Estimating Historical Yield Curves With Sparse Data*

Jonathan Payne† Bálint Szőke‡ George Hall§ Thomas J. Sargent¶

July 3, 2023

Abstract

Estimating 19th century US federal bond yield curves involves challenges because few bonds were traded, bonds had peculiar features, government policies changed often, and there were wars. This paper compares statistical approaches for confronting these difficulties and shows that a dynamic Nelson-Siegel model with stochastic volatility and bond-specific pricing errors does a good job for historical US bond prices. This model is flexible enough to interpolate data across periods in a time-varying way without over-fitting. We exploit new computational techniques to deploy our model and estimate yield curves for US federal debt from 1790-1933.

JEL CLASSIFICATION: E31, E43, G12, N21, N41

KEY WORDS: Big data, yield curve, gold standard, government debt, Hamilton Monte Carlo, pricing errors, specification analysis

*We thank Clemens Lehner for outstanding research assistance and Refet Gürkaynak, Lars Peter Hansen, and Min Wei for suggestions. The views expressed here are those of the authors and do not necessarily represent the views of the Federal Reserve Board or its staff.

†Princeton University, Department of Economics. Email: jepayne@princeton.edu

‡Federal Reserve Board, Division of Monetary Affairs. Email: balint.szoke@frb.gov

§Brandeis University, Department of Economics. Email: ghall@brandeis.edu

¶New York University, Department of Economics and Hoover Institution, Stanford University. Email: thomas.sargent@nyu.edu

1 Introduction

Historical yield curves contain information about financial implications of political, institutional, and technological changes. However, estimating yield curves using historical bond data requires confronting several challenges: few bonds were issued, some bonds had discretionary features, and macroeconomic data are unreliable. For a new data set containing prices, quantities, and descriptions of all securities issued by the US Treasury between 1776 and 1960, this paper uses modern sampling techniques to estimate and compare different approaches for inferring term structures of yields on US federal bonds. An information criterion selects a non-linear state space model with drifting parameters and stochastic volatility as a parsimonious model with enough sufficient flexibility to pool data across periods in a time-dependent way. Exploiting new computational techniques to handle non-linear models with time varying parameters lets us work with very long time series that span different institutional arrangements. We thereby build bridges between macroeconomics and US economic history.

A first challenge is that our data set is sparse along the cross-section dimension. We tackle this problem by adopting a time-varying version of a statistical model proposed by Nelson and Siegel (1987). Economists at policy institutions use a similar parameterization, but in inferring a yield curve from observed prices and quantities they face a different challenge than we do. Because they have a superabundance of *cross-section* data on prices and quantities at each date, they solve an *overdetermined* inference problem. Our data are too sparse along the cross-section dimension to allow us to use even a *just-identified* version of the commonly used procedure. To confront this data deficiency, we enlist a "prejudice" or "induction bias" in the form of a parameterized statistical model of a panel having scattered missing observations. The data and statistical model tell us how much *smoothing* across time to do.

A second challenge is that 19th century US federal bonds often gave lenders and the Treasury discretion over maturity dates, conversions, and other features. Our inference procedure assumes that agents priced bonds under perfect foresight about those discretionary contract features. To prevent such assumptions from influencing our inferences too much, we introduce bond-specific idiosyncratic pricing errors. This decreases the influence of peculiar bonds on our yield estimates while alerting us to situations when our assumptions prevent our pricing formulas from consistently pricing our *cross-section* of bonds.

A third challenge is that 19th century macroeconomic data are unreliable. This prevents us from directly estimating a stochastic discount factor process that prices macroeconomic risks, especially at high frequencies. For this reason, we adopt a flexible approach that specifies a general discount function process with a law-of-one price restriction across maturities for each date, but that does not explicitly impose the absence of arbitrage. Our specification captures various models ranging from affine asset pricing models to preferred habitat models; but using it restricts us to estimating yield curves that bundle haircut risk and convenience premia into a single time-varying pricing kernel.

Another challenge is to infer posterior distributions for parameters of a complicated non-linear statistical model without relying on the particle filter or Gibbs sampling. We approximate posterior probabilities by deploying Hamiltonian Monte Carlo and No U-Turn sampling (HMC-NUTS). Our data set has many peculiarities—such as changing numbers of observed assets, bonds that have payoff streams of varying lengths, missing price observations, and relevant sets of bond-specific pricing errors changing over time in complicated ways—that prevent us from applying a "standard" `Stan` toolkit and force us to code our log posterior functions from scratch. Our application of the DynamicHMC.jl package by Papp et al. (2021) can be used for other economic models with tractable likelihood functions that do not easily fit into the `Stan` framework.

Related Work Our work is related to Svensson (1995), Dahlquist and Svensson (1996), Cecchetti (1988), Annaert et al. (2013), Andreasen et al. (2019), Diebold and Li (2006) and Diebold et al. (2008) who, like Gürkaynak et al. (2007) and ourselves, implement versions of the parametric yield curve model of Nelson and Siegel (1987).

Computing posterior distributions implied by our data and our statistical model is a formidable task that we accomplish by using the HMC-NUTS algorithm of Hoffman and Gelman (2014) and Betancourt (2018). While this estimator has been used extensively in statistics, economic applications are scarce. Prominent exceptions are Bouscasse et al. (2021), who use it to study the evolution of productivity in England from 1250 to 1870 and Farkas and Tatár (2021), who estimate DSGE models with ill-behaved posterior densities.

Outline: Section 2 describes the dataset. Section 3 describes and compares a range of statistical models. Section 4 conducts a "laboratory" experiment as a robustness check for our procedure. Section 5 discusses the model fit. Section 6 concludes.

2 Data Set and Its Limitations

2.1 Data Description

We have assembled prices, quantities, and descriptions of all securities issued by the US Treasury between 1776 and 1960. We combined existing historical databases with transcription from the digital archives of newspapers and government reports. Table 3 summarizes the different data sources we have used in constructing the dataset. The data set for bond prices and quantities is available on Github repository[1] and construction methods are explained in Hall et al. (2018).Our bond price data are monthly. When available, we use the closing price at the end of each month. However, if a closing price is not available, then we use an average of high and low prices or an

[1]Our data are posted at `https://github.com/jepayne/US-Federal-Debt-Public`. Only data from publicly available data sets are posted on the GitHub page.

average of bid and ask prices. The quantity data are quarterly from 1776 to 1871 and monthly thereafter. All quantity entries record the quantity outstanding on the last business day of the period.

In order to estimate yield curves, we need to construct the currency flows promised by each bond. For many of the early bonds in the sample, both the coupon dates and the maturity date are ambiguous because we lack detailed information on each bond issue and because it is unclear whether newspaper prices are ex or cum dividend. For the coupon dates, we used the following rule. If Bayley (1882) lists exact coupon dates, then we use those dates. Otherwise, we identify the coupon dates from cyclical decreases in the price series at the frequency of coupon payment. For the maturity dates, we used the following rules. For bonds with explicit maturity dates, we set the maturity to that date. Otherwise, we impose that investors had perfect foresight about the early redemption and set the maturity date to be the date at which greater than 90% of the outstanding bonds had been redeemed.

From 1862-1878, two currencies circulated: gold coins and non-convertible "greenback" dollars. In this paper, we restrict attention to gold denominated bonds and exclude all bonds of other denominations. We extend our approach to estimate greenback denominated yield curves and exchange rate expectations in Payne et al. (2023).

