ARTICLE IN PRESS

Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Solving Euler equations via two-stage nonparametric penalized splines

Liyuan Cui [a], Yongmiao Hong [b,c], Yingxing Li [d,*]

[a] Department of Economics and Finance, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Hong Kong
[b] Department of Economics and Department of Statistics & Data Science, Cornell University, 424 Uris Hall, Ithaca, N.Y. 14853, USA
[c] Gregory and Paula Chow Center for Economic Research, Xiamen University, China
[d] The Wang Yanan Institute of Studies in Economics, and MOE Key Laboratory of Econometrics, Xiamen University, China

ARTICLE INFO

Article history:
Received 23 November 2018
Received in revised form 24 March 2020
Accepted 26 April 2020
Available online xxxx

JEL classification:
C1
C3
C4
C5
E1
G12

Keywords:
Euler equation
Nonparametric penalized splines
Two-stage regression
Return predictability
Implied price–dividend ratios

ABSTRACT

This study proposes a novel estimation-based approach to solving asset pricing models for both stationary and time-varying observations. Our method is robust to misspecification errors while inheriting a closed-form solution. By representing the Euler equation into a well-posed integral equation of the second kind, we propose a penalized two-stage nonparametric estimation method and establish its optimal convergence under mild conditions. With the merit of penalized splines, our estimate is less sensitive to the spline setting and we also design a fast data-driven algorithm to effectively tune the key smoother, i.e. the penalty amount. Our approach exhibits excellent finite sample performance. Using the US data from 1947 to 2017, we reinvestigate the return predictability and find that the estimated implied dividend yield significantly predicts lower future cash flows and higher interest rates at short horizons.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Since the seminal work by Lucas (1978), Euler equations have been widely adopted as a main vehicle in finance and macroeconomics to investigate the connection between agent preferences, asset prices, and economic fundamentals (Cochrane, 2009). Enormous efforts have been seen on the development of asset pricing models and there is a pressing interest in evaluating different models' explanation powers on market anomalies. Whereas, the price–dividend (P/D) ratio function, which is a function of state variables and recursively specified in Euler equations, plays an indispensable role in constructing test statistics and moment conditions. An unsatisfactory solution of the P/D ratio function from the Euler equations may lead to incomplete or misleading information about its accounting identity, and may thereby contaminate the predictability of the dividend yields on stock returns, cash flows, and short-term interest rates (Ang and Bekaert, 2006; Elliott et al., 2015; Xu, 2018). Therefore, among others, the P/D function from a given Euler equation, is one of the central quantities to solve (Mehra and Prescott, 1985; Campbell and Cochrane, 2000). Given the fact that economic theory usually does not suggest a concrete functional form of the state dynamics and unknown P/D ratios, one goal of this strand

* Corresponding author.
E-mail addresses: liyuan.cui@cityu.edu.hk (L. Cui), yh20@cornell.edu (Y. Hong), yxli@xmu.edu.cn (Y. Li).

https://doi.org/10.1016/j.jeconom.2020.04.042
0304-4076/© 2020 Elsevier B.V. All rights reserved.

Please cite this article as: L. Cui, Y. Hong and Y. Li, Solving Euler equations via two-stage nonparametric penalized splines. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.04.042.

of literature is to strike a balance between solution accuracy and robustness of the distributional assumption imposed on the state variables.

The literature on the solution of Euler equations begins with numerical approaches, which impose auxiliary assumptions on the unknown functions, or state dynamics, or both for ease of computation. Traditional perturbation methods might miss nonlinear patterns or curvature for the P/D ratio function (Pohl et al., 2018). Given a pre-specified data generating process (DGP), global approximations of the P/D ratio function are considered, but few theoretical derivations on the approximation bounds exist, and the practical choice of the series order is considerably subjective (Calin et al., 2005). By assuming autoregressive (AR) processes with Gaussian shocks on the state variables, discretization methods do not rely on the functional-form assumption, but may induce interpolation biases. Meanwhile, the widely used AR(1) assumption on the consumption growth rates might be inadequate, since Cecchetti et al. (2000) find strong evidence for thresholding models in reflecting different evolutions of the state dynamics in different states of the economy, and Bansal and Yaron (2004) and Andrei and Hasler (2014) show that the equity premium puzzle may be better explained when considering autoregressive conditional heteroscedasticity (ARCH) and time-varying processes. Traditional regression-based methods, such as the parametrized expectation algorithm (PEA), might be among the few methods that do not require distributional assumptions on the state variables (Den Haan and Marcet, 1990). However, they require the pre-specified functional form of the P/D ratio, and the approach may not converge during iterations as they attempt to treat the Euler equation as the integral equation of the first kind (Type I) (Canay et al., 2013). Calin et al. (2005) directly estimate the P/D ratio by assuming that it is an analytic nonparametric function, but the existence of the solution is not always guaranteed for some chosen stochastic discount factors (SDFs).

In this paper, we propose a nonparametric estimation-based solution method for Euler equations, which enjoys a closed-form solution without numerical integrations and avoids misspecification errors from the traditionally imposed auxiliary assumptions made on latent transition densities or unknown functional forms. We add insights of our novel regression method by representing the Euler equation as a well-posed integral equation of the second kind (Type II) and view it from a regression perspective. Due to the recursive occurrences of the unknown function over two time periods, we adopt the two-stage nonparametric regression to solve the endogeneity issue. By establishing an equivalence between our estimation-based procedure and the projection method widely used in solving Type II equations, we demonstrate that our proposed estimator can achieve the optimal convergence rate, as opposed to the slower one of a general ill-posed Type I problem. By adopting series estimation and regression techniques,[1] our estimation method inherits a closed-form solution and we avoid solving the more involved Perron–Frobenius eigenfunction representation as that in Escanciano et al. (2020) and Christensen (2017). In particular, we recommend the use of the B-splines basis rather than other global basis functions, such as the power series and the Chebyshev series, because the B-splines basis has an appealing local modification scheme[2] and is more numerically stable.

Our paper further contributes to the literature by establishing the large sample properties of our proposed two-stage estimation with and without the penalization smoother to control the roughness of the fitted curve. A closely related work is Chen and Pouzo (2015), who propose estimating a general nonparametric conditional moment model via penalized sieve minimum distance. However, our work differs essentially from theirs as we focus more on adopting the penalization to play the key role of smoothing. Since we do not rely on the number of the spline basis for regularization, the impact from the spline setting is substantially reduced, which frees us from worrying about the choice of the spline basis and the placement of knots. There are relatively few other studies that investigate how to tune the penalty so that the estimate could achieve the optimal convergence rate. Exceptions include Li and Ruppert's (2008) and Claeskens et al.' (2009) studies on simple univariate nonparametric regressions, and Chen et al. (2015)'s work on varying coefficient models with an integrated regressor. However, to the best of our knowledge, we are the first to establish theoretical results for nonparametric regressions with endogeneity. We enjoy the advantages of our proposed smoothing via a combined regularization scheme, where we can achieve undersmoothing in the first stage while maintaining the optimal smoothing in the second stage.

Our theoretical contributions also shed light on its practical implementation. As Newey and Powell (2003) point out, it is very important to choose the smoothing parameters in two-stage nonparametric regression. To ensure the reliable performance of our new methodology, we propose a fast-implemented generalized cross-validation (GCV) algorithm to select the key smoother, namely, the penalty parameter in the two-stage regression. Normally, the calculation of GCV for each given penalty requires the matrix inversion, which, in turn, requires a computation cost of K^3, where K is equivalent to the total number of spline basis. However, our implemented algorithm can further reduce this computation to $O(K^2)$, thus allowing for a faster search for the optimal smoother. It is worth mentioning that our newly proposed GCV algorithm is not limited to asset pricing modeling. Moreover, such an approach can be extended to allow for multivariate state variables with the use of tensor product B-splines.

Simulation studies confirm the excellent performance of our estimation even when the sample size is not large, say 100 in univariate cases or 500 in bivariate cases. Our simulation are carried out based on three main objectives. First, given true P/D ratio functions, which are designed to be periodic, non-periodic or multivariate, we examine the solution accuracy of our proposed estimator under an extensive class of stationary and time-varying state dynamics. Second, under Mehra and

1 Unlike local smoothing methods, we do not require a proper choice of steady states that might be difficult to determine (Juillard, 2011).

2 When adding or changing very few observations, one could modify the fitted curve locally without globally changing the whole shape.

Prescott's (1985) path-breaking model wherein an analytic solution may exist under certain circumstances, we conduct a set of comparison studies with current popular numerical methods. Specifically, when the imposed DGP specification of the state dynamics is correct, we find that our method still demonstrates competitive performance, in terms of accuracy, computational time, and number of iterations, with existing numerical methods. When allowing possible misspecifications on the state dynamics, simulation studies confirm that our method is more robust against misspecification errors. Third, in line with Mehra and Prescott (1985) and Campbell and Cochrane (1999), we relax all auxiliary assumptions on the dynamics of state variables, and reinvestigate these models' economic implications and economically significant benefits that the real data could offer in general equilibrium asset pricing models beyond these models' current findings. By doing so, we find that the correlation between the model-implied risk-free and risky returns is largely reduced and document negatively skewed equity premiums, which are both closer to real data counterparts.

