UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Cambridge Working Papers in Economics

Cambridge Working Papers in Economics: 1916

NONPARAMETRIC RECOVERY OF THE YIELD CURVE EVOLUTION FROM CROSS-SECTION AND TIME SERIES INFORMATION

Bonsoo Koo   Davide La Vecchia   Oliver Linton

26 February 2019

We develop estimation methodology for an additive nonparametric panel model that is suitable for capturing the pricing of coupon-paying government bonds followed over many time periods. We use our model to estimate the discount function and yield curve of nominally riskless government bonds. The novelty of our approach is the combination of two different techniques: cross-sectional nonparametric methods and kernel estimation for time varying dynamics in the time series context. The resulting estimator is able to capture the yield curve shapes and dynamics commonly observed in the fixed income markets. We establish the consistency, the rate of convergence, and the asymptotic normality of the proposed estimator. A Monte Carlo exercise illustrates the good performance of the method under different scenarios. We apply our methodology to the daily CRSP bond dataset, and compare with the popular Diebold and Li (2006) method.

Nonparametric Recovery of the Yield Curve Evolution from Cross-section and Time series Information*

Bonsoo Koo†
Monash University

Davide La Vecchia‡
University of Geneva

Oliver Linton§
Cambridge University

February 2019

Abstract

We develop estimation methodology for an additive nonparametric panel model that is suitable for capturing the pricing of coupon-paying government bonds followed over many time periods. We use our model to estimate the discount function and yield curve of nominally riskless government bonds. The novelty of our approach is the combination of two different techniques: cross-sectional nonparametric methods and kernel estimation for time varying dynamics in the time series context. The resulting estimator is able to capture the yield curve shapes and dynamics commonly observed in the fixed income markets. We establish the consistency, the rate of convergence, and the asymptotic normality of the proposed estimator. A Monte Carlo exercise illustrates the good performance of the method under different scenarios. We apply our methodology to the daily CRSP bond dataset, and compare with the popular Diebold and Li (2006) method.

Key words: nonparametric inference, panel data, time varying, yield curve dynamics,
Journal of Economic Literature Classification: C13, C14, C22, G12

---

*We thank Tom Wansbeek, two referees, and seminar and conference participants for helpful comments and discussions. All errors are ours.

†Corresponding author, Department of Econometrics and Business Statistics, Monash Business School, Monash University, PO Box 11E, Clayton Campus, VIC 3800, Australia; e-mail: bonsoo.koo@monash.edu

‡Research Center for Statistics and Geneva School of Economics and Management, University of Geneva, Blv. Pont D'Arve 40, CH-1211 Geneva, Switzerland; e-mail: Davide.LaVecchia@unige.ch

§Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD. Email: obl20@cam.ac.uk

1 Introduction

The yield curve, or the term structure of interest rates, describes how interest rates vary with maturity. The curve reveals the level of current interest rates, but it also carries critical information about the expectations of market participants regarding future movements of interest rates. It is one of the main factors driving financial and economic decisions, at both the private and public level. The yield curve is instrumental in monetary policy decisions, because it serves as an indicator of the future interest rate level formed by current market expectations. Furthermore, it can serve as a touchstone for financial market participants to identify any abnormality in newly issued and traded bond prices. We refer to Piazzesi (2005), Ang et al. (2006), Dai et al. (2007) and Björk (2004) for more detailed discussion. The statistical analysis of the yield curve lies at the heart of empirical finance.

Estimation and inference about the yield curve poses several issues. The yield curve is not directly observable in practice due to the lack of zero coupon bonds for time-to-maturities longer than one year; it needs to be estimated from the prices of traded coupon paying bonds. There are two general approaches to estimating the yield curve at a point in time: parametric and nonparametric. Nelson and Siegel (1987) propose a parametric model for the forward curve that is nonlinear in one parameter; estimation is based on nonlinear least squares for the implied discount function (the forward, yield, and discount functions are in one to one relation, see (Anderson et al., 1996, pp 12-13) compared with observed coupon bond prices. Svensson (1994) extends this model by adding an extra term. In the parametric case, standard asymptotic theory can be used to derive the limiting distribution of the estimator and to justify confidence intervals obtained from this. Following the seminal work of McCulloch (1971), an alternative approach is to use nonparametric curve fitting methods. McCulloch (1971, 1975) uses regression splines for the discount function, Chambers et al. (1984) use polynomials for the yield curve, Vasicek and Fong (1982) uses exponential splines for the discount function, and Fisher et al. (1995) use a general variable roughness penalty approach (VRP). Tanggaard (1997) uses kernel smoothing, which was extended in Linton et al. (2001). Both classes of methods are well represented in central bank toolkits: the nonparametric spline approach is used for example by the Bank of England and the Federal Reserve Board, whereas the ECB primarily uses Nelson-Siegel, see BIS (2005) and Nymand-Andersen et al. (2018). Diebold and Li (2006) and Härdle and Majer (2014) provide good discussions of the shortcomings of existing approaches.

There is an almost separate literature that focuses on time series modelling of yield curves under continuous time equilibrium models. The equilibrium approach derives the yield curve from the dynamic evolution of the *time series* of instantaneous spot rates and perhaps other state variables under the equilibrium condition. In the empirical literature usually a small dimensional vector of "pre-calculated" yields are taken as given. The fact that these are estimated from data is not usually acknowledged. In this literature, the spot rate modelling has been a focus of particular attention: Ait-Sahalia (1996) rejects the usual parametric framework and advocates a semiparametric model. Recently, there has been some work modelling the cross-sectional quantities of the yield curve evolving dynamically over time, utilizing both cross-sectional and time series data; see e.g., Diebold and Li (2006), Kargin and Onatski (2008), Christensen et al. (2011), Hays et al. (2012), Härdle and Majer (2014) and reference therein.

We focus on the panel data framework that combines both cross-sectional and time series information. In particular, we develop new econometric methodology to estimate time varying yield curves nonparametrically from discrete time (daily) panel data on bond prices and cash flows; we also provide the means to conduct inference about the yield curves. Our model incorporates the standard present discounted value assumption from the cross section of bonds at a point in time (it is static arbitrage free), and allows the discount function to change over time either deterministically or driven by a vector of observed covariates. We suppose that the observed prices are subject to some measurement error. This could be associated with market microstructure issues or with irrational behavioural errors; either way, this generates sampling error in our estimation procedures. We compute our estimator from quoted prices and cash-flow information of both zero-coupon and coupon bonds issued by a government at a series of discrete time points. Our model is nonparametric so that the point in time discount curve is not restricted by functional form, and the way the discount curve evolves over time or state is also not restricted, except that we require the drivers to be observed. In some sense the dynamic parametric models of Diebold and Li (2006) are a special case of our approach.

Our main goal is to derive a robust inferential procedure for the curve: starting from traded assets we estimate the yield curve time period-by-time period. We do not impose a particular functional form nor do we explicitly impose the dynamic no-arbitrage condition in the time dimension (even if this were possible in our general class it would require making a number of additional strong assumptions on the state variable dynamics for example, which we do not feel are justified). Our purpose is different from Ang and Piazzesi (2003) and Ang et al. (2006), whose aim is forecasting the yield curve inside the class of affine models. They explicitly enforce the dynamic no-arbitrage restriction inside their continuous time model. We do not make any such assumption about the continuous time evolution of the curve. Of course, ignoring this aspect might entail efficiency loss when the restriction is valid—e.g. when the underlying data generating process has a specific affine parametric form, when the stochastic discount factor is an affine exponential martingale and the time-varying risk-premium has an affine structure. Nevertheless, we emphasize that, from our stand point, the potential efficiency gain obtained imposing the modelling restriction needed to satisfy the no-arbitrage condition must be weighed against the possibility of inducing an estimation bias, due to model misspecification. Our model is very general in terms of functional form of the yield curve and unlike the models considered in the above-mentioned papers, our inferential procedure does not involve the affine structure. We flag that if the no-arbitrage condition is approximately satisfied in the analyzed data set- some have argued that this is likely the case for the U.S. Treasury bill market, as remarked by Diebold et al. (2006) (although the treasury flash rally episode of 20141015 surely suggests some caution on this) - it will also likely be approximately satisfied by our estimates, as our model is quite flexible and provides a good fit to the data.[1] It is an interesting topic for our future research how to impose the dynamic no arbitrage restrictions on our model, Cheng and Tehranchi (2018) provide some work in this direction. Our approach has an advantage in comparison with other methods in terms of residual analysis and model diagnostics; we illustrate this aspect in Section 6.4.

We use kernel-based nonparametric methodology to estimate the unknown discount function, generalizing Linton et al. (2001). We establish the properties of our procedure when both the cross-section and time series dimension are large. In particular, we establish consistency, the rate of convergence, and the asymptotic normality of our yield curve estimator. We also propose a statistical procedure to define confidence intervals, deriving a consistent estimator for the asymptotic covariance of our estimator. In our view, this is a key aspect of our estimation method: extant procedures have not clearly addressed how to construct confidence intervals for proposed estimators of the discount/yield curve and its dynamics; see, e.g. Diebold and Li (2006) and Yallup (2012). Indeed, in practice, our estimated yield curve can be applied to evaluate the *meaningful* impact of implemented monetary policies on the economy and identify any *significant* discrepancy between observed bond prices and the estimated prices so as to price new issues or establish and evaluate investors' trading strategies. Given that the yield curve needs to be estimated, such purposes can be accomplished in a fully informative way only when a theoretical device is available to quantify the uncertainty in the estimates. Confidence intervals provide helpful tools to achieve this goal.

The development of this theory encounters several challenges. We use a panel dataset of daily prices of bonds along with their nominal cashflows, and there seems to be no good reason to restrict the cross-sectional or time series dependence of the error process. Bonds traded in the same market are likely to experience common factors in their liquidity and hence pricing errors. Existing asymptotic results for yield curve estimation are available either for the (independent observations) cross-sectional set-up only, e.g. Linton et al. (2001) or for the time series set-up only, e.g. Koo and Linton (2012). We develop new asymptotic techniques that allow a general correlation structure over the cross-sectional units and over time in a unified framework.

The remainder of this paper is organized as follows. In Section 2 we introduce the model and the related framework along with our estimation procedure for the yield curve dynamics. Section 4 develops the large sample distribution theory. The Monte Carlo exercises reported in Section 5 provide numerical evidence that our nonparametric estimation method ensures the degree of flexibility required to match several time-changing shapes of the yield curve. Application results are reported in Section 6. Section 7 concludes. The mathematical proofs are provided in the supplementary Appendix.

[1] For a discussion on this point, we refer to Diebold et al. (2005), page 417. See also the report on the treasury flash rally, Treasury (2015).

2 The Model

Most bond markets do not possess zero coupon bonds for useful spans of maturities. The discount function must therefore be extracted from the prices of coupon bonds instead. Coupon bonds generate several payments at future dates, and in an efficient bond market, the present value of these future payments should, apart from a small error, be equal to the trading price. We observe a panel of Government bonds, in particular their prices p_{it} and their promised cashflows for $i = 1, \ldots, n_t$ and $t = 1, \ldots, T$. We consider the following discrete time semiparametric model for the bond prices:

$$p_{it} = \sum_{j=1}^{m_{it}} c_{it}(\tau_{ij}) d(\tau_{ij}, X_t) + \varepsilon_{it}, \tag{1}$$

where $c_{it}(\tau_{ij})$ are the cash flows at future payment dates (known at time t) attributable by holding the bond i to maturity, where payment dates are $\{\tau_{ij}\}_{j=1}^{m_{it}}$ with $\tau_{i1} < \cdots < \tau_{im_{it}}$ for each bond i. We treat the cash flows as nonstochastic; each bond may have a different expiry date and different cashflows. We suppose that $X_t \in \mathbb{R}^L$, with $L \ll n(= \min_t n_t)$, are some (possibly stochastic) covariates or factors that are observable at time t and are known to drive the evolution of the yield curve over time, while ε_{it} is a conditional (on X_t) mean zero pricing error associated with the bond i at time t.[2] The function d is unknown but is assumed to be a smooth function of its arguments, it is the object of interest. For each value of x, $d(\tau, x)$ is the discount function corresponding to maturity τ when the covariate equals x (this allows counterfactual analysis or extrapolation to different x). For the realized value X_t, the function $d(\cdot, X_t)$ in (1) is the discount function $d_t(\cdot)$ pertaining at time t that reflects the idealized arbitrage free world where cash flows are known and bond prices are the present discounted value of future payments, i.e., $p_{it} = \sum_{j=1}^{m_{it}} c_{it}(\tau_{ij}) d_t(\tau_{ij})$. In this idealized world (under some restrictions on d such as monotonicity) there are no "static arbitrage opportunities" that are available for trading bonds at time t, since all bonds are perfectly priced reflecting their coming cash flows. Unless there are further restrictions on d, there may be dynamic arbitrage opportunities available for trading bonds over time (in the absence of ε). We will discuss this issue further below.

The factors X_t could include deterministic time and stochastic determinants of the yield curve dynamics such as the short rate or macroeconomic variables. The special case where $X_t \equiv t/T \in [0, 1]$ implies that d is deterministic and the yield curve changes slowly over time in an orderly fashion. While this special case provides useful (and effective) statistical descriptions of the yield curve dynamics in the high frequency case, some other factors such as macroeconomic variables give some insights into the nature of the underlying economic forces that drive the movements of yield curves over the longer horizon. This type of modelling is common in financial econometrics. For instance, Engle and Rangel (2008) incorporate macroeconomic factors to explain the underlying long run trend in a multiplicative volatility model, while Koo and Linton

[2]Note however that d is not the regression function of prices on maturities and state variables. The model is a type of weighted additive regression model with varying number of components.

(2015) consider a deterministic time component for the long run trend. We allow for both deterministic and stochastic X_t.

We suppose that the pricing error ε_{it} is mean zero, stationary, and affects price in an additive fashion. The error term is needed to reconcile the different observed bond prices corresponding to identical cash flows. This type of model for microstructure noise is common in equity modelling using high frequency data, see Foucault et al. (2013). The noise represents microstructure features such as price discreteness, bid-ask bounce, and liquidity issues in the trading process. The pricing error can also reflect short term irrational behavioural pricing errors, Lamont and Thaler (2003); we do not distinguish or need to distinguish between these two sources of pricing errors. We emphasize that we adjoin the errors to the price process rather than to yields as has been assumed in some studies. We allow the error process to be heterogeneous and correlated across bonds and over time in a general fashion. We may suppose that the pricing error is quite small so that the no-arbitrage condition is not violated after transactions costs and market impact costs are taken into account, although this is not necessary for the statistical theory.

Finally, we remark that our dataset is an unbalanced panel since we observe a variety of bonds over time but for a given day t, the number of bonds in the cross-section varies. Moreover, various bonds for different payment structure and maturities are issued or expire on the different date. Nevertheless, as will be shown later, our estimation method will accommodate this general data structure.

2.1 Connection with Diebold and Li (2006)

Recall that the discount function, yield curve, and forward curve satisfy the relations: $d(\tau, x) = \exp(-\tau y(\tau, x))$ and $f(\tau, x) = -d'(\tau, x)/d(\tau, x)$, where $d'(\tau, x) = \partial d(\tau, x)/\partial\tau$, so that one can obtain from one of these curves any of the others, in particular, given $d(\tau, x)$ we obtain $y(\tau, x) = -(\log d(\tau, x))/\tau$. In our case since d is unrestricted in its functional form, so is y, and the yield curve can be upward sloping or downward sloping for different values of x.

Nelson and Siegel (1987) specify a parametric model for the static forward curve. Let

$$f_\theta(\tau) = \beta_0 + \left(\beta_1 + \beta_2\left(\frac{\tau}{\tau_0}\right)\right)\exp\left(-\frac{\tau}{\tau_0}\right),$$

where $\theta = (\beta_0, \beta_1, \beta_2, \tau_0)^\intercal$ are unknown parameters. This implies that the static discount function and yield curve are:

$$d_\theta(\tau) = \exp\left(-\tau\left(\beta_0 + (\beta_1 + \beta_2)\left(1 - \exp(-\tau/\tau_0)\right) \times \frac{\tau_0}{\tau} - \beta_2 \exp(-\tau/\tau_0)\right)\right) \tag{2}$$

$$y_\theta(\tau) = \sum_{j=0}^{2} \beta_j \varphi_j(\tau; \tau_0),$$

where $\varphi_0(\tau; \tau_0) = 1$, $\varphi_1(\tau; \tau_0) = (1 - \exp(-\tau/\tau_0))(\tau_0/\tau)$, and $\varphi_2(\tau; \tau_0) = (1 - \exp(-\tau/\tau_0))(\tau_0/\tau) - \exp(-\tau/\tau_0)$. This functional form is quite general and allows a combination of shapes for the long end and the short end of the yield curve. Exploiting the linearity of the static yield curve $y_\theta(\tau)$ in $\beta = (\beta_0, \beta_1, \beta_2)^\intercal$, Diebold and Li (2006) allow the unknown parameters/latent factors β_t to evolve in a dynamic fashion according to a VAR process, thereby generating a dynamic yield curve model in discrete time. They develop estimation methodology to uncover the parameters of these processes. First, based on seventeen yields (these were estimated by the Fama-Bliss method using a larger sample of bonds) at time t, they estimate the parameters β_t (they peg τ_0 at some plausible value) by least squares. Second, they fit a time series model to these β_t. We argue that their model is a special case of our model if we allow X_t to be the unobserved dynamic parameters β_t. In the case where X_t are observed covariates, then we may consider the following alternative but related model for y (or equivalently d)

$$y(\tau, X_t) = \sum_{j=1}^{L} \beta_j(X_{jt})\varphi_j(\tau), \tag{3}$$

where $\beta_j(.), \varphi_j(.)$ are smooth but unknown functions. This corresponds to a special case of the d function in (1) and is closely related to the Diebold and Li (2006) specification if further $\varphi_j(\tau) = \varphi_j(\tau; \tau_0)$ and $\beta_j(X_{jt}) = X_{jt}$, see Huse (2011) and Härdle and Majer (2014). Incidentally, the general nonparametric version of (3) is related to Breiman's PIMPLE, Breiman (1991).

We do not impose the particular structure (3) and allow the function d in (1) to be unrestricted in its second argument. We allow for arbitrary non-separability between maturity effects and calendar time or the covariates driving evolution. The general model can be used as a starting point for testing or further specification especially in the case where L is large where models like (3) would be attractive.