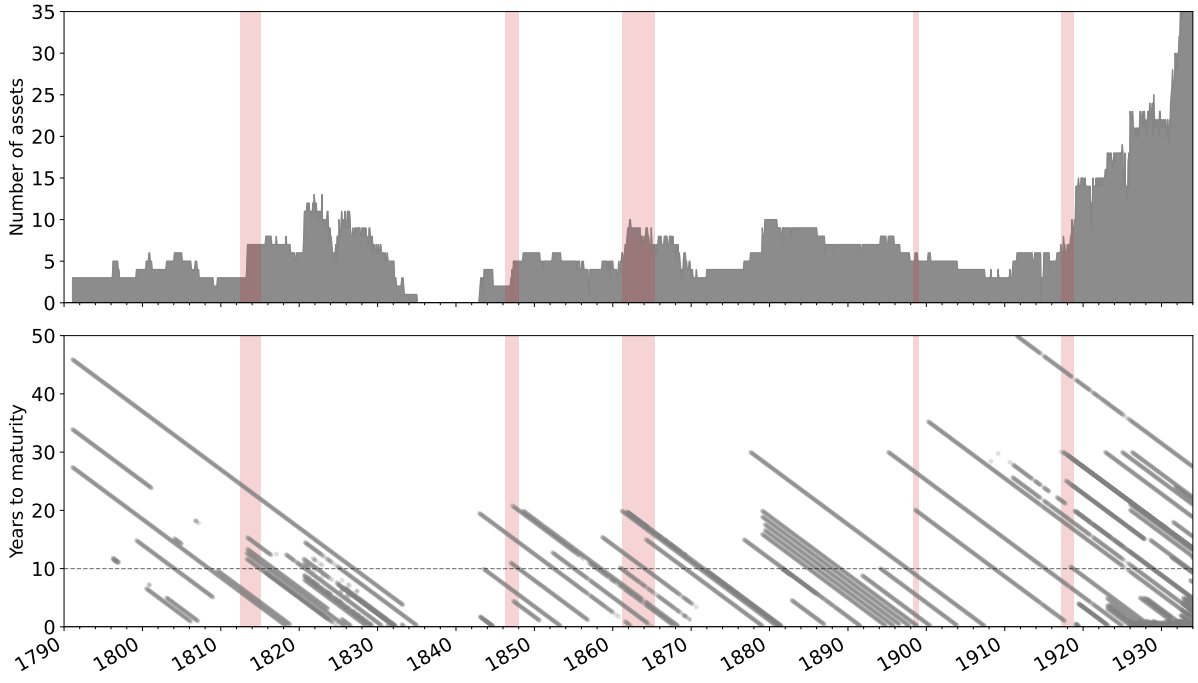

Figure 1: Our Dataset

The top panel depicts the number of securities with observed prices each month. The bottom panel depicts maturities (in years) of observed securities. Darker lines indicate overlapping securities. Red bars correspond to wars.

2.2 Inference Challenges

Skilled researchers have estimated yield curves on US federal debt for the post-WW2 period, by which time federal debts had become standardized and government bonds for sale had become plentiful. We estimate yield curves starting in 1791 and so have to confront challenges posed by peculiar structures of US federal bond markets before 1920. This requires us to address the following questions.

Q1. How should we handle periods with sparse bond data? Figure 1 depicts a monthly time series for the number of securities with observed prices and times to maturity of *all* outstanding bonds. Often there were fewer than five price observations at a given date and no price observations in the late 1830s when the federal government had no outstanding debt. This means that while we have "big data," our unbalanced sample prevents us from applying commonly used techniques from the yield curve estimation literature. Instead, we must posit a statistical model that lets us learn about yields at all dates simultaneously by pooling information across time periods.

Q2. How should we handle peculiar bonds? Throughout our sample, many US Treasury securities had special features such as indefinite maturities associated with call or conversion options. We start by *ex post* imputing perfect foresight about call dates and other discretionary components of the contracts. We then look for bond-specific pricing errors and refine these assumptions.

Q3. How should we handle haircut risk and convenience yields? There are many reasons to think that different maturities of US federal debt carried different haircut risks and "convenience" (or "liquidity") yields at different times during the 19th century. We address this by packaging haircut risk and convenience benefits into a single time varying pricing kernel that imposes that haircut risk and convenience benefits can vary across maturities but not across bonds. This allows us to estimate the prices of risky government promises.

Q4. How should we handle periods that provide sparse or inaccurate macroeconomic data? In principle, we could attempt to use historical macroeconomic data to estimate a model of the stochastic discount factor that prices macroeconomic risks. However, we are skeptical about the quality of 19th century macroeconomic data, especially at high frequencies. For this reason, we estimate a model that doesn't directly specify a pricing kernel process.

3 Statistical Models

We consider the observed pattern in Figure 1 as representative of a prototypical historical sample: sparse cross-sectional coverage at some periods with some maturities observed continuously for a long time. Positing that the yield curves at consecutive dates are correlated with each other allows us to fill the gaps we do not know in the cross-section with information we do know in the time dimension. To this end, we implement Dynamic Nelson-Siegel (DNS) and Dynamics

Nelson-Siegel-Svensson (DNSS) models as in Diebold and Rudebusch (2013). These models have two characteristics that are important for us: (1) tight parametric structures in the cross-section that come from the parsimonious 3-factor Nelson and Siegel (1987) or 4-factor Svensson (1994) yield curve parameterizations, (2) utilization of information in the time dimension. While the original motivation behind the DNS and DNSS models was to provide good yield curve forecasts, we are primarily interested in their attractive information pooling properties.

In a departure from the literature, we add *bond-specific pricing errors* to the observation equations of these models. We use this device because historical bonds, unlike Treasuries issued after the 1920s, were not standardized, so the usual procedure of homogenizing the sample before estimation is impractical. In subsections 3.1-3.4 we discuss a set of plausible assumptions that provide alternative ways of parameterizing and pooling information about the yield curve. In subsection 3.5 we evaluate these models based on their predictive accuracy.

3.1 Tight parameterization across maturities

Suppose that at time t we observe prices on an integer number M_t of *coupon-bearing* government bonds. A given bond, i, promises a sequence of gold dollar coupon and principal payments $\overline{\mathsf{m}}_t^{(i)} := \{\overline{m}_{t+j}^{(i)}\}_{j=1}^{\infty}$. We let $p_t^{(i)}$ denote the price of such a coupon-bearing gold dollar bond in terms of gold. Let $q_t^{(j)}$ denote the gold price of a government promise to one gold dollar at time $t+j$. We call the sequence $\mathsf{q}_t := \{q_t^{(j)}\}_{j=0}^{\infty}$ a *discount function*.

As is standard in the yield curve estimation literature, we start by assuming that the law of one price holds for a more or less homogeneous set of bonds. For each $t \geq 0$ there exists a discount function q_t such that

$$p_t^{(i)} = \sum_{j=1}^{\infty} q_t^{(j)} \overline{m}_{t+j}^{(i)} = \left\langle \mathsf{q}_t,\ \overline{\mathsf{m}}_t^{(i)} \right\rangle.$$

This is a key identifying restriction: within each time period, there is a common discount function that prices all bonds in our sample, i.e., there is no cross-sectional variation in how promises of bond repayment are priced. Note that q_t implicitly includes compensations for haircut risks, convenience benefits or inflation risks so it should be thought of as the price of a *risky* promise. Our specification allows these components to vary with the maturity j and time t, just not by individual bond.

We parameterize the discount function q_t by parameterizing the corresponding j-period zero-coupon yields defined as $y_t^{(j)} := -\log q_t^{(j)}/j$. We use a parametric family first proposed by Nelson and Siegel (1987). As Diebold and Li (2006) argued, this family is flexible enough to generate "typical yield curve shapes" (e.g., monotonic, humped, and S-shaped curves). To us, a particularly attractive feature of this family is that it is compatible with estimates of recent yield curves.[2]

[2]For example, Gürkaynak et al. (2007) use this form for the period 1961-1980. After 1980, they use an extension

(I). Nelson and Siegel (1987): The j-period gold dollar zero-coupon yield is

$$y_t^{(j)} = L_t + S_t\left(\frac{1-\exp(-j\tau)}{j\tau}\right) + C_t\left(\frac{1-\exp(-j\tau)}{j\tau} - \exp(-j\tau)\right)$$

where L_t, S_t, and C_t are hidden factors that characterize the level, slope, and curvature of the yield curve at time t and τ is a fixed parameter that identifies the location of a potential hump in the forward yield curve.

(II). Svensson (1994): The j-period gold dollar zero-coupon yield is

$$\begin{aligned} y_t^{(j)} = L_t + S_t\left(\frac{1-\exp(-j\tau_1)}{j\tau_1}\right) + C_{1,t}\left(\frac{1-\exp(-j\tau_1)}{j\tau_1} - \exp(-j\tau_1)\right) \\ + C_{2,t}\left(\frac{1-\exp(-j\tau_2)}{j\tau_2} - \exp(-j\tau_2)\right) \end{aligned}$$

where L_t, S_t, $C_{1,t}$, and $C_{2,t}$ are hidden factors that characterize the level, slope, first curvature, and second curvature of the yield curve at time t and τ_1 and τ_2 are fixed parameters that identify locations of the first and second hump in the forward yield curve, respectively.