In the real data analysis, we obtain a data-driven nonparametric estimate of the P/D function and its reciprocal, called implied dividend yields under Mehra and Prescott's (1985) and Campbell and Cochrane' (1999) models. This implied dividend yield offers a clean reflection of the rational model-implied measure for discounted future cash flows without misspecifications. Using this measure, we re-examine and reinvestigate the predictability of dividend yields on stock returns. Our main empirical findings are as follows. First, in line with Ang and Bekaert (2006), we confirm that the rational model-implied dividend yield is not a direct predictor for excess returns, but we do document significant predictability for dividend growth at short horizons. At short horizons, high implied dividend yields predict low future cash flows and high short-term interest rates significantly. Second, implied and observed dividend yields have opposite impacts on cash flow predictions. We think these two predictors might contain different sets of information; specifically, the former reflects discounted future cash flows in a fully rational model setting, whereas the latter may contain information about limited rationality and disagreement in beliefs. Note that Ang and Bekaert (2006) and Lettau and Ludvigson (2005) obtain a positive predictability relationship between dividend growth and dividend yields. The authors attribute this theory-contradictory result to the omitted nonlinearity feature of the present value model. Our findings confirm the existence of this positive predictability, but we further explain the importance of solving Euler equations without misspecification errors.

The remaining study is organized as follows. In Section 2, we establish the model and estimation method. We also present the asymptotic results and the data-driven implementation procedure. In Section 3, we report the simulation results, and in Section 4, we present the empirical study. In Section 5, we conclude our study. All mathematical proofs are presented in Appendix A.

2. Methodology

2.1. Model setup

We first begin by reviewing a general Euler equation framework in an exchange economy that has been extensively studied by Lucas (1978), Hansen and Singleton (1982), Grossman and Shiller (1981), and Borovička et al. (2016) in the econometric and theoretical asset pricing literature: in arbitrage-free environments, there exists a positive SDF process such that

$$f(x) = E\left\{m(X_{t+1})[f(X_{t+1}) + 1]\middle|X_t = x\right\}, \tag{1}$$

where X_t is a Markov state vector that summarizes the law of motions,[3] $m(X_{t+1})$ is a function of SDF based on state variables,[4] P_t is the date-t price, D_t is the date-t dividend, $E\{\cdot\}$ denotes rational expectations,[5] and $f(X_t) \equiv P_t/D_t$ is the unknown price–dividend (P/D) ratio function to be solved such that Eq. (1) is satisfied.

We assume that the Markov state vector, being either stationary or time-varying non-stationary, is observable to econometricians. Let $\pi(X_{t+1}|X_t)$ and $\pi(X_t)$ denote the underlying transition density function and probability density function, whose data generating processes are not known or constrained to any parametric from. Following Gagliardini et al. (2011) and Christensen (2017), we also assume that $m(\cdot)$ is an observable process given a completely known SDF, but our method still could also shed light on more complicated cases when SDF is a given function of unknown model parameters that can be calibrated or first estimated from empirical data. For example, we can consider Mehra and Prescott's (1985) rational asset pricing model, where we obtain $m(X_{t+1}) = \beta e^{(1-\gamma)X_{t+1}}$ with $X_t = \log(C_t/C_{t-1})$, the logarithm consumption growth rate, as the Markov state variable. β, the time discount factor, and γ, the risk aversion level, are often first estimated by generalized method of moments (GMM) using empirical asset returns (Hansen and Singleton, 1982). Note that the parametric estimator has a faster convergence rate than the nonparametric part. Hence, we expect that the theoretical conclusions of this paper to hold even when we have a parametric model for the SDF.[6]

3 Examples of X_t often include log consumption-, dividend-, and income growth rates.

4 Note that $m(X_{t+1})$ in our paper is not a SDF, but a SDF multiplied by the gross dividend growth rate, which is often a state variable.

5 Rational expectations coincide with mathematical ones. Irrational expectations occur when subjective expectations differ from objective expectations. One can convert the subjective expectation back to the mathematical expectation using the Radon–Nikodym theory.

6 Even though the last two decades have witnessed many studies on the development of consumption-based asset pricing theory, all the current asset pricing models, such as habit persistence proposed by Campbell and Cochrane (1999) and recursive preferences adopted by Epstein and Zin (1989), could be derived as specifications of (1) instead of as alternatives to it (Campbell and Cochrane, 2000).

The general Euler equation (1) can be rewritten as an integral equation of the second kind (Type II), for which identification has been carefully studied by Kress (1989). Specifically, we have:

$$f(X_t) - (Af)(X_t) = \varrho(X_t), \tag{2}$$

where A is a linear operator that $(Af)(X_t) \equiv \int m(X_{t+1})f(X_{t+1})\pi(X_{t+1}|X_t)dX_{t+1}$ and $\varrho(X_t) = E[m(X_{t+1})|X_t]$. For identification purposes, we first review some notations in the linear integral equation literature: A is a bounded operator if A maps bounded sets into bounded sets, and A is compact if A maps bounded sets into relatively compact sets.

Assumption 2.1. Assume (A1) There exists a unique nonzero f satisfying Eq. (2); (A2) A is a compact linear operator.

Assumption 2.1(A1) is an essential condition for the existence and uniqueness of a solution f to (2). (A1) holds under some primitive conditions, such as $I - A$ is injective, or 1 is not an eigenvalue of A given that A is a self-adjoint compact operator,[7] both of which imply that the homogeneous equation $g - Ag = 0$ only has the trivial solution $g = 0$. In the asset pricing literature, A is an integral operator defined according to some positive SDF and plays a discounting role. If A is a contraction, then (A1) is also satisfied.

Assumption 2.1(A2) further implies that $(I - A)^{-1}$ is bounded by Theorem 3.4 in Kress (1989), and thus the solution f depends continuously on ϱ, which frees us from the ill-posed problem. There are several ways to guarantee (A2). For example, if the domain of interest has finite dimensional range, or, if the Hilbert–Schmidt condition is satisfied (Christensen, 2017).

2.2. Two-stage B-splines estimation-based solution

Existing methods that solve f often require a correct specification and precise estimation of the conditional density function $\pi(X_{t+1}|X_t)$ (Carrasco et al., 2007). To avoid numerical integrations and adverse impacts from these potentially misspecified auxiliary assumptions, we propose an estimation-based solution for f via our newly proposed two-stage penalized estimation method.

To consistently estimate the unknown function f in the Euler equation, we view Eq. (2) from a regression framework and connect it with the following nonlinear time series model:

$$y_{t+1} = f(X_t) - y_{t+1}f(X_{t+1}) + \varepsilon_{t+1}, \tag{3}$$

where $y_{t+1} = m(X_{t+1})$ and $E(\varepsilon_{t+1}|X_t) = 0$. Note that Eq. (3) does not require specifying the conditional distribution, and can thus accommodate the flexible dependence structure of X_t. However, due to the occurrence of the dependent variable on the right hand side of (3) and the recursive specification of the unknown function f over two time periods, t and $t+1$, we encounter endogeneity issue as a by-product. In other words, suppose we approximate $f(x)$ by $f_a(x) = \sum_{j=1}^{q} \phi_j(x)b_j$ using some basis functions, and transform (3) into a linear regression as

$$y_{t+1} = \sum_{j=1}^{q} [\phi_j(X_t) - y_{t+1}\phi_j(X_{t+1})]b_j + \varepsilon_{t+1}. \tag{4}$$

However, the coefficients $b_j, j = 1, \ldots, q$, cannot be directly estimated by regressing y_{t+1} on $\psi_j(X_{t+1}) = \phi_j(X_t) - y_{t+1}\phi_j(X_{t+1}), j = 1, \ldots, q$, because $\psi_j(X_{t+1})$ is correlated with ε_{t+1}.

To eliminate endogeneity biases, we need instrumental variables (IV) to conduct a two-stage regression. Note that $E(\varepsilon_{t+1}|X_t) = 0$ and $cov(X_t, \psi_j(X_{t+1})) \neq 0$, therefore basis functions of X_t, which form the vector $\phi(X_t) = \left(\phi_1(X_t), \ldots, \phi_q(X_t)\right)'$, can be employed as valid IVs. Correspondingly, we can estimate the unknown function f that uniquely satisfies (1) via a nonparametric two-stage regression in a similar way in which the traditional two-stage least-square regression is carried out: in the first stage, for each of the endogenous variable $\psi_j(X_{t+1})$, we regress it on $\phi(X_t)$ to obtain the fitted value, denoted as $\hat{\psi}_j(X_{t+1})$; in the second stage, we regress y_{t+1} on $\hat{\psi}_j(X_{t+1}), j = 1, \ldots, q$, to obtain estimates of coefficients b_j's. Therefore, an estimation-based solution for the Euler equation can be constructed as $\hat{f}(x) = \sum_{j=1}^{q} \phi_j(x)\hat{b}_j$. A detailed Algorithm is included at the end of this section.

The above two-stage regression procedure is also popularly employed in the nonparametric instrumental variable (NPIV) literature, however, we emphasize that our two-stage regression can be proven equivalent to solving a well-posed Type II equation via projection instead of being a potentially ill-posed Type I problem in the general NPIV setup. Particularly, we let Π_q be a projection based on some basis functions $\{\phi_j(X_t)\}_{j=1}^{q}$, and we have

$$\Pi_q Af(X_t) = E[y_{t+1}f(X_{t+1})|\phi_1(X_t), \ldots, \phi_q(X_t)] = \phi'(X_t)\gamma,$$

[7] The self-adjoint assumption ensures that the eigenvalues of A are real and eigenelements to different eigenvalues are different and orthogonal. A self-adjoint compact operator is widely used in the econometric literature, such as Horowitz (2011), Darolles et al. (2011) and Christensen (2017), etc.

where $\gamma = [E\phi(X_t)\phi'(X_t)]^{-1}[E\phi(X_t)y_{t+1}\phi'(X_{t+1})]b$. By replacing the expectation with its sample analogue, the left-hand side of the projected Type II equation becomes

$$f(X_t) - \Pi_q Af(X_t) = \left\{\phi'(X_t) - \phi'(X_t)[\sum_t \phi(X_t)\phi'(X_t)]^{-1}[\sum_t \phi(X_t)y_{t+1}\phi'(X_{t+1})]\right\}b =: \tilde{\psi}'(X_{t+1})b.$$

A close investigation shows that $\tilde{\psi}'(X_{t+1})$ is identical to $\hat{\psi}'(X_{t+1})$, whose jth component $\hat{\psi}_j(X_{t+1})$ is the fitted value in the first stage obtained by regressing $\psi_j(X_{t+1})$ on $\phi(X_t)$. Therefore, due to such an equivalence, we point out that our described two-stage regression approach can achieve the well-posed convergence rate under Assumption 2.1 rather than the well-acknowledged slower convergence rate in a general NPIV framework which typically solves a Type I equation.