The parameters β_t in Diebold and Li (2006) are interpreted as "level, slope, and curvature" aspects of the yield curve. These aspects are usually measured by time series averages of yields or differenced yields, Diebold and Li (2006, Table 1). In our framework we may define these quantities as functionals of $y(\tau, x)$ in a model-free fashion. In particular, the average yield curve can be defined as $l(x) = \int y(\tau, x)w(\tau, x)d\tau$, where w is a weighting function. Likewise, we define the slope as the average slope of the yield curve for a given covariate value x or averaged over the covariate, and likewise the curvature:

$$s(x) = \int \frac{\partial y(\tau, x)}{\partial \tau} w(\tau, x)d\tau, \quad s = \int \frac{\partial y(\tau, x)}{\partial \tau} w(\tau, x)d\tau dx. \tag{4}$$

$$c(x) = \int \frac{\partial^2 y(\tau, x)}{\partial \tau^2} w(\tau, x)d\tau, \quad c = \int \frac{\partial^2 y(\tau, x)}{\partial \tau^2} w(\tau, x)d\tau dx. \tag{5}$$

2.2 Dynamic Arbitrage Restrictions

Gouriéroux et al. (2002) consider a discrete time model of the term structure. They show how to impose the absence of dynamic arbitrage on their model. They assume the compound autoregressive process (affine Markov conditional characteristic function) for the observed state variables, which include both endogenous state variables such as the short rate and exogenous state variables such as macroeconomic variables. As they remark, in the discrete time framework the stochastic discount factor is not unique. They assume that the stochastic discount factor is log linear in the current and past state variables, so that after imposing the economic restrictions on it, they obtain $M_{t,t+1} = \exp(\gamma_0 + \gamma_1^\intercal X_t + \gamma_2^\intercal X_{t+1})$ for parameters γ_j. In this case, in our notation, $d_t(\tau) = \exp(a_\tau^\intercal X_t + b_\tau)$, where a_τ and b_τ satisfy nonlinear first order difference equations that are determined by the parameter functions in the conditional characteristic function of the state variables and the parameters of the stochastic discount function. In their framework, both time and maturity occupy the same time grid and so their d function is defined only on integers apparently, but one can interpolate to $d_t(\tau) = \exp(a(\tau)^\intercal X_t + b(\tau))$ for continuous functions $a(\tau), b(\tau)$, which is a special case of our model (3). Duffie and Kan (1996) work in a continuous time continuous maturity framework; they show that the absence of arbitrage implies that $a(\cdot)$ and $b(\cdot)$ must obey certain differential equations. Diebold and Rudebusch (2013, Ch.3) show how to adjust the Diebold and Li (2006) model to satisfy the dynamic arbitrage restrictions. We could impose the Duffie and Kan (1996) restrictions on the functions $a(\cdot)$ and $b(\cdot)$ in the special case of our model $d_t(\tau) = \exp(a(\tau)^\intercal X_t + b(\tau))$, but this is a major undertaking computationally and merits a new paper.

In our general framework it is not clear how to impose the absence of dynamic arbitrage restrictions. Cheng and Tehranchi (2018) show how to impose these restrictions in a class of polynomial models, but it remains an open question whether this can be extended to a "nonparametric class of functions". In any case, we have not assumed a particular form for the stochastic discount function, we have not assumed that the yield curve is affine, and we have not even so far assumed that the state variables are Markov, let alone having an affine log conditional characteristic function. We have not yet embedded our discrete time model into a continuous time process that assumes continuous costless trading is possible. All of these assumptions are questionable, and there is empirical evidence against them.

3 Estimation

We introduce our estimation methodology based on kernel smoothing or localized least squares. The advantage of this method is that it requires only low dimensional optimization and its theoretical properties are well understood allowing us to propose inference methods that are valid under quite general conditions. We could apply the method in Linton et al. (2001) to each cross-section to extract a yield curve or discount function $d_t(\tau)$ for each time t, and then estimate the relationship between $d_t(\tau)$ and the covariates X_t using the time series information. Instead, we propose to use both cross-section and time series information simultaneously in our joint smoothing method. There are reasons to think that this is more efficient than doing first cross-section estimation and then time series estimation when the structure is correct, see Andreasen et al. (2017). Kernel methods are discussed in Härdle and Linton (1994) and Li and Racine (2007).

3.1 Local constant smoothing

The model (1) is a kind of structural nonparametric regression model with a varying number of additive components that are identical apart from known scaling factors due to the cash flows. We employ a nonparametric kernel estimation method that generalizes Linton et al. (2001). The idea is to start with a smoothed objective function and then to derive feasible estimating equations from that. Mammen et al. (1999) introduced this class of methods for the estimation of additive nonparametric models and argued why it has superior theoretical performance to methods that start from approximating first order conditions. Define the following smoothed sample least squares objective function for any smooth function $d(.)$

$$Q_{nT}(d) = \sum_{t=1}^{T}\sum_{i=1}^{n_t}\int\left\{p_{it} - \sum_{j=1}^{m_{it}} c_{it}(\tau_{ij})d(\tau_{ik},x)\right\}^2 \prod_{k=1}^{m_{it}}\left\{K_h(\tau_{ik}-\tau_{ij})d\tau_{ik}\right\}\mathcal{K}_h(x-X_t)dx, \tag{6}$$

where $\mathcal{K}_h(s) = \prod_{i=1}^{L} K_{h_i}(s_i)$ with a bandwidth parameter h_i and $K_{h_i}(\cdot) = K(\cdot/h_i)$. For the sake of notational simplicity, we henceforth set $h_1 = ... = h_i = h$. The estimator of the discount function is the minimizer of $Q_{nT}(\cdot)$ such that

$$\hat{d}(.) = \arg\min_{d(\cdot)\in\mathcal{D}} Q_{nT}(d) \tag{7}$$

where $\mathcal{D}$ is the class of all functions for which $Q(\cdot)$ is well defined. The objective function $Q_{nT}(d)$ can be thought of as expectation, with respect to the kernel density estimate of the "density" of τ, X, of the conditional least squares error brought about by a candidate function d.

We next obtain the first order condition that defines $\widehat{d}(.)$. Let $\delta_{(\tau,x)}(\cdot,\cdot)$ be the two-dimensional Dirac delta function at (τ,x) such that $\int_{r\in\mathbb{R}^L}\int_{y\in\mathbb{R}^+}\delta_{(\tau,x)}(y,r)g(y,r)dydr = g(\tau,x)$ for any generic function $g(\cdot,\cdot)$ continuous at the point (τ,x). Note that by definition, $\delta_{(\tau,x)}(\cdot,\cdot) = \delta_{(\tau)}(\cdot)\prod_{j=1}^{L}\delta_{(x_j)}(\cdot)$, where $x = (x_1,\ldots,x_L)$. Let $d(\cdot,\cdot) = \hat{d}(\cdot,\cdot) + \eta\delta_{(\tau,x)}(\cdot,\cdot)$, and differentiate $Q_{nT}(d)$ with respect to η at the point of $\eta = 0$. The first order condition for (6) yields $\hat{d}(\tau,x)$ which is obtained by

$$\begin{aligned}
&\sum_{t=1}^{T}\sum_{i=1}^{n_t}\sum_{j=1}^{m_{it}} c_{it}(\tau_{ij})^2 K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)\hat{d}(\tau,x) - \sum_{t=1}^{T}\sum_{i=1}^{n_t}\sum_{j=1}^{m_{it}} p_{it}c_{it}(\tau_{ij})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t) \quad (8)\\
= &-\sum_{t=1}^{T}\sum_{i=1}^{n_t}\sum_{j=1}^{m_{it}}\sum_{\substack{p=1\\p\neq j}}^{m_{it}} c_{it}(\tau_{ij})c_{it}(\tau_{ip})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)\int\hat{d}(\tau',x)K_h(\tau'-\tau_{ip})d\tau'.
\end{aligned}$$

Therefore, the first order condition (8) can be further simplified as the following integral equation,

$$\hat{d}(\tau,x) = \bar{d}(\tau,x) + \int\hat{H}(\tau,\tau',x)\hat{d}(\tau',x)d\tau', \tag{9}$$

where:

$$\bar{d}(\tau, x) = \frac{\sum_{t=1}^{T}\sum_{i=1}^{n_t}\sum_{j=1}^{m_{it}} p_{it} c_{it}(\tau_{ij}) K_h(\tau - \tau_{ij}) \mathcal{K}_h(x - X_t)}{\sum_{t=1}^{T}\sum_{i=1}^{n_t}\sum_{j=1}^{m_{it}} c_{it}(\tau_{ij})^2 K_h(\tau - \tau_{ij}) \mathcal{K}_h(x - X_t)} \tag{10}$$

$$\hat{H}(\tau, \tau', x) = -\frac{\sum_{t=1}^{T}\sum_{i=1}^{n_t}\sum_{j=1}^{m_{it}}\sum_{p=1,p\neq j}^{m_{it}} c_{it}(\tau_{ij}) c_{it}(\tau_{ip}) K_h(\tau - \tau_{ij}) K_h(\tau' - \tau_{ip}) \mathcal{K}_h(x - X_t)}{\sum_{t=1}^{T}\sum_{i=1}^{n_t}\sum_{j=1}^{m_{it}} c_{it}(\tau_{ij})^2 K_h(\tau - \tau_{ij}) \mathcal{K}_h(x - X_t)}. \tag{11}$$

Details regarding the derivation of (8) and (9) are provided as Lemma 2 in the Appendix. It is important to note that $\bar{d}(\tau, x)$ is a linear function of the observed prices.

In the special case where all bonds are zero-coupon bonds, this equation simplifies considerably and one has a closed form solution for $\hat{d}(\tau, x)$. In general, however, (9) involves the unknown function $\hat{d}(\cdot, x)$ in the left hand side and inside of the integral, which prevents exact solution and renders computation tricky. This is also known as a Fredholm integral equation. See Riesz and Nagy (1990) for more details. One way of solving this type of integral equation is the backfitting approach or successive approximation, see Mammen et al. (1999). We define our estimator for the general case as follows. For $a = 0, 1, \ldots$ let

$$\hat{d}^{[a+1]}(\tau, x) = \bar{d}(\tau, x) + \int \hat{H}(\tau, \tau', x)\hat{d}^{[a]}(\tau', x) d\tau', \tag{12}$$

where the starting value $\hat{d}^{[0]}(\cdot)$ is equal to $\bar{d}(\cdot)$. We iterate this until it converges. Each step of the update involves simple linear operations. Some conditions that ensure convergence are given in Section 4 and Appendix.

Following Mammen et al. (1999), we also interpret the estimator obtained by (9) as a projection of the data in a suitable functional space. Indeed, as seen from (12), to ensure convergence of the successive approximation method, the key set of regularity conditions is concerned with the operator $\hat{\mathcal{H}}$ defined, for every $g \in L^2$, as

$$\hat{\mathcal{H}} g(\tau, x) = \int \hat{H}(\tau, \tau', x) g(\tau', x) d\tau'. \tag{13}$$

We also introduce the operator sup-norm of $\hat{\mathcal{H}} g(\cdot)$ such that

$$\|\hat{\mathcal{H}}\| = \sup_{\|g\|=1} \left\|\hat{\mathcal{H}} g\right\|.$$

In Lemma 1 in the Appendix, we show that an application of the Neumann series yields:

$$\hat{d} = (1 - \hat{\mathcal{H}})^{-1}\bar{d}, \tag{14}$$

which is well-defined if $(1 - \hat{\mathcal{H}})$ is invertible, namely provided that $\|\hat{\mathcal{H}}\| < 1$.

Once the estimate of the discount curve is available, we can easily estimate other related quantities, exploiting the one-to-one mapping between the discount function $d(\tau, x)$, the yield curve $y(\tau, x)$ and the forward rate $f(\tau, x)$, where τ and x denote time-to-maturity and time. Thus, from the relationship mentioned in Section 2.1, estimates of the yield curve and of the forward curve are defined as

$$\hat{y}(\tau, x) = -\ln(\hat{d}(\tau, x))/\tau \tag{15}$$

$$\hat{f}(\tau, x) = -\frac{\int K_h'(t-s)\hat{d}(s,x)ds}{\int K_h(t-s)\hat{d}(s,x)ds}. \tag{16}$$

The level, slope, and curvature functionals can also be defined based on this.

3.2 Shape Restrictions

We can incorporate restrictions on the shape of the discount curve. For example, it is natural to assume that the discount function be monotonically decreasing in the maturity. To impose this restriction we define our estimator as the minimizer of $Q_{nT}(d)$ in (6), where the minimization runs now over a suitably constrained class of functions. Let us denote by $\hat{d}_{\text{constr}}$ the estimate resulting from the solution to (6). This solution has an interpretation in terms of Hilbert spaces theory: it represents the projection of $\hat{d}$ in (14) onto the constrained class with respect to the norm

$$\|g\|^2 = \sum_{t=1}^{T}\sum_{i=1}^{n_t}\int_{\tau_{ik}}\int_x \left\{\sum_{j=1}^{m_{it}} c_{it}(\tau_{ij}) g_i(\tau_{ik}, x)\right\}^2 \prod_{k=1}^{m_{it}} \{K_h(\tau_{ik}-\tau_{ij})d\tau_{ik}\}\,\mathcal{K}_h(x-X_t)dx$$

for the n-tuples of functions $g = (g_1, g_2, \ldots, g_n)$. Namely, we have:

$$\begin{aligned}
\hat{d}_{\text{constr}} &= \arg\min_{d(\cdot)} \|d - \hat{d}\|^2 \\
&= \arg\min_{d(\cdot)} \sum_{t=1}^{T}\sum_{i=1}^{n_t}\int_{\tau_{ik}}\int_x \left\{\sum_{j=1}^{m_{it}} c_{it}(\tau_{ij})\left(d(\tau_{ik}, x) - \hat{d}(\tau_{ik}, x)\right)\right\}^2 \\
&\times \prod_{k=1}^{m_{it}} \{K_h(\tau_{ik}-\tau_{ij})d\tau_{ik}\}\,\mathcal{K}_h(x-X_t)dx,
\end{aligned}$$

where the minimization runs over the constrained class of functions d.

3.3 Local constant exponential smoothing

Many empirical studies try to directly model the relationships between bond yields and macro variables; see e.g. Ang and Piazzesi (2003) and Ang et al. (2006) for applications in the case of affine term structure models. Rather than using local constant smoothing and then defining a yield curve estimates as in (15), here we propose to fit directly local exponential functions, which directly extracts the yield curve.

We define the estimator of the yield curve, by the minimization, with respect to $y(\cdot)$, of the following objective function:

$$
\begin{aligned}
\mathcal{Q}_{nT}(y) &= \sum_{t=1}^{T}\sum_{i=1}^{n_t}\int \left\{ p_{it} - \sum_{j=1}^{m_{it}} c_{it}(\tau_{ij}) \exp\{-\tau_{ik} y(\tau_{ik}, x)\} \right\}^2 \\
&\quad \times \prod_{k=1}^{m_{it}} \{K_h(\tau_{ik} - \tau_{ij}) d\tau_{ik}\} \mathcal{K}_h(x - X_t) dx. \qquad (17)
\end{aligned}
$$

Theoretically this is hard to analyse as the first order condition is a nonlinear integral equation.

3.3.1 Imposing Affine Yield Curve

A special case of our model is consistent with the affine structure. We may further assume the affine structure $y(\tau, x) = a(\tau) + b(\tau)x$ for some unknown functions $a(\cdot), b(\cdot)$, see Gouriéroux et al. (2002). In this case, we do not need to smooth against X_t, that is, we minimize the smoothed objective function

$$
\mathcal{Q}_{nT}(a, b) = \sum_{t=1}^{T}\sum_{i=1}^{n_t}\int_{\tau_{ik}} \left\{ p_{it} - \sum_{j=1}^{m_{it}} c_{it}(\tau_{ij}) \exp\{-\tau_{ik}(a(\tau_{ik}) + b(\tau_{ik})X_t)\} \right\}^2 \prod_{k=1}^{m_{it}} K_h(\tau_{ik} - \tau_{ij}) d\tau_{ik}
$$

with respect to $a(\cdot)$, $b(\cdot)$, similar to Gouriéroux et al. (2002). An alternative and simpler way of imposing affinity is to proceed in two steps, i.e., find the closest affine yield curve to our unrestricted nonparametric yield curve (15) based on some simple objective function. For example, $\int \left(\widehat{y}(\tau, x) - a(\tau) - b(\tau)x\right)^2 dw(\tau, x)$, where w is a weighting function. One could use the minimized objective function to deliver a statistical test of affinity.

3.4 Forecasting Future Bond Prices and Yields

We next discuss forecasting of future yield curves within our model framework. There are several approaches here depending on the choice of covariates, which itself is partly determined by the sampling frequency of the data, and the treatment of the error term. We first consider the case where the error term is considered unforecastable given the past state and prices.

We first consider the case where $X_t = t/T$, in which case the yield curve evolves deterministically, although the form of d in this case is unknown to the econometrician. We have by Taylor expansion

$$
d\left(\tau, \frac{T+k}{T}\right) = d(\tau, 1) + d_2(\tau, 1)\frac{k}{T} + \frac{1}{2} d_{22}(\tau, 1)\frac{k^2}{T^2} + o(k^2 T^{-2}),
$$

where $d_2(\tau, x) = \partial d(\tau, x)/\partial x$ and $d_{22}(\tau, x) = \partial^2 d(\tau, x)/\partial x^2$. This suggests that we can forecast the discount function (and hence the yield curve) at period $T+k$ by $d(\tau, 1)$, by $d(\tau, 1) + d_2(\tau, 1)\frac{k}{T}$, or even by $d(\tau, 1) + d_2(\tau, 1)\frac{k}{T} + \frac{1}{2}d_{22}(\tau, 1)\frac{k^2}{T^2}$, depending on the horizon k and the accuracy required, Robinson

(2012). This amounts to: constant, linear, or quadratic interpolation of the curve. In practice one has to replace $d(\tau,1)$, $d_2(\tau,1)$, and d_{22} by estimates. The special case as in (3) corresponds to

$$y(\tau,t/T)=\sum_{j=1}^{L}\beta_{jt}\varphi_j(\tau), \tag{18}$$

where the quantities $\beta_{jt}=\beta_j(t/T)$ are deterministic and change slowly over time, in particular, $\beta_{jt}=\beta_{j,t-1}+\delta_{jt}$, where $\beta_{jt}=\beta(t/T)$ and δ_{jt} is a small deterministic update factor or "error term". This should be compared with the Diebold and Li (2006) model in which β_t has an autoregressive form with stochastic error term. This kind of "local trend model" in (18) is an alternative to dynamic specification, see Starica (2003) and Vogt and Linton (2014). Engle and Rangel (2008), Hafner and Linton (2010) and Koo and Linton (2015) combine the slowly varying nonparametric trend with a short run dynamic model.

A second way of forecasting uses predictive regression, whereby the stochastic X_t that enter d, and hence y, are taken to be lagged values. This changes the interpretation of the discount function somewhat, but is otherwise an approach that is widely used.