Introducing a low dimensional parameterization of the yield curve in the maturity dimension enables us to handle periods in which few bonds were traded. An alternative low dimensional characterization of the yield curve would be a macroeconomic factor model, but 19th century macroeconomic data are not reliable enough for this.

3.2 Bond specific measurement errors

Researchers estimating the modern yield curve typically undertake a pre-selection exercise to restrict their sample to a collection of bonds with relatively homogeneous characteristics. Because our sample is sparse in the cross section, we cannot do that. Instead, we start by including our full sample of bonds and introducing *bond specific* measurement errors, as described in Assumption 1.

Assumption 1. Each bond i has a pricing error that is statistically independent from errors on other bonds and has a time-invariant Gaussian distribution with mean 0 and standard deviation $\sigma_m^{(i)}$. The observation equation becomes:

$$\tilde{p}_t^{(i)} = \left\langle \mathsf{q}(\lambda_t, \tau),\ \overline{\mathsf{m}}_t^{(i)} \right\rangle + d_t^{(i)} \sigma_m^{(i)} \varepsilon_t^{(i)}$$

where $\tilde{p}_t^{(i)}$ denote the *observed* period-t price of bond i in terms of gold and $d_t^{(i)}$ is the Macaulay duration of bond i in period t.

proposed by Svensson (1994) to allow for a second hump in the yield curve.

Introducing these measurement errors serves two purposes. First, it lets our model decrease the influences of peculiar bonds on our yield estimates. This means that bonds that violate our assumption that all bonds can be priced with a common discount function are given less weight in interpolating across maturities. Second, bond specific measurement errors inform us about situations in which our collection of assumptions prevents us from consistently pricing a *cross-section* of bonds. Starting from a presumption that all bonds can be priced with a common discount function, we look for patterns in estimated pricing errors, the idea being that misjudgments in our bond classification will show up as large, cluster-specific relative pricing errors.

3.3 Flexible parameterization across time

Because prior to World War I, price data are sparse and coverage varies over time, we use a multilevel (a.k.a. an hierarchical) statistical model to efficiently pool information over time. Let λ_t be a collection of time-varying yield curve factors: (L_t, S_t, C_t) for the Nelson-Siegel yield curve, $(L_t, S_t, C_{1,t}, C_{2,t})$ for the Svensson yield curve. We assume that vector λ_t follows the flexible stochastic process described in Assumption 2. This is in contrast with Gürkaynak et al. (2007) who—for the years after 1960—estimate yield curves period-by-period, assuming no intertemporal dependence among the elements of λ_t. But, it is similar in spirit to Diebold and Li (2006) who introduce mean-reverting factor dynamics to evaluate the Nelson-Siegel model's forecasting ability.

Assumption 2. Parameter τ is time-invariant. Parameter vector λ_t follows:

$$\lambda_{t+1} = \bar{\lambda}_t + \varrho(\lambda_t - \bar{\lambda}_t) + \Sigma_t^{\frac{1}{2}} \varepsilon_{\lambda,t+1}$$

where Σ_t is a covariance matrix with $\Sigma_t = \Xi_t \Omega \Xi_t$, Ω is the time-invariant correlation matrix and Ξ_t is a diagonal matrix containing marginal standard deviations σ_t that follow:

$$\log \sigma_{t+1} = \log \sigma_t + \Xi_\sigma \varepsilon_{\sigma,t+1}$$

where Ξ_σ is a positive definite diagonal matrix. In addition

$$\bar{\lambda}_{t+1} = \begin{cases} \bar{\lambda}_t + \overline{\Xi} \varepsilon_{\bar{\lambda},t+1} & \text{if } \; t = k\Delta \quad \text{for } k \in \mathbb{N} \\ \bar{\lambda}_t & \text{otherwise} \end{cases}$$

where $\overline{\Xi}$ is a positive definite diagonal matrix and $\Delta \geq 1$ is the frequency at which $\bar{\lambda}_t$ updates. Shocks $\varepsilon_{\lambda,t}$, $\varepsilon_{\sigma,t}$, and $\varepsilon_{\bar{\lambda},t}$ are Standard Normal for $t \geq 1$.

Four features of this model characterize how information is pooled across time:

(i) Parameter matrix Σ_t governs how evidence about a yield curve at one date affects inferences about yield curves at other dates. The closer are two dates to each other, the more correlated are the associated yield curves, with Σ_t capturing what "close" means.[3] The limit $\Sigma \to 0$ corresponds to *complete pooling*: here the yield curve is assumed to be fixed over time so that each observation has equal influence with all other dates. Contrary situations in which $\Sigma \to \infty$ call for *no pooling*: so there is no relationship between adjacent parameter estimates and we end up using only period t information to estimate period t yield curve parameters as in Gürkaynak et al. (2007). By inferring Σ from the data, we learn how much pooling across time we should do to improve estimates in light of intertemporal imbalances in data availability. In our context, "stochastic volatility" means that the amount of pooling can be time varying throughout the sample.

(ii) We allow shocks to different components of λ_t to be correlated. This enables us to infer relatively precise estimates of short ends of yield curves throughout our sample period. Assuming that different parts of the yield curve follow correlated but time-invariant dynamics allows us to transmit what we learn about co-movements between short- and long-term yields from times when many maturities are outstanding (as in the second half of the 19th century) to times when data about short-term yields are scarce (as in the early 20th century).

(iii) Yield curve parameter processes are decomposed into permanent and temporary components: the vector $\bar{\lambda}_t$ denotes a slow moving "long-run mean" to which λ_t reverts, and the matrix ϱ governs the rate at which this mean reversion occurs. We refer to $\bar{\lambda}_t$ as a "low-frequency" component and $\lambda_t - \bar{\lambda}_t$ as a "temporary" component of λ_t. We impose this structure to allow for potential mean reversion in the yield curve without imposing a common mean across the entire period from 1791-1933.

(iv) The long-run mean $\bar{\lambda}_t$ follows a random walk with updates at frequency Δ. As $\Delta \to \infty$, the frequency of parameter updates goes to zero, providing a state-space model with time-invariant long-run mean $\bar{\lambda}$. Setting $\Delta > 1$ to low values is a compromise between identifying the long-run mean with high accuracy, on the one hand, and letting it move over time, on the other hand. In effect, we divide the period of interest into subperiods of equal length Δ and assume complete pooling within subperiods and partial pooling across subperiods. We set $\Delta = 24$ months as a compromise between identifying the long-run mean with high accuracy and letting it move over time.

[3] One might be inclined to call this procedure "stochastic smoothing" because consecutive λ_t vectors are linked by a sequence of random variables $\{\varepsilon_{\lambda,t}\}$. Alternatively, one could define a deterministic smoothing function that specifies the sequence $\{\lambda_t\}$ in terms of parameters λ_0 and Σ, mimicking frequently used averaging techniques such as a moving-average. Modeling the sequence $\{\lambda_t\}$ as a stochastic process allows our algorithm to deploy a much richer set of smoothing functions.

