returns. In the next section, we investigate model selection and averaging.

TABLE 3
MODEL POSTERIOR PROBABILITIES.

	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_4
Inflation	0.00	0.00	0.00	1.00
Short rate	0.00	0.00	0.00	1.00
Term structure	0.00	0.00	0.00	1.00
Dividends	0.00	0.00	0.00	1.00
Stock index returns	0.00	0.00	0.00	1.00
Complete framework	0.00	0.00	0.00	1.00

This table reports the posterior model probabilities for each component and the entire framework. These are obtained via the reversible jump MCMC methodology introduced by Green (1995). For details on the method, see Section SM.B.3 of the Supplementary Material.

5. PROVIDING A COHERENT MECHANISM FOR MODEL SELECTION

This section first aims at finding the model posterior probabilities and then discusses Bayesian model averaging. Again, by using Bayes' theorem and combining the likelihood function with prior distributions, we obtain the following model posterior probabilities:

$$\underbrace{\pi(\mathcal{M} | \mathbf{X})}_{\text{Model Posterior}} \propto \int \underbrace{\mathcal{L}(\mathbf{X} | \boldsymbol{\Theta}, \mathcal{M})}_{\text{Likelihood}} \underbrace{\pi(\boldsymbol{\Theta} | \mathcal{M})}_{\text{Parameter Prior}} \underbrace{\pi(\mathcal{M})}_{\text{Model Prior}} d\boldsymbol{\Theta}. \quad (5.1)$$

The latter equation requires the modeller's initial beliefs about the models or a model prior probability distribution $\pi(\mathcal{M})$, in other words.¹³ We first consider uniform probabilities such that $\pi(\mathcal{M}_k) = \frac{1}{4}$ for all models.

We rely on a reversible jump MCMC scheme to find the model posterior probabilities of Equation (5.1). This methodology, introduced by Green (1995), allows us to implement Bayesian model selection and estimate model posterior probabilities readily by running a Markov chain similar to the adaptive Metropolis method used so far. Green's method allows the chain to move from one model to another, and the model posterior probability can therefore be approximated by the proportion of time spent in each given model. The implementation details of the methodology are given in Section SM.B.3 of the Supplementary Material.

Table 3 reports the posterior model probabilities for each component as well as the entire framework. The table's results are unequivocal: the new ESG is preferred with a posterior probability of one for each component and for the complete framework (i.e., estimated probability of exactly 100%, compared to 0% for simpler models). The data give very weak evidence in favour of the simpler models \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 .

Note, however, that a formal Bayesian treatment of this problem would not discuss model selection directly but would combine the models based on

their posterior probabilities. Bayesian model averaging, in fact, provides a coherent way to account for model uncertainty by averaging the different outputs according to their posterior probabilities (see Hoeting *et al.*, 1999, for an excellent tutorial on Bayesian model averaging). Using the model-dependent posterior-prediction density of Equation (4.1), we can derive the following posterior-prediction density that accounts for both model and parameter uncertainty:

$$f(\mathbf{Y} | \mathbf{X}) = \sum_{k=1}^4 \int f(\mathbf{Y} | \boldsymbol{\Theta}, \mathcal{M}_k) \pi(\boldsymbol{\Theta} | \mathbf{X}, \mathcal{M}_k) \pi(\mathcal{M}_k | \mathbf{X}) d\boldsymbol{\Theta}.$$

In this case, it is synonymous with using the new ESG to perform all predictions as $\pi(\mathcal{M}_4 | \mathbf{X}) = 1$.

To assess whether the model posterior probability is robust to other choices of model prior, we rerun the reversible jump MCMC scheme based on a different prior. Given that a more complicated model nesting simpler models as special cases can always fit the data better than the nested models, we penalize less parsimonious models by giving them smaller prior probabilities. Inspired by the Akaike information criterion (Akaike, 1974), we assume that the model prior for model \mathcal{M}_k is related to the number of parameters:

$$\pi(\mathcal{M}_k) \propto e^{-d_k},$$

where d_k is the number of parameters considered in model \mathcal{M}_k .

Interestingly, the results based on the new prior are identical to those reported in Table 3, meaning that the conclusion mentioned above is robust to the use of priors penalizing for model complexity. Again, the posterior probabilities associated with the most complex models are virtually 100%, favouring \mathcal{M}_4 to the other simpler models.

6. OUT-OF-SAMPLE PERFORMANCE

The previous section showed that the most complex framework fits the data better and is preferred. This result is important and interesting but should not come as a surprise: a general model like \mathcal{M}_4 captures more stylized facts than \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 .

Nonetheless, as ESGs are used to generate future paths, the in-sample fit might not be the only dimension on which one should focus. In fact, we should be careful when drawing conclusions from these in-sample comparisons. Models may pass statistical goodness-of-fit tests, but they may not perform well *ex ante*.

In this section, we therefore assess the four economic scenario generators using different criteria:

1. **Conservatism.** Whether the framework yields conservative figures.

2. **Forecasting performance.** Whether the framework yields median values that are close to past realized values.
3. **Coverage.** Whether the framework yields a range of outcomes consistent with past realized values.

Depending on the situation at hand, some of these criteria might not be of interest to the end user. We let the end user determine which criteria make more sense for their application.