A third approach is to specify a dynamic model for the covariates or rather restrict the dynamics somewhat. Specifically, suppose that X_t follows a first order Markov process with time invariant transition density $\pi(x,x')$ (it is common practice to assume a VAR for X_t). We can forecast the future coupon bond price $P_{i,T+1}$ given the data set $\{P_{it},X_t,i=1,\ldots,n_t,\ t=1,\ldots,T\}$ as follows. First, estimate $d(\tau,x)$ by our method. Second, estimate $g(\tau,x)=\int d(\tau,x')\pi(x',x)dx'$ by smoothing the estimated discount function against lagged values of X_t, i.e., let

$$\widehat{g}(\tau,x)=\frac{\sum_{t=1}^{T-1}\sum_{i=1}^{n_t}K_h(\tau-\tau_i)K_h(x-X_t)\widehat{d}(\tau_i,X_{t+1})}{\sum_{t=1}^{T-1}\sum_{i=1}^{n_t}K_h(\tau-\tau_i)K_h(x-X_t)}.$$

Then our forecast of $P_{i,T+1}$ is

$$\widehat{P}_{i,T+1|T}=\sum_{j=1}^{m_{i,T+1}}c_{ij,T+1}\widehat{g}(\tau_{ij},X_T), \tag{19}$$

where we use the known payment scheme for bond i prevailing at time $T+1$. Because the model is linear in the discount function, this gives an approximately unbiased forecast of future prices. Likewise we could forecast future yields by smoothing our estimated $y(\tau,X_{t+1})$ against lagged X. A prediction interval for $d(\tau,X_{t+1})$ or $y(\tau,X_{t+1})$ can be obtained by estimating the corresponding conditional c.d.f. given X_t under the Markov assumption.

Finally, if one is willing to make further restrictions such as linear covariate effect in the yield curve then one can obtain much simpler forecasts. Suppose that

$$y(\tau,X_t)=\sum_{j=1}^{L}X_{jt}\varphi_j(\tau),$$

where $\varphi_j(\tau)$ are smooth but unknown functions of maturity only, which corresponds to the Affine class. Further suppose that for some $L \times L$ matrix A, $X_t = AX_{t-1} + \eta_t$, as in Diebold and Li (2006). This is a special case of our model. We may identify $\varphi_j(\tau)$ nonparametrically by $\varphi_j(\tau) = \partial y(\tau, x)/\partial x_j$, $j = 1, \ldots, L$. Then we would forecast $y(\tau, X_{T+k})$ by

$$\widehat{y}(\tau, X_T) = \sum_{j=1}^{L} \left(e_j A^k X_T\right) \varphi_j(\tau), \tag{20}$$

where e_j is the j^{th} unit vector. In practice we have to replace A and $\varphi_j(\tau)$ by estimates. In this case, the persistence in the yield curves is driven by the persistence in the observed factors.

We have so far supposed that there is no information about future prices in the current and past values of the error terms. In fact, pricing errors are serially correlated and it is possible in principle to use this information to better predict future bond prices. Suppose that $\varepsilon_{it} = \rho_i \varepsilon_{i,t-1} + \eta_{it}$, where η_{it} is i.i.d. with mean zero. Then we may add to (19) a term such as $\widehat{\rho}_i \widehat{\varepsilon}_{iT}$, where hats indicate estimated quantities.

4 Large Sample Properties

In this section, we provide the Central Limit Theorem for the local constant estimator of the discount function. As described in the previous section, our approach involves nonparametric estimation over two dimensions, i.e., time-to-maturity and covariate, e.g., time. This pooling of information allows for more precise estimation of the common structure.

Our estimator is the solution of a linear integral equation (12), and successive approximation such as the backfitting projection algorithm is required for solution. Therefore, we require regularity conditions for the solution to (12) to converge, which we will discuss in detail in this section. We impose *high level* assumptions on the error structure $\{\varepsilon_{it}\}$, the operator $\hat{\mathcal{H}}$ and its norm $\|\mathcal{H}^k\|$ for $k \geq 1$. In the following, c is a generic constant, which may be different at different places.

4.1 Asymptotic distribution

Assumption A

(A1) *Model (1) holds. Furthermore, the discount function d is twice boundedly continuously differentiable in τ_{ij} (time-to-maturity) and in each component of the L-dimensional X_t (factors). Each factor X_{it}, for $i = 1, \ldots, L$, is stationary and has finite second moment.*

(A2) *For each time point t, the error variables $\{\varepsilon_{it}\}_{i=1}^{n_t}$ represent bond specific errors. They are heteroskedastic and possibly cross-sectionally correlated in a sense that their variance changes proportionally with time to maturity and common sources for pricing errors are accommodated. Furthermore, for a given bond type i, there exists time dependence, which exponentially decays. In particular,* *the error variables* $\{\varepsilon_{it}\}_{i=1}^{n_t}$ *have the following structure: for* $i = 1, \ldots, n_t$ *and* $t = 1, \ldots, T$,

$$\varepsilon_{it} = \sum_{s=1}^{T} \sum_{j=1}^{n_t} W_{ijts} \epsilon_{js} \tag{21}$$

where $\{\epsilon_{js}\}_{j=1,\ldots,n_t; s=1,\ldots,T}$ *are independent of* $\{X_t\}$, *independent and identically distributed random variables, with mean zero and* $E|\epsilon_{js}|^{2+\zeta} < \infty$ *for any arbitrary positive value* ζ. *Each* W_{ijts} *remains general and unspecified: it captures the cross-sectional correlation between the bond types* i *and* j *and their temporal dependence at time point* t *and* s.

(A3) *Let the operator* $\hat{\mathcal{H}}$ *be as in (13). There exist a constantc and* $\rho < 1$ *such that, for* $k \geq 1$ *and* n *large enough,* $\|\hat{\mathcal{H}}\|_p < c$ *and* $\|\hat{\mathcal{H}}^k\|_p < \rho$ *when* $p = 2$.

(A4) *The following conditions hold:*

$$\max_{1 \leq k \leq n_t; 1 \leq s \leq T} \int \left(\int \hat{H}(\tau, \tau', x) R_{k,s}(\tau', x) d\tau' \right)^2 d\tau = o_p\left(h^{-(L+1)}\right) \tag{22}$$

$$\int \hat{H}^2(\tau, \tau', x) d\tau' \leq c \tag{23}$$

where $X_t \in \mathbb{R}^L$ *and*

$$R_{k,s}(\tau, x) = \frac{\sum_{t=1}^{T} \sum_{i=1}^{n_t} W_{ikts} \sum_{j=1}^{m_{it}} c_{it}(\tau_{ij}) K_h(\tau - \tau_{ij}) \mathcal{K}_h(x - X_t)}{(nT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{n_t} \sum_{j=1}^{m_{it}} c_{it}(\tau_{ij})^2 K_h(\tau - \tau_{ij}) \mathcal{K}_h(x - X_t)}.$$

For all t, $\tau \in \chi$ *and* $x \in \mathbb{R}^L$,

$$\sum_{k=1}^{n_t} \int \left| \frac{\partial}{\partial \tau} \hat{H}(\tau, \tau', x) R_{k,s}(\tau', x) d\tau' \right| \leq C, \tag{24}$$

$$\sum_{s=1}^{T} \int \left\| \frac{\partial}{\partial x} \hat{H}(\tau, \tau', x) R_{k,s}(\tau', x) d\tau' \right\| \leq C, \tag{25}$$

where $C < \infty$.

(A5) *The quantity*

$$\omega_{it}(\tau, x) = \frac{\sum_{j=1}^{m_{it}} c_{it}(\tau_{ij}) K_h(\tau - \tau_{ij}) \mathcal{K}_h(x - X_t)}{\sum_{t=1}^{T} \sum_{i=1}^{n_t} \sum_{j=1}^{m_{it}} c_{it}(\tau_{ij})^2 K_h(\tau - \tau_{ij}) \mathcal{K}_h(x - X_t)}, \tag{26}$$

is uniformly bounded in $(\tau, x) \in \mathbb{R}^{L+1}$.

Let us comment briefly on these assumptions. Assumption (A1) is concerned with the model and the main object of interest in our estimation. It also ensures that the discount function is smooth enough for us to employ a nonparametric kernel estimation method. We maintain here that the factor is stationary but the theory can easily be extended to allow the more general local stationarity property, Vogt and Linton (2014).

Assumption (A2) involves the error structure. The error in our model involves two indexes: one for the bond type and one for time. Note that the error structure implicitly assumes the double array, $i = 1, \ldots, n_t$ and $t = 1, \ldots, T$. Although (22), (24), and (25) are affected by the structure of W_{ijts} via $R_{\cdot,\cdot}(\cdot, \cdot)$, the exact form of W_{ijts} is not required for the derivation of the asymptotic theory. However, for the development of inference tools it is helpful to impose additional structure, as shall do in Section 4.2.

In what follows, for the sake of simplicity, we treat W_{ijts} as a deterministic function of the relative distance between the ordered bonds at different time points. Assumptions (A4) and (A5) is needed to prove the consistency and asymptotic normality of the proposed estimator, with (A5) guaranteeing the existence of the estimator variance.

Additionally, for the validity of our kernel estimation and its asymptotics, we assume:

Assumption B

(B1) *The kernel $K(\cdot)$ is a bounded symmetric around zero function such that: (i) it is continuously differentiable up to order r on R with $2 \leq r$; (ii) it belongs to L^2, $\int |K(u)|du < \infty$, $\int K(u)du = 1$, and the support of K is contained in $[-1, 1]$; (iii) $\mu_i(K) = \int u^i K(u)du = 0$, $i = 1, \ldots, r-1$, and: $\int u^r K(u)du \neq 0$, $\int |u|^r |K(u)|du < \infty$, $\lim_{||u|| \to \infty} ||u|| K(u) = 0$; (iv) $K(\cdot)$ is Lipschitz continuous, i.e. $|K(u) - K(u')| \leq C\|u - u'\|$ for all $u, u' \in R$.*

(B2) *Let h be the bandwidth and $n = \min_t n_t$. ; (i) $T = O(n)$; (ii) as $T \to \infty$, $h \to 0$, $Th \to \infty$ and $nTh^{L+1} \to \infty$.*

Define the following quantities:

$$\mathscr{B}^*_{\hat{d}}(\tau, x) = \tilde{\mathscr{B}}^*_{\hat{d}}(\tau, x) + \tilde{\tilde{\mathscr{B}}}^*_{\hat{d}}(\tau, x),$$

$$\tilde{\mathscr{B}}^*_{\hat{d}}(\tau, x) = \frac{\sum_{t=1}^T \sum_{i=1}^{n_t} \sum_{j=1}^{m_{it}} c_{it}^2(\tau_{ij})[d(\tau_{ij}, X_t) - d(\tau, x)] K_h(\tau - \tau_{ij}) \mathcal{K}_h(x - X_t)}{\sum_{t=1}^T \sum_{i=1}^{n_t} \sum_{j=1}^{m_{it}} c_{it}^2(\tau_{ij}) K_h(\tau - \tau_{ij}) \mathcal{K}_h(x - X_t)},$$

$$\tilde{\tilde{\mathscr{B}}}^*_{\hat{d}}(\tau, x) = \frac{\sum_{t=1}^T \sum_{i=1}^{n_t} \sum_{j=1}^{m_{it}} \sum_{p=1, p\neq j}^{m_{it}} c_{it}(\tau_{ij}) c_{it}(\tau_{ip}) K_h(\tau - \tau_{ij}) \mathcal{K}_h(x - X_t) \int K_h(\tau' - \tau_{ip}) \left[d(\tau_{ip}, x) - d(\tau', x)\right] d\tau'}{\sum_{t=1}^T \sum_{i=1}^{n_t} \sum_{j=1}^{m_{it}} c_{it}^2(\tau_{ij}) K_h(\tau - \tau_{ij}) \mathcal{K}_h(x - X_t)},$$

$$\mathscr{V}_{\hat{d}}(\tau, x) = (nTh^{L+1}) \text{var}\left[\sum_{t=1}^T \sum_{i=1}^{n_t} \omega_{it}(\tau, x) \varepsilon_{it}\right], \tag{27}$$

where $\omega_{it}(\tau, x)$ is given in (26).

Theorem 1. *Suppose that assumptions (A1)-(A4) and (B1)-(B2) hold. Then,*

$$\sqrt{nTh^{L+1}}\mathscr{V}_{\hat{d}}(\tau,x)^{-1/2}\left(\hat{d}(\tau,x)-d(\tau,x)-\mathscr{B}^*_{\hat{d}}(\tau,x)\right)\xrightarrow{d}\mathscr{N}(0,1). \tag{28}$$

Following (15) and (16), the asymptotic properties of the related estimators ($\hat{y}(\tau,x)$ for the yield curve dynamics and $\hat{f}(\tau,x)$ for the forward curve dynamics) can be obtained by applying the delta-method to the results in Theorem 1. For instance, regarding the estimator for the yield curve dynamics, $\hat{y}(\tau,x)$, the following asymptotic properties can be derived.

Corollary 1. *Suppose that all assumptions for Theorem 1 hold. Then,*

$$\sqrt{nTh^{L+1}}\mathscr{V}_{\hat{y}}(\tau,x)^{-1/2}\left(\hat{y}(\tau,x)-y(\tau,x)-\mathscr{B}^*_{\hat{y}}(\tau,x)\right)\xrightarrow{d}\mathscr{N}(0,1),$$

where $\mathscr{B}^*_{\hat{y}}(\tau,x)=(\tau d(\tau,x))^{-1}\mathscr{B}^*_{\hat{d}}(\tau,x)$ *and* $\mathscr{V}_{\hat{y}}=(\tau d(\tau,x))^{-2}\mathscr{V}_{\hat{d}}(\tau,x)$ *with* $\mathscr{V}_{\hat{d}}(\tau,x)$ *specified in* (27).

The rate of convergence is fast, upto order $(nT)^{-1/3}$ for the case $L=1$ and twice continuously differentiable d, reflecting the fact that both time series and cross-sectional information are used to estimate the function d. The estimated yields $y_{jt}=\widehat{y}(\tau_j,X_t)=y(\tau_j,X_t)+\xi_{jt,T}$, where the error terms $\xi_{jt,T}$ are correlated over time and yield, under our assumptions, and are small.

4.2 Asymptotic variance estimation

The availability of an asymptotic distribution as in Theorem 1 represents one of the advantages of our method, since $\sqrt{nTh^{L+1}}\left(\hat{d}(\tau,x)-d(\tau,x)-\mathscr{B}^*_{\hat{d}}(\tau,x)\right)\xrightarrow{d}\mathscr{N}(0,\Sigma_{\hat{d}}(\tau,x))$, where

$$\Sigma_{\hat{d}}(\tau,x)=\lim_{n,T\to\infty}\ \mathrm{var}\left[\frac{1}{\sqrt{nTh^{L+1}}}\sum_{t=1}^{T}\sum_{i=1}^{n_t}\omega_{it}(\tau,x)\varepsilon_{it}\right]. \tag{29}$$

where $\lim_{n_t,T\to\infty}$ represents the short hand notation for the limit taken as n and T diverge.

Thus, if a consistent estimator of the asymptotic variance in (29) is available, then we can compute consistent pointwise confidence intervals for the estimated discount curve. To achieve this inferential goal, we propose the following two steps: first, we obtain $\hat{\varepsilon}_{it}=p_{it}-\sum_{j=1}^{m_{it}}c_{it}(\tau_{ij})\hat{d}(\tau_{ij},X_t)$; second, we define

$$\hat{V}_{\hat{d}}(\tau,x)=\frac{1}{nTh^{L+1}}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{j=1}^{n_t}\sum_{i=1}^{n_t}\omega_{it}(\tau,x)\omega_{js}(\tau,x)\hat{\varepsilon}_{it}\hat{\varepsilon}_{js}.$$

Because the error structure specified in (A2) and (A4) is quite general, it is difficult (if not impossible) to obtain a consistent estimator of $\Sigma_{\hat{d}}(\tau,x)$ without any further restriction, either on the bandwidth or on the weight $\{W_{itjs}\}$ as in (21). Therefore, we need to impose more structure on $\{W_{itjs}\}$. Among the possible solutions, we choose to impose a restriction on both cross-sectional dependence and temporal dependence. Thanks to this additional assumption, we can prove that $\hat{V}_{\hat{d}}(\tau, x) \xrightarrow{p} \Sigma_{\hat{d}}(\tau, x)$, as n and T diverge.

We suppose that there exists some relabelling of the cross-sectional bonds, whose generic index we denote by i, such that the cross sectional dependence decays with the bond economic distance, which is a function of $|i - j|$. This approach has been adopted, e.g., in Lee and Robinson (2016) and in Ma et al. (2017) in the context of spatial panel data models. In our setting, it is natural to obtain the data ordering according to the maturity: two bonds are similar and close if they have similar string of cash flows. To elaborate further, we introduce the ϕ-mixing coefficient on the error term. For $S_1, S_2 \subset [1, \dots, n] \times [1, \dots, T]$, let

$$\phi(S_1, S_2) \equiv \sup\{|P(A|B) - P(A)| : A \in \sigma(\varepsilon_{it}, (i,t) \in S_1), B \in \sigma(\varepsilon_{js}, (j,s) \in S_2)\},$$

where $\sigma(\cdot)$ denotes a σ-field. Then, for any $k \in \mathbb{N}$, the ϕ-mixing coefficient of $\{\varepsilon_{it}\}$ is defined as

$$\phi(k) \equiv \sup\{\phi(S_1, S_2) : d_B(S_1, S_2) \geq k\},$$

where

$$d_B(S_1, S_2) \equiv \min\{\sqrt{|t-s|^2 + |i-j|^2} : (i,t) \in S_1, (j,s) \in S_2\},$$

represents the economic distance between ordered bonds.

Then, we introduce the following:

Assumption C

(C1) *The stochastic process* $\{\varepsilon_{it}\}$ *is a random field of* ϕ*-mixing random variables, where for each* i *and* t, $E[|\varepsilon_{it}|^{4+\zeta}] < \infty$, *for* $\zeta > 0$. *The* ϕ*-mixing coefficient of* $\{\varepsilon_{it}\}$ *satisfies* $\phi(k) \leq ce^{-\lambda_1 k}$ *for* $c, \lambda_1 > 0$. *For each given* i, $\{\varepsilon_{it}\}$ *is a strictly stationary sequence.*

(C2) *The asymptotic variance resulting from Theorem 1 and as in (29) is such that* $\Sigma_{\hat{d}}(\tau, x) \neq 0$, *for every* $(\tau, x) \in \mathbb{R}^{L+1}$.

Assumption (C1) complements the error structure given in Assumption (A2), ensuring that errors are marginally cross-sectional mixing, and marginally temporally mixing. This type of dependence is assumed in much literature on spatial and temporal processes; see e.g. Connor and Korajczyk (1993), Conley (1999) and Lee and Robinson (2016), among many others. Loosely speaking, we assume that, as the distance between points in time gets larger, the temporal dependence decays at the exponential rate. Likewise, as the types of bonds under consideration become more heterogeneous, say a short-term treasury bill and a long-term treasury bond, the cross-sectional dependence between them gets smaller at the exponential rate. Thus, to account for dependence in the time-to-maturity space, we may set $W_{ijts} \propto \exp^{-b_1|i-j|}$, for some $b_1 > 0$, as suggested by Rebonato and Cooper (1997) (see Table 1 in that paper), whilst we may set $W_{ijts} \propto \exp^{-b_2|t-s|}$, for some $b_2 > 0$, to account for temporal dependence, as it is customary in the time series literature. This is a mathematically convenient choice which helps in proof of the consistency of the proposed estimator of $\Sigma_{\hat{d}}$. We emphasize that this additional specification is needed only for the estimation of the asymptotic variance: some other specifications can be considered at the cost of a more complicated proof for the consistency of the proposed asymptotic variance estimator; see Linton et al. (2001), page 202 for a related discussion. Assumption (C2) is standard and it guarantees that the asymptotic variance admit an inverse; see, e.g., Assumption (B5) in Conley (1999).