3.4 A Nonlinear State Space Model of Bond Prices

Putting everything together, we can write our generic nonlinear state space model as:

$$
\begin{aligned}
\tilde{p}_t^{(i)} &= \left\langle \mathsf{q}(\lambda_t, \tau),\ \overline{\mathsf{m}}_t^{(i)} \right\rangle + d_t^{(i)} \sigma_m^{(i)} \varepsilon_t^{(i)} && \text{bonds} \\
\lambda_{t+1} &= \bar{\lambda}_t + \varrho(\lambda_t - \bar{\lambda}_t) + \Sigma_t^{\frac{1}{2}} \varepsilon_{\lambda,t+1} && \text{yield curve parameters} \\
\log \sigma_{t+1} &= \log \sigma_t + \Xi_\sigma \varepsilon_{\sigma,t+1} && \text{stochastic volatility} \\
\bar{\lambda}_{t+1} &= \begin{cases} \bar{\lambda}_t + \overline{\Xi} \varepsilon_{\bar{\lambda},t+1}, & \text{if } \ t = k\Delta \quad \text{for } k \in \mathbb{N} \\ \bar{\lambda}_t & \text{otherwise} \end{cases} && \text{long-run mean} \\
\text{with} \quad & \varepsilon_t^{(i)} \sim \mathcal{N}(0,1) \ \ \forall i, \quad \varepsilon_{\lambda,t} \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_3) \quad \varepsilon_{\bar{\lambda},t} \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_3) \quad \varepsilon_{\sigma,t} \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_3), \forall t \geq 1
\end{aligned}
$$

where $\tilde{p}_t^{(i)}$ denotes the *observed* period-t price of bond i in terms of gold. We estimate versions of this state space model using Bayesian methods. In particular, we approximate posterior probabilities by deploying Hamiltonian Markov Chain and No U-Turn sampler (HMC-NUTS). The posterior distribution is constructed by adding up Gaussian log-likelihoods associated with the independent shocks and combining them with priors, as described in Appendix A. We specify weakly informative prior distributions for the model's hyper-parameters for the specific purpose of *regularizing* our estimator and facilitating smooth operation of the sampling algorithm.

3.5 Horse Race

Our subsection 3.4 state space model combines a collection of potential ways to manage the deficiencies of our historical data-set. In this subsection, we investigate which are important. We study the following models that combine different parameterizations of the discount function q with different assumptions on the way we pool information over time.

Name	Parameterization	λ-dynamics	stoch. vol.	correlated shocks
Model A	Nelson-Siegel	random walk	No	No
Model B	Nelson-Siegel	random walk	No	Yes
Model C	Nelson-Siegel	random walk	Yes	Yes
Model D	Nelson-Siegel	mean-reversion	Yes	Yes
Model S	Svensson	random walk	No	Yes

Table 1: Models For Comparison

Conditional on q, the observation equations of these different models identical. The models differ in their state equations, but all are special cases of the non-linear state space model in Section 3.4. The first four rows of Table 1 are ordered to be increasing in complexity. We do not

include the Svensson model with stochastic volatility or mean reversion because the complexity of this model appears prohibitively large relative to our limited data.

Comparison Approach: We want to compare these models based on their predictive performances, not on their fits. This prompts us to use various cross-validation and information criteria that are meant to approximate models' predictive accuracies. In what follows, we will use the Watanabe-Akaike Information Criterion (WAIC) proposed by Watanabe (2010), which provides an approximation of the out-of-sample deviance, and the Pareto-Smoothed Importance Sampling Cross-Validation (PSIS) of Vehtari et al. (2017), which provides an approximation of the model's cross-validation score.[4] Both of these criteria are *point-wise*, i.e., prediction is considered observation-by-observation, which means that they come with approximate standard errors.

Model	WAIC	s.e.	ΔWAIC	s.e.	PSIS	s.e.	ΔPSIS	s.e.
Model A	26156	233	5553	207	12922	109	2773	100
Model B	23107	225	2504	180	11417	125	1268	88
Model C	20602	289	0	-	10149	141	0	-
Model D	22282	258	1680	226	10997	111	847	108
Model S	20582	260	-20	196	10154	125	6	91

Table 2: Model Comparison

The first column provides the model label. The second and third columns depict the WAIC and associated standard error. The fourth and fifth columns depict the difference from the WAIC of Model C and its associated standard error. The sixth and seventh columns depict the PSIS and associated standard error. The eighth and ninth columns depict a difference from the minimum PSIS and its associated standard error. We highlighted in red the models that have the lowest WAIC and PSIS, within standard error.

Model selection: Table 2 compares performance of the models. Evidently, both criteria agree that model C (Nelson-Siegel with random walk λ-dynamics, stochastic volatility, and correlated shocks) and model S (Svensson model with random walk λ-dynamics, no stochastic volatility, and correlated shocks) are preferred to the other models. This indicates that adding stochastic volatility to the standard Dynamic Nelson-Siegel improves its predictive accuracy because it lets the degree of information pooling vary over time. In contrast, mean reverting parameters are not called for in light of the additional complexity they introduce.

However, the criteria are unable to discriminate between models C and S because they exhibit very similar predictive accuracy. We view the two models as capturing different features of

[4]These criteria are complementary to regularizing priors. Regularization reduces overfitting while predictive criteria measure it. See also Gelman et al. (2014).

historical data. The Dynamic Nelson-Siegel model *with* stochastic volatility provides flexibility in how to pool information over time, which turns out to be particularly important during wars. The Dynamic Nelson-Siegel-Svensson model *without* stochastic volatility brings more flexibility in fitting long maturities by allowing for a second hump that is largely independent of the short end of the yield curve. From observed yields-to-maturity, we do see evidence to suggest that there might be periods with a second hump at long maturities (greater than 15-20 years). However, the estimate of the Svensson model gives humps at approximately 1.4 and 7.2 years, where the first hump is activated during the War of 1812 and the Civil War. This makes us worry that the estimate of the Svensson model is overfitting short maturity bonds during wars because the model does not allow sufficient flexibility in the amount of information pooling during wars. An additional concern is that the second hump does not actually help to fit the long end of the yield curve, which is its purpose. For these reasons, we choose the Dynamic Nelson-Siegel model *with* stochastic volatility but *without* a second hump.

Fitting prices vs fitting yields: We estimate parameters directly from bond prices adjusted by the bond's duration as in Gürkaynak et al. (2007). An alternative approach in the literature is to minimize yield errors. The two approaches are conceptually equivalent but for the following reasons we found it more practical to minimize price errors. Since zero-coupon yields are not directly observed, minimizing yield errors involves first producing approximate zero-coupon yield observations. To resolve this issue, Diebold and Li (2006) and Diebold et al. (2006) use the approximate zero-coupon yields calculated by Fama and Bliss (1987) with their proprietary "bootstrap" method. In principle, we could extend this methodology back through our historical sample. However, we are concerned that the cross-sectional sparsity of our data-set will impair the procedure's accuracy. We see little benefit from introducing these complications since, unlike Diebold and Li (2006) and Diebold et al. (2006), we include stochastic volatility. That would make our model non-linear even if were to minimize yield errors. For these reasons we minimize price errors.

As a robustness check, in Appendix C, we minimize the difference between model implied and observed yields-to-maturities. We show that, as we had anticipated, results are broadly similar to our estimates and have somewhat higher (within 2.5 standard error) WAIC and PSIS than our chosen model.

4 Laboratory Experiment

The Nelson-Siegel parameterisation can capture a wide range of yield curve shapes. However, as was shown in Figure 1, we want to infer yield curve parameters from relatively few price observations, with most observed prices being for long term bonds. How can we recover short yields? To show how pooling information over time can help with this matter, we conduct