We wish to do our performance testing on an *ex ante* basis. We therefore reestimate the four models at the end of each year based on an expanding window that starts in 1982; the expanding window must contain at least 15 years of data so that our first estimation uses the 1982–1996 data, the second one the 1982–1997 data, and so forth. Then, we select the most recent estimation results for a given month and generate 5 years of monthly variables (for instance, scenarios starting in June 1997 depend on parameters estimated with the 1982–1996 data and on variable values observed in May 1997). This process yields multiple measures—confidence interval, forecasts, coverage errors—that are out of sample by construction.

6.1. Conservatism

In this section, we assess our first criterion and verify that the *ex ante* views provided by our ESGs are suitably conservative, which are essential to preserve the profitability of a line of business and the solvency of a pension plan or an insurance company.

We rely on individual variables as a proxy to understand how conservative each model is. Typically, one would need a specific example to assess the interaction between the different variables (and their impact on the example at hand). It is very hard, unfortunately, to assess this properly without the introduction of multiple examples mimicking typical problems ESG end users encounter. We leave these complex applications for future research.¹⁴

Table 4 reports the average 5th (lower bound) and 95th (upper bound) percentiles used to construct 90% confidence intervals for the cumulative inflation rate, the short rate, the dividend yield, the cumulative stock index returns and key interest rates over different horizons. Generally speaking, the new ESG leads to more conservative numbers, followed by the Ahlgrim, D’Arcy, and Gorvett-like framework, the extended Wilkie-like framework, and then the Wilkie-like framework.

Similar to the results of Section 4, cumulative inflation rates are larger for \mathcal{M}_4 , leading to more significant lower and upper bounds for this model (see Panel A of Table 4). As both very low and very high inflation rates can be problematic for insurers (Bégin, 2016), it is unclear whether one model would be more conservative than the others from an inflation perspective.

The average bounds of the short rate are given in Panel B. In this case, the new ESG leads to more conservative figures, whereas \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3

TABLE 4
AVERAGE LOWER AND UPPER BOUNDS FOR 90% CONFIDENCE INTERVALS OVER DIFFERENT HORIZONS.

Panel A: Cumulative inflation rate																
	Lower bound								Upper bound							
	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year
\mathcal{M}_1	−0.001	−0.001	0.002	0.010	0.031	0.054	0.079	0.104	0.006	0.015	0.027	0.048	0.087	0.125	0.162	0.198
\mathcal{M}_2	−0.001	−0.001	0.002	0.010	0.031	0.054	0.079	0.104	0.006	0.015	0.027	0.048	0.087	0.125	0.162	0.198
\mathcal{M}_3	−0.001	−0.001	0.002	0.010	0.031	0.054	0.079	0.104	0.006	0.015	0.027	0.048	0.087	0.125	0.162	0.198
\mathcal{M}_4	−0.001	0.000	0.004	0.014	0.038	0.063	0.089	0.117	0.006	0.015	0.028	0.051	0.093	0.132	0.172	0.210
Panel B: Short rate																
	Lower bound								Upper bound							
	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year
\mathcal{M}_1	0.018	0.016	0.014	0.012	0.009	0.008	0.007	0.007	0.026	0.030	0.034	0.039	0.046	0.052	0.057	0.061
\mathcal{M}_2	0.018	0.016	0.014	0.012	0.009	0.008	0.007	0.007	0.026	0.030	0.034	0.039	0.046	0.052	0.057	0.061
\mathcal{M}_3	0.018	0.016	0.014	0.012	0.009	0.008	0.007	0.007	0.026	0.030	0.034	0.039	0.046	0.052	0.057	0.061
\mathcal{M}_4	0.018	0.015	0.012	0.008	0.004	0.002	0.001	0.000	0.026	0.031	0.036	0.045	0.058	0.067	0.074	0.080
Panel C: Dividend yield.																
	Lower bound								Upper bound							
	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year
\mathcal{M}_1	0.017	0.016	0.015	0.014	0.013	0.012	0.011	0.010	0.020	0.021	0.022	0.023	0.025	0.027	0.028	0.028
\mathcal{M}_2	0.017	0.016	0.015	0.014	0.013	0.012	0.011	0.010	0.020	0.021	0.022	0.023	0.025	0.027	0.028	0.028
\mathcal{M}_3	0.017	0.016	0.015	0.014	0.013	0.012	0.011	0.010	0.020	0.021	0.022	0.023	0.025	0.027	0.028	0.028
\mathcal{M}_4	0.017	0.016	0.015	0.014	0.012	0.011	0.010	0.010	0.020	0.021	0.022	0.024	0.026	0.028	0.029	0.030

TABLE 4
CONTINUED.