Despite the theoretical feasibility to attain the optimal large sample properties via two-stage nonparametric regression, the literature has been aware of several subtle issues from a practical perspective. For example, it may be difficult to find a proper choice of basis functions that can guard stable numerical performance in finite samples, though a large set of basis functions could ensure the optimal convergence rate in general. Moreover, in the two-stage nonparametric regression literature, it remains unclear how to conduct the appropriate smoothing in both stages. If we consider the classical two-stage approach (without penalty), the smoothing parameter is the order of basis functions q, which controls the flexibility of a projection/approximation procedure. However, the order of basis functions in the first stage could not exceed that in the second for parameter estimation identification purposes. Therefore, if solely relying on the order of basis functions q for regularization, the bias resulted from the first stage could not be negligible compared to that in the second stage.

In this paper, we are motivated to propose a new penalized two-stage nonparametric regression procedure, where we use a large number of basis for undersmoothing in the first stage, but let the penalty term conduct the appropriate smoothing in the second stage. Hence our estimate could achieve the optimal rate without worrying about the bias in the first stage. In particular, we recommend using the B-splines with the difference penalty. The B-splines have excellent numerical stability and they are piecewise polynomials defined by a set of control points, also called knots, $\tilde{c}_i$, such that $\tilde{c}_0 \leq \tilde{c}_1 \leq \cdots \leq \tilde{c}_K$. The B-splines basis can be completely determined once the degree of the polynomial p and the knots points are given, and it leads to a total of $q = K + p$ basis functions.[8] A relatively large K may result in a complicated model with excess variability. To better control for the roughness, we include a penalty term on the total variation defined by the difference operator $\Delta b_j = b_j - b_{j-1}$ and $\Delta^\nu = \Delta(\Delta^{\nu-1})$ for some positive integer, ν. Then our penalized two-stage regression estimate the B-splines coefficients b by minimizing

$$\sum_{t=1}^{T-1}\{y_{t+1} - \sum_{j=1}^{q}\hat{\psi}(X_{t+1})b_j\}^2 + \lambda^* \sum_{k=\nu+1}^{q}\{\Delta^\nu(b_k)\}^2,\ \lambda^* \geq 0. \tag{5}$$

Our new estimation procedure above can be summarized using the matrix form, and it has an appealing closed-form solution. Let $y = (y_2, \ldots, y_T)'$, $\varepsilon = (\varepsilon_2, \ldots, \varepsilon_T)'$ and $b = (b_1, \ldots, b_q)'$. Define the control variable matrix $\Psi = (\Psi_2, \ldots, \Psi_T)'$ with $\Psi_t = \psi(X_t) = \left(\psi_1(X_t), \ldots, \psi_q(X_t)\right)'$ and the IV matrix $\Phi = (\Phi_2, \ldots, \Phi_T)'$ with $\Phi_t = \phi(X_t) = \left(\phi_1(X_t), \ldots, \phi_q(X_t)\right)'$, where in both cases we suppress the dependence of Φ_t and Ψ_t on X for ease of notations. Then the auxiliary regression (without penalization in the first stage) yields the fitted values while the penalized coefficients obtained in the second stage satisfies

$$\hat{b} = (\hat{\Psi}'\hat{\Psi} + \lambda^* D_\nu^T D_\nu)^{-1}\hat{\Psi}'Y,$$

where D_ν is the νth order differencing matrix with dimension $(p + K - \nu) \times (p + K)$. Then the estimated P/D ratio f function based on our newly proposed penalized two-stage regression procedure is $\hat{f}(x) = \sum_{j=1}^{q}\phi_j(x)\hat{b}_j$.

We formally summarize our new nonparametric penalized B-splines method in Algorithm 1.

Algorithm 1: The nonparametric penalized spline method for solving Euler equations

1 Start with the Euler equation:
$$f(X_t) = E\{m(X_{t+1})[f(X_{t+1}) + 1]|X_t\}.$$
2 Define $y_{t+1} = m(X_{t+1})$ and conduct the transformation:
$$y_{t+1} = f(X_t) - y_{t+1}f(X_{t+1}) + \varepsilon_{t+1}.$$
3 Implement the two-stage penalized B-splines regression:
3.1 Stage 1: $\hat{\Psi} = \Phi(\Phi'\Phi)^{-1}\Phi'\Psi$,
3.2 Stage 2: $\hat{f}_\lambda(x) = \sum_{j=1}^{q}\phi_j(x)\hat{b}_j$, where $\hat{b} = (\hat{b}_1, \ldots, \hat{b}_q)' = (\hat{\Psi}'\hat{\Psi} + \lambda^* D_\nu^T D_\nu)^{-1}\hat{\Psi}'Y$.

Notes: λ^* is the penalty parameter, Δ is the difference operator that $\Delta b_j = b_j - b_{j-1}$ and $\Delta^\nu = \Delta(\Delta^{\nu-1})$ for any positive integer ν. $\Psi = (\Psi_2, \ldots, \Psi_T)'$ with $\Psi_t = \left(\psi_1(X_t), \ldots, \psi_q(X_t)\right)'$ and $\Phi = (\Phi_2, \ldots, \Phi_T)'$ with $\Phi_t = \left(\phi_1(X_t), \ldots, \phi_q(X_t)\right)'$. $\{\phi_i\}_{i=1}^q$ are the B-splines basis and $\psi_j(X_t) = \phi_j(X_t) - y_{t+1}\phi_j(X_{t+1})$ for $j = 1, \ldots, q$.

[8] The formula and property for the pth degree B-splines can be found in Eilers and Marx (1996).

As demonstrated in our theoretical investigation and simulation analysis, our penalized two-stage splines approach is robust against the choice of the spline degree and the placement of knots if we let the penalty amount play the optimal smoothing in the second stage. We also provide thorough discussions, theoretically and empirically, on how to effectively select the appropriate penalty parameter in the rest of this paper, which makes our approach more operational in practice.

3. Large sample properties

In this section, we discuss the asymptotics for both the unpenalized and penalized two-stage B-splines estimators. The first nonparametric estimation strategy falls into the traditional series estimation literature, in which the performance of the estimator crucially depends on the order of series expansion, equivalently the number of knots K in our setup. The other strategy is the one our paper promotes: instead of relying on the order of series expansion, we introduce the penalty parameter λ^*, which takes the leading role in balancing model complexity and goodness of fit. Besides theoretical investigation, we also provide guidance for practical implementation of our methods in the end of this section.

First, let us consider the state variables, X_t, and the unobservable aggregate pricing shock, ε_{t+1}.

Assumption 3.1. (i) The state variable X_t is stationary and has a positive and continuous density function $\pi(x)$ which is bounded away from 0 and ∞. (ii) $\{\pi(x)\}$ has a bounded support $\mathbf{X}$.

The boundedness assumption is applicable to some state variables in economics and finance. Such a boundedness assumption is also used in Newey and Powell (2003) and Ai and Chen (2003), based on which they study nonparametric regression models with instrumental variables and conditional moment estimation. It is possible to further relax this assumption for unbounded state variables using a simple monotonic transformation. Specifically, suppose $\tilde{X}_t$ is the original time series with unbounded support, then one could instead consider X_t defined as $X_t = 1/[1+\exp(-\tilde{X}_t)]$. Such a transformation is recommended as it will not lose information as shown in Hong and White (2005). Darolles et al. (2011) also explain the generality of a bounded support up to some monotone transformations in nonparametric regression analysis. Alternatively, one may also consider other different basis functions like Hermite polynomials as used in Aït-Sahalia (2002) and Xiu (2014) on state variables with unbounded support, but it is beyond the scope of this paper.

Assumption 3.2. For all t and j, there exists some $\delta > 0$ and $0 < \tilde{\Delta} < \infty$ such that $\{\varepsilon_t\}$ is an α-mixing sequence with mixing coefficients $\alpha(j)$ so that $\sum_{j=1}^{\infty} \alpha(j)^{\frac{\delta}{4+\delta}} < \tilde{\Delta}$, and $E|\varepsilon_t|^{4+\delta} < \tilde{\Delta}$.

Note that the assumptions imposed on the behavior of state variables are flexible, and they could further help embrace situations involving conditional heteroskedasticity, nonlinear time series processes and so forth.

Next, we impose conditions that hold under both smoothing strategies, namely the unpenalized and penalized two-stage B-splines estimation, respectively.

3.1. The unpenalized two-stage B-splines estimator

Assumption 3.3. The actual solution f^o is differentiable up to order 2ν. We use the pth degree B-splines series with equi-distanced K knots to approximate f^o.

The smoothness assumption is a typical condition in the nonparametric literature. Note that we are flexible in choosing the B-splines basis, since its degree need not be higher than the smooth order of f^o. Together with properties of B-splines, one can conclude that the best approximated function of f^o, denoted as $f_a(x)$, in the space spanned by the B-splines basis assumed in Assumption 3.3 must satisfy $f_a(x) - f^o(x) = O(K^{-\min(1+p,2\nu)})$. This conclusion could be extended for unequally spaced knots if the maximum knot distance is $O(K^{-1})$.