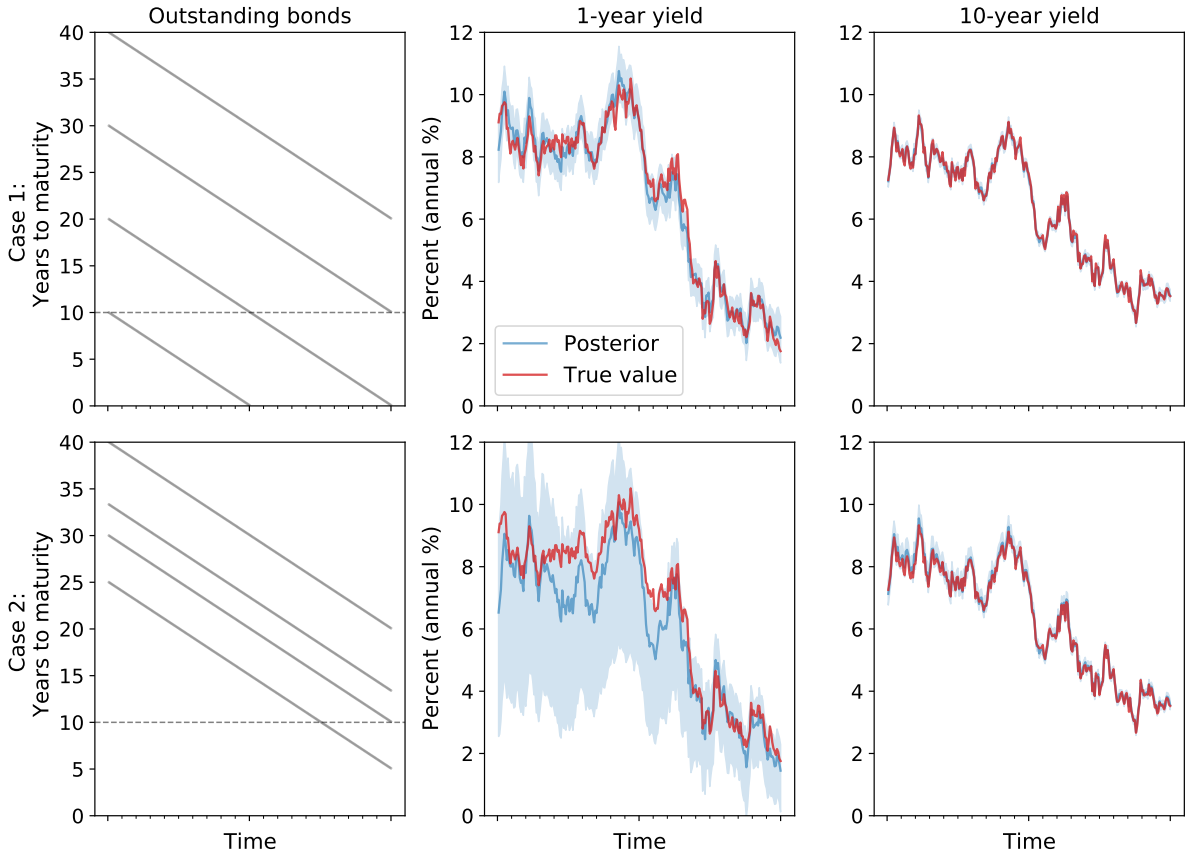

Figure 2: Comparison of Posteriors to True Values.

Artificial samples with 4 bonds ($T = 20$ year). **Case 1:** [**top row**] (i) 6 % (semi-annual), 10 year maturity, $\sigma_m^{(i)} = 3$; (ii) 3 % (semi-annual), 20 year maturity, $\sigma_m^{(i)} = 2$; (iii) 5 % (semi-annual), 30 year maturity, $\sigma_m^{(i)} = 1$; (iv) 2 % (semi-annual), 40 year maturity, $\sigma_m^{(i)} = 4$. **Case 2:** [**bottom row**] (i) 6 % (semi-annual), 25 year maturity, $\sigma_m^{(i)} = 3$; (ii) 3 % (semi-annual), 33 year maturity, $\sigma_m^{(i)} = 2$; (iii) 5 % (semi-annual), 30 year maturity, $\sigma_m^{(i)} = 1$; (iv) 2 % (semi-annual), 40 year maturity, $\sigma_m^{(i)} = 4$.

the following "laboratory experiment": taking a particular yield curve process (in line with our Subsection 3.4 state space model) as given, we use it to price four bonds with known characteristics (maturity, coupons, pricing error), then perform our econometric procedure, and compare our posterior yield estimates to the "true" values used to generate our artificial data. We investigate two situations:

Case 1: long term bonds with maturity dates that are distributed relatively evenly over the sample period

Case 2: there is an extended period without bonds that mature in less than 10 years

We create bonds that are "representative" of our sample in the sense that originally they are all

long term bonds. Here information about short yields must be recovered from prices of bonds that were originally long term but are now approaching maturity.

Rows of Figure 2 depict the outcomes of the two scenarios. The red lines are the true 1-year (middle column) and 10-year yields (right column) that were used to generate prices of the four bonds, the characteristics of which are depicted in the left column. The blue lines depict the posterior median and the shaded blue area depicts the 90% interquantile range of the posterior distribution. Even though we have few price observations for bonds with short maturity, the algorithm still does a good job of recovering the true 1-year yield under the first scenario (Case 1). Thus, at least when the common pricing kernel assumption is a good description of the data, observing a few long term bonds is sufficient to recover the short end of the yield curve as long as the maturity dates of the observed bonds are distributed relatively uniformly over time. This is what our model's ability to pool information buys us.

To illustrate this point, Case 2 represents a situation in which all four bonds mature after 20 years and shorter term securities are not issued in the meantime,[5] so our model has little chance to utilize information about short yields. The result is depicted in the bottom row of Figure 2. Evidently, the algorithm can still recover the true 10-year yield (it can observe bonds close to 10-years in the second half of the sample) but it has much more trouble trying to recover the 1-year yield. The posterior 90% interquantile range is large, and the posterior median departs significantly from the true value for many periods. This illustrates that the structure of our Nelson-Siegel parameterisation does not automatically generate tight posteriors. We do need some observations of prices for short maturity bonds to recover the yield curve.

5 Fits

To demonstrate why we believe that our chosen yield curve model provides a reasonable summary of the available bond price data, we now show that: (1) duration-weighted mean absolute price errors are generally small for all bonds that we include in the estimation of yield curves, (2) differences between observed and model-implied yields-to-maturities are small over time and across maturities, and (3) yields-to-maturity of observed bonds concentrate around our estimated par yield curves.

5.1 Small pricing errors across bonds

An important part of our approach is the assumption of bond-specific pricing errors. This allows the algorithm to decide whether particular bonds are likely to violate the common discount function assumption. The black crosses in Figure 3 depict *duration-weighted mean absolute pricing errors* for each bond included in the analysis. They are computed as time averages of

[5] This situation describes the last decade of the eighteenth century well, during which we observe only the three "Hamilton bonds."

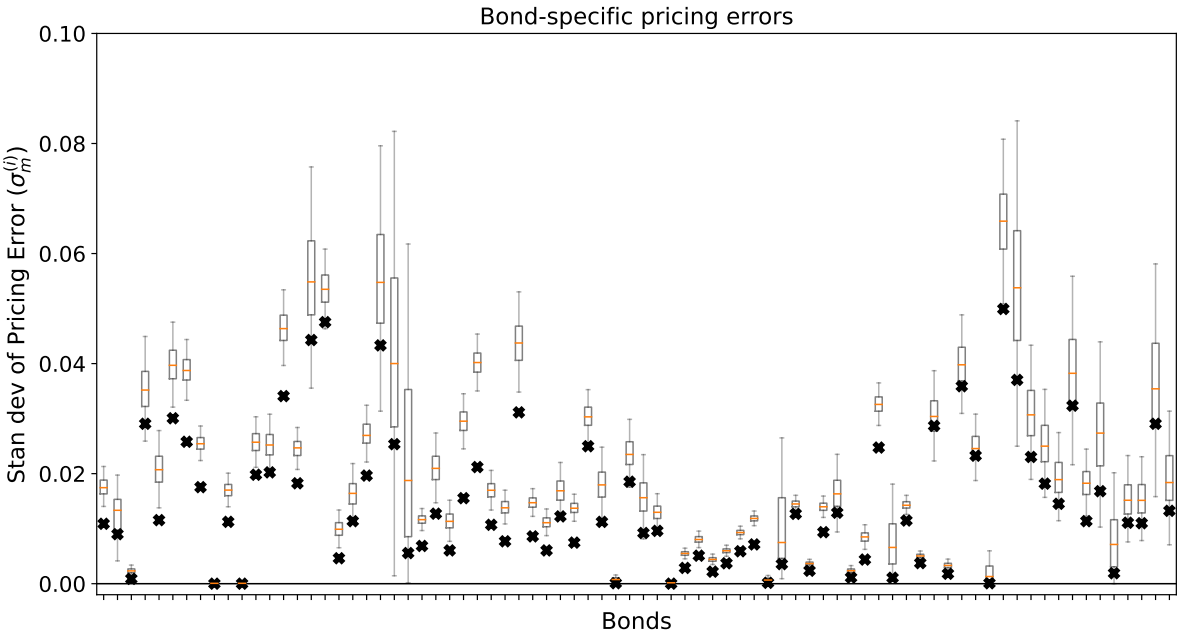

Figure 3: Posterior Distributions of Bond-Specific Pricing Error Variances

Each boxplot represents the interquartile range (IQR) and median (orange line) of the posterior distribution of $\sigma_m^{(i)}$—that is the standard deviation of the bond-specific pricing error—for each bond used in the estimation. Black crosses represent *mean absolute price errors* computed from the difference between observed and model-implied prices for each bond.

absolute differences between observed prices and posterior median price forecasts weighted by the inverse Macaulay duration. Evidently, our yield curves estimate prices of included bonds fairly well, with similar errors across different bonds. This indicates good in-sample fit and also that imposing a common discount function provides a good description of the gold dollar bonds with maturities larger than 1 year.