Panel D: Cumulative stock index returns																
	Lower bound								Upper bound							
	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year
\mathcal{M}_1	-0.066	-0.106	-0.141	-0.181	-0.220	-0.234	-0.236	-0.228	0.078	0.142	0.213	0.326	0.513	0.684	0.850	1.014
\mathcal{M}_2	-0.066	-0.106	-0.141	-0.181	-0.220	-0.234	-0.236	-0.228	0.078	0.142	0.213	0.326	0.513	0.684	0.850	1.014
\mathcal{M}_3	-0.066	-0.108	-0.145	-0.189	-0.237	-0.258	-0.265	-0.261	0.079	0.145	0.217	0.333	0.524	0.697	0.865	1.031
\mathcal{M}_4	-0.067	-0.124	-0.173	-0.235	-0.310	-0.355	-0.384	-0.403	0.078	0.134	0.202	0.314	0.504	0.676	0.842	1.004
Panel E: 1-year interest rate																
	Lower bound								Upper bound							
	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year
\mathcal{M}_1	0.019	0.017	0.015	0.013	0.011	0.010	0.009	0.008	0.033	0.036	0.039	0.044	0.051	0.057	0.061	0.066
\mathcal{M}_2	0.019	0.017	0.015	0.013	0.011	0.010	0.009	0.008	0.033	0.036	0.039	0.044	0.051	0.057	0.062	0.066
\mathcal{M}_3	0.019	0.017	0.015	0.013	0.011	0.010	0.009	0.008	0.033	0.036	0.039	0.044	0.051	0.057	0.062	0.066
\mathcal{M}_4	0.018	0.016	0.013	0.009	0.005	0.003	0.002	0.001	0.031	0.035	0.041	0.050	0.062	0.071	0.078	0.084
Panel F: 5-year interest rate																
	Lower bound								Upper bound							
	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year
\mathcal{M}_1	0.025	0.023	0.021	0.019	0.016	0.014	0.014	0.013	0.046	0.048	0.051	0.055	0.062	0.067	0.072	0.076
\mathcal{M}_2	0.029	0.027	0.024	0.021	0.016	0.014	0.013	0.012	0.045	0.049	0.052	0.058	0.066	0.072	0.077	0.082
\mathcal{M}_3	0.029	0.027	0.024	0.021	0.016	0.014	0.013	0.012	0.045	0.049	0.052	0.058	0.066	0.072	0.077	0.082
\mathcal{M}_4	0.025	0.022	0.018	0.013	0.007	0.004	0.003	0.002	0.041	0.047	0.054	0.063	0.077	0.086	0.093	0.099

TABLE 4
CONTINUED.

Panel G: 30-year interest rate																
	Lower bound								Upper bound							
	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year
\mathcal{M}_1	0.026	0.025	0.024	0.021	0.019	0.017	0.016	0.016	0.059	0.060	0.063	0.066	0.072	0.077	0.081	0.085
\mathcal{M}_2	0.038	0.034	0.031	0.026	0.020	0.016	0.014	0.012	0.055	0.060	0.064	0.071	0.080	0.087	0.093	0.098
\mathcal{M}_3	0.038	0.034	0.031	0.026	0.020	0.016	0.014	0.012	0.055	0.060	0.064	0.071	0.080	0.087	0.093	0.098
\mathcal{M}_4	0.035	0.031	0.026	0.018	0.009	0.005	0.003	0.002	0.053	0.059	0.065	0.076	0.089	0.099	0.107	0.113

This table reports the average lower and upper bounds of 90% confidence intervals for the cumulative inflation rate, the short rate, the dividend yield, the cumulative stock index returns, the 1-, 5- and 30-year interest rates. We take the averages over *ex ante* estimates of the 5th (lower bound) and 95th quantiles (upper bound) obtained by reestimating the four models at the end of each year. For a given month, we select the most recent estimation results and generate 5 years of monthly variables. These economic and financial quantities are taken over horizons of 1, 3 and 6 months, and 1, 2, 3, 4 and 5 years.

lead to less conservative estimates, on average. This contrasts with the average dividend yield quantiles (Panel C of Table 4); the bounds are relatively similar for all models.

Stock index returns are generally more conservative for \mathcal{M}_4 , as reported in Panel D of Table 4. Lower bounds are smaller for \mathcal{M}_3 when compared to \mathcal{M}_1 and \mathcal{M}_2 .

Panels E–G of Table 4 show the average lower and upper bounds of 90% confidence interval for three key interest rates: 1-year, 5-year and 30-year. Overall, the new ESG leads to more conservative estimates, followed by the Ahlgrim, D’Arcy, and Gorvett-like and the extended Wilkie-like frameworks, and then the Wilkie-like framework, except for very short horizons. In fact, \mathcal{M}_2 and \mathcal{M}_3 yield more conservative upper bounds than \mathcal{M}_4 over 1- and 3-month horizons.

Notice that “this does not imply that the model generating worst-case results is the most credible. However, it does suggest that inappropriate [...] advice may be offered [...] based on outputs from overly simplistic models” (Collins *et al.*, 2015). Indeed, interest rate models—short rate and term structure—and the stock index return model benefited from more complex models. Conservatism is not the only criterion, however.

6.2. Forecasting performance

We now investigate the out-of-sample forecasting accuracy of the median forecast for the four models by means of RMSEs. We define the prediction error as the difference between the median scenario and the observed value. The errors are computed for 1-, 3- and 6-month, and 1-, 2-, 3-, 4- and 5-year horizons.

Table 5 reports the RMSEs (multiplied by 100) for each model, horizon and economic series considered; that is, cumulative inflation rate, short rate, dividend yield, stock index returns, 1-, 2-, 3-, 5-, 7-, 10-, and 30-year interest rates.

The new ESG yields the most accurate forecasts, on average, for the short rate, stock index returns, and term structure of interest rates. Indeed, the average short rate forecast RMSEs for the new ESG are 5% lower than for \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 (see Panel B of Table 5). We obtain similar figures for the term structure of interest rates; across all tenors and maturities, \mathcal{M}_4 decreases the RMSEs by 17% when compared to \mathcal{M}_2 and \mathcal{M}_3 , and by 24% when compared to \mathcal{M}_1 . For the cumulative stock index returns, the RMSEs are large, generally speaking, as it is more challenging to forecast returns. Nonetheless, as presented in Panel D of Table 5, \mathcal{M}_4 performs better than the other models, with average decreases in the RMSEs of about 4%.