Assumption 3.4. $K \to \infty$ and $K = o(T)$.

Assumption 3.4 implies that the order of basis should grow with the sample size so that the approximation errors converge to 0. As we can show later, the asymptotic variance of the unpenalized estimate is of order K/T. Hence it is necessary to assume $K = o(T)$.

Assumption 3.5. $T = o(K^{2\min(1+p,2\nu)+1})$.

It will be shown in the rest of this paper that the approximation errors are $O(K^{-\min(1+p,2\nu)})$. Hence Assumption 3.5 further implies that $K^{-2\min(1+p,2\nu)} = o(K/T)$, i.e. the square of the maximum rate of the modeling bias is negligible compared with K/T, which is the asymptotic variance in the unpenalized case. Therefore, the unpenalized estimate is undersmoothing if K grows fast enough such that Assumption 3.5 is satisfied.

Assumption 3.6. Denote $M = T/K$. We assume that both stages use the same number of spline basis and use the notations of Φ and Ψ to denote the design matrices as described in Section 2.2. Assume that there exists some finite positive constant C such that $d_{\min}\left(\Phi'\Psi/M\right) \geq C(1+o_p(1))$, where $d_{\min}(A)$ denotes the minimum singular value of A.

Note that Assumption 3.6 implies that the minimum singular value of $\Phi'\Psi/M$ is bounded away from 0, and hence Φ is a valid instrument matrix. It may be possible to further relax Assumption 3.6 to allow for different numbers of basis in each stage. However, we do not pursue this as we recommend using the same number of basis functions and let the penalty term play a key role of smoothing. More discussions will be provided after we derive the asymptotics for the penalized case in Theorem 3.3.

Theorem 3.1 (*Consistency*)**.** *Suppose Assumptions 2.1–3.6 hold. Then there exists a unique solution $f^o(x)$ to Eq. (1), and the nonparametric 2SLS series estimator $\hat{f}_a(x)$ satisfies:*

$E_{\mathcal{X}}[\hat{f}_a(x) - f^o(x)]^2 = O_p(K/T)$ for any $x \in \mathbf{X}$, where $E_{\mathcal{X}}$ is the conditional expectation when the observed data $\mathcal{X} = (x_1, \ldots, x_T)$ are given.

Theorem 3.1 implies that our nonparametric estimator is consistent without specifying a particular model for the DGP of state variables. This appealing property cannot be attained by existing numerical solution methods which may thus suffer from model misspecification.

For rigorous statistical inference, such as confidence interval estimation and hypothesis testing, we shall derive the asymptotic distribution of the series estimator, $\hat{f}_a(x)$.

Theorem 3.2 (*Asymptotic Normality*)**.** *Suppose Assumptions 2.1–3.6 hold. Conditioning on the observed $\mathcal{X}$, we have*

$$\frac{\hat{f}_a(x) - f^o(x)}{\sqrt{var(\hat{f}_a(x)|\mathcal{X})}} \xrightarrow{d} N(0, 1). \tag{6}$$

Because of Assumption 3.5, the bias of the series estimator is negligible compared with the standard deviation. Therefore, the estimate is undersmoothing and does not achieve the optimal convergence rate.

Our method is also applicable to hidden Markov processes. Suppose the state variables, X_t, are not directly observable, but can be estimated by methods such as Kalman filters. Intuitively, the estimated state variables, $\hat{x}$, converge in probability to the point x at a parametric rate, $T^{-1/2}$, which is faster than the convergence rate of the nonparametric series estimator, $\hat{f}_a(x)$ to $f^o(x)$. As a result, the sampling errors of the estimator, $\hat{x}$, of x do not have an impact on the asymptotic distribution of $\hat{f}_a(\hat{x})$.

For conducting inference, we need to estimate the asymptotic variance $Mvar(\hat{f}_a(x))$ in practice. To take into account the dependence among the error terms, we propose the following estimator given the observed data:

$$V_T = (\frac{\hat{\Psi}'\hat{\Psi}}{M})^{-1} W_T (\frac{\hat{\Psi}'\hat{\Psi}}{M})^{-1},$$

where W_T is the Newey and West (1987) estimator

$$W_T = \frac{1}{T}\sum_{t=1}^{T} \hat{\Psi}_t'\hat{\Psi}_t \hat{u}_t^2 + \frac{1}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T}(1 - \frac{l}{L+1})\hat{u}_t\hat{u}_{t-l}(\hat{\Psi}_t'\hat{\Psi}_{t-l} + \hat{\Psi}_{t-l}'\hat{\Psi}_t),$$

$\hat{u}_t = \hat{\Psi}_t(y_t - \varphi(x_t)'\hat{b})$, and $\hat{b}$ is the unpenalized estimate of the coefficients. Note that for each element of W_T, $w_{ij} - \tilde{w}_{ij} = o_p(1)$, where $\tilde{w}_{ij}$ is the (i, j)th element of $\tilde{W}_T = \frac{\hat{\Psi}'E(\varepsilon\varepsilon')\hat{\Psi}}{M}$. It is straight forward to show that each element of V_T also satisfies $v_{ij} - \tilde{v}_{ij} = o_p(1)$, where $\tilde{V}_T = (\frac{\hat{\Psi}'\hat{\Psi}}{M})^{-1}\tilde{W}_T(\frac{\hat{\Psi}'\hat{\Psi}}{M})^{-1}$. The B-splines basis $\phi(x)$ has only a finite number of nonzero elements. Therefore, $\phi(x)'V_T\phi(x) - Mvar(\hat{f}_a(x)|\mathcal{X}) = o_p(1)$.

We have established the asymptotics for the unpenalized two-stage B-splines estimator and the results indicate that the number of the spline basis K plays the key role of smoothing. However, such an approach could encounter several difficulties in practical implementation, as the degree of spline, the number of spline basis/knots as well as the placement of the knots specified in each stage will affect the finite sample performance (Eilers and Marx, 1996). In the next subsection, we introduce a solution to resolve these empirical difficulties.

3.2. The penalized two-stage B-splines estimator

Now we consider the penalized splines case, which adopts a penalty term to control overfitting. When the penalty term λ, rather than the number of knots K, plays the key role of smoothing, the estimate could be more robust against the choice of the spline setting.

Assumption 3.7. The penalty matrix is of the form $\lambda^* D_\nu' D_\nu$. Define $\lambda = \lambda^*/M$ and let $h = \lambda^{1/(2\nu)}/K$. Then λ^* is chosen such that $h \to 0$ and $Th \to \infty$.

The term h defined in Assumption 3.7 is the equivalent bandwidth in the nonparametric smoothing procedure. Assumption 3.7 is used to prove the consistency of the estimator, since it guarantees that the bias and the variance of the estimator both shrink to zero.

Theorem 3.3 (*Consistency*)**.** *Suppose Assumptions 2.1–3.6 and 3.7 hold. Then there exists a unique solution $f^o(x)$ to Eq. (1). The 2SLS estimator with penalization $\hat{f}_\lambda(x)$ satisfies:*

$E_{\mathcal{X}}[\hat{f}_\lambda(x) - f^o(x)]^2 = O_p(\frac{1}{Th} + h^{4\nu})$, *for any $x \in \mathbf{X}$.*

In particular, the optimal convergence rate is achieved when $h \sim T^{-1/(4\nu+1)}$, or equivalently, $\lambda = \lambda^*/M$ grows exactly at the rate of $K^{2\nu}(T^{-2\nu/(4\nu+1)})$.

When K is beyond some minimum bounds such that Assumption 3.5 holds, it is no longer crucial for smoothing because the penalty parameter will play the key role in avoiding overfitting. Note that there is no penalization in the first stage. A large choice of K without penalty implicitly conducts undersmoothing. Hence, the modeling bias in the first stage is negligible and will not affect the regression in the second stage.

3.3. Discussion on the time-varying nonstationary dynamics

So far, we have carefully examined the theoretical properties of our proposed estimation-based solution method under the stationary framework, which is commonly assumed in the asset pricing literature (Christensen, 2017). However, we are aware that the widely imposed stationarity assumption may be strong for macroeconomic variables that often span a longer time period (Hansen, 2001). Therefore, we consider establishing the consistency of our estimators under a further relaxed condition.

Assumption 3.8. (i) The state variable X_t has a positive and continuous density function $\pi_t(x)$, which may depend on t. (ii) $\{\pi_t(x)\}$ share a common bounded support $\mathbf{X}$. (iii) For all t and $x \in \mathbf{X}$, there exist two positive continuous functions $\pi_L(x)$ and $\pi_U(x)$ such that $\pi_L(x) \le \pi_t(x) \le \pi_U(x)$, and $\pi_L(x)$ and $\pi_U(x)$ are bounded away from 0 and ∞.

Assumption 3.8 allows the state variable to exhibit time varying patterns, especially when the mean and variance function slowly change along time, thus being more suitable to empirical analysis. It also embraces cases when there is a potential structural break in the state variable, though it rules out situations when the state variable is integrated. In such a case, we recommend a transformation, such as differencing, to make it stationary.