Finally, we state the following.

Theorem 2. *Suppose that assumptions (A1)-(A4), (B1)-(B2) and (C1)-(C2) hold. Then, as* $n, T \to \infty$

$$\hat{V}_{\hat{d}}(\tau, x) \xrightarrow{p} \Sigma_{\hat{d}}(\tau, x).$$

Theorem 2 justifies the use of $\hat{V}_{\hat{d}}(\tau, x)$ in the definition of asymptotic confidence intervals, pointwise in $(\tau, x) \in \mathbb{R}^{L+1}$, for the estimated yield curve. The confidence intervals are useful when considering scenarios for the yield curve based on different values of the state variables and maturities, since we should take account of the estimation uncertainty in the presentation.

5 Monte Carlo Study

5.1 Simulation design

We illustrate the performance of our nonparametric estimator in a simulation study. We consider a time horizon of about 10 years of bi-weekly data (namely, 3750 days), We set the number of bonds to $n = 23$ and for each bond (either zero coupon or coupon bearing) the face value (paid at the expiration date) is set to one. We consider zero coupon bonds issued with original time to maturity of 1-12 months, while coupon bonds have 2, 3, 5, and 10 years maturity. We keep the number of bonds constant over time: to achieve this goal, we adopt a rotating panel structure. Specifically, the bonds are repeatedly observed over time, until their expiry. Then, we replace each expired bond with a new one, having the same maturity and the same cash flow payment schedule (just for the sake of simplicity) as the expired one, but with a different identification number.

For a given day t, bond prices are generated by the model in (1) with ε_{it} having Gaussian distribution, zero mean and variance increasing over duration (therefore, bonds having longer maturities display larger variability). The noise term is uncorrelated between bonds.

We generate the initial yield curve using the Nelson & Siegel model (Nelson and Siegel, 1987), a model which has been applied in several empirical studies of term structure estimation (see, e.g., Bliss (1996); Dahlquist and Svensson (1996)). We choose the extended version

$$y(\tau) = \beta_0 + \beta_1 \left[\frac{1 - \exp(-\tau/\tau_1)}{\tau/\tau_1} \right] + \beta_2 \left[\frac{1 - \exp(-\tau/\tau_2)}{\tau/\tau_2} - \exp(-\tau/\tau_2) \right] \tag{30}$$

and we derive the corresponding discount curve. We select the parameterization: $\beta_0 = 0$, $\beta_1 = 0.05$, $\beta_2 = 2, \tau_1 = 0.75$ and $\tau_2 = 125$, which is such that the curve has a stationary point; see Linton et al. (2001). The yields generated by (30) define the initial (at time τ_0) yield curve. To incorporate the time (in t) dynamics, we introduce two deterministic trends for the yield curve: the first is a cubic trend having equation $-0.55t^3 + 0.55t^2 - 0.55t + 0.1$, while the second is a quadratic trend having equation $-0.85t^2 + 0.85t - 0.1$. Both trends affect the initial Nelson & Siegel yield curve and induce, over time, a cubic or parabolic pattern in the simulated discount curve, respectively. Within these two different settings, we obtain bond prices using two coupon rates: we set $c_{it}(\tau_{ij}) = 2.5\%$ (annualized), $\forall i, j, t$ for the cubic trend, while $c_{it}(\tau_{ij}) = 7\%$ (annualized), $\forall i, j, t$ for the quadratic trend.

The selected design resembles closely the structure of the US Treasury market, over the last 50 years. Indeed, T-bills (zero coupon bonds) are issued with original time to maturity of 1-12 months. US Treasury notes (coupon bearing bonds) are issued with original maturities of 2, 3, 5, and 10 years. When a bond expires, a new one with similar characteristics is issued to replace it. Moreover a cubic/parabolic dynamics is observable in the US market in the period 1976-1989, which includes the so called Volcker's period, characterized by a sharp interest rates increase (from 1979-1981) and a subsequent down trend.

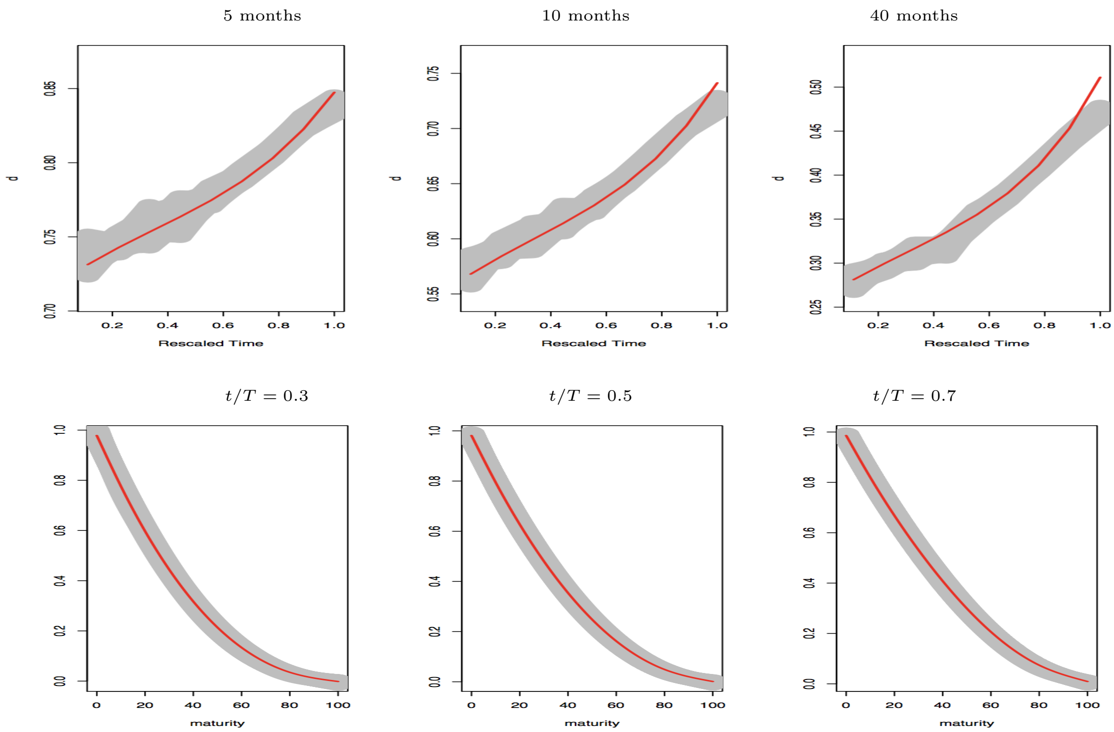


Figure 1: Cubic time-dynamics of the discount curve.

Top panels: in each plot, we display the true cubic dynamics of the discount rate (continuous line) and the estimated cubic dynamics of the discount rate in each of the 1000 Monte Carlo runs (to each run corresponds a continuous light grey line), over rescaled time (x-axis). Bottom panels: actual (central thick line) and estimated discount curve (each grey line corresponds to a Monte Carlo run), over all considered maturities (x-axes), for three different time rescaled points.

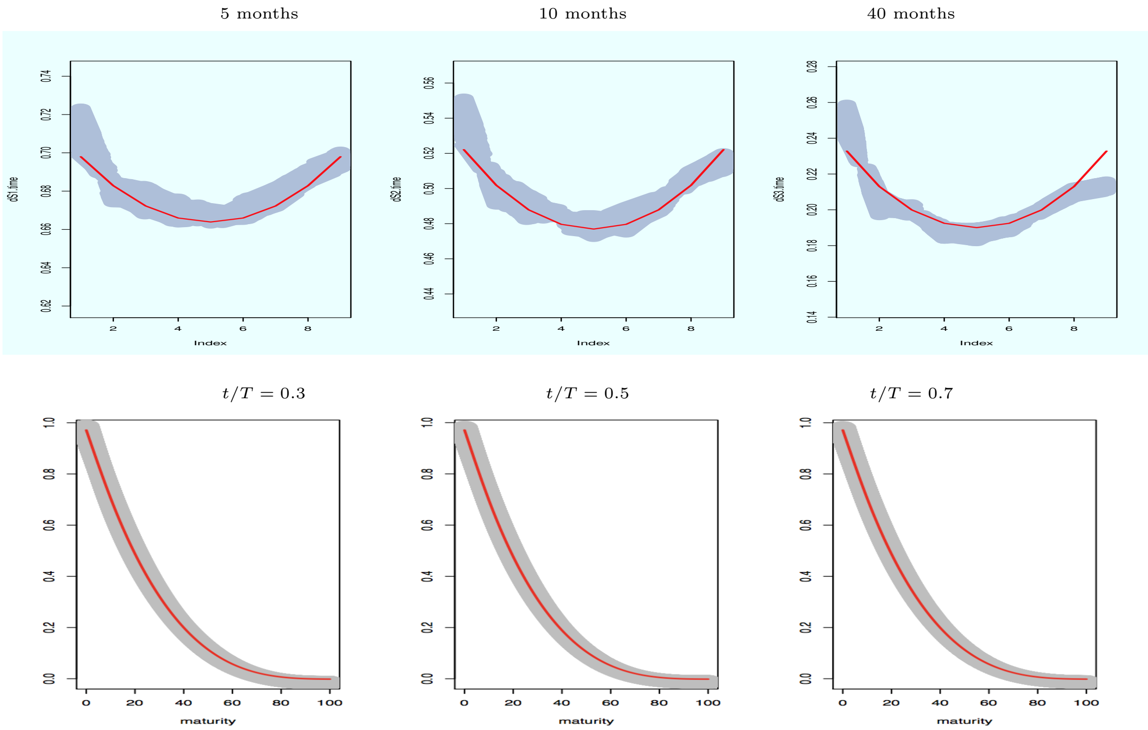


Figure 2: Parabolic time-dynamics of the discount curve.
Top panels: in each plot, we display the true parabolic dynamics of the discount rate (continuous line) and the estimated parabolic dynamics of the discount rate in each of the 1000 Monte Carlo runs (to each run corresponds a continuous light grey line), over rescaled time (x-axis). Bottom panels: Each plot displays actual (central thick line) and estimated discount curve (each grey line corresponds to a Monte Carlo run), over all considered maturities (x-axes), for three different time rescaled points.

5.2 Outputs

Before commenting on the results, a few implementation details are in order. In general, a constant bandwidth over the time-to-maturity space does not perform well. This is due to the simulation design, which is such that the payments distribution is far from uniform over the considered maturities. Accordingly, we use a nonconstant bandwidth over the time-to-maturity space, for both the cubic and parabolic dynamics. To be more precise, for a given trading day t, the selected bandwidth is proportional to the duration of each considered bond. The Epanechnikov kernel is employed.

In each Monte Carlo simulation and for each day t, we estimate the discount curve using the method of Section 3. In Figure 1, we display the results for the cubic dynamics. In the top panels we focus on the time evolution. Specifically, we select three time-to-maturities (5, 10 and 40 months) and we compare over rescaled time (t/T, on the x-axes) the actual discount rate dynamics (continuous line) with the estimated discount rate. For each maturity, we plot the time evolution of the discount rate, as estimated in the 1000 Monte Carlo runs (to each estimate correspond a grey continuous line and the collection of whole grey lines creates the shaded area). The plots show that our method captures the discount curve cubic (upward) dynamics over time.

To validate our method further, we investigate the ability of our method to capture also the shape of the discount curve, on a given day. To this end, in the bottom panels of Figure 1, we select three rescaled times and for every considered maturity (expressed in months, on the x-axes), we plot (for 1000 Monte Carlo runs) the actual and the estimated discount curve. The central thick line represents the true discount curve, while, for each Monte Carlo run, a grey line corresponds to the estimated discount curve: the collection of whole grey lines creates the shaded area. The plots illustrate that the model captures the shape of the actual discount curve. Similar considerations hold also for the parabolic dynamics. In Figure 2, we show that the method captures the behaviour of the yield curve, for the selected rescaled times.

6 US yield curve evolution

We apply our methodology to US bond data. We establish the confidence interval for the US discount function when other factors including time and three-month treasury bill rates are its drivers. We discuss the in-sample fit and the out-of-sample forecasting performance of our estimation in comparison with the widely used method of Diebold and Li (2006). This section starts with our data and real-data implementation issues before we report our estimation results and comparison studies.[3]

6.1 Data description

Our dataset consists of daily US treasury data obtained from the Centre for Research in Security Prices (CRSP) US Treasury Database. CRSP provides extensive historical data related to all US treasury bonds including each bond's issue date, its type (say, treasury bill, note, bond), its maturity, its historical bid and ask price quotes with corresponding accrued interests, a history of its coupon payments and payment dates. Note that when a bond is sold, the actual price the buyer pays is the quoted price plus the interest amount accrued until the transaction day. Each bond is characterized by its unique identifier. There are nine different types of issues depending on whether it is callable or not and tax related issues. Note that callable bonds contain an intrinsic option of redemption to the issuer and tax would affect yield curves, which are not captured by our model. Therefore, for our analysis, we select noncallable bonds and notes, treasury bills. Almost all of coupon payments are made biannually except for a few bonds that pay the first coupon in less than a half year and paid biannually thereafter. The data covers the seven-year-period from Jan. 2001 to Dec. 2007.[4] We provide several statistics in relation to the data for our estimation in Table 1.

[3]All our computations are conducted with the r programming language. Codes are available upon request.

[4]The entire period ranges from Jan. 1970 to Dec. 2007. The results of the same analyses for the other periods are not different from those reported in this paper. Results for the other periods are available upon request.

Table 1: Summary Statistics of CRSP Bonds Data

	2001	2002	2003	2004	2005	2006	2007
Average No. of Daily Quotes	162	149	143	150	160	171	182
Time to Maturity$^{(1)}$	10963	10636	10271	9906	9539	10956	10964
Ratio of Treasury bills$^{(2)}$	0.34	0.38	0.38	0.38	0.36	0.33	0.32
Ratio of Treasury notes$^{(2)}$	0.44	0.42	0.42	0.45	0.48	0.52	0.53
Ratio of Treasury bonds$^{(2)}$	0.21	0.21	0.2	0.18	0.16	0.15	0.15

Notes: (1) Longest days from the quote date to maturity for each year.

(2) calculated as the number of each type over the number of entire set of bonds for each year.

We generate a price data matrix p_{it} for a bond i at time t using the average of bid and ask quotes of bond i and corresponding accrued interest for each quotation day. We calculate time-to-maturity of each bond as a calendar-day between the quotation day and its maturity for each time, t. Similarly, we calculate payment dates of each bond, τ_{ij} for $j = 1, \ldots, m_{it}$ as the calendar day between the quotation day and payment dates. Based on these, we calculate the cash flows $c_{it}(\tau_{i\cdot})$ for the bond i by combining a history of coupon payments and their payment dates over different quotation days.

We use deterministic time ($u = t/T$) and three-month treasury bill rates (r) as factors ($x = (u, r)^\intercal$) for our estimation. Three month treasury bill rates (secondary market rate) are obtained from FRED at the Federal Reserve Bank of St. Louis. We plot them in Figure 3 for each time grid in 2007 with Treasury Bill factor grids in the dotted lines.

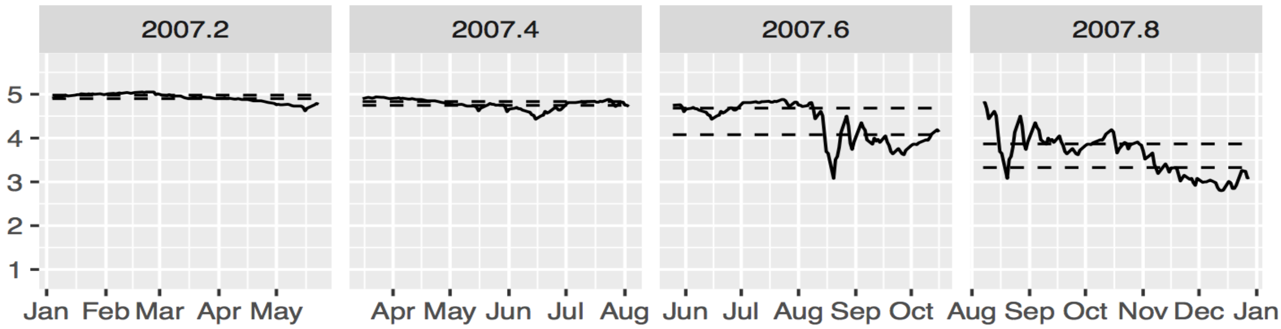


Figure 3: Three-month Treasury Bill Rates (Secondary Market)
Three-month Treasury bill rates (2007) used for one of factors are plotted in solid line in each estimation window. Dotted lines denote the grids set for each estimation window (33% and 66%).

6.2 Implementation of nonparametric estimation

We estimate $\bar{d}(\tau, x)$ and $\hat{H}(\tau, \tau', x)$ given as in (10) and (11). The estimate for the discount function, $\hat{d}(\tau, x)$ is obtained via (9). The corresponding yield curve, $\hat{y}(\tau, x)$ is estimated indirectly from the one-to-one relationship between the discount function and the yield curve as in (15). The variances of $\bar{d}(\tau, x)$ and $\hat{y}(\tau, x)$ are calculated from Theorem 2 and Corollary 1. In doing so, there are several issues worth noting in the empirical implementation. First, although the discount function of interest is infinite-dimensional and is assumed continuous over an interval in $\mathbb{R}^+$, discrete grids for factors and time-to-maturities need to be selected for the estimation in practice. Moreover, intrinsically, there exists the data sparsity for the longer time-to-maturities because bonds with more than ten-year maturity are scarce. This incurs significant estimation challenge so that we pay close attention to bonds with longer maturities so as to ensure the data for longer time-to-maturities are enough for reliable estimation. That is, we cannot choose equally-spaced grids for the time-to-maturity. Rather, it is crucial to choose grids for time-to-maturities in such a way that a decent number of data points are available in the window around each selected grid while the window around each selected grid covers the decent local coverage of the corresponding time-to-maturity.

Second, since we obtain estimates of the function of interest only for finite trivariate discrete grids (time, treasury bill rate, time-to-maturity), interpolation of estimates for between grid points is required. Furthermore, the three dimensional grids chosen for time, treasury bill rate and time-to-maturity are not necessarily equally spaced but rather irregularly or/and regularly spaced for the maximum precision and efficiency, which renders interpolation tricky. Many interpolation techniques are available including multilinear interpolation, cubic spline interpolation and local polynomials interpolation. We use equally spaced grids for time and treasury bill rate whereas we use irregularly spaced grids for time-to-maturity due to the first issue mentioned in the previous paragraph. For the interpolation method, we use the local polynomials interpolation for time-to-maturity and linear interpolation for time and treasury bill rate.