The tale is different when it comes to inflation and the dividend yield. The new ESG performs as well as \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 when it comes to the dividend yield (see Panel C of Table 5). This result is consistent with the in-sample

TABLE 5

OUT-OF-SAMPLE FORECAST ROOT-MEAN-SQUARE ERRORS OVER DIFFERENT HORIZONS.

Panel A: Cumulative inflation rate									
	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	Average
\mathcal{M}_1	0.256	0.611	0.930	1.409	2.294	3.144	4.097	5.085	2.228
\mathcal{M}_2	0.256	0.611	0.930	1.409	2.294	3.144	4.097	5.085	2.228
\mathcal{M}_3	0.256	0.611	0.930	1.409	2.294	3.144	4.097	5.085	2.228
\mathcal{M}_4	0.263	0.617	0.950	1.474	2.504	3.549	4.698	5.877	2.492
Panel B: Short rate									
	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	Average
\mathcal{M}_1	0.211	0.436	0.735	1.274	2.032	2.449	2.609	2.606	1.544
\mathcal{M}_2	0.211	0.436	0.735	1.274	2.032	2.449	2.609	2.606	1.544
\mathcal{M}_3	0.211	0.436	0.735	1.274	2.032	2.449	2.609	2.606	1.544
\mathcal{M}_4	0.195	0.362	0.599	1.127	1.988	2.467	2.603	2.442	1.473
Panel C: Dividend yield									
	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	Average
\mathcal{M}_1	0.092	0.171	0.265	0.365	0.449	0.449	0.456	0.477	0.340
\mathcal{M}_2	0.092	0.171	0.265	0.365	0.449	0.449	0.456	0.477	0.340
\mathcal{M}_3	0.092	0.171	0.265	0.365	0.449	0.449	0.456	0.477	0.340
\mathcal{M}_4	0.092	0.171	0.265	0.366	0.450	0.445	0.463	0.503	0.344
Panel D: Cumulative stock index returns									
	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	Average
\mathcal{M}_1	4.305	7.755	11.809	18.030	28.870	36.629	43.139	49.009	24.943
\mathcal{M}_2	4.305	7.755	11.809	18.030	28.870	36.629	43.139	49.009	24.943
\mathcal{M}_3	4.296	7.731	11.759	17.929	28.765	36.518	43.101	48.985	24.886
\mathcal{M}_4	4.288	7.741	11.713	17.705	28.087	35.262	41.075	46.115	23.998

evidence presented in Section 4. Model \mathcal{M}_4 is worse for inflation with an average increase in the RMSEs of about 12%.

6.3. Coverage

Our last criterion is coverage; it assesses whether the model-implied range of outcomes is consistent with past realized values. To do so, we count how many times the realized values were within the 90% confidence interval for each series and horizon (based on the out-of-sample 5th and 95th percentiles). This number divided by the total number of realized values—the coverage probability—should be close to 90% by construction.

Table 6 shows coverage errors (defined as the absolute value of the difference between the coverage probability and 0.9) for each model, horizon,

TABLE 5
CONTINUED.

Panel E: \mathcal{M}_1										
		Horizon								
		1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	Average
Tenor	1-year	0.384	0.580	0.842	1.316	2.039	2.471	2.661	2.691	1.623
	2-year	0.628	0.774	0.974	1.360	2.006	2.427	2.651	2.749	1.696
	3-year	0.741	0.857	1.016	1.333	1.906	2.302	2.547	2.684	1.673
	5-year	0.921	0.986	1.080	1.271	1.694	2.020	2.278	2.466	1.589
	7-year	1.055	1.098	1.157	1.269	1.560	1.822	2.078	2.294	1.542
	10-year	1.194	1.216	1.245	1.292	1.464	1.640	1.868	2.076	1.499
	30-year	1.444	1.443	1.434	1.392	1.360	1.367	1.501	1.692	1.454
Panel F: \mathcal{M}_2 and \mathcal{M}_3										
		Horizon								
		1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	Average
Tenor	1-year	0.401	0.594	0.854	1.322	2.037	2.463	2.653	2.694	1.627
	2-year	0.661	0.804	1.001	1.368	1.979	2.378	2.610	2.742	1.693
	3-year	0.754	0.874	1.038	1.339	1.868	2.231	2.488	2.675	1.659
	5-year	0.808	0.901	1.031	1.248	1.652	1.936	2.216	2.468	1.533
	7-year	0.798	0.886	1.006	1.188	1.517	1.753	2.039	2.319	1.438
	10-year	0.744	0.826	0.942	1.105	1.385	1.569	1.848	2.129	1.318
	30-year	0.579	0.667	0.786	0.928	1.141	1.263	1.523	1.794	1.085
Panel G: \mathcal{M}_4										
		Horizon								
		1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	Average
Tenor	1-year	0.299	0.476	0.720	1.184	1.966	2.452	2.602	2.482	1.523
	2-year	0.395	0.583	0.819	1.201	1.849	2.287	2.441	2.405	1.498
	3-year	0.411	0.608	0.841	1.165	1.712	2.095	2.253	2.281	1.421
	5-year	0.431	0.619	0.846	1.086	1.464	1.731	1.880	1.998	1.257
	7-year	0.440	0.629	0.856	1.061	1.330	1.511	1.643	1.808	1.160
	10-year	0.464	0.631	0.844	1.020	1.208	1.307	1.419	1.600	1.062
	30-year	0.421	0.581	0.781	0.926	1.007	0.991	1.051	1.239	0.875

This table reports RMSEs for each model, horizon, and economic series considered, that is, cumulative inflation rate, short rate, dividend yield, cumulative stock index returns, 1-, 2-, 3-, 5-, 7, 10- and 30-year interest rates. We define the prediction error as the difference between the median scenario and the observed value, which is computed for 1-, 3- and 6-month, and 1-, 2-, 3-, 4- and 5-year horizons. RMSEs are multiplied by 100 in the table.

and economic series considered. The table also reports whether these errors are statistically different from zero at a significance level of 5% by means of asterisks.¹⁵ Similar to the results presented above, the most complex short rate

TABLE 6
OUT-OF-SAMPLE COVERAGE ERRORS OVER DIFFERENT HORIZONS.