Similarly, estimated standard deviations of bond-specific pricing errors, $\sigma_m^{(i)}$, are also small. The boxplots in Figure 3 depict summary statistics of corresponding posterior distributions. The relative magnitudes of these estimates indicate how much particular bonds influence estimated yield curves. Our algorithm assigns relatively less weight to bonds with large estimated $\sigma_m^{(i)}$ values. Figure 3 shows that the set of bonds with relatively little influence more or less coincides with bonds having the highest duration-weighted mean absolute pricing errors.

5.2 Small yield errors over time and across maturities

Figure 4 depicts cross-sectional averages (over bonds for each month) of pricing errors, as measured by the absolute difference between observed prices and posterior median price forecasts. The largest errors are associated with the War of 1812, the Civil War, and the First World War.

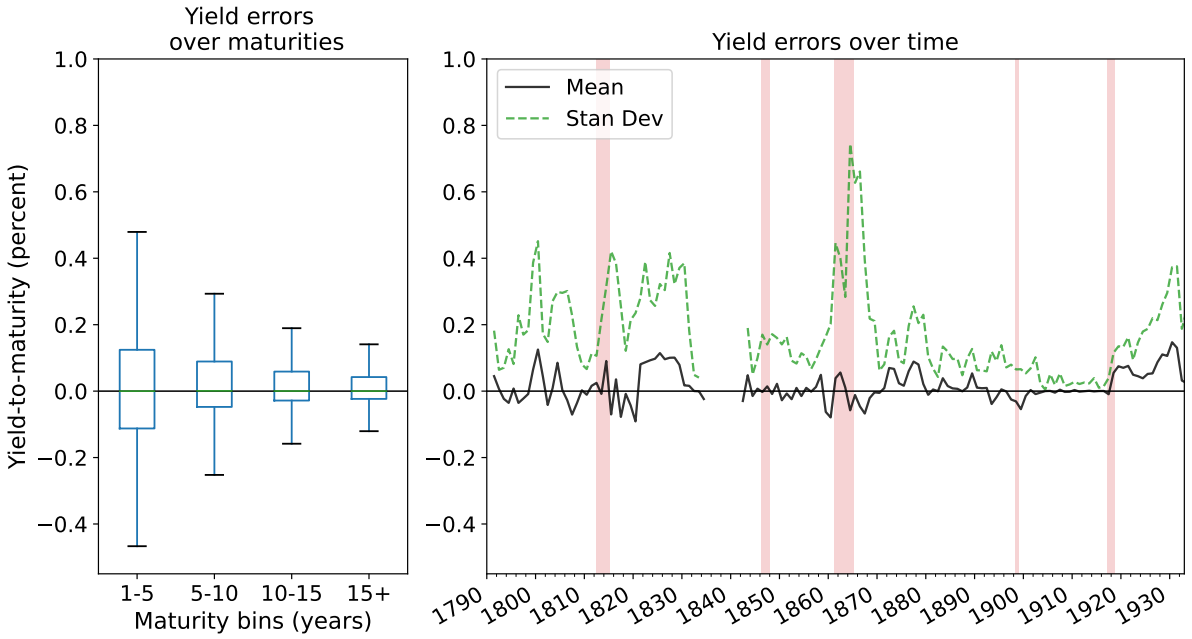

Figure 4: Mean Absolute Pricing Errors

The black line depicts the cross-sectional average (over bonds for each month) of the absolute difference between observed prices and posterior median price forecasts. The light gray intervals depict recessions as dated by Davis (2006) for the 1796-1914 period and NBER recessions thereafter. The light red intervals depict wars (from left to right: the War of 1812, the Mexican-American War, the Civil War, the Spanish-American War, and World War I).

This suggests that our model struggles to price cross-sections of bonds during wars.

To obtain a measure of fit with a more interpretable scale, we take the posterior median of our zero-coupon yield estimates, compute the implied yields-to-maturity for each bond at each month, and compare them to the observed yields-to-maturity. The panels of Figure 4 report different aspects of these *yield errors*. The left panel depicts distributions of yield errors for specific maturity bins. We see that on average our parametric yield curve specification fits observed yield-to-maturities well and without systematic differences across maturities larger than 1 year. The right panel depicts cross-sectional means and standard deviations (over bonds for each calendar year) of yield errors. The mean error stays close to zero and its variation is also typically small but becomes relatively large during the early 19th century and the Civil War, indicating that we have the most difficulty pricing the cross section of bonds during those years.

5.3 Observed yields-to-maturity are close to estimated par yield curves

Another argument favoring the plausibility of our estimated yield curves is that Congress and the Treasury often aimed to set coupon rates on new bonds so that initially they would sell

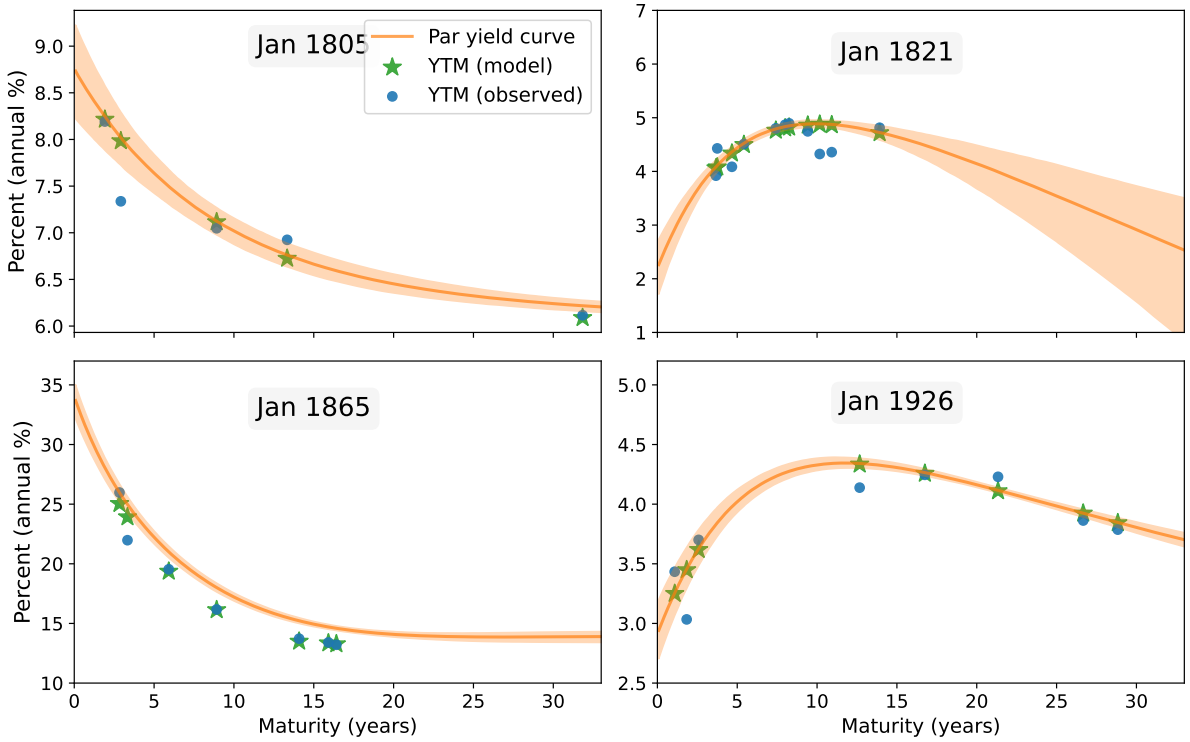

Figure 5: Par Yield Curve Estimates vs. Yield-to-Maturities.