Third, there are three different bandwidths required for the estimation: ones for the time-to-maturity h_τ, the treasury bill rate h_r and the time h_u. As is the usual case with nonparametric kernel estimation, choosing appropriate bandwidths, in particular one for the time-to-maturities leads to better estimation results. This is very much so in that the structure of bond data is very diverse according to the time-to-maturities. A lot of data are available for the shorter maturity, in particular zero coupon bonds and coupon paying bonds about to expire (less than 6 months) account for more than 50% of price data for any given day whereas the number of middle- and long-term bonds with more than twice coupon payments left are relatively very small. Consequently, we set smaller bandwidths for the shorter time-to-maturities while we set larger bandwidths for the longer time-to-maturities, which ensures reliable and robust estimation results. For h_r and h_u, we employ the rule of thumb bandwidth selection.

Here, we choose equally spaced time grids such that $u_{i+1} - u_i = 0.2$ where $u_i = t_i/T$. That is, for the entire period of k years, we have $N(= 5k - 2)$ number of grids since we have five time grid points for each year except for the beginning and the end of the entire period. The bandwidth for each chosen time grid is $h_u = 0.2$ with the Epanechnikov kernel. For the treasury bill rate, we choose r grids as the ones corresponding to 33 percentile and 66 percentile of treasury bill rates in the window in the vicinity of time point u. As mentioned, the bandwidths for the chosen r grids are selected following the rules of thumb.[5] On the other hand, we employ an automated selection of grids for the time-to-maturities in such a way

[5]Although the choice for h_r and h_u follows the rules of thumb, we tried finer or wider u and r grids with different bandwidths but results are quite similar if not worse.

that we choose a grid with a maximum number of data while there exist enough price and cash flow data in the windows around the selected grids where each window is determined by the corresponding bandwidth. For instance, when the grid in the vicinity of 10 years in time-to-maturity domain needs to be chosen, we check the number of data available for each candidate of time-to-maturities in the window where enough data reside say at least 5, say from 3600 days to 3700 days and select the grid, say 3640 if that time-to-maturity day has the maximum number of data available for the estimation. Based on the selected trivariate grids, we calculate $\bar{d}(\tau, x)$ and $\hat{H}(\tau, \tau', x)$ and finally $\hat{d}(\tau, x)$. Regarding the interpolation, we employ an r-package called `loess` for the interpolation of estimates for our application although other interpolation methods readily available could be used without much issue.

6.3 Estimation results

Estimation results for the discount function and yield curve for the year 2007 are reported in Tables 2 and 3, respectively. We report the curve at five different days in 2007 for a range of maturities while the treasury bill rates (r) is fixed at 33 percentile in each estimation window, along with the pointwise standard error in parentheses. We also provide the corresponding graphical representation of the estimated discount function and yield curve along with their corresponding pointwise confidence bands via interpolation in Figure 4 and Figure 5 respectively. Figures 6a and 6b report three-dimensional shapes of discount functions and yield curves over time-to-maturity and time from Jan. 2004 to Dec. 2007 respectively while r is fixed at 50 percentile of three month treasury bill rates in each estimation window.

Table 2: Estimates of Discount Function (2007)

Grid (u,r)	3 Months	6 Months	1 Year	3 Years	5 Years	10 Years	15 Years
2007	0.984	0.973	0.952	0.871	0.796	0.626	0.475
u:0.0, r:4.94	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
2007	0.984	0.973	0.953	0.872	0.797	0.624	0.471
u:0.2, r:4.90	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
2007	0.985	0.974	0.952	0.868	0.789	0.610	0.454
u:0.4, r:4.75	(0.006)	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)
2007	0.986	0.976	0.955	0.875	0.797	0.616	0.456
u:0.6, r:4.08	(0.008)	(0.01)	(0.01)	(0.008)	(0.008)	(0.008)	(0.008)
2007	0.990	0.981	0.963	0.893	0.822	0.641	0.475
u:0.8, r:3.33	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)

Notes: (1) Values in parentheses denote standard errors.

(2) The following calendars correspond to the time point u we consider: 2007-01-02 ($u = 0.0$), 2007-03-15 ($u = 0.2$), 2007-05-24 ($u = 0.4$), 2007-08-06 ($u = 0.6$), 2007-10-17 ($u = 0.8$).

(3) Three month Treasury bill rates (r) is fixed at 33 percentile in each estimation window.

Table 3: Estimates of Yield Curve (2007)

Period	3 Months	6 Months	1 Year	3 Years	5 Years	10 Years	15 Years
2007	0.043	0.045	0.046	0.048	0.049	0.053	0.058
u:0.0, r:4.94	(0.016)	(0.008)	(0.004)	(0.002)	(0.002)	(0.001)	(0.001)
2007	0.018	0.033	0.041	0.048	0.05	0.055	0.06
u:0.2, r:4.90	(0.016)	(0.008)	(0.004)	(0.002)	(0.002)	(0.001)	(0.001)
2007	0.022	0.034	0.04	0.045	0.047	0.052	0.058
u:0.4, r:4.75	(0.02)	(0.01)	(0.006)	(0.002)	(0.002)	(0.001)	(0.001)
2007	-0.023	0.009	0.026	0.039	0.043	0.051	0.057
u:0.6, r:4.08	(0.036)	(0.018)	(0.01)	(0.004)	(0.002)	(0.002)	(0.002)
2007	-0.023	0.011	0.028	0.041	0.045	0.053	0.059
u:0.8, r:3.33	(0.034)	(0.018)	(0.01)	(0.002)	(0.002)	(0.002)	(0.002)

Notes: (1) Values in parentheses denote standard errors.

(2) The following calendars correspond to the time point u we consider: 2007-01-02 ($u = 0.0$), 2007-03-15 ($u = 0.2$), 2007-05-24 ($u = 0.4$), 2007-08-06 ($u = 0.6$), 2007-10-17 ($u = 0.8$).

(3) Three month Treasury bill rates (r) are fixed at 33 percentile in each estimation window.

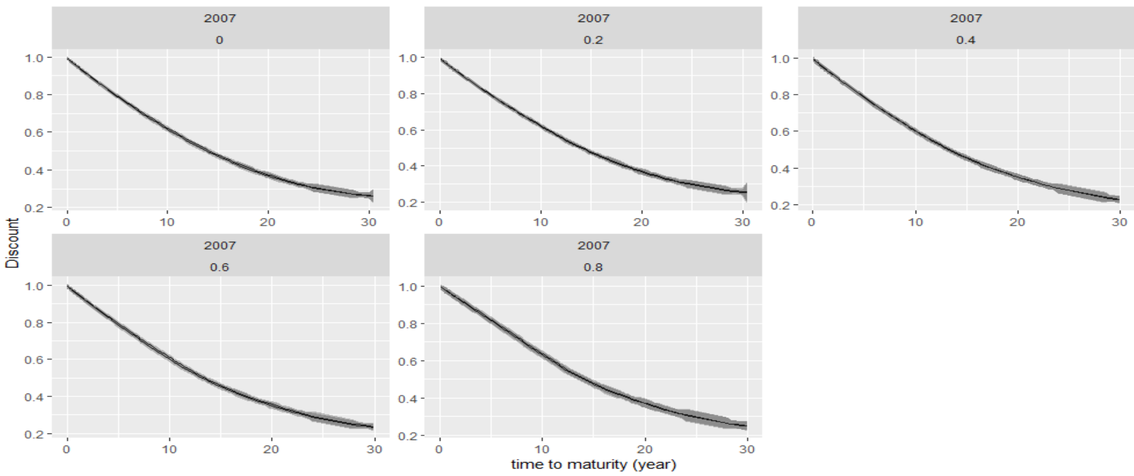


Figure 4: Estimates for the discount function $\hat{d}(\tau, x)$ with pointwise confidence intervals

Discount function given a time point with 95% and 5% confidence intervals. Given time grids ($u = (0.0, 0.2, 0.4, 0.6, 0.8)$ for a sample year, 2007) and time-to-maturities grids, we plot our estimates of the discount functions while r is fixed at 50 percentile of treasury bill rates in each estimation window for 2007 with the shaded pointwise confidence interval.

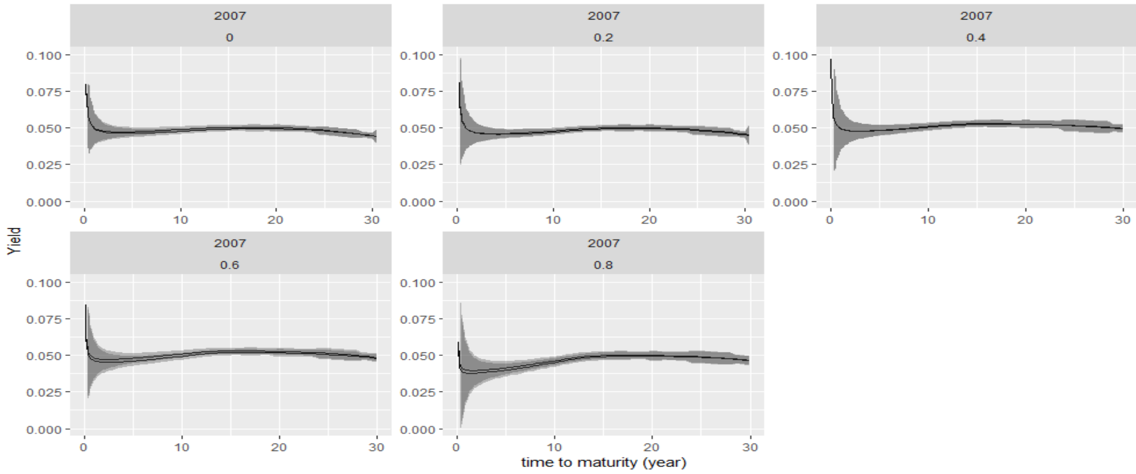
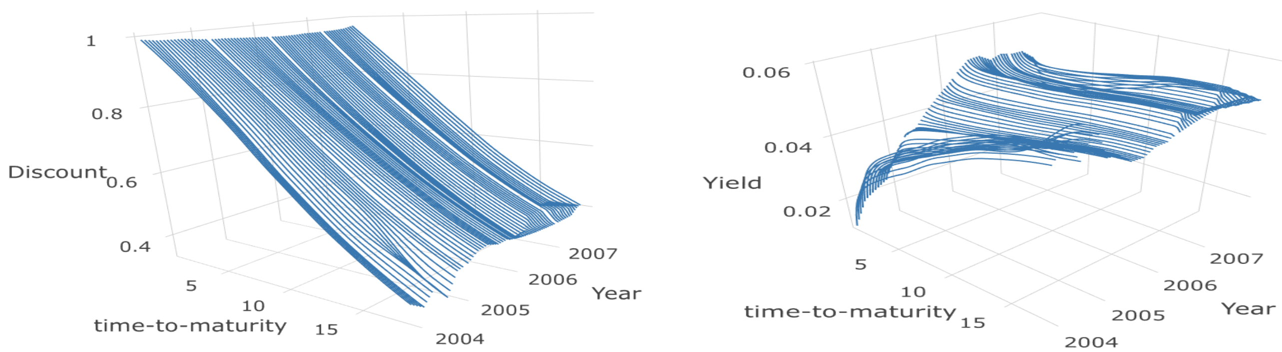


Figure 5: Estimates for the yield curve $\hat{y}(\tau, x)$ with pointwise confidence intervals

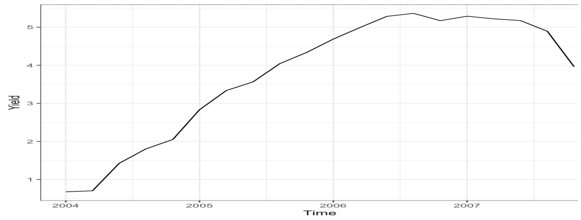
Yield curve given a time point with 95% and 5% confidence intervals. Given time grids ($u = (0.0, 0.2, 0.4, 0.6, 0.8)$ for a sample year, 2007) and time-to-maturities grids, we plot our estimates of the yield curves while r is fixed at 50 percentile of treasury bill rates in each estimation window for 2007 with the shaded pointwise confidence interval.



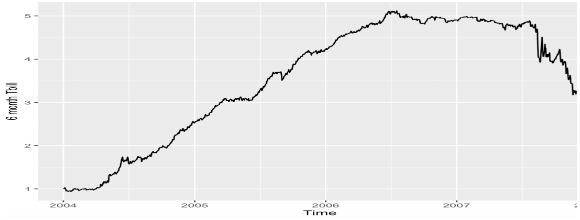
(a) Discount function 3-dim

(b) Yield Curve Dynamics 3-dim

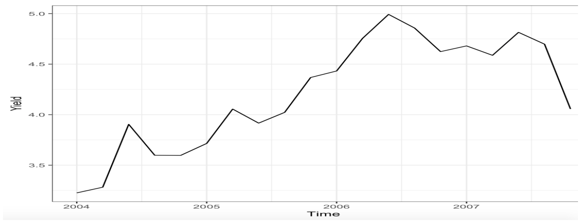
Figure 6: 3-dimensional shapes of $\hat{d}(\tau, x)$ and $\hat{y}(\tau, x)$



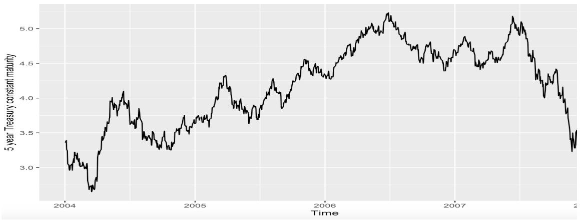
(a) Cross-section of 3-dim Yield Curve at $\tau = 0.5yrs$



(b) Six-month T-Bill Rates



(c) Cross-section of 3-dim Yield Curve at $\tau = 5yrs$



(d) 5 year Treasury Constant Maturity Rates

Figure 7: Cross-sections of $\hat{y}(\tau, x)$ at 6 month and 5 year time-to-maturities

Generally, as seen from Figures 4 and 5, discount functions seem to display smoothly declining shapes over time-to-maturity but yield curves display various shapes over time-to-maturity depending on the economic conditions in the neighbourhood of the time point. Meanwhile, the shape of yield curve over time given a time-to-maturity is compatible with the movement of the other market interest rates, e.g. dynamics of six-month treasury bill rates and five-year treasury constant maturity rates are comparable with cross-sections of three-dimensional yield curve estimates at corresponding time-to-maturities, i.e. $\tau = 0.5yrs$ and $\tau = 5yrs$ respectively from Jan. 2004 to Dec. 2007 as seen from Figure 7. The confidence bands for the discount curve and the yield curve show the degree of uncertainty about the actual value the curve takes given the data we have used. The confidence interval generally widens with maturity for both discount function and yield curve; there is also quite a bit of uncertainty regarding the spot yield.

6.4 Evaluation of estimates of the US discount function

This section discusses the performance of the proposed nonparametric estimation of the discount function via its in-sample fit and out-of-sample forecasting. We calculate the residuals as the difference between the observed price (quoted price plus accumulated interest) and the corresponding theoretical price constructed from the first principle that the asset price is the present value of all the cash flows generated from the asset $\widehat{\varepsilon}_{it} = p_{it} - \hat{p}_{it}$, where $\hat{p}_{it} = \sum_{j=1}^{m_{it}} c_{it}(\tau_{ij})\widehat{d}(\tau_{ij}, X_t)$ with X_t as factors. We calculate the mean squared errors (MSE) by summing the squared values of all residuals.

Likewise, forecasting errors are calculated via the difference between the h-period ahead observed price, $p_{i,t+h}$ and the price forecast constructed from $\hat{p}_{it}(h) = \sum_{j=1}^{m_{it}} c_{it}(\tau_{ij})\widehat{d}(\tau_{ij}, X_{t+h})$, where X_{t+h} is the forecast of factors. The forecasting mean squared errors (FMSE) are calculated accordingly.

We provide the empirical distribution of the residuals and discuss some notable features of residuals in line with some of the assumptions we made for the asymptotic analysis. We also use the MSE criterion for the in-sample fit and the FMSE criterion for the out-of-sample forecasting performance. Since we observe price quotes of an array of different types of bonds per day, it is natural for us to calculate the MSE and the FMSE on a daily basis. Then, for each year, we provide a summary of distribution of daily MSE and FMSE for concise presentation.

6.4.1 In sample fit: residual analysis

We next plot the empirical distribution of the residuals in Figure 8: different colours are used for different types to highlight the magnitude of residuals over type. As seen from Figure 8, residuals for the long-term bonds and for highly volatile periods are larger. This is consistent with our assumption that the variances of discount functions and yield curves vary over time-to-maturity and time. We also report quantiles and mean of residuals for each year in Table 4.

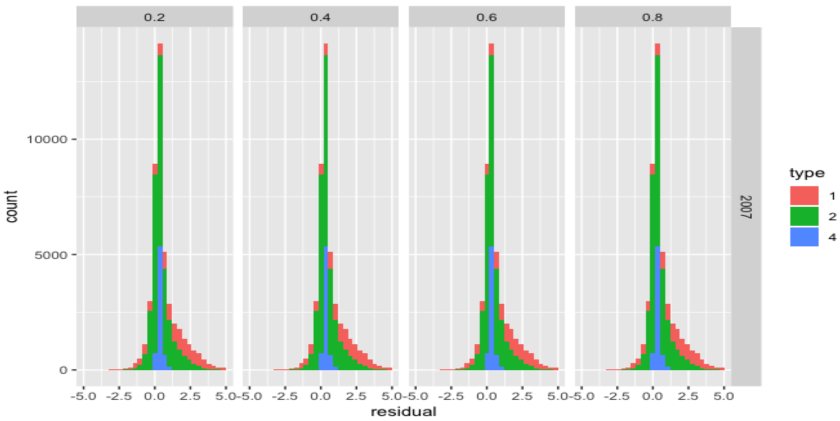


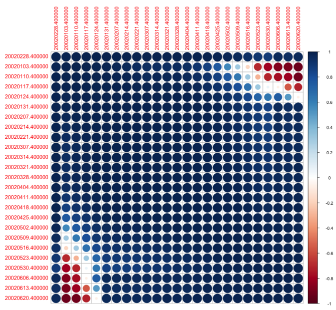
Figure 8: Histograms of residuals for the sample year

Different colours are used according to the bond type. Type 1, 2, and 4 denote treasury bonds, treasury notes and treasury bills respectively.

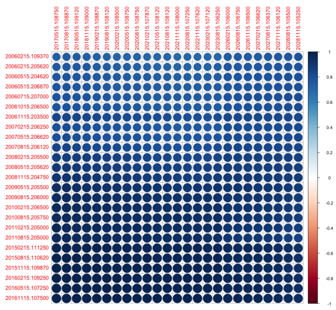
Table 4: Percentiles of Residuals

	2001	2002	2003	2004	2005	2006	2007
min	-3.5686	-4.2374	-5.5529	-6.3623	-3.7558	-3.5268	-4.0874
25%	-0.1994	-0.6942	-0.8825	-0.4851	0.0361	0.0826	0.1028
median	0.2703	0.3626	0.1841	0.3190	0.2151	0.2609	0.3448
mean	0.5164	0.4056	0.4418	0.4463	0.4017	0.4465	0.6578
75%	0.9355	1.2126	1.2385	1.0053	0.5162	0.5694	0.9251
max	10.278	7.8678	10.658	8.2870	6.9790	5.0412	8.7145

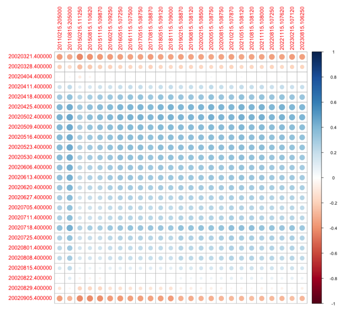
In addition, we look at the correlations of residuals across different time-to-maturities and across different points in time. Recall that our asymptotic theory is based on the several assumptions related to random pricing errors ε_{it}. The error structure assumed for the method is quite general in that it could be allowed to be heterogeneous and correlated across the bond type given the fixed time point t and it might be dependent over time.