Panel A: Cumulative inflation rate									
	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	Average
\mathcal{M}_1	0.034	0.020	0.007	0.065	0.136	0.235	0.287	0.382*	0.146
\mathcal{M}_2	0.034	0.020	0.007	0.065	0.136	0.235	0.287	0.382*	0.146
\mathcal{M}_3	0.034	0.020	0.007	0.065	0.136	0.235	0.287	0.382*	0.146
\mathcal{M}_4	0.037	0.067	0.088	0.148	0.286	0.334	0.543*	0.597*	0.263
Panel B: Short rate									
	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	Average
\mathcal{M}_1	0.106	0.085	0.054	0.111	0.183	0.247	0.230	0.184	0.150
\mathcal{M}_2	0.106	0.085	0.054	0.111	0.183	0.247	0.230	0.184	0.150
\mathcal{M}_3	0.106	0.085	0.054	0.111	0.183	0.247	0.230	0.184	0.150
\mathcal{M}_4	0.106	0.031	0.026	0.021	0.038	0.057	0.043	0.086	0.051
Panel C: Dividend yield									
	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	Average
\mathcal{M}_1	0.003	0.005	0.010	0.028	0.057	0.003	0.048	0.059	0.027
\mathcal{M}_2	0.003	0.005	0.010	0.028	0.057	0.003	0.048	0.059	0.027
\mathcal{M}_3	0.003	0.005	0.010	0.028	0.057	0.003	0.048	0.059	0.027
\mathcal{M}_4	0.028	0.009	0.012	0.024	0.026	0.050	0.065	0.068	0.035
Panel D: Cumulative stock index returns									
	1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	Average
\mathcal{M}_1	0.005	0.009	0.008	0.050	0.077	0.123	0.109	0.111	0.062
\mathcal{M}_2	0.005	0.009	0.008	0.050	0.077	0.123	0.109	0.111	0.062
\mathcal{M}_3	0.010	0.009	0.026	0.035	0.069	0.098	0.091	0.079	0.052
\mathcal{M}_4	0.031	0.020	0.030	0.010	0.030	0.028	0.017	0.013	0.022

and the stock index return models work better: across all horizons, we have an average coverage error of 5.1% for the short rate and 2.2% for the stock index returns when considering \mathcal{M}_4 . These numbers are higher when considering other models: for the short rate, the coverage error is 15.0% when using \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 . It is 6.2% and 5.2% for the cumulative stock index returns when employing both Wilkie-like frameworks and the Ahlgrim, D'Arcy, and Gorvett-like framework, respectively.

The coverage errors are comparable when considering the dividend yield dynamics (Panel C of Table 6): the average is close to zero for all four models. This contrasts with the cumulative inflation rate: the coverage errors are higher for \mathcal{M}_4 , with an average of 26.3%. It is 14.6% for \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 .

Panels E–G of Table 6 report the errors for the yield curve series over different horizons. Unlike the other models, the new ESG has very good coverage

TABLE 6
CONTINUED.

Panel E: \mathcal{M}_1										
		Horizon								
		1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	Average
Tenor	1-year	0.046	0.031	0.003	0.054	0.160	0.198	0.209	0.161	0.108
	2-year	0.041	0.035	0.043	0.054	0.128	0.127	0.213	0.166	0.101
	3-year	0.135*	0.071	0.065	0.047	0.109	0.119	0.187	0.152	0.111
	5-year	0.167	0.104	0.069	0.054	0.069	0.094	0.135	0.083	0.097
	7-year	0.261*	0.184*	0.113	0.062	0.057	0.078	0.070	0.042	0.108
	10-year	0.333*	0.249*	0.146	0.062	0.034	0.049	0.013	0.001	0.111
	30-year	0.250*	0.184*	0.135*	0.010	0.057	0.088*	0.100	0.050	0.109
Panel F: \mathcal{M}_2 and \mathcal{M}_3										
		Horizon								
		1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	Average
Tenor	1-year	0.013	0.031	0.001	0.054	0.160	0.198	0.213	0.157	0.103
	2-year	0.178*	0.078	0.054	0.058	0.124	0.131	0.204	0.161	0.124
	3-year	0.257*	0.195*	0.146*	0.099	0.101	0.119	0.183	0.143	0.155
	5-year	0.247*	0.162*	0.113	0.069	0.042	0.082	0.043	0.028	0.098
	7-year	0.257*	0.129*	0.084	0.005	0.017	0.030	0.057	0.017	0.075
	10-year	0.257*	0.085	0.014	0.040	0.065	0.100	0.100	0.082*	0.093
	30-year	0.037	0.042	0.060	0.092*	0.100	0.100	0.100	0.100	0.079
Panel G: \mathcal{M}_4										
		Horizon								
		1-month	3-month	6-month	1-year	2-year	3-year	4-year	5-year	Average
Tenor	1-year	0.031	0.056	0.045	0.029	0.026	0.032	0.070	0.100	0.049
	2-year	0.017	0.038	0.038	0.029	0.014	0.005	0.078*	0.100	0.040
	3-year	0.010	0.042	0.038	0.044	0.041	0.055	0.096*	0.100	0.053
	5-year	0.010	0.060	0.049	0.074*	0.084	0.092*	0.100	0.100	0.071
	7-year	0.003	0.056	0.071*	0.092*	0.100	0.100	0.100	0.100	0.078
	10-year	0.001	0.053	0.082*	0.096*	0.100	0.100	0.100	0.100	0.079
	30-year	0.039	0.085*	0.093*	0.100	0.100	0.100	0.100	0.100	0.090