Theorem 3.4 (*Consistency*)**.** *Suppose Assumptions 2.1, 3.2–3.6, and 3.8 hold. Then there exists a unique solution $f^o(x)$ to Eq. (1).*

(i) Define the nonparametric 2SLS estimator without penalty as $\hat{f}_a(x)$. Then for any given $x \in \mathbf{X}$,

$$E_{\mathcal{X}}[\hat{f}_a(x) - f^o(x)]^2 = O_p(K/T);$$

ii) Further suppose Assumption 3.7 holds and defines the 2SLS estimator with penalization as $\hat{f}_\lambda(x)$. Then for any given $x \in \mathbf{X}$,

$$E_{\mathcal{X}}[\hat{f}_a(x) - f^o(x)]^2 = O_p(\frac{1}{Th} + h^{4\nu}),$$

where in the latter case, the optimal convergence rate is achieved when $h \sim T^{-1/(4\nu+1)}$, or equivalently, $\lambda = \lambda^*/M$ grows exactly at the rate of $K^{2\nu}(T^{-2\nu/(4\nu+1)})$.

In the simulation study, we examine the finite sample performance for both the unpenalized and penalized two-stage B-splines estimator in the presence of abrupt structural breaks and smooth changes. The robust performance obtained confirms the superiority of our method in the time-varying framework.

3.4. Data-driven implementation

Section 3.3 presents the theoretical recommendations for the selection of the smoothing parameter. In practice, a data-driven procedure might be more useful. Based on the theoretical investigation, we shall let the number of knots, K, grow sufficiently large, and then use the generalized cross-validation (GCV) method to select the penalty amount in order to prevent overfitting. Since K is not the crucial smoothing parameter, the choice of the degree of the B-splines basis, as well as the placement of the knots, is not important as is confirmed in our simulation studies below. Any choice of a lower degree p, say 0, 1, 2, or 3, together with a relatively large number of equidistance knots, will suffice. The penalty order ν reflects one's belief about the smoothness of the estimated function. The common choices of ν are 1 or 2, though one may further increase ν if higher order derivatives exist.

The penalty parameter, λ^*, is important. To reduce the burden of computation, we propose using the GCV approach to determine its value.

Theorem 3.5. *Let Y, Φ, Ψ, $\hat{\Psi}$ and D_ν be defined as in sub Section 2.2. Define $\Sigma = \Psi'\Phi(\Phi'\Phi)^{-1}\Phi'\Psi$. Denote r_j as the jth eigenvalues of $\Sigma^{-\frac{1}{2}}D_\nu'D_\nu\Sigma^{-\frac{1}{2}}$. Then the GCV value equals*

$$GCV(\lambda^*) = \frac{\sum_{i=1}^{T} y_i^2 - 2\sum_{i=1}^{K+p} \frac{z_{1,i}z_{2,i}}{1+\lambda^* r_i} + \sum_{i=1}^{K+p} z_{3,i}^2(\lambda^*)}{(T-tr)^2}, \tag{7}$$

where $tr = \sum_{j=1}^{K+p} \frac{1}{1+\lambda^* r_j}$, $z_{1,i}, z_{2,i}$ are provided under Eq. (A.8), and $z_{3,i}(\lambda^*)$ are provided under Eq. (A.9).

The derivation of the GCV value is slightly more involved than that in standard nonparametric regressions, since the predicted values are calculated by $\Psi\hat{b}$ rather than $\hat{\Psi}\hat{b}$, where $\hat{b} = (\hat{\Psi}'\hat{\Psi} + \lambda^* D_\nu' D_\nu)^{-1}\hat{\Psi}'Y$. Note that we can pre-calculate terms such as $\sum_{i=1}^{T} y_i^2$, $z_{1,i}$, $z_{2,i}$, r_i, and $\tilde{r}_i$. Then, in each evaluation, we could directly calculate $GCV(\lambda^*)$ using Eq. (7). Since we need not calculate the inverse of a matrix of order $q \times q$, where $q = K + p$, we reduce the computation from, $O(K^3)$, to $O(K^2)$.

Our approach could be extended to multivariate state variables by using tensor product B-splines. In the spirit of penalized splines, appropriate penalties can be imposed by putting a separate difference penalty along each dimension of the state variables. Taking the two-dimensional smoothing as an example, we could approximate the P/D ratio by

$$f(x_1, x_2) = \sum_{k_1=1}^{q_1}\sum_{k_2=1}^{q_2} \phi_{k_1}(x_1)\tilde{\phi}_{k_2}(x_2) b_{k_1,k_2},$$

where all b_{k_1,k_2}'s form the $(K_1 + p_1)(K_2 + p_2)$ dimensional coefficient vector b. We still use the notation Φ to denote the design matrix, but its element now is defined by $\phi_{k_1}(x_1)\tilde{\phi}_{k_2}(x_2)$ and its dimension is $T \times (K_1 + p_1)(K_2 + p_2)$. Note that we could similarly define Ψ and we still have $Y = \Psi b + \varepsilon$. In terms of penalization, we could define the row and column penalties as $P_1 = I \otimes (D_{\nu_1}' D_{\nu_1})$ and $P_2 = (D_{\nu_2}' D_{\nu_2}) \otimes I$. The solution satisfies that $\hat{b} = (\hat{\Psi}'\hat{\Psi} + \lambda_1^* P_1 + \lambda_2^* P_2)^{-1}\hat{\Psi}'Y$. To reach the appropriate amount of smoothing, one could adopt an iterative approach, where in each step, we fix one of the penalty term, and reduce the problem into searching the optimal choice of the other penalty. A similar fast calculation of GCV could also be derived.

4. Simulation studies

In this section, we evaluate the finite sample performance of our estimation. Our simulation studies are based on three main objectives: Monte Carlo simulations when the true P/D function is known; comparison with numerical and analytic solutions; and evaluation of general equilibrium asset pricing models. Codes are available upon request.

4.1. Monte Carlo simulations

We now examine the finite sample performance of our proposed B-splines estimation method in solving functions with different smoothness under different conditional distributions of state variables. Designs of the unknown function and state dynamics are illustrated as follows:

DGP F.1 [a periodic case]: $f(X_t) = 3 + 0.5sin(20X_t + 1.2) + cos(10X_t + 2)$.

DGP F.2 [a non-periodic case]: $f(X_t) = exp(X_t)$.

DGP F.3 [a multiplicative case]: $f(X_t) = 3 + (X_{t,1} + 1) * cos(X_{t,2} + 0.2)$.

DGP F.4 [an additive case]: $f(X_t) = 3 + sin(5 * X_{t,1}) + cos(X_{t,2} + 0.2)$.

The first designed function DGP F.1 is a periodically non-monotonic function with changing curvatures. The second designed function, DGP F.2, is non-periodically monotone in its domain. DGPs F.3 and F.4 are used to examine the performance of our method with multivariate state variables. The dynamics of state variables X_t is crucial in shaping the Euler equation. Since the empirical data of X_t often span a relatively long period of time, to examine the applicability of our method, we consider both stationary and time-varying non-stationary dynamics of X_t as follows.

DGP S.1 [a stationary homoscedastic state variable]: We have

$$X_{t+1} = \Gamma X_t + \epsilon_{t+1},\ \epsilon_t \sim i.i.d.N(0, \sigma_s^2).$$

where in DGP S.1.1, we assume $\Gamma = 0.1$ with $\sigma_s = 0.1$; in DGP S.1.2, we assume $\Gamma = 0.8$ with $\sigma_s = 0.5$.

DGP S.2 [a stationary conditional heteroskedastic state variable]: We have

$$\begin{cases} X_t &= z_t\sqrt{h_t} \\ h_t &= \omega_0 + \omega_1 X_{t-1}^2 + \omega_2 h_{t-1} \\ z_t &\sim i.i.d.N(0, 1) \end{cases}$$

where in DGP S.2.1, $h_t = 0.05 + 0.2X_{t-1}^2 + 0.6h_{t-1}$; in DGP S.2.2, $h_t = 0.05 + 0.5X_{t-1}^2 + 0.3h_{t-1}$.

DGP S.3 [a nonstationary state variable with multiple structural breaks]: We have

$$X_{t+1} = \begin{cases} 0.1 + 0.5X_t + \epsilon_t, & \text{if } t \le 0.4T, \\ X_t + \epsilon_t, & \text{otherwise}, \end{cases}$$

where $\epsilon_t \sim i.i.d.N(0, \sigma_s^2)$ with $\sigma_s = 0.5$.

DGP S.4 [a nonstationary state variable with smooth structural changes]: We have

$$X_t = F(\tau_t)(1 + 0.5\bar{X}_t) + \epsilon_t,$$

where $\tau_t = t/T$, $F(\tau_t) = 1.5 - 1.5\exp[-3(\tau_t - 0.5)^2]$, $\bar{X}_t \sim i.i.d.N(0, 1)$, and $\epsilon_t \sim i.i.d.N(0, 0.5^2)$.

DGP S.5 [multivariate state variables]: Define $X_t = (X_{1,t}, X_{2,t})$ such that $X_{1,t}$ is a GARCH(1,1) process and $X_{2,t}$ is a AR(1) process. In particular, we have

$$\begin{cases} X_{1,t} &= h_t^{1/2} z_{1,t} \\ h_t &= 0.01 + 0.64 h_{t-1} + 0.05 z_{1,t-1}^2 \end{cases}$$

and

$$X_{2,t} = 0.01 + 0.5 X_{2,t-1} + z_{2,t},$$

where $(z_{1t}, z_{2,t})$ are bivariate normally distributed with parameters (0, 0, 1, 1, 0.2).

In DGP S.1, we first explore the finite sample performance under a popularly used AR(1) process with two scenarios: specifically in DGP S.1.1, the serial dependency fades out exponentially fast as the time distance between state variables increases; in DGP S.1.2, the numerical solution methods, especially the projection method, face challenges. However, a stronger dependency between two consecutive state observations is favorable for our method because it leads to a tightened relationship between instruments and endogenous variables in the nonparametric two-stage regression procedure. In DGP S.2, we examine finite sample performance under a general conditional heteroskedastic data generating process for the state variable X_t using a GARCH specification. In DGP S.3, we consider a scenario with an abrupt structural break, which makes X_t a nonstationary process. In DGP S.4, we examine the finite sample performance of our method when the state variable exhibits smooth structural changes in time. Finally, in DGP S.5, we provide an evaluation of the solution accuracy given there are multiple state variables.