(a) Correlations between short term bonds



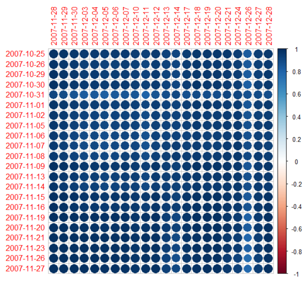
(b) Correlations between long term bonds



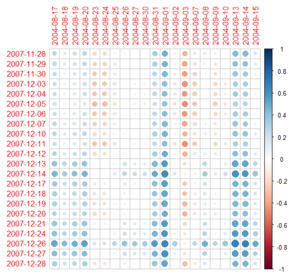
(c) Correlations between short and long term bonds

Figure 9: Cross-sectional correlations

Labels denote identifiers of various bonds. The first eight digit (year-month-day) denotes the maturity of the bond and the number following a dot denotes its type (1: bond; 2: note; 3: bill) and the remainder denotes its coupon rate.



(a) Correlations between close time points



(b) Correlations between times far away each other

Figure 10: Temporal correlations

Labels denote the date of quote for various bonds.

Nevertheless, in Assumption (C1) for the asymptotic variance estimation, we assume that as difference in the maturities between two different bonds i and j gets bigger, say for a long term bond and a short term bond, the cross-sectional dependence gets smaller and smaller and is expected to tend to zero when the distance tends to infinity. Also, a similar type of assumption for the time series temporal dependence is made in that as time points, t and s are far away each other, the temporal dependence gets smaller, for instance, mixing type of dependence.

It is difficult to identify the exact structure of random pricing errors but we can check whether there is a tendency by graphically examining correlations over time-to-maturity and time. Correlations for cross-sectional dependence and temporal dependence are plotted in Figure 9 and Figure 10 respectively. More specifically, as seen from Figures 9a and 9b, the correlations between residuals with respect to short term bonds (type 4) or long term bonds (type 1 or 2) are high, whether it is positive or negative. However, Figure 9c shows that the correlation between short term and long term bonds is close to zero. This is compatible with our assumption of cross-sectional dependence. Meanwhile, Figure 10 shows the temporal dependence of residuals. Congruent with our assumption of the temporal dependence of random pricing errors, high correlation between residuals with respect to the same bonds with close points in time (close quote dates) is reported in Figure 10a, whereas lower correlation is reported when the distance between time points gets larger, say 3 years in Figure 10b.

6.4.2 Out of sample forecasting

Recall that we can forecast the discount function or yield curve based on the dynamics of factors incorporated in our model. That is, $d(\tau, X_{t+h})$ can be forecast by $d(\tau, X_{t+h})$ and the dynamics of X_{t+h}.

We implement our forecasting study using the horizon, $h = 30$ days. Also, for the dynamics of factors, X_t, in our experiment (the three month Treasury Bill rates) an autoregressive time series model is used. The empirical distribution of forecasting residuals for 2007 are reported in Figure 11. Table 5 reports the summary of forecasting residuals for each year from 2001 to 2007.

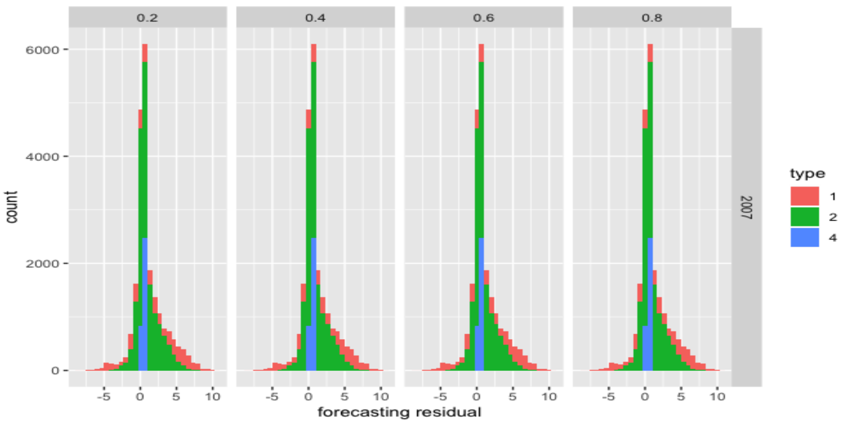


Figure 11: Histograms of forecasting residuals for the sample year

Different colours are used according to the bond type. Type 1, 2, and 4 denote treasury bonds, treasury notes and treasury bills respectively.

Table 5: Percentiles of Forecasting Residuals

	2001	2002	2003	2004	2005	2006	2007
min	-5.00	-5.19	-12.8	-8.49	-6.21	-11.5	-8.85
25%	0.090	-0.783	-0.891	-0.926	-0.0599	-0.360	0.114
median	0.903	0.300	0.125	-0.058	-0.197	0.199	0.520
mean	1.59	1.18	0.146	0.408	0.212	0.270	1.11
75%	2.40	2.22	1.24	1.80	0.241	0.649	1.88
max	13.3	14.5	13.3	9.48	13.3	7.70	10.4

In the following section, we compare the estimation in sample fit and out of sample forecasting performance of our method with those in Diebold and Li (2006).

6.5 Comparison studies

This section compares our results with those of Diebold and Li (2006). Using the same CRSP US data set, we estimate discount functions and yield curves based on the following parametric specification:

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right)$$

where $y_t(\tau)$ is the yield of bonds with the time-to-maturity τ, β_i for $i = 1, 2, 3$ are three common factors and λ_t reflects the exponential decay rate.

The three common factors $\{\beta_i\}_{i=1,2,3}$ can be interpreted in terms of level, slope, and curvature respectively, as how their respective loadings change represents the characteristics of long-term, short-term and medium-term: the loading on β_1 is a constant, the loading on β_2 starts at one and decays to zero, and the loading on β_3 starts at 0, increases then drops to 0. Once λ_t is fixed at a prespecified value, then two factor loadings are computed and from those computed factor loadings, three common factors are obtained via ordinary least squares. For more details, see Diebold and Li (2006).

Once all of βs are estimated, the h-step ahead forecast of yield curve is achieved by

$$\hat{y}_{t+h/t}(\tau) = \hat{\beta}_{1,t+h/t} + \hat{\beta}_{2,t+h/t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \hat{\beta}_{3,t+h/t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right)$$

where the factors have underlying AR(1) specification such that

$$\hat{\beta}_{i,t+h/t} = \hat{c}_i + \hat{\gamma}_i\hat{\beta}_{it}$$

for $i = 1, 2, 3$.

These results are reported in Tables 6 and 7. Table 6 reports in sample fit of Diebold and Li (2006) and our nonparametric method. Overall, our method compares favourably against the method proposed by Diebold and Li (2006) Meanwhile, Table 7 reports the out-of-sample forecasting performance of both methods. Again, our method performs competitively with the alternative.

7 Concluding Remarks

We propose a nonparametric estimation method that estimates slowly time varying yield curves from panel data on bond prices and payments without imposing parametric restrictions. The method is straightforward to implement and we illustrated our approach with a simulation study and an application to US data. We established the consistency and asymptotic normality of our procedures and developed consistent confidence intervals based on our theory. We allow throughout for general cross-sectional and temporal weak dependence in the error terms, and show that our standard errors are consistent under these general conditions.

Table 6: Summary of Distribution of Daily Prediction MSE
(In Sample Fit)

year	method	min	25%	median	mean	75%	max
2001	NP	0.0445	0.257	0.548	1.22	1.30	15.0
	DL	0.193	1.257	2.077	3.471	3.698	32.14
2002	NP	0.460	0.709	1.010	1.51	1.60	8.20
	DL	0.330	3.638	7.83	10.31	14.43	37.97
2003	NP	0.513	0.818	1.21	2.49	2.90	18.1
	DL	0.304	0.817	3.137	6.912	8.090	50.90
2004	NP	0.081	0.537	0.862	1.53	1.88	7.93
	DL	0.325	1.136	2.359	4.125	5.200	17.80
2005	NP	0.024	0.092	0.252	0.548	0.624	4.38
	DL	0.175	0.792	1.645	2.196	3.052	8.514
2006	NP	0.022	0.127	0.303	0.433	0.611	1.83
	DL	0.084	0.573	2.068	2.498	3.727	8.763
2007	NP	0.045	0.158	0.389	0.677	0.856	5.39
	DL	0.170	0.512	1.104	3.351	4.594	23.69

Note: NP and DL denote nonparametric and Diebold and Li (2006) estimations respectively.

Table 7: Summary of Distribution of Daily FMSE
(Out of Sample Forecasting)

year	method	min	25%	median	mean	75%	max
2001	NP	0.326	1.11	2.86	4.44	5.68	17.8
	DL	1.918	4.56	10.83	21.1	24.7	97.2
2002	NP	0.510	0.789	2.21	6.15	5.43	32.1
	DL	1.631	3.290	6.84	8.03	10.7	38.9
2003	NP	0.537	0.826	1.47	7.26	14.1	27.6
	DL	1.997	8.200	17.43	23.61	36.5	81.0
2004	NP	0.752	1.99	4.26	4.15	5.87	9.94
	DL	1.182	2.66	4.39	28.37	43.9	89.06
2005	NP	0.197	0.837	2.41	4.49	8.96	15.1
	DL	0.757	1.899	4.17	5.66	8.11	19.6
2006	NP	0.078	0.718	1.75	1.88	2.73	5.89
	DL	0.77	1.324	2.775	7.301	10.52	59.4
2007	NP	0.078	0.273	0.981	2.11	3.57	11.0
	DL	0.83	1.61	3.58	6.25	9.74	27.21

Note: NP and DL denote nonparametric and Diebold and Li (2006) estimations respectively. (2) $h = 30$ days.

References

Ait-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *The review of financial studies*, 9(2):385–426.

Anderson, N., Breedon, F., Deacon, M., Derry, A., and Murphy, G. (1996). Estimating and interpreting the yield curve, series in financial economics and quantitative analysis.

Andreasen, M. M., Christensen, J. H., and Rudebusch, G. D. (2017). Term structure analysis with big data. Federal Reserve Bank of San Francisco.

Ang, A. and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary economics*, 50(4):745–787.

Ang, A., Piazzesi, M., and Wei, M. (2006). What does the yield curve tell us about GDP growth? *Journal of Econometrics*, 131(1):359–403.

BIS, M. (2005). Zero-coupon yield curves estimated by central banks. *BIS papers*, 25.

Björk, T. (2004). *Arbitrage theory in continuous time*. Oxford University Press.

Bliss, R. R. (1996). Testing term structure estimation methods. Technical report, Working Paper, Federal Reserve Bank of Atlanta.

Breiman, L. (1991). The ii method for estimating multivariate functions from noisy data. *Technometrics*, 33(2):125–143.

Chambers, D. R., Carleton, W. T., and Waldman, D. W. (1984). A new approach to estimation of the term structure of interest rates. *Journal of financial and quantitative analysis*, 19(3):233–252.

Cheng, S. and Tehranchi, M. R. (2018). Polynomial term structure models. *mimeo*.

Christensen, J. H., Diebold, F. X., and Rudebusch, G. D. (2011). The affine arbitrage-free class of nelson–siegel term structure models. *Journal of Econometrics*, 164(1):4–20.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of econometrics*, 92(1):1–45.

Connor, G. and Korajczyk, R. A. (1993). A test for the number of factors in an approximate factor model. *the Journal of Finance*, 48(4):1263–1291.

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

Dai, Q., Singleton, K. J., and Yang, W. (2007). Regime shifts in a dynamic term structure model of US treasury bond yields. *Review of Financial Studies*, 20(5):1669–1706.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians: An Introduction for Econometricians*. Oxford university press.

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., Piazzesi, M., and Rudebusch, G. D. (2005). Modeling bond yields in finance and macroeconomics. *American Economic Review*, 95(2):415–420.

Diebold, F. X. and Rudebusch, G. D. (2013). *Yield curve modeling and forecasting: the dynamic Nelson-Siegel approach*. Princeton University Press.

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

Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates. *Mathematical finance*, 6(4):379–406.

Engle, R. and Rangel, J. (2008). The spline-garch model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies*, 21(3):1187–1222.

Fisher, M., Nychka, D., and Zervos, D. (1995). Fitting the term structure of interest rates with smoothing splines.

Foucault, T., Pagano, M., Roell, A., and Röell, A. (2013). *Market liquidity: theory, evidence, and policy.* Oxford University Press.

Gouriéroux, C., Monfort, A., Polimenis, V., et al. (2002). *Affine Term Structure Model.* INSEE.

Hafner, C. M. and Linton, O. (2010). Efficient estimation of a multivariate multiplicative volatility model. *Journal of econometrics*, 159(1):55–73.

Härdle, W. and Linton, O. (1994). Applied nonparametric methods. *Handbook of econometrics*, 4:2295–2339.

Härdle, W. K. and Majer, P. (2014). Yield curve modeling and forecasting using semiparametric factor dynamics. *The European Journal of Finance*, (ahead-of-print):1–21.

Hays, S., Shen, H., Huang, J. Z., et al. (2012). Functional dynamic factor models with application to yield curve forecasting. *The Annals of Applied Statistics*, 6(3):870–894.

Huse, C. (2011). Term structure modelling with observable state variables. *Journal of Banking & Finance*, 35(12):3240–3252.

Kargin, V. and Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, 99(10):2508–2526.

Koo, B. and Linton, O. (2012). Estimation of semiparametric locally stationary diffusion models. *Journal of Econometrics*, 170(1):210–233.

Koo, B. and Linton, O. (2015). Let's get lade: Robust estimation of semiparametric multiplicative volatility models. *Econometric Theory*, 31(04):671–702.

Lamont, O. A. and Thaler, R. H. (2003). Anomalies: The law of one price in financial markets. *Journal of Economic Perspectives*, 17(4):191–202.

Lee, J. and Robinson, P. M. (2016). Series estimation under cross-sectional dependence. *Journal of Econometrics*, 190(1):1–17.

Li, Q. and Racine, J. S. (2007). *Nonparametric econometrics: theory and practice.* Princeton University Press.

Linton, O., Mammen, E., Nielsen, J. P., and Tanggaard, C. (2001). Yield curve estimation by kernel smoothing methods. *Journal of Econometrics*, 105(1):185–223.

Ma, S., Linton, O. B., and Gao, J. (2017). Estimation in semiparametric quantile factor models.

Mammen, E., Linton, O., Nielsen, J., et al. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *The Annals of Statistics*, 27(5):1443–1490.

McCulloch, J. H. (1971). Measuring the term structure of interest rates. *Journal of Business*, 44:19–31.

McCulloch, J. H. (1975). The tax-adjusted yield curve. *The Journal of Finance*, 30(3):811–830.

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

Nymand-Andersen, P. et al. (2018). Yield curve modelling and a conceptual framework for estimating yield curves: evidence from the european central banks yield curves. Technical report, European Central Bank.

Piazzesi, M. (2005). Bond yields and the federal reserve. *Journal of Political Economy*, 113(2):311–344.

Rebonato, R. and Cooper, I. A. (1997). The limitations of simple two-factor interest rate models. *Journal of Financial Enginneering*, 5:1–16.

Riesz, F. and Nagy, S. (1990). Functional analysis. *Ungar, New York*.

Robinson, P. M. (2012). Inference on power law spatial trends. *Bernoulli*, 18(2):644–677.

Starica, C. (2003). Is garch (1, 1) as good a model as the nobel prize accolades would imply. *Preprint*.

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

Tanggaard, C. (1997). Nonparametric smoothing of yield curves. *Review of Quantitative Finance and Accounting*, 9(3):251–267.

Treasury (2015). The us treasury market on october 15, 2014. Technical report, Joint Staff Report, July.

Vasicek, O. A. and Fong, H. G. (1982). Term structure modeling using exponential splines. *The Journal of Finance*, 37(2):339–348.

Vogt, M. and Linton, O. (2014). Nonparametric estimation of a periodic sequence in the presence of a smooth trend. *Biometrika*, 101(1):121–140.

Yallup, P. J. (2012). Models of the yield curve and the curvature of the implied forward rate function. *Journal of Banking & Finance*, 36(1):121–135.