This table reports coverage errors for each model, horizon, and economic series considered; that is, cumulative inflation rate, short rate, dividend yield, cumulative stock index returns, 1-, 2-, 3-, 5-, 7, 10- and 30-year interest rates. Coverage errors are computed by taking the absolute value of the difference between the amount of time the realized values is within the 90% confidence interval for each series and horizon and 0.9. They are computed for 1-, 3- and 6-month, and 1-, 2-, 3-, 4- and 5-year horizons. An asterisk indicates a coverage error that is statistically different from zero at a significance level of 5%. The significance is assessed with robust Newey–West t -statistics as given in Equation (39) of Engle *et al.* (2017).

over short horizons and for longer tenor. Its coverage gets closer to 100% for longer horizons and tenors, which is too high.

The other models— \mathcal{M}_1 , \mathcal{M}_2 and \mathcal{M}_3 —lead to higher coverage errors, generally speaking. The Wilkie-like framework has coverage errors that are well above zero in all cases. For the extended Wilkie-like and the Ahlgrim, D’Arcy, and Gorvett-like frameworks, on the other hand, it depends on the tenor: the coverage error is decent for short and long tenors but tends to be higher for midterm tenors.

7. CONCLUDING REMARKS

The contributions of this study are threefold. First, we introduced a new state-of-the-art ESG for the actuarial profession. The latter nests versions of well-known actuarial models. Second, we proposed an MCMC-based estimation methodology for this class of models; it relied on the adaptive Metropolis algorithm and Green’s (1995) reversible jump MCMC method. Third, we investigated the framework empirically using US data.

This article also focused on model risk and the tradeoff between model parsimony and complexity. In the context of ESGs, model risk could be significant because the data generating process is complex and unknown. We looked at the in-sample fit of various models and concluded that the most complex ESG performed better. From an out-of-sample perspective, however, it is less clear-cut: although models for interest rates and stock index returns benefited from more sophisticated dynamics, simple inflation and dividend yield models performed better or similar to complex models.

As the saying goes, it is difficult to make predictions, especially about the future. Perhaps a better way to construct ESGs is by combining them to create ensemble models. This strategy has been used extensively in data-heavy domains such as weather forecasting (e.g., Gneiting and Raftery, 2005).¹⁶ A range of methodology already exists to combine models; it goes from simple unweighted averages to more sophisticated methods such as principal component-based methods, trimmed means, performance-based weighting, optimal least squared estimates, and Bayesian shrinkage (Graefe *et al.*, 2015).

Preliminary (unweighted average) results obtained by simply combining the four different frameworks investigated in this study are presented in the Supplementary Material. Overall, this strategy provides a good compromise; it allows for correcting overly or insufficiently conservative estimates. Moreover, if one model is off at times, the other models compensate and bring the ensemble to decent levels. This ultimately reduces model risk, as the effect of inaccurate models is averaged out by considering other, possibly better models.

Combining models is therefore a promising approach. Hopefully, this first brief attempt will not be the last. Specifically, considering better weights (e.g., based on the Akaike information criterion or the Bayesian information

criterion) is an interesting generalization of our initial results. Another way to construct ensemble models is by including more complex models, potentially from different families of ESGs. We leave these exciting questions for future research.

ACKNOWLEDGEMENTS

The author would like to thank Mario Wüthrich (the Editor-in-Chief), three anonymous referees, David Cantor, Yaqi Chen and Barbara Sanders for their helpful suggestions and comments. The author wishes to acknowledge the financial support of the National Science and Engineering Research Council of Canada (NSERC), Simon Fraser University and the NVIDIA Corporation.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/asb.2021.21>.

NOTES

1. The proposed ESG can be extended in different directions by adding more variables such as the gross domestic product, the unemployment rate and foreign exchange rates.

2. If m_t is in both upward and downward regimes at time t based on Equation (2.1), we assume that it is in the *status quo* regime.

3. Note that if we let \bar{r} be very small, we recover a model that allows for negative rates, as in Ahlgrim *et al.* (2005). On the other hand, if \bar{r} is very large, then we recover a function that behaves like a logarithmic transformation—similar to that of Wilkie (1986). In this study, we use $c_r = 0$ and $\bar{r} = 0.005$.

4. Using Chokesly's decomposition, it is easy to rewrite this model so that the short rate model depends explicitly on the inflation innovation, as in Wilkie (1986); the two approaches are therefore similar.

5. Similar to the short rate, we rely on $c_f = 0$ and $\bar{f} = 0.005$.

6. While the usual factors derived from forward rates have a natural economic interpretation, our factors are the slope and curvature of transformed forward rates and may not have the same interpretation.