For each Monte Carlo study in the univariate case, we generate samples with sizes of 100, 250 and 500, and implement our new two-stage B-splines regression with and without penalization. In each study, we also examine the estimation performance under 35 knots and 50 knots using piecewise-linear, piecewise-quadratic and piecewise-cubic B-splines, namely $p = 1, 2, 3$. Overall, our newly proposed penalized two-stage B-splines estimator works uniformly well for conditional homoscedastic, heteroscedastic and time-varying non-stationary processes with different levels of serial dependencies. Comparing with the true P/D ratio functions, we can visualize the estimation results with and without penalties in Figs. 1–4. The penalized B-splines will generally enhance the estimation results, especially when solving smoother functions. The outstanding accuracy of small samples demonstrates how our method offers significant advantages for current asset pricing and macroeconomic general equilibrium modeling, where state variables are mostly available at quarterly frequency.

In addition, we also examine the bivariate case with a larger sample size, 500 or 1000. We also consider spline degrees from 1 to 3. In each direction, we consider imposing 10 or 15 equally spaced knots. The integrated mean squared error (IMSE) has been widely used in nonparametric series estimations to evaluate finite sample performance (Hansen, 2015). Thus, the IMSE of the estimator, $\hat{f}(x)$, is

$$IMSE = \int E[\hat{f}(x) - f(x)]^2 dF(x), \tag{8}$$

where $F(x)$ is the cumulative distribution function for the state variable. To evaluate the goodness of fit, we calculate the IMSE for each Monte Carlo study by implementing the estimation method 100 times for both the univariate case and the multivariate case. The mechanism can be clearly seen from Tables 1–4. For the unpenalized case, the IMSE varies among different choices of the degree of the splines p and the number of the knots K. In contrast, the choice of the degree of the splines does not affect the penalized estimate much. When a larger number of knots are used, our fast GCV algorithm will automatically generate a larger penalty to correct for the potential overfitting problem. This simulation evidence further demonstrates the superiority of our new method.

4.2. Comparison with numerical and analytic solutions

We confine the scope of this section to scenarios wherein analytic solutions of the P/D ratios exist (Burnside, 1998). We provide a detailed comparison of our nonparametric penalized B-splines series regression method with some representative and popularly used numerical solution methods. For the numerical solution methods, we consider perturbation, projection, and discretization, as well as the PEA and the improved PEA algorithms. Except for the original PEA algorithm, all these numerical solution methods require complete knowledge of the dynamics of state variables, whereas the true DGP of state variables in the real world is not completely known by empirical practitioners, possibly because of limited skill, time, or noisy observations. In this section, we examine how asset pricing models could be affected when DGPs of state variables are misspecified.

Analytic solutions based on Mehra and Prescott's (1985) model are obtainable under special circumstances (Burnside, 1998). In this exchange economy, there is an infinitely lived representative agent who wishes to maximize his or her expected lifetime utility at time zero:

$$\max_{\{C_t\}} E \sum_{t=0}^{\infty} \beta^{t-1} \frac{C_t^{1-\gamma}}{1-\gamma}$$
$$\text{s.t.}\, C_t + P_{t+1}\theta_{t+1} + Q_t b_{t+1} = b_t + (D_t + P_t)\theta_t,$$

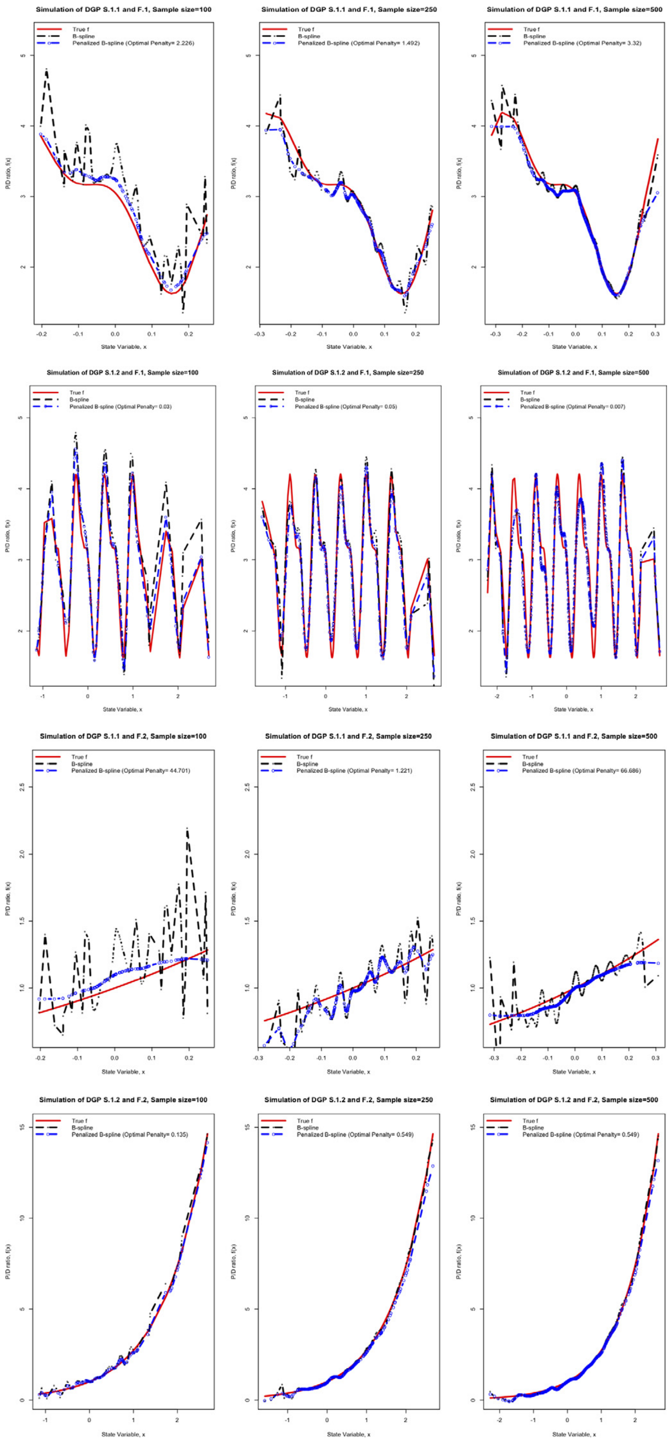

Fig. 1. Solution accuracy of P/D ratios with homoscedastic state dynamics. This figure plots the estimated P/D ratio with and without a penalty when the state variable exhibits conditional homoskedasticity. In the first two rows, the true P/D ratio function is assumed to follow F.1 (a periodic function) and the state variable follows DGP S.1.1 in the top row and DGP S.1.2 in the second row. In the last two rows, the true P/D ratio function is assumed to follow F.2 (a non-periodic function) and the state variable follows DGP S.1.1 in the third row and DGP S.1.2 in the fourth row. For each case, the sample size for the left, middle and right ones are 100, 250 and 500 respectively. In each plot, the red solid line represents the true unknown function to be estimated, the dotted black line represents the estimation from the two-stage B-splines method without any penalty and the blue dotted line represents the estimation from the penalized two-stage B-splines method, where the associated optimal penalty is reported in the parenthesis. For simplicity, our nonparametric two-stage B-splines estimation results have $p = 1$, $K = 35$ with equally spaced knots and the optimal penalty is determined via the proposed fast GCV algorithm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