The solid orange lines depict the median of our posterior for the gold dollar par yield curve at four specific dates (in gray boxes). The light orange bands around the posterior median depict the 95% interquantile ranges. Blue dots represent observed yield-to-maturities for bonds that are outstanding at the given period. Green stars depict model implied yield-to-maturities for the same bonds–computed from the posterior median price forecasts.

at par. That outcome would make their yields-to-maturities equal their coupon rates. This practice implies that we should expect observed yields-to-maturities to be close to a so called *par yield curve*: one that shows the required coupon rate for a bond with maturity j to sell at par. This object is a non-linear, one-to-one function of the zero-coupon yield curve, so we can use our estimated model to see how well observed yields-to-maturities line up with estimated par yield curves, at least in "non-emergency" periods when issuing new bonds at par was feasible.

The subplots of Figure 5 depict estimated par yield curves (orange lines) at dates that are more or less representative of some sub-periods of our sample. Observed and model-implied yields-to-maturities for the outstanding bonds are represented by blue dots and green stars, respectively. The close proximity of dots and stars indicates that the fit of our model is quite good across the whole maturity spectrum: our model is able to replicate a variety of yield curve shapes and succeeds in capturing that yields at the long end of the maturity spectrum are often lower than yields at medium horizons irrespective of how short-term yields behave.[6] Moreover,

[6]In other words, allowing for a "hump" in the yield curve is often necessary.

comparing the blue dots to the estimated par yield curves illustrates that Congress's objective to sell bonds at par was often achieved (for example, see the subplots for 1805, 1821 or 1926).

6 Concluding Remarks

We have compared applications of alternative statistical models of yield curves to an historical data set that is sparse in the cross section, has bonds with peculiar features, and covers periods with major policy changes and wars. Information and cross validation criteria suggest that a Dynamic Nelson Siegel model with stochastic volatility and correlated shocks is a good model. The model fits the data with small errors and performs well in a "laboratory experiment" in which we generate data designed to mimic our historical sample of US federal bond prices.

References

Andreasen, M. M., Christensen, J. H., and Rudebusch, G. D. (2019). Term structure analysis with big data: One-step estimation using bond prices. *Journal of Econometrics*, 212:26–46.

Annaert, J., Claes, A. G., Ceuster, M. J. D., and Zhang, H. (2013). Estimating the spot rate curve using the nelson–siegel model: A ridge regression approach. *International Review of Economics and Finance*, 27:482–496.

Bayley, R. A. (1882). *The National Loans of the United States, from July 4, 1776, to June 30, 1880*. Number 246. US Government Printing Office.

Betancourt, M. (2018). A conceptual intro to Hamiltonian Monte Carlo. Technical report.

Bouscasse, P., Nakamura, E., and Steinsson, J. (2021). When did growth begin? New estimates of productivity growth in England from 1250 to 1870. Technical report, National Bureau of Economic Research.

Cecchetti, S. G. (1988). The case of the negative nominal interest rates: New estimates of the term structure of interest rates during the great depression. *Journal of Political Economy*, 96(6):1111–1141.

Dahlquist, M. and Svensson, L. E. O. (1996). Estimating the term structure of interest rates for monetary policy analysis. *Scandinavian Journal of Economics*, 2:163–183.

Davis, J. H. (2006). An improved annual chronology of U.S. business cycles since the 1790s. *The Journal of Economic History*, 66(1):103–121.

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364.

Diebold, F. X., Li, C., and Yue, V. Z. (2008). Global yield curve dynamics and interactions: A dynamic Nelson-Siegel approach. *Journal of Econometrics*, 146(2):351–363.

Diebold, F. X. and Rudebusch, G. D. (2013). *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*. Princeton University Press, Princeton, New Jersey.

Diebold, F. X., Rudebusch, G. D., and Borağan Aruoba, S. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, 131(1):309–338.

Fama, E. F. and Bliss, R. R. (1987). The information in long-maturity forward rates. *The American Economic Review*, 77(4):680–692.

Farkas, M. and Tatár, B. (2021). Bayesian estimation of DSGE models with Hamiltonian Monte Carlo. Technical report, IMFS Working Paper Series.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and Computing*, 24(6):997–1016.

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The U.S. treasury yield curve: 1961 to present. *Journal of Monetary Economics*, 54(8):2291–2304.

Hall, G. J., Payne, J., Sargent, T. J., and Szőke, B. (2018). US federal debt 1776-1940: Prices and quantities. https://github.com/jepayne/US-Federal-Debt-Public.

Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(47):1593–1623.

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.

Martin, J. G. (1886). *Martin's Boston Stock Market: Eighty-eight Years, from January 1, 1798, to January, 1886.* The author.

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious modeling of yield curves. *The Journal of Business*, 60(4):473–489.

Officer, L. H. and Williamson, S. H. (2021). Explaining the measures of worth. *Measuring Worth.*

Papp, T. K., Aluthge, D., JackRab, Widmann, D., TagBot, J., and Piibeleht, M. (2021). tpapp/dynamichmc.jl: v3.1.0. Julia package version 3.1.0.

Payne, J., Szőke, B., Hall, G. J., and Sargent, T. J. (2023). Costs of financing US federal debt under a gold standard: 1791-1933. Technical report, Princeton University.

Razaghian, R. (2002). Financial credibility in the United States: The impact of institutions, 1789-1860. Data made available through the International Center for Finance at Yale University.

Svensson, L. E. (1994). Estimating and interpreting forward interest rates: Sweden 1992 - 1994. Working Paper 4871, National Bureau of Economic Research.

Svensson, L. E. O. (1995). Estimating forward interest rates with the extended Nelson and Siegel method. *Sveriges Riksbank Quarterly Review*, 3.

Sylla, R. E., Wilson, J., and Wright, R. E. (2006). Early U.S. security prices. *http://eh.net/databases/early-us-securities-prices.*

U.S. Department of the Treasury (1869-2015). *Monthly Statement of the Public Debt.* U.S. Treasury Department.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, 27:1413–1432.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(12).

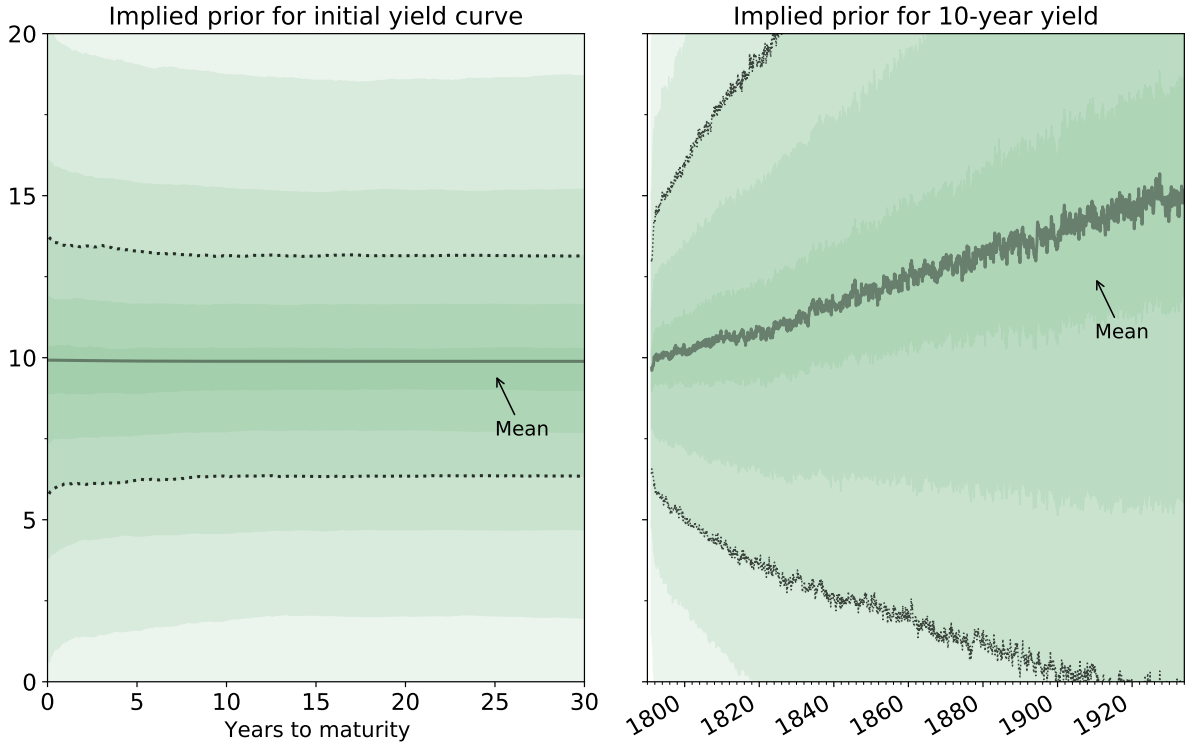

Figure 6: Implied Prior Distribution of the Initial Yield Curve and the 10-year Zero-Coupon Yield.