7. The factor model is reminiscent of the vector autoregressive (VAR) model commonly used in econometrics to model the interaction between economic variables. One of their main advantages is that they do not require a detailed theory to estimate the relationship between economic variables (Sims, 1980). They were used by Campbell and Viceira (2002) in the investment context and Chan (2002) in the ESG context.

8. Note that σ_ϵ^2 could be generalized to a regime-dependent variance level, similar to Ahlgrim *et al.* (2005). Preliminary evidence shows that the GARCH model is sufficient to capture the heteroscedasticity.

9. We consider the short rate and the slope, whereas Ahlgrim, D'Arcy, and Gorvett model the short and long rates. Indeed, the slope is defined as a longer rate minus a shorter rate, so these two modelling choices are equivalent to some extent.

10. Most economies have a similar rate; for example, the Reserve Bank of Australia's cash rate target for Australia, the Bank of Canada's target overnight rate for Canada, the Bank of

England's bank rate for England and the European Central Bank's key interest rate for the Eurozone.

11. Since 2009, the Fed gives a target range (lower and upper limits) instead of a target rate.
12. We rely on Monte Carlo simulation to obtain a sample of 100,000 paths of \mathbf{Y} over the next 30 years (i.e., $\tau = 360$).
13. In this section, we consider the \mathbb{M} -closed view, where the true data generating process is assumed to be one of the models; that is, $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4$ (see Chapter 6 of Bernardo and Smith, 2001).
14. We also encourage end users to test conservatism in their specific situations. Accordingly, the code used to build and estimate these models is available on the author's personal website.
15. The Newey–West t -statistics are computed as in Equation (39) of Engle *et al.* (2017). More details can be found in Section 6 of their paper.
16. This rationale is consistent with the \mathbb{M} -open view, where the true data generating process is not assumed to be one of the models because this would be too difficult or we do not have the expertise (see Chapter 6 of Bernardo and Smith, 2001). Under this paradigm, $\pi(\mathcal{M}_k) = 0$, meaning that we cannot find posterior probabilities as we did in Section 5. The so-called weights, therefore, become variables that need to be optimized.

REFERENCES

- AHLGRIM, K.C., D'ARCY, S.P. and GORVETT, R.W. (2005) Modeling financial scenarios: A framework for the actuarial profession. *Proceedings of the Casualty Actuarial Society*, vol. 92, pp. 177–238. Arlington, VA, USA: Casualty Actuarial Society.
- AKAIKE, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic C*, **19**(6), 716–723.
- BÉGIN, J.-F. (2016) Deflation risk and implications for life insurers. *Risks*, **4**(4), 46.
- BÉGIN, J.-F. (2019) Economic scenario generator and parameter uncertainty: A Bayesian approach. *ASTIN Bulletin*, **49**(2), 335–372.
- BÉGIN, J.-F. (2020) Levelling the playing field: A VIX-linked structure for funded pension schemes. *Insurance: Mathematics and Economics*, **94**, 58–78.
- BÉGIN, J.-F. and BOUDREAU, M. (forthcoming) Do jumps matter in the long term? A tale of two horizons. *North American Actuarial Journal*.
- BERNARDO, J.M. and SMITH, A.F. (2001) *Bayesian Theory*. New York, NY, USA: Wiley.
- BLACK, F. (1976). Studies of stock price volatility changes. *Proceedings of the 1976 Meeting of the American Statistical Association, Business and Economic Statistics Section*, pp. 177–181. Washington, DC, USA: American Statistical Association.
- BLACK, F. and SCHOLES, M. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**(3), 637–654.
- BOLLERSLEV, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**(3), 307–327.
- BRIGO, D. and MERCURIO, F. (2007) *Interest Rate Models—Theory and Practice: With Smile, Inflation and Credit*. Berlin, Germany: Springer.
- CAIRNS, A.J. (2000) A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics*, **27**(3), 313–330.
- CAMPBELL, J.Y., LO, A.W. and MACKINLAY, A.C. (1997) *The Econometrics of Financial Markets*. Princeton, NJ, USA: Princeton University Press.
- CAMPBELL, J.Y. and VICEIRA, L.M. (2002) *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. New York, NY, USA: Oxford University Press.
- CANADIAN INSTITUTE OF ACTUARIES (2002) Report of the task force on segregated fund investment guarantees. Technical report, Canadian Institute of Actuaries.
- CHAN, T. (1998) Some applications of Lévy processes to stochastic investment models for actuarial use. *ASTIN Bulletin*, **28**(1), 77–93.
- CHAN, W.-S. (2002) Stochastic investment modelling: A multiple time-series approach. *British Actuarial Journal*, **8**(3), 545–591.