Supplementary Appendix

A Appendix

A.1 Proofs of Theorems

Proof of Theorem 1. By triangular inequality,

$$|\hat{d}(\tau,x)-d(\tau,x)| \leq \underbrace{|\hat{d}(\tau,x)-\mathrm{E}\hat{d}(\tau,x)|}_{(A.1.1)} + \underbrace{|\mathrm{E}\hat{d}(\tau,x)-d(\tau,x)|}_{(A.1.2)} \tag{31}$$

Let us start with (A.1.2). Recall that the sample criterion function for the estimation of the discount function is given by

$$Q_{nT}(d) = \frac{1}{n_t T}\sum_{t=1}^{T}\sum_{i=1}^{n_t}\int_{\tau_{ik;k=1,\ldots,m_{it}}}\int_x \left\{p_{it} - \sum_{j=1}^{m_{it}} c_{it}(\tau_{ij})d(\tau_{ik},x)\right\}^2 \prod_{k=1}^{m_{it}}\{K_h(\tau_{ik}-\tau_{ij})d\tau_{ik}\}\mathcal{K}_h(x-X_t)dx. \tag{32}$$

We henceforth set $n = \min_t n_t$ and $m_{it} = m_i$ for notational simplicity without loss of much generality. Given that $\hat{d}(\cdot,\cdot)$ can be interpreted as least squares projection of the data in an appropriate function space affiliated with (32), due to the model construction,

$$\mathrm{E}\hat{d}(\tau,x) = \arg\min_v \sum_{t=1}^{T}\sum_{i=1}^{n}\int\int_x \left\{\sum_{j=1}^{m_i} c_{it}(\tau_{ij})\left[d(\tau_{ij},X_t)-v(\tau_{ik},x)\right]\right\}^2 \prod_{k=1}^{m_i}\{K_h(\tau_{ik}-\tau_{ij})d\tau_{ik}\}\mathcal{K}_h(x-X_t)dx$$

This implies that the bias component has the following representation,

$$\begin{aligned}
& \mathrm{E}\hat{d}(\tau,x)-d(\tau,x) \\
= \; & \arg\min_v \sum_{t=1}^{T}\sum_{i=1}^{n}\int\int_x \left(\sum_{j=1}^{m_i} c_{it}(\tau_{ij})\left[d(\tau_{ij},X_t)-v(\tau_{ik},x)-d(\tau_{ik},x)\right]\right)^2 \\
& \times \prod_{k=1}^{m_i}\{K_h(\tau_{ik}-\tau_{ij})d\tau_{ik}\}K_h(x-X_t)dx
\end{aligned} \tag{33}$$

Due to Lemma 3, the first order condition for (33) yields

$$\mathrm{E}\hat{d}(\tau,x)-d(\tau,x) = \mathscr{B}^*_{\hat{d}}(\tau,x) + \int \hat{H}(\tau,\tau',x)\left[\mathrm{E}\hat{d}(\tau',x)-d(\tau',x)\right]d\tau'. \tag{34}$$

where

$$\begin{aligned}
\mathscr{B}^*_{\hat{d}}(x,u) = & \frac{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i} c_{it}^2(\tau_{ij})[d(\tau_{ij},X_t)-d(\tau,x)]K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)}{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i} c_{it}(\tau_{ij})^2 K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)} \\
& + \frac{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{p=1,p\neq j}^{m_i} c_{it}(\tau_{ij})c_{it}(\tau_{ip})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)\int_{\tau'} K_h(\tau'-\tau_{ip})\left[d(\tau_{ip},x)-d(\tau',x)\right]d\tau'}{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i} c_{it}(\tau_{ij})^2 K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)},
\end{aligned}$$

and

$$\hat{H}(\tau,\tau',x) = -\frac{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{p=1,p\neq j}^{m_i} c_{it}(\tau_{ij})c_i(\tau_{ip})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)K_h(\tau'-\tau_{ip})}{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i} c_{it}(\tau_{ij})^2 K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)}.$$

Also, let the bias component $(\mathrm{E}\hat{d}(\tau,x)-d(\tau,x))$ be $\mathscr{B}_{\hat{d}}(\tau,x)$. From Lemma 1 it follows

$$\mathscr{B}_{\hat{d}}(\tau,x) = \sum_{k=0}^{\infty} \hat{\mathcal{H}}^k \mathscr{B}^*_{\hat{d}}(\tau,x). \tag{35}$$

For the stochastic component (A.1.1),

$$\hat{d}(\tau,x) - \mathrm{E}\hat{d}(\tau,x) = \arg\min_v \sum_{t=1}^{T}\sum_{i=1}^{n} \int\int_x \left\{\varepsilon_{it} - \sum_{j=1}^{m_i} c_{it}(\tau_{ij})v(\tau_{ik},x)\right\}^2 \prod_{k=1}^{m_i} \{K_h(\tau_{ik}-\tau_{ij})dx_{ik}\}\,\mathcal{K}_h(x-X_t)dx \tag{36}$$

Due to Lemma 4, the first order condition for (36) yields

$$\hat{d}(\tau,x) - \mathrm{E}\hat{d}(\tau,x) = \mathscr{S}^*_{\hat{d}}(\tau,x) + \int \hat{H}(\tau,\tau',x)\left[\hat{d}(\tau',x) - \mathrm{E}\hat{d}(\tau',x)\right]dx.$$

where

$$\mathscr{S}^*_{\hat{d}}(\tau,x) = \frac{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \varepsilon_{it}c_{it}(\tau_{ij})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)}{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i} c_{it}(\tau_{ij})^2K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)}$$

Again, due to Lemma 1, for the stochastic component $\mathscr{S}_{\hat{d}}(\tau,x) = \hat{d}(\tau,x) - \mathrm{E}\hat{d}(\tau,x)$, we have

$$\mathscr{S}_{\hat{d}}(\tau,x) = \sum_{k=0}^{\infty} \hat{\mathcal{H}}^k \mathscr{S}^*_{\hat{d}}(\tau,x). \tag{37}$$

We need to show that

$$\sqrt{nTh^{L+1}}\mathscr{S}_{\hat{d}}(\tau,x)\mathscr{V}_{\hat{d}}^{-1/2} \xrightarrow{d} \mathscr{N}(0,1) \tag{38}$$

where $\mathscr{V}_{\hat{d}} = (nTh^{L+1})\mathrm{var}\left[\sum_{t=1}^{T}\sum_{i=1}^{n}\omega_{it}(\tau,x)\varepsilon_{it}\right]$. To begin with, we will show that $\mathscr{S}_{\hat{d}}(\tau,x)$ and $\mathscr{S}^*_{\hat{d}}(\tau,x)$ are of the same asymptotic order and this relationship is uniform in τ and x. This allows us to focus on $\mathscr{S}^*_{\hat{d}}(\tau,x)$ instead of $\mathscr{S}_{\hat{d}}(\tau,x)$. Note that

$$\sup_{\substack{\tau\in\chi \\ x\in\mathbb{R}^{L+1}}} |\hat{\mathcal{H}}^k\mathscr{S}^*_{\hat{d}}(\tau,x)| \;\leq\; \sup\left|\int_{\mathbb{R}^+}\hat{H}(\tau,\tau',x)d\tau'\right|^{1/2} \|\hat{\mathcal{H}}^{k-1}\mathscr{S}^*_{\hat{d}}(\tau,x)\|_2 \leq c\rho^k U \tag{39}$$

where $U = \|\hat{\mathcal{H}}\mathscr{S}^*_{\hat{d}}(\tau,x)\|_2/\rho^2$ for $\rho<1$. The first inequality comes from the Cauchy-Schwarz inequality with (23) in Assumption (A4). The second inequality comes from the fact that

$$\|\hat{\mathcal{H}}^k\mathscr{S}^*_{\hat{d}}(\tau,x)\|_2 \leq \rho^{k-1}\|\hat{\mathcal{H}}\mathscr{S}^*_{\hat{d}}(\tau,x)\|_2$$

due to Assumption (A3). Equation (39) implies that

$$\sup_{\substack{\tau \\ x}} |\mathscr{S}_{\hat{d}}(\tau,x)| = \sup\left|\sum_{k=0}^{\infty}\hat{\mathcal{H}}^k\mathscr{S}^*_{\hat{d}}(\tau,x)\right| \leq \sum_{k=0}^{\infty} c\rho^k U = cU/(1-\rho) = o_p\left((nTh^{L+1})^{-1/2}\right) \tag{40}$$

due to (57) in Lemma 5. Equation (40) implies

$$\sup_{\substack{\tau\\x}} \left|\mathscr{S}_{\hat{d}}(\tau,x) - \mathscr{S}^*_{\hat{d}}(\tau,x)\right| = o_p\left((nTh^{L+1})^{-1/2}\right) \tag{41}$$

due to Lemma 1. Because of (41), we have that

$$\sqrt{nTh^{L+1}}\mathscr{S}^*_{\hat{d}}(\tau,x)\mathscr{V}_{\hat{d}}^{-1} \xrightarrow{d} \mathscr{N}(0,1), \tag{42}$$

and

$$\mathscr{S}^*_{\hat{d}}(\tau,x) = \sqrt{h^{L+1}/nT}\sum_{s=1}^{T}\sum_{k=1}^{n} R_{k,s}(\tau,x)\epsilon_{ks},$$

where

$$R_{k,s}(\tau,x) = \frac{\sum_{t=1}^{T}\sum_{i=1}^{n} W_{ikts}\sum_{j=1}^{m_i} c_{it}(\tau_{ij})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)}{(nT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i} c_{it}(\tau_{ij})^2 K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)}.$$

Therefore, the standard central limit theorem can be applied due to assumptions on the error structure and normalisation with rescaling variance, $\mathscr{V}_{\hat{d}}^{-1}$. Combining (34) and (42) completes the proof. □

Proof of Corollary 1. For the asymptotic distribution of the estimator for the yield curve dynamics, $\hat{y}(\tau,x)$, notice that, based on the usual Δ-method, for $\tau > 0$,

$$\hat{y}(\tau,x) - y(\tau,x) = -\frac{\hat{d}(\tau,x) - d(\tau,x)}{\tau d(\tau,x)} + o_p(\|\hat{d}(\tau,x) - d(\tau,x)\|)$$

Then, applying Theorem 1 yields the result. □

Proof of Theorem 2. We modify proofs of Proposition 3 in Conley (1999) under our setup. To begin with, we define

$$V_{\hat{d}} = \left(nTh^{L+1}\right)^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{it}(\tau,x)\omega_{js}(\tau,x)\varepsilon_{it}\varepsilon_{js}. \tag{43}$$

This contrived quantity represents a device that we apply in our proof, whose strategy goes through three steps. Step 1, we prove that $V_{\hat{d}} \xrightarrow{p} E[V_{\hat{d}}]$, as n and T diverge; Step 2, we prove that $E[V_{\hat{d}}] \to \Sigma_{\hat{d}}$; Step 3, we show that $\hat{V}_{\hat{d}} \xrightarrow{p} V_{\hat{d}}$. For the sake of readability of the lengthy proof, we separate each step. In what follows we set $M_1 = o((h^{L+1}n)^{1/3})$ and $M_2 = o((h^{L+1}T)^{1/3})$, while C is a generic constant, which may be different at different places.

Step 1. We have to prove that $V_{\hat{d}} \xrightarrow{p} E[V_{\hat{d}}]$. First, we show that

$$\left(nTh^{L+1}\right)^{-1}\left\{\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_{it}(\tau,x)\omega_{js}(\tau,x)\varepsilon_{it}\varepsilon_{js} - E[\omega_{it}(\tau,x)\omega_{js}(\tau,x)\varepsilon_{it}\varepsilon_{js}]\right\} \xrightarrow{p} 0,$$

via Chebychev's inequality. For fix $(\tau,x) \in \mathbb{R}^{L+1}$, we set

$$G_{it} = \sum_{s=1}^{T}\sum_{j=1}^{n}\{\omega_{it}(\tau,x)\varepsilon_{it}\omega_{js}(\tau,x)\varepsilon_{js} - \mathrm{E}[\omega_{it}(\tau,x)\varepsilon_{it}\omega_{js}(\tau,x)\varepsilon_{js}]\}$$

and we have to prove that $\text{var}(\sum_{t=1}^{T}\sum_{i=1}^{n} G_{it})$ is bounded. To achieve this goal, let us notice that the triangle inequality yields:

$$
\begin{aligned}
\text{var}\left(\sum_{t=1}^{T}\sum_{i=1}^{n} G_{it}\right) \leq & \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{|i-i'|\leq M_1}\sum_{|t-t'|\leq M_2} |\mathrm{E}G_{it}G_{i't'}| \\
& + \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{|i-i'|\geq M_1}\sum_{|t-t'|\leq M_2} |\mathrm{E}G_{it}G_{i't'}| \\
& + \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{|i-i'|\leq M_1}\sum_{|t-t'|\geq M_2} |\mathrm{E}G_{it}G_{i't'}| \\
& + \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{|i-i'|\geq M_1}\sum_{|t-t'|\geq M_2} |\mathrm{E}G_{it}G_{i't'}|,
\end{aligned}
$$

where we split the sum into four terms, according to close/far economic distance and close/far time lag separation[6].

The first term contains the quantities with close economic distance and close time lag separation, as obtained by summing over the indexes $|i-i'| \leq M_1$ and $|t-t'| \leq M_2$. The maximum number of terms within the considered region is $(2M_1+1)(2M_2+1)$. Therefore, letting $\|\cdot\|_q$ indicate the L_q-norm, we have

$$
\begin{aligned}
\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{|i-i'|\leq M_1}\sum_{|t-t'|\leq M_2} |\mathrm{E}G_{it}G_{i't'}| \leq & CnT(2M_1+1)(2M_2+1)\sup_{\substack{i\\t}} \|G_{it}\|_2^2 \\
\leq & CnT(2M_1+1)(2M_2+1)L_{nT}
\end{aligned}
\tag{44}
$$

where $L_{nT} = (M_1+1)^2(M_2+1)^2$ the last inequality follows from the fact that Assumption (C1) and Minkowski's inequality imply that, for $\zeta > 0$,

$$
\sup_{i,t} \|G_{it}\|_{2+\zeta}^2 \leq CL_{nT}. \tag{45}
$$

Thus, Holder's inequality implies $\sup_{i,t} \|G_{it}\|_2^2 \leq CL_{nT}$.

For the second term, where $|i-i'| \geq M_1$ and $|t-t'| \leq M_2$, we have

$$
\begin{aligned}
\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{|i-i'|\geq M_1}\sum_{|t-t'|\leq M_2} |\mathrm{E}G_{it}G_{i't'}| \leq & Cn(2M_2+1)\sum_{i,i'} \phi(|i-i'|)^{1/2}\mathrm{E}|G_{it}|^{2+\zeta}\mathrm{E}|G_{i't'}|^{2+\zeta} \\
\leq & Cn(2M_2+1)\sum_{i,i'} e^{-\lambda_1(|i-i'|)/2}\mathrm{E}|G_{it}|^{2+\zeta}\mathrm{E}|G_{i't'}|^{2+\zeta} \\
\leq & Cn(2M_2+1),
\end{aligned}
\tag{46}
$$

where the first inequality follows from the mixing inequality as in Corollary 14.5 in Davidson (1994) and third inequality follows from (C1), as in (45).

[6] We refer to Conley (1999) and in particular to the remark after Eq. (A.5) in that paper.

The third term is similar to the second term (but for the fact that we consider far time lag distances rather than far economic distances) and hence the following relationship holds:

$$
\begin{aligned}
\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{|i-i'|\leq M_1}\sum_{|t-t'|\geq M_2}|\mathrm{E}G_{it}G_{i't'}| \leq & CT(2M_1+1)\sum_{t,t'}\phi(|t-t'|)^{1/2}\mathrm{E}|G_{it}|^{2+\zeta}\mathrm{E}|G_{i't}|^{2+\zeta} \\
\leq & CT(2M_1+1)\sum_{t,t'}e^{-\lambda_1(|t-t'|)/2}\mathrm{E}|G_{it}|^{2+\zeta}\mathrm{E}|G_{i't'}|^{2+\zeta} \\
\leq & CT(2M_1+1).
\end{aligned}
$$

The last term involves the quantities with far distances, say $|i-i'|\geq M_1$ and $|t-t'|\geq M_2$, so:

$$
\begin{aligned}
&\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{|i-i'|\geq M_1}\sum_{|t-t'|\geq M_2}|\mathrm{E}G_{it}G_{i't'}| \\
\leq & CL_{nT}\left[\sum_{i,i'}\phi(|i-i'|)^{1/2}\mathrm{E}|G_{it}|^{2+\zeta}\mathrm{E}G_{i't}|^{2+\zeta}\right]\left[\sum_{t,s}\phi(|t-t'|)^{1/2}\mathrm{E}|G_{it}|^{2+\zeta}\mathrm{E}|G_{it'}|^{2+\zeta}\right] \\
\leq & CL_{nT}\left[\sum_{i,i'}e^{-\lambda_1(|i-j|)/2}\mathrm{E}|G_{it}|^{2+\zeta}\mathrm{E}|G_{jt}|^{2+\zeta}\right]\left[\sum_{t,s}e^{-\lambda_1(|t-t'|)/2}\mathrm{E}|G_{it}|^{2+\zeta}\mathrm{E}|G_{it'}|^{2+\zeta}\right] \\
\leq & CL_{nT}\max\left(\sum_{i,i'}e^{-\lambda_1(|i-i'|)},\sum_{t,t'}e^{-\lambda_1(|t-t'|)}\right)\sup_{i,t}\mathrm{E}|G_{it}|^{4+\zeta}\leq CL_{nT},
\end{aligned}
\tag{47}
$$

with the last inequality following from (C1). Combining all the above results, Chebychev's inequality yields, for any positive number ξ,

$$
\Pr\left\{\left|\frac{1}{nTh^{L+1}}\sum_{t=1}^{T}\sum_{i=1}^{n}G_{it}\right|\geq\xi\right\}\leq\frac{1}{\xi^2}\frac{1}{(nTh^{L+1})^2}\mathrm{var}\left(\sum_{t=1}^{T}\sum_{i=1}^{n}G_{it}\right)
$$
$$
\leq\frac{1}{\xi^2}\left[\frac{CnT(2M_1+1)(2M_2+1)L_{nT}+Cn(2M_2+1)+CT(2M_1+1)+CL_{nT}}{(nTh^{L+1})^2}\right].
$$

The summands in the RHS of the last inequality go all to zero and this implies

$$
\frac{1}{nTh^{L+1}}\sum_{t=1}^{T}\sum_{i=1}^{n}G_{it}\xrightarrow{p}0.
\tag{48}
$$

Now remark that

$$
V_{\hat{d}}-E[V_{\hat{d}}]=\frac{1}{nTh^{L+1}}\sum_{t=1}^{T}\sum_{i=1}^{n}G_{it},
$$

thus (48) yields $V_{\hat{d}}\xrightarrow{p}E[V_{\hat{d}}]$, as n and T diverge.

Step 2. We have to prove that $E[V_{\hat{d}}]\to\Sigma_{\hat{d}}$, as n and T diverge. Notice that, from (43), it follows

$$
E[V_{\hat{d}}]=\left(nTh^{L+1}\right)^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n}E[\omega_{it}(\tau,x)\varepsilon_{it}\omega_{js}(\tau,x)\varepsilon_{js}],
\tag{49}
$$

thus, moving along the same lines as in Step 1, we have:

$$
\begin{aligned}
E[V_{\hat{d}}] - \text{var}\left(\frac{1}{\sqrt{nTh^{L+1}}}\sum_{t=1}^{T}\sum_{i=1}^{n}\omega_{it}(\tau,x)\varepsilon_{it}\right) \leq\ & \tfrac{1}{nTh^{L+1}}\Big[\textstyle\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{|i-i'|\leq M_1}\sum_{|t-t'|\leq M_2}|\mathrm{E}\varepsilon_{it}\varepsilon_{i't'}| \\
& +\textstyle\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{|i-i'|\geq M_1}\sum_{|t-t'|\leq M_2}|\mathrm{E}\varepsilon_{it}\varepsilon_{i't'}| \\
& +\textstyle\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{|i-i'|\leq M_1}\sum_{|t-t'|\geq M_2}|\mathrm{E}\varepsilon_{it}\varepsilon_{i't'}| \\
& +\textstyle\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{|i-i'|\geq M_1}\sum_{|t-t'|\geq M_2}|\mathrm{E}\varepsilon_{it}\varepsilon_{i't'}|\Big].
\end{aligned} \tag{50}
$$

We remark that Assumption (C1) and Minkowski's inequality imply that, for $\zeta > 0$, $\sup_{i,t}\|\varepsilon_{it}\|_{2+\zeta}^2 \leq CL_{nT}$; thus, Holder's inequality implies $\sup_{i,t}\|\varepsilon_{it}\|_2^2 \leq CL_{nT}$. Therefore, moving along the same lines as in Step 1, we apply the same arguments as in (44)-(47) to prove that the four summands in the RHS of the inequality (50) converges to zero, as n and T diverge. This yields $E[V_{\hat{d}}] \to \Sigma_{\hat{d}}$, as n and T diverge.

Step 3. Finally, we have show that $\hat{V}_{\hat{d}} \xrightarrow{p} V_{\hat{d}}$, as n and T diverge. This result follows from Theorem 1, in tandem with the continuous mapping theorem.

The combination of the results in the three steps above concludes the proof.