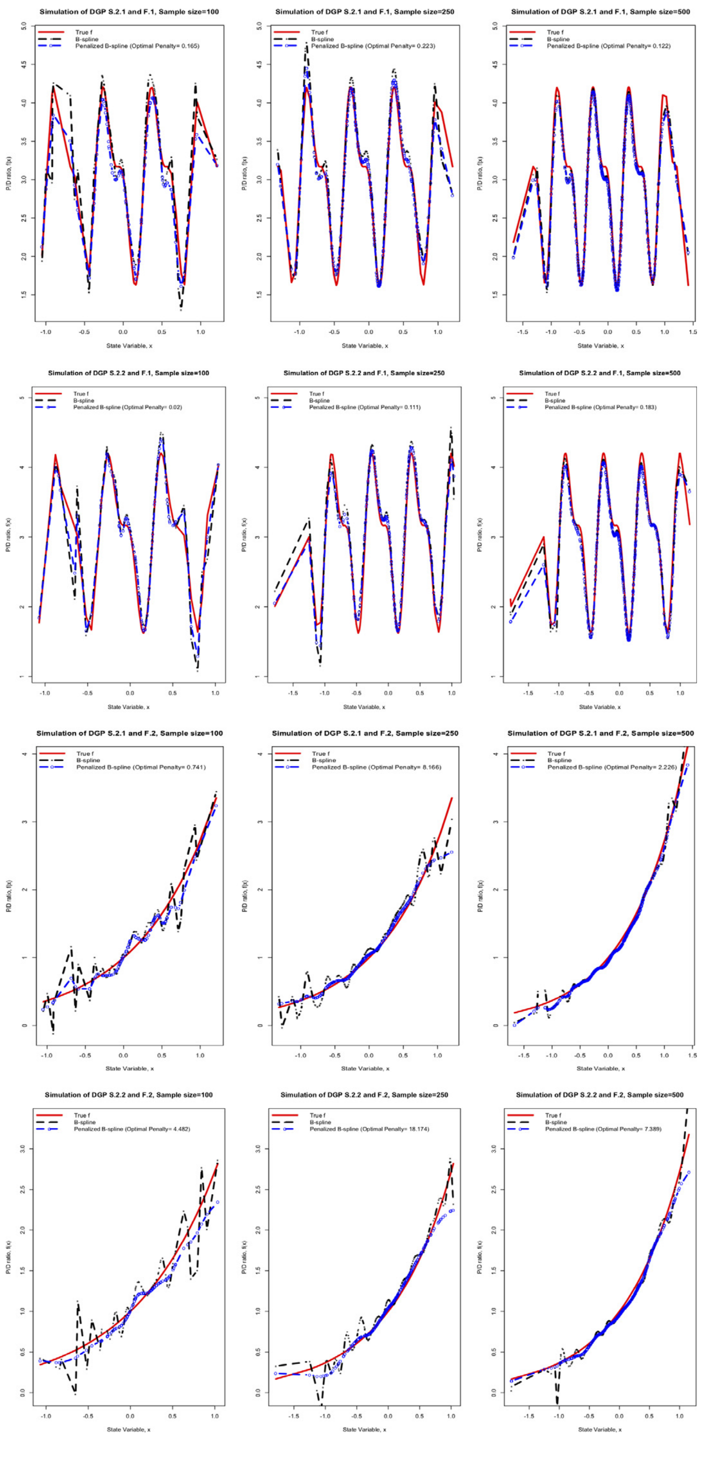

Fig. 2. Solution accuracy of P/D ratios with heteroscedastic state dynamics. This figure plots the estimated P/D ratio with and without a penalty when the state variable exhibits conditional heteroskedasticity. In the first two rows, the true P/D ratio function is assumed to follow F.1 (a periodic function) and the state variable follows DGP S.2.1 in the top row and DGP S.2.2 in the second row. In the last two rows, the true P/D ratio function is assumed to follow F.2 (a non-periodic function) and the state variable follows DGP S.2.1 in the third row and DGP S.2.2 in the fourth row. For each case, the sample size for the left, middle and right ones are 100, 250 and 500 respectively. In each plot, the red solid line represents the true unknown function to be estimated, the dotted black line represents the estimation from the two-stage B-splines method without any penalty and the blue dotted line represents the estimation from the penalized two-stage B-splines method, where the associated optimal penalty is reported in the parenthesis. For simplicity, our nonparametric two-stage B-splines estimation results have $p = 1$, $K = 35$ with equally spaced knots and the optimal penalty is determined via the proposed fast GCV algorithm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

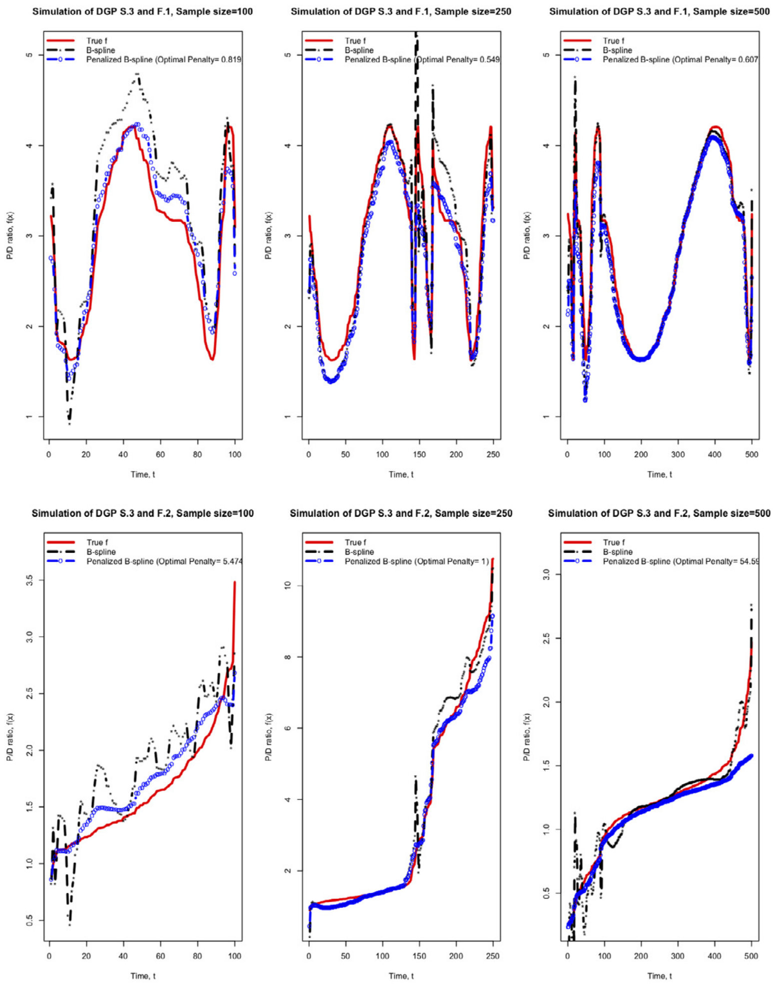

Fig. 3. Solution accuracy of P/D ratios for state dynamics with abrupt structural breaks. This figure plots the estimated P/D ratio with and without a penalty based on our penalized two-stage B-splines approach, where $p = 1$, $K = 35$ with equally spaced knots and the optimal penalty is determined via the proposed fast GCV algorithm. The sample size for the left, middle and right ones are 100, 250 and 500 respectively. The true P/D ratio function and the state variable are assumed to follow F.2 (a nonperiodic function) under DGP S.1.1 (graphs in the top row) or DGP S.1.2 (graphs in the second row) or DGP S.2.1 (graphs in the third row) or DGP S.2.2 (graphs in the bottom row), respectively. For each case, the red solid line represents the true unknown function to be estimated, the dotted black line represents the estimation from the two-stage B-splines method without any penalty and the blue dotted line represents the estimation from the penalized two-stage B-splines method, where the associated optimal penalty is reported in the parenthesis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where $X_t = ln(C_t/C_{t-1})$, C_t is the consumption at time t; D_t is the dividend payment at time t; P_t is the current value that reflects future dividend payments; Q_t is the price of a risk-free asset that pays 1 in period $t+1$; b_t and θ_t are the holdings of the risky and risk-free asset at time t. In this simple economy, the dividend payment, D_t, is equal to the optimal consumption, C_t, in equilibrium. Let $f_t = P_t/D_t$, and the Euler equation can be derived as follows:

$$f_t = \beta E[e^{(1-\gamma)X_{t+1}}(f_{t+1} + 1)|X_t].$$

We consider two numerical studies to examine the performance of our newly proposed estimation method for finite and small samples. The first scenario is one where the DGP of state variable, X_t, is known and correctly specified. In the second, economists only have empirical observations of state variables, X_t, and do not know their conditional distributions. These DGPs are given below:

DGP B.1 [a correctly specified DGP]: We have

$$X_{t+1} - \mu = \Gamma(X_t - \mu) + \epsilon_{t+1},$$

where $\epsilon_t \sim IIDN(0, \sigma^2)$. For ease of comparisons, we set $\beta = 0.96$, $\gamma = 2.5$, $E(X_t) = 0.0179$, and $\sigma = 0.0379$ with $\Gamma = -0.139$ in DGP B.1.1 and $\Gamma = 0.8$ in DGP B.1.2.

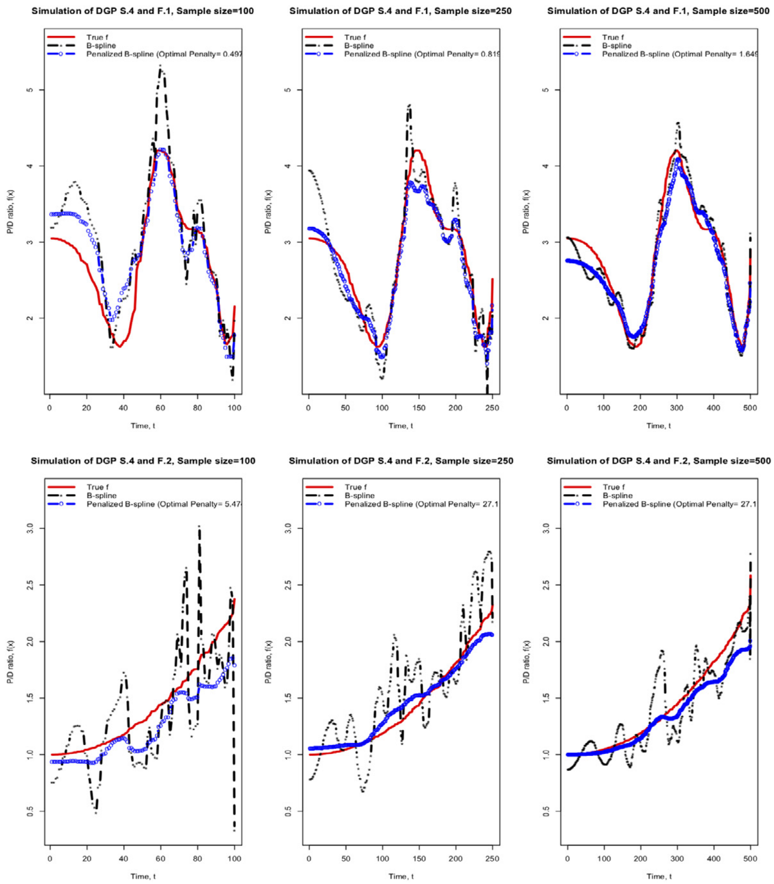

Fig. 4. Solution accuracy of P/D ratios for state dynamics with smooth structural changes. This figure plots the estimated P/D ratio with and without a penalty based on our penalized two-stage B-splines approach, where $p = 1$, $K = 35$ with equally spaced knots and the optimal penalty is determined via the proposed fast GCV algorithm. The sample size for the left, middle and right ones are 100, 250 and 500 respectively. The true P/D ratio function and the state variable are assumed to follow F.2 (a nonperiodic function) under DGP S.1.1 (graphs in the top row) or DGP S.1.2 (graphs in the second row) or DGP S.2.1 (graphs in the third row)) or DGP S.2.2 (graphs in the bottom row), respectively. For each case, the red solid line represents the true unknown function to be estimated, the dotted black line represents the estimation from the two-stage B-splines method without any penalty and the blue dotted line represents the estimation from the penalized two-stage B-splines method, where the associated optimal penalty is reported in the parenthesis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