- CHAN, W.-S., WONG, A.C. and TONG, H. (2004) Some nonlinear threshold autoregressive time series models for actuarial use. *North American Actuarial Journal*, **8**(4), 37–61.
- COLLINS, P.J., LAM, H. and STAMPFLI, J. (2015) How risky is your retirement income risk model? *Financial Services Review*, **24**(3), 193–216.
- COX, J., INGERSOLL, J. and ROSS, S. (1985) A theory of the term structure of interest rates. *Econometrica*, **53**(2), 385–408.
- CUI, Z., FENG, R. and MACKAY, A. (2017) Variable annuities with VIX-linked fee structure under a Heston-type stochastic volatility model. *North American Actuarial Journal*, **21**(3), 458–483.
- DUFFIE, D., PAN, J. and SINGLETON, K. (2000) Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, **68**(6), 1343–1376.
- ENGLE, R., ROUSSELLET, G. and SIRIWARDANE, E. (2017) Scenario generation for long run interest rate risk assessment. *Journal of Econometrics*, **201**(2), 333–347.
- ENGLE, R.F. and NG, V.K. (1993) Measuring and testing the impact of news on volatility. *Journal of Finance* **48**(5), 1749–1778.
- GEOGHEGAN, T., CLARKSON, R., FELDMAN, K., GREEN, S., KITTS, A., LAVECKY, J., ROSS, F., SMITH, W. and TOUTOUNCHI, A. (1992) Report on the Wilkie stochastic investment model. *Journal of the Institute of Actuaries*, **119**(2), 173–228.
- GNEITING, T. and RAFTERY, A.E. (2005) Weather forecasting with ensemble methods. *Science*, **310**(5746), 248–249.
- GRAEFE, A., KÜCHENHOFF, H., STIERLE, V. and RIEDL, B. (2015) Limitations of ensemble Bayesian model averaging for forecasting social science problems. *International Journal of Forecasting*, **31**(3), 943–951.
- GREEN, P.J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**(4), 711–732.
- HAARIO, H., SAKSMAN, E. and TAMMINEN, J. (2001) An adaptive Metropolis algorithm. *Bernoulli*, **7**(2), 223–242.
- HAMILTON, J.D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**(2), 357–384.
- HARDY, M.R. (2001) A regime-switching model of long-term stock returns. *North American Actuarial Journal*, **5**(2), 41–53.
- HARTMAN, B.M., RICHARDSON, R. and BATEMAN, R. (2017) Parameter uncertainty. Technical report, Casualty Actuarial Society, Canadian Institute of Actuaries, and Society of Actuaries Report.
- HESTON, S. (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, **6**(2), 327.
- HIBBERT, J., MOWBRAY, P. and TURNBULL, C. (2001) A stochastic asset model and calibration for long-term financial planning. Technical report, Barrie & Hibbert Limited.
- HOETING, J.A., MADIGAN, D., RAFTERY, A.E. and VOLINSKY, C.T. (1999) Bayesian model averaging: A tutorial. *Statistical Science*, **14**(4), 382–401.
- HUBER, P. (1997) A review of Wilkie's stochastic asset model. *British Actuarial Journal*, **3**(1), 181–210.
- IOANNIDIS, C. and KONTONIKAS, A. (2008) The impact of monetary policy on stock prices. *Journal of Policy Modeling*, **30**(1), 33–53.
- KILIAN, L. and MANGANELLI, S. (2007) Quantifying the risk of deflation. *Journal of Money, Credit and Banking*, **39**(2–3), 561–590.
- LEWANDOWSKI, D., KUROWICKA, D. and JOE, H. (2009) Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, **100**(9), 1989–2001.
- LITTERMAN, R. and SCHEINKMAN, J. (1991) Common factors affecting bond returns. *Journal of Fixed Income*, **1**(1), 54–61.
- MISHKIN, F.S. (1992) Is the Fisher effect for real?: A reexamination of the relationship between inflation and interest rates. *Journal of Monetary Economics*, **30**(2), 195–215.
- PEDERSEN, H., CAMPBELL, M.P., CHRISTIANSEN, S.L., COX, S.H., FINN, D., GRIFFIN, K., HOOKER, N., LIGHTWOOD, M., SONLIN, S.M. and SUCHAR, C. (2016) Economic scenario generators—A practical guide. Technical report, Society of Actuaries.

- RENDLEMAN, R.J. and BARTTER, B.J. (1980) The pricing of options on debt securities. *Journal of Financial and Quantitative Analysis*, **15**(1), 11–24.
- RENNE, J.-P. (2017) A model of the Euro-area yield curve with discrete policy rates. *Studies in Nonlinear Dynamics & Econometrics*, **21**(1), 99–116.
- ŞAHİN, Ş., CAIRNS, A. and KLEINOW, T. (2008) Revisiting the Wilkie investment model. *18th International AFIR Colloquium, Rome*, pp. 1–24.
- SIMS, C.A. (1980) Macroeconomics and reality. *Econometrica*, **48**(1), 1–48.
- SNEDDON, T., ZHU, Z. and O'HARE, C. (2016) Modelling defined contribution retirement outcomes: A stochastic approach using Australia as a case study. *Australian Journal of Actuarial Practice*, **4**, 5–19.
- UPTON, G. and COOK, I. (2014) *A Dictionary of Statistics*. Oxford, UK: Oxford University Press.
- VASICEK, O. (1977) An equilibrium characterization of the term structure. *Journal of Financial Economics*, **5**(2), 177–188.
- WHITTEN, S. and THOMAS, R.G. (1999) A non-linear stochastic asset model for actuarial use. *British Actuarial Journal*, **5**(5), 919–953.
- WILKIE, A.D. (1986) A stochastic investment model for actuarial use. *Transactions of the Faculty of Actuaries*, **39**, 341–403.
- WILKIE, A.D. (1995) More on a stochastic asset model for actuarial use. *British Actuarial Journal*, **1**(5), 777–964.
- ZHANG, S., HARDY, M. and SAUNDERS, D. (2018) Updating Wilkie's economic scenario generator for US applications. *North American Actuarial Journal*, **22**(4), 600–622.

JEAN-FRANÇOIS BÉGIN (Corresponding author)

Department of Statistics and Actuarial Science

Simon Fraser University

8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada

E-Mail: jbegins@sfu.ca