The solid grey lines depict the mean, and dotted lines depict the 25% and 75% percentiles of the prior distribution. Shaded areas represent interquantile ranges so dark areas are indicative of concentrated prior probability.

A Priors

Priors: Assumptions 2 give rise to a flexible model of the gold denominated yield curve process that is pinned down by a small set of hyper-parameters. Aa prior on τ and the initial (time 0) λ vector that effectively determines an "average yield curve" for the whole sample period. We use log-normal prior for τ and independent log-normal priors for the three entries of the initial λ vector that implies the prior distribution for the initial yield curve shown in the left panel of Figure 6. Our prior imposes a flat "average yield curve," i.e., for all maturities the prior mean is 10% with standard deviation of around 5%. Underlying priors are:

$$\lambda_{0,0} \sim \log\mathcal{N}\left(10-\underline{\beta},\ 6\right), \quad \lambda_{1,0} \sim \log\mathcal{N}\left(10-\underline{\beta},\ 6\right),$$
$$\lambda_{2,0} \sim \log\mathcal{N}\left(10-\underline{\beta},\ 15\right), \quad \tau \sim \log\mathcal{N}(60,\ 60).$$

While the "average yield curve" influences our posterior distribution in the early part of the

Table 3: Summary of Data Sources

Series	Period	Frequency	Source
Bond prices	1776-1839	M	Razaghian (2002), Sylla et al. (2006) and *Global Financial Data*.
	1840-1859	M	Razaghian (2002), *The New York Times*, and *Global Financial Data*
	1860-1925	M	*Commercial & Financial Chronicle*, *Global Financial Data*, *Merchant's Magazine*, *The New York Times*, *US Treasury Circulars*, and Martin (1886).
	1925-1960	M	*CRSP US Treasury Database*.
Quantities	1790-1871	Q	Bayley (1882).
	1872-1960	M	U.S. Department of the Treasury (2015).
Contract Info.	1790-1960		1790-1871 from Bayley (1882).
			1872-1960 from U.S. Department of the Treasury (2015).
Gold/Goods	1800-1860	M	Wholesale Price Index (Warren/Pearson)
Exchange Rate	1860-1913	M	U.S. Index of the General Price Level (NBER Macrohistory: Series NBER 04051)
	1913-2020	M	CPI (BLS)
GDP	1790-2020	A	Officer and Williamson (2021)
Gold/Greenbacks	1862-1878	M	Yale SOM ICF dataset
Exchange Rate			

[1] Repository for bond time series: `https://github.com/jepayne/US-Federal-Debt-Public`

sample, it is much less influential later due to the random walk component in λ_t. The right panel of Figure 6 illustrates how the prior mean and "prior coverage bands" for the 10-year yield grow over time. How much our prior for λ_0 affects the posterior distribution for later periods depends mainly on priors for $\{\bar{\lambda}_t\}$, ϱ, and $\{\Sigma_t\}$ that we specify as follows:

- For the correlation matrix Ω we use the LKJ prior with a concentration parameter $\eta = 5$, which is a unimodal but fairly vague distribution over the space of correlation matrices. For η values larger than 1, the LKJ density increasingly concentrates mass around the unit matrix, i.e., favoring less correlation.[7]

- For the initial standard deviations σ_0 we use independent log-normal priors: $\sigma_{i,0} \sim \log\mathcal{N}(0.05,\ 0.1)$.

- We use *common* exponential priors on the standard deviation in the diagonal of Ξ_σ, with the rate parameter tuned so that *a priori* the probability that $\sigma_\sigma^{(i)} > 0.15$ is lower than 5%. The prior mean is 0.05.

- We use independent normal priors on the entries of ϱ. The prior mean is chosen as a diagonal matrix with diagonal entries $[0.8, 0.8, 0.8]$ while we set the standard deviation to 0.3 for all 9 entries of ϱ.

- We use independent log-normal priors for the three entries of the initial $\overline{\lambda}_0$ (permanent component of λ):

$$\overline{\lambda}_{0,0} \sim \log\mathcal{N}\left(10 - \underline{\beta}, 6\right), \quad \overline{\lambda}_{1,0} \sim \log\mathcal{N}\left(10 - \underline{\beta}, 6\right), \quad \overline{\lambda}_{2,0} \sim \log\mathcal{N}\left(10 - \underline{\beta}, 15\right)$$

- We use *common* exponential priors on the standard deviation in the diagonal of $\overline{\Xi}$, with the rate parameter tuned so that *a priori* the probability that $\bar{\sigma}^{(i)} > 0.15$ is lower than 5%. The prior mean is 0.05.

We use *common* exponential priors on the standard deviation of pricing errors, $\sigma_m^{(i)}$, with the rate parameter tuned so that *a priori* the probability that $\sigma_m^{(i)} > 30$ is lower than 5%. Prior mean is 10.

B Estimation Details

Alternative to Particle Filtering: Estimating the model in Section 3.4 involves a complicated filtering problem due to the non-linear nature of bond prices and the existence of stochastic volatility. A standard approach to such non-linear filtering problems is to use some version of particle filtering. However, thanks to the length and other complexities of our data set, well-known drawbacks of particle filters, such as sample degeneracy and impoverishment, become

[7] See Lewandowski et al. (2009). The LKJ distribution is defined by $p(\Omega|\eta) \propto \det(\Omega)^{\eta-1}$. For $\eta = 1$, this is a uniform distribution.

particularly acute in our case. We deploy an alternative strategy and approach the problem as a high-dimensional statistical model that "treats latent variables as parameters."[8] From this viewpoint, the model has more than $7,500$ parameters. To cope with such a high-dimensional parameter space, we use Hamiltonian Monte Carlo with a "No-U-Turn Sampler" of Hoffman and Gelman (2014), along with subsequent developments described in Betancourt (2018). The basic idea is to use slope information about the log-likelihood to devise an efficient Markov Chain Monte Carlo sampler. This method can attain a nearly i.i.d. sample from the posterior by proposing moves to distant points in the parameter space through (an approximately) energy conserving simulated Hamiltonian dynamic.

Computational issues: While `Stan` might seem an obvious choice for the task at hand—it is a well-developed software that efficiently implements the HMC-NUTS sampler—non-trivial features of our data set make it inconvenient for our purposes. Some of our main technical difficulties are: (1) the number of observed assets changes over time, (2) each bond has a payoff stream of varying length, (3) there are many periods without price observations, (4) the set of bond-specific pricing errors that are relevant at a given period t changes over time in a complicated fashion, etc. To tackle these difficulties, we code the log posterior function of our model from scratch and feed it into the DynamicHMC.jl package by Papp et al. (2021) which is a robust implementation of the HMC-NUTS sampler mimicking many aspects of `Stan`. An important advantage of this package is that it allows the user to provide the Jacobian of the log-posterior manually. Not relying on automatic differentiation for a model with $7,500+$ parameters cuts running time by several orders of magnitude. In most cases, we use the recommended (default) tuning parameters for the NUTS algorithm.

C Minimizing Price vs Yield Errors

Model	WAIC	s.e.	ΔWAIC	s.e.	PSIS	s.e.	ΔPSIS	s.e.
Model C (price)	20602	289	0	-	10149	141	0	-
Model C (yield)	21053	271	451	213	10373	134	224	105

Table 4: Estimating Model C (Nelson-Siegel with random walk λ-dynamics, stochastic volatility, and correlated shocks) by minimizing duration-adjusted price errors (first row) vs by minimizing yields-to-maturities (second row).

[8] We use quotation marks because in the Bayesian paradigm there is no clear distinction between latent variables and parameters.