□

A.2 Lemmata

Lemma 1 (Neumann series). *Let $\hat{\mathcal{H}}$ be an operator such that, for any square summable function $g(\cdot) \in L^2$, $\hat{\mathcal{H}}g(\cdot) = \int \hat{H}(\cdot,t)g(t)dt$ with a square-summable kernel $\hat{H}(\cdot,\cdot) \in L^2$. Also, we have an integral equation of the following form.*

$$\hat{g}(\tau) = \bar{g}(\tau) + \int \hat{H}(\tau,t)\hat{g}(t)dt$$

Then, provided that $\|\hat{\mathcal{H}}\| < 1$, the above integral equation has the solution such that

$$\hat{g}(\tau) = (I-\hat{\mathcal{H}})^{-1}\bar{g}(\tau) = \sum_{k=1}^{\infty}\hat{\mathcal{H}}^k\bar{g}(\tau)$$

where $\hat{\mathcal{H}}^k$ is defined by the recurrence relation such that $\hat{\mathcal{H}}^k g(\cdot) = \hat{\mathcal{H}}(\hat{\mathcal{H}}^{k-1}g(\cdot))$.

Proof.

$$
\begin{aligned}
\hat{g}(x) &= \bar{g}(x) + \int \hat{H}(x,t)\hat{g}(t)dt \\
\hat{g}(x) - \int \hat{H}(x,t)\hat{g}(t)dt &= \bar{g}(x) \\
(I-\hat{\mathcal{H}})\hat{g}(x) &= \bar{g}(x)
\end{aligned}
$$

The desired result follows once you notice that the inverse of $I-\hat{\mathcal{H}}$ has an infinite series of the polynomial of $\hat{\mathcal{H}}$ such that

$$\left(I-\hat{\mathcal{H}}\right)^{-1} = \sum_{k=0}^{\infty}\hat{\mathcal{H}}^k.$$

See Sections 66 and 67 in Riesz and Nagy (1990) for more details. $\square$

Lemma 2. *Suppose*

$$\hat{d}(\tau, x) = \arg\min_{d \in \mathcal{D}} Q(d(\tau, x))$$

where

$$Q(d) = \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \int_{\tau_{ik;k=1,\dots,m_i}} \int_x \left\{ p_{it} - \sum_{j=1}^{m_i} c_{it}(\tau_{ij}) d(\tau_{ik}, x) \right\}^2 \prod_{k=1}^{m_i} \{K_h(\tau_{ik} - \tau_{ij}) d\tau_{ik}\} \mathcal{K}_h(x - X_t) dx. \tag{51}$$

Then, $\hat{d}(\tau, x)$ has the following representation.

$$\hat{d}(\tau, x) = \bar{d}(\tau, x) + \int \hat{H}(\tau, \tau', x) \hat{d}(\tau', x) d\tau'$$

where

$$\bar{d}(\tau, x) = \frac{\sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j=1}^{m_{it}} p_{it} c_{it}(\tau_{ij}) K_h(\tau - \tau_{ij}) \mathcal{K}_h(x - X_t)}{\sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j=1}^{m_i} c_{it}(\tau_{ij})^2 K_h(\tau - \tau_{ij}) \mathcal{K}_h(x - X_t)} \tag{52}$$

$$\hat{H}(\tau, \tau', x) = -\frac{\sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{p=1, p \neq j}^{m_i} c_{it}(\tau_{ij}) c_{it}(\tau_{ip}) K_h(\tau - \tau_{ij}) K_h(\tau' - \tau_{ip}) \mathcal{K}_h(x - X_t)}{\sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j=1}^{m_i} c_{it}(\tau_{ij})^2 K_h(\tau - \tau_{ij}) \mathcal{K}_h(x - X_t)} \tag{53}$$

Proof. We can think of the minimizer of $Q(d)$ in a way that any slight deviation from the minimizer in any direction would increase the value of $Q(d)$. That is, the derivative of $Q(d)$ with respect to any slight deviation from the minimizer, say η must be zero in all direction. To formalize this, we consider a Gateaux derivative. Let $d(\tau_{ij}, x) = \hat{d}(\tau_{ij}, x) + \eta \delta_{(\tau,x)}(\tau_{ij}, x)$, where η is any arbitrarily small positive number. Then, solving (51) is equivalent to solving the problem

$$\arg\min_{\eta} \sum_{t=1}^{T} \sum_{i=1}^{n} \int_{\tau_{i\cdot}} \int_x \left\{ p_{it} - \sum_{j=1}^{m_i} c_{it}(\tau_{ij}) [\hat{d}(\tau_{ik}, x) + \eta \delta_{(\tau,x)}(\tau_{ik}, x)] \right\}^2 \prod_{k=1}^{m_i} \{K_h(\tau_{ik} - \tau_{ij}) d\tau_{ik}\} \mathcal{K}_h(x - X_t) dx \tag{54}$$

where $\int_{\tau_{i\cdot}} = \int_{\tau_{ik;k=1,\dots,m_i}}$. For the ease of notation, in the following, we drop $\tau_{i\cdot}$. Take the derivative of equation (54) with respect to η at the point $\eta = 0$ yields the first order condition

$$\begin{aligned} & \sum_{t=1}^{T} \sum_{i=1}^{n} \int \int_x p_{it} \left[\sum_{j=1}^{m_i} c_{it}(\tau_{ij}) \delta_{(\tau,x)}(\tau_{ik}, x) \right] \prod_{k=1}^{m_i} \{K_h(\tau_{ik} - \tau_{ij}) d\tau_{ik}\} \mathcal{K}_h(x - X_t) dr \\ = & \sum_{t=1}^{T} \sum_{i=1}^{n} \int \int_x \left[\sum_{j=1}^{m_i} c_{it}(\tau_{ij}) \hat{v}(\tau_{ik}, r) \right] \left[\sum_{p=1}^{m_i} c_{it}(\tau_{ip}) \delta_{(\tau,x)}(\tau_{ip}, x) \right] \prod_{k=1}^{m_i} \{K_h(\tau_{ik} - \tau_{ij}) d\tau_{ik}\} \mathcal{K}_h(x - X_t) dx. \end{aligned} \tag{55}$$

Note that $\int K_h(\tau_{ik}-\tau_{ij})d\tau_{ik}=1$, $\int \delta_x(u)f(u)du=f(x)$ for any generic function, which is continuous at x and $\delta_{(x,u)}(\cdot,\cdot)=\delta_{(x)}(\cdot)\delta_{(u)}(\cdot)$. From these facts, (55) becomes

$$\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i} c_{it}(\tau_{ij})^2 K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)\hat{d}(\tau,x)-\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i} p_{it}c_{it}(\tau_{ij})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)$$
$$=-\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{\substack{p=1\\p\neq j}}^{m_i} c_{it}(\tau_{ij})c_{it}(\tau_{ip})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)\int_{\tau'}\hat{d}(\tau',x)K_h(\tau'-\tau_{ip})d\tau'$$

Rearranging it completes the proof. □

Lemma 3. *The first order condition for (33) is given as*

$$\mathrm{E}\hat{d}(\tau,x)-d(\tau,x)=\mathscr{B}^*_{\hat{d}}(\tau,x)+\int \hat{H}(\tau,\tau',x)\left[\mathrm{E}\hat{d}(\tau',x)-d(\tau',x)\right]d\tau'.$$

where

$$\mathscr{B}^*_{\hat{d}}(\tau,x)=\frac{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i} c_{it}(\tau_{ij})^2[d(\tau_{ij},X_t)-d(\tau,x)]K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)}{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i} c_{it}(\tau_{ij})^2K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)}$$
$$+\frac{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{p=1,p\neq j}^{m_i} c_{it}(\tau_{ij})c_{it}(\tau_{ij})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)\int K_h(\tau'-\tau_{ip})\left[d(\tau_{ip},x)-d(\tau',x)\right]d\tau'}{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i} c_{it}^2(\tau_{ij})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)}$$

and

$$\hat{H}(\tau,\tau',x)=-\frac{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{p=1,p\neq j}^{m_i} c_{it}(\tau_{ij})c_{it}(\tau_{ip})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)K_h(\tau'-\tau_{ip})}{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i} c_{it}(\tau_{ij})^2K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)}.$$

Proof. Note that

$$\begin{aligned}&\mathrm{E}\hat{d}(\tau,x)-d(\tau,x)\\=\ &\arg\min_v\sum_{t=1}^{T}\sum_{i=1}^{n}\int\int_x\left(\sum_{j=1}^{m_i} c_{it}(\tau_{ij})\left[d(\tau_{ij},X_t)-v(\tau_{ik},x)-d(\tau_{ik},x)\right]\right)^2\\\times\ &\prod_{k=1}^{m_i}\{K_h(\tau_{ik}-\tau_{ij})d\tau_{ik}\}\mathcal{K}_h(x-X_t)dx.\end{aligned}\tag{56}$$

Replace $v(\tau_{ik},x)$ with $\hat{v}(\tau_{ik},x)+\eta\delta_{(\tau,x)}(\tau_{ik},x)$ and differentiate (56) with respect to η at $\eta=0$.

$$\sum_{t=1}^{T}\sum_{i=1}^{n}\int\int_x\left[\sum_{j=1}^{m_i} c_{it}(\tau_{ij})\hat{v}(\tau_{ik},x)\right]\left[\sum_{p=1}^{m_i} c_{it}(\tau_{ip},X_t)\delta_{(\tau,x)}(\tau_{ip},x)\right]\prod_{k=1}^{m_i}\{K_h(\tau_{ik}-\tau_{ij})d\tau_{ik}\}\mathcal{K}_h(x-X_t)dr$$
$$=\sum_{t=1}^{T}\sum_{i=1}^{n}\int\int_x\left[\sum_{j=1}^{m_i} c_{it}(\tau_{ij})[d(\tau_{ij},X_t)-d(\tau_{ik},x)]\right]\left[\sum_{p=1}^{m_i} c_{it}(\tau_{ip})\delta_{(\tau,x)}(\tau_{ip},x)\right]$$
$$\times\prod_{k=1}^{m_i}\{K_h(\tau_{ik}-\tau_{ij})d\tau_{ik}\}\mathcal{K}_h(x-X_t)dx.$$

Using $\int K_h(\tau_{ik}-\tau_{ij})d\tau_{ik}=1$ and $\int \delta_x(u)f(u)du = f(x)$ for any generic function, which is continuous at x yields

$$
\begin{aligned}
& \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i} c_{it}(\tau_{ij})^2 \hat{v}(x,u) K_h(\tau-\tau_{ij})K_h(x-X_t) \\
+ & \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{\substack{p=1\\p\neq j}}^{m_i} c_{it}(\tau_{ij})c_{it}(\tau_{ip})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)\int_{\tau'}\hat{v}(\tau',x)K_h(\tau'-\tau_{ip})d\tau' \\
= & \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i} c_{it}(\tau_{ij})^2[d(\tau_{ij},X_t)-d(\tau,x)]K_h(\tau-\tau_{ij})\mathcal{K}_h(u-u_t) \\
+ & \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{\substack{p=1\\p\neq j}}^{m_i} c_{it}(\tau_{ij})c_{it}(\tau_{ip})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)\int_{\tau'}[d(\tau_{ij},x)-d(\tau',x)]K_h(\tau'-\tau_{ip})d\tau'
\end{aligned}
$$

where $\hat{v}(x,u)=\mathrm{E}\hat{d}(x,u)-d(x,u)$ by construction. Arranging the above equation completes the proof. □

Lemma 4. *The first order condition for (36) is given as*

$$\hat{d}(\tau,x)-\mathrm{E}\hat{d}(\tau,x)=\mathscr{S}^*_{\hat{d}}(\tau,x)+\int \hat{H}(\tau,\tau',x)\left[\hat{d}(\tau',x)-\mathrm{E}\hat{d}(\tau',x)\right]d\tau'.$$

where

$$\mathscr{S}^*_{\hat{d}}(\tau,x)=\frac{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\varepsilon_{it}c_{it}(\tau_{ij})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)}{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i}c_{it}(\tau_{ij})^2K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)}$$

and

$$\hat{H}(\tau,\tau',x)=-\frac{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{j=1,p\neq j}^{m_i}c_{it}(\tau_{ij})c_{it}(\tau_{ip})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)K_h(\tau'-\tau_{ip})}{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i}c_{it}(\tau_{ij})^2K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)}$$

Proof. The proof is identical to the one for Lemma 2 once you notice that the only difference between the objective functions (51) and (36) is $p_i(u_t)$ for (51) and ε_{it} (36) respectively. □

Lemma 5. *Consider $\mathscr{S}_{\hat{d}}(\tau,x)=\hat{d}(\tau,x)-\mathrm{E}\hat{d}(\tau,x)$ and $\mathscr{S}^*_{\hat{d}}(\tau,x)$ given in Lemma 4. Then, for any arbitrarily small $\varsigma>0$,*

$$\|\hat{\mathcal{H}}\mathscr{S}^*_{\hat{d}}\|_2 = o_p\left((nTh^{L+1})^{-1/2}\right) \tag{57}$$

$$\sup_{\tau\in\chi,x\in[\varsigma,1-\varsigma]}|\hat{\mathcal{H}}\mathscr{S}^*_{\hat{d}}(\tau,x)| = o_p\left((nTh^{L+1})^{-1/2}\right) \tag{58}$$

where $\|f\|_2=\left(\int f(\tau)^2d\tau\right)^{1/2}$.

Proof. Recall that $\varepsilon_{it}=\sum_{s=1}^{T}\sum_{k=1}^{n}W_{ik,ts}\epsilon_{ks}$ where ϵ_{ks} is an i.i.d random variable, and that

$$R_{k,s}(\tau,x)=\frac{\sum_{t=1}^{T}\sum_{i=1}^{n}W_{ikts}\sum_{j=1}^{m_i}c_{it}(\tau_{ij})K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)}{(nT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m_i}c_{it}(\tau_{ij})^2K_h(\tau-\tau_{ij})\mathcal{K}_h(x-X_t)}.$$

Then, by the definition of the operator $\hat{\mathcal{H}}$ as in (13), we have

$$\begin{aligned}
\mathrm{E}\|\hat{\mathcal{H}}\mathscr{S}_{\hat{d}}^*(\tau,x)\|_2^2 &= \mathrm{E}\frac{1}{(nT)^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{k=1}^{n}\sum_{m=1}^{n}\int_\tau\int_q\int_p \hat{H}(\tau,q,x)\hat{H}(\tau,p,x)R_{k,t}(q,x)R_{m,s}(p,x)dp\ dq\ d\tau\ \epsilon_{kt}\epsilon_{ms} \\
&\leq \frac{1}{nT}\sum_{t=1}^{T}\sum_{k=1}^{n}\int_\tau\int_q\int_p \hat{H}(\tau,q,x)\hat{H}(\tau,p,x)R_{k,t}(q,x)R_{k,t}(p,x)dp\ dq\ d\tau\ \mathrm{E}[\epsilon_{kt}^2] \\
&\leq \frac{C}{nT}\int_\tau\left(\int_q \hat{H}(\tau,q,x)R_{k,t}(q,x)dq\right)^2 d\tau = o_p\left(\left(nTh^{L+1}\right)^{-1}\right)
\end{aligned}$$

The first equality comes from the deterministic τ_{ij}, W_{itjs}, time, m_i and cash flow[7]. The second inequality comes from Cauchy-Schwartz inequality. Finally, the use of (22) (as in Assumption (A4)) completes the proof for (57).

For (58), note that

$$\hat{\mathcal{H}}\mathscr{S}_{\hat{d}}^*(\tau,x) = \frac{1}{nT}\sum_{t=1}^{T}\sum_{k=1}^{n}\mathfrak{R}_{k,t}\epsilon_{kt},$$

where $\mathfrak{R}_{k,t} = \int \hat{H}(\tau,q,x)R_{k,t}(q,x)dq$. Note that $\mathrm{E}\hat{\mathcal{H}}\mathscr{S}_{\hat{d}}^*(\tau,x) = 0$.

Let $S_{n,T} = \sum_{t=1}^{T}\sum_{k=1}^{n}\mathfrak{R}_{k,t}$, $S_{n-1,T} = \sum_{t=1}^{T}\sum_{k=1}^{n-1}\mathfrak{R}_{k,t}$, and $S_{n,T-1} = \sum_{t=1}^{T-1}\sum_{k=1}^{n}\mathfrak{R}_{k,t}$ and so on. Then, due to (A4) and Chernoff's bounding method, for a positive value c and ξ,

$$\begin{aligned}
\Pr\left((nTh^{L+1})^{1/2}\hat{\mathcal{H}}\mathscr{S}_{\hat{d}}^*(\tau,x) > \xi\right) &\leq \mathrm{E}\left(\exp\left\{\lambda\left[(nTh^{L+1})^{1/2}\hat{\mathcal{H}}\mathscr{S}_{\hat{d}}^*(\tau,x) - \xi\right]\right\}\right) \\
&\leq e^{-\lambda\xi}\mathrm{E}\left(\exp\left(\lambda(h^{L+1}/nT)^{1/2}\sum_{t=1}^{T}\sum_{k=1}^{n}\mathfrak{R}_{k,t}\epsilon_{kt}\right)\right) \\
&\leq e^{-\lambda\xi}\mathrm{E}\left(\exp\left(\lambda(h^{L+1}/nT)^{1/2}\sum_{t=1}^{T}\sum_{k=1}^{n-1}\mathfrak{R}_{k,t}\epsilon_{kt}\right)\right)e^{\lambda^2c_\alpha^2/8} \\
&\leq e^{-\lambda\xi}\mathrm{E}\left(\exp\left(\lambda(h^{L+1}/nT)^{1/2}\sum_{t=1}^{T-1}\sum_{k=1}^{n-1}\mathfrak{R}_{k,t}\epsilon_{kt}\right)\right)e^{\lambda^2c_\alpha^2/8}e^{\lambda^2c_\beta^2/8} \\
&\leq e^{-\lambda\xi}e^{\lambda^2(nc_\alpha^2+Tc_\beta^2)/8} = e^{-2\xi^2/(nc_\alpha^2+Tc_\beta^2)} = o_p(1),
\end{aligned}$$

where we set $\lambda = 4\xi/(nc_\alpha^2+Tc_\beta^2)$ and the feature of independent ϵ_{kt} has been applied. Meanwhile, replacing $(nTh^{L+1})^{1/2}\hat{\mathcal{H}}\mathscr{S}_{\hat{d}}^*(\tau,x)$ by $-(nTh^{L+1})^{1/2}\hat{\mathcal{H}}\mathscr{S}_{\hat{d}}^*(\tau,x)$ yields the following result

$$\Pr\left((nTh^{L+1})^{1/2}\hat{\mathcal{H}}\mathscr{S}_{\hat{d}}^*(\tau,x) < -\xi\right) = o_p(1).$$

Combining results gives

$$\Pr\left(|\hat{\mathcal{H}}\mathscr{S}_{\hat{d}}^*(\tau,x)| > \xi\right) = o_p\left((nTh^{L+1})^{-1/2}\right). \tag{59}$$

Therefore, (58) follows. □

[7]This can be relaxed with relative ease to include also stochastic quantities. However the generalization will entail a much more cumbersome nation, without yielding any additional insights. Therefore, we do not cover it here.