DGP B.2 [a misspecified DGP]: We have

$$X_{t+1} = \begin{cases} \mu + \Gamma_1 X_t + u^*_{1,t+1},\ u^*_{1,t} \sim i.i.d.N(0, \sigma_1^2) & \text{if } X_t > 0, \\ \mu + \Gamma_2 X_t + u^*_{2,t+1},\ u^*_{2,t} \sim i.i.d.N(0, \sigma_2^2) & \text{if } X_t \leq 0 \end{cases}$$

$$\tilde{X}_{t+1} - \mu = \bar{\Gamma}(\tilde{X}_t - \mu) + v_{t+1}, \text{ and } v_{t+1} \sim i.i.d.N(0, \sigma_v^2),$$

where X_t is the true underlying DGP with $\sigma_1 = 3.48\%$, $\sigma_2 = 2\sigma_1$, $\beta = 0.96$, $\gamma = 2.5$, $\Gamma_1 = 0.8$, and $\Gamma_2 = -0.139$, but not known by econometricians, who fit the data with $\tilde{X}_t$ with $\bar{\Gamma}$ and σ_v^2 chosen to match the autocorrelation with the true process. DGP B.2 explores a threshold structure whose importance has been widely acknowledged in many economic studies (Hong and Li, 2005).

Fig. 5 compares the approximated P/D ratio function from different solution methods under DGPs B.1 and DGP B.2 for sample sizes of 100, 250 and 500 respectively. Specifically, DGP B.1.1 shows small serial correlation of state variables in absolute values and the true P/D ratio behaves as a linear function. Therefore, as shown in Fig. 5, all studied numerical methods can provide accurate approximations. However, the performance of our regression-based nonparametric penalized B-splines method is enhanced dramatically when the sample size increases. In DGP B.1.2, where the serial dependency is large, we first confirm that the projection method with low-order serial expansions fails (Calin et al., 2005). The perturbation method faces problems when approximating tails. Our penalized B-splines method always exhibits superior performance compared to the PEA method. It is worth mentioning that discretization and improved PEA methods work well when the DGP of state variables is correctly specified.

Table 1
Two-stage B-splines solution accuracy under AR distributed state variable.

	Sample size = 100			Sample size = 250			Sample size = 500		
	IMSE		λ^*	*IMSE*		λ^*	*IMSE*		λ^*
	Unpenalized	Penalized	penalty	Unpenalized	Penalized	penalty	Unpenalized	Penalized	penalty
DGP F.1 DGP S.1.1									
p = 1, K = 35	0.049	0.029	2.225	0.018	0.013	2.532	0.009	0.007	2.877
p = 1, K = 50	0.108	0.034	2.837	0.024	0.015	3.528	0.012	0.008	4.269
p = 2, K = 35	0.067	0.029	2.135	0.022	0.013	2.200	0.010	0.007	2.493
p = 2, K = 50	0.173	0.035	2.699	0.034	0.015	3.252	0.012	0.007	3.842
p = 3, K = 35	0.055	0.029	1.982	0.020	0.013	2.026	0.010	0.007	2.285
p = 3, K = 50	0.104	0.035	2.664	0.043	0.015	3.095	0.013	0.007	3.568
DGP F.1 DGP S.1.2									
p = 1, K = 35	0.090	0.077	0.084	0.067	0.083	0.208	0.062	0.063	0.208
p = 1, K = 50	0.097	0.075	0.140	0.069	0.063	0.274	0.030	0.039	0.336
p = 2, K = 35	0.253	0.065	0.049	0.083	0.065	0.100	0.075	0.063	0.161
p = 2, K = 50	0.112	0.069	0.108	0.056	0.055	0.199	0.025	0.031	0.223
p = 3, K = 35	0.083	0.060	0.033	0.042	0.039	0.026	0.043	0.042	0.036
p = 3, K = 50	0.437	0.069	0.086	0.196	0.029	0.084	0.016	0.016	0.078
DGP F.2 DGP S.1.1									
p = 1, K = 35	0.034	0.011	19.362	0.013	0.004	28.854	0.007	0.002	37.307
p = 1, K = 50	0.062	0.014	21.030	0.018	0.005	35.058	0.009	0.003	45.771
p = 2, K = 35	0.054	0.011	20.258	0.015	0.005	28.920	0.007	0.002	38.375
p = 2, K = 50	0.095	0.014	22.439	0.020	0.0053	34.832	0.009	0.003	48.047
p = 3, K = 35	0.039	0.011	20.199	0.014	0.004	29.976	0.007	0.002	38.320
p = 3, K = 50	0.090	0.014	23.001	0.022	0.005	38.085	0.010	0.003	50.074
DGP F.2 DGP S.1.2									
p = 1, K = 35	0.053	0.041	0.879	0.034	0.073	2.239	0.015	0.051	2.376
p = 1, K = 50	0.071	0.048	1.102	0.044	0.081	3.070	0.019	0.058	3.514
p = 2, K = 35	0.082	0.040	0.807	0.035	0.071	2.154	0.030	0.050	2.309
p = 2, K = 50	0.175	0.044	0.932	0.045	0.079	2.924	0.020	0.057	3.349
p = 3, K = 35	0.073	0.036	0.698	0.024	0.020	0.578	0.011	0.011	0.483
p = 3, K = 50	0.137	0.042	0.913	0.061	0.024	0.873	0.017	0.013	0.799

Notes: DGPs of state variables are described in DGP S.1.1 and DGP S.1.2. The unknown function satisfies DGP F.1 and DGP F.2. K is the number of knots and these knots are equally spaced on the entire range of the state variable. p is the degree of B-splines. Specifically, $p = 1$ is piecewise linear B-spline, and $p = 2$ is piecewise quadratic B-splines and $p = 3$ is a cubic B-spline. We calculate the IMSE for each Monte Carlo study by implementing the estimation without and with penalty 100 times and λ^* is the average optimal penalty for the penalized case.

Therefore, in DGP B.2, where the analytic solution does not exist (Burnside, 1998), we first assume that the DGP is given, and use discretization methods with sufficiently fine grids to generate an accurate proxy for the true unknown P/D ratios for the purpose of comparison. As Fig. 5 shows, except for the PEA algorithm, all other numerical solution methods fail in the presence of a misspecified DGP. A critical issue with the PEA algorithm is that the optimal order for the parametric series expansion is unknown, and a large order may result in an ill-conditioned problem. Therefore, the PEA algorithm provides improved, but sub-optimal, approximations. In contrast, by using our penalized B-splines regression, we can obtain a consistent, unbiased, and efficient estimation of the unknown P/D ratios in the presence of an unknown DGP.

In Table 5, we further report and compare the number of iterations, real computational time, and mean squared errors for the simulation studies above. Our method is a speedy one-step procedure with an optimized level of penalization. It has stable performance when the state variable has both small and large serial dependencies. We find that, under correctly specified cases, the penalized B-splines regression performs reasonably well compared with this large set of representative numerical solution methods. Under the misspecified case, only our new method can provide accurate model solutions because it does not depend on distributional assumptions.

Using these two empirically relevant situations, we find that, in the presence of misspecified dynamics of state variables, current numerical solution methods can lead researchers to incorrectly interpret model implications. Therefore, when the DGP of state variables is not fully specified, the newly proposed method will become an indispensable approach for obtaining a consistent estimate of the P/D ratio function and for constructing the most reliable and accurate model implications.

4.3. Model implied equity returns

In this section, we investigate the role that the misspecification plays on model implied asset returns by removing the distributional assumptions on the state dynamics in two prevailing asset pricing models, namely Mehra and Prescott's (1985) and Campbell and Cochrane's (1999) models. We first quickly review the model setup of Campbell and Cochrane

Table 2
Two-stage B-splines solution accuracy under GARCH distributed state variable.

	Sample size = 100			Sample size = 250			Sample size = 500		
	IMSE		λ^*	IMSE		λ^*	IMSE		λ^*
	Unpenalized	Penalized	Penalty	Unpenalized	Penalized	Penalty	Unpenalized	Penalized	Penalty
DGP F.1 DGP S.2.1									
p = 1, K = 35	0.052	0.051	0.174	0.026	0.025	0.169	0.022	0.021	0.197
p = 1, K = 50	0.069	0.058	0.284	0.026	0.025	0.313	0.015	0.014	0.327
p = 2, K = 35	0.072	0.047	0.134	0.022	0.021	0.121	0.015	0.015	0.120
p = 2, K = 50	0.166	0.055	0.237	0.027	0.024	0.264	0.013	0.013	0.274
p = 3, K = 35	0.072	0.045	0.112	0.040	0.020	0.094	0.013	0.013	0.086
p = 3, K = 50	0.133	0.053	0.208	0.026	0.023	0.223	0.014	0.012	0.232
DGP F.1 DGP S.2.2									
p = 1, K = 35	0.066	0.056	0.171	0.045	0.044	0.162	0.60	0.059	0.245
p = 1, K = 50	0.083	0.052	0.260	0.029	0.028	0.256	0.023	0.023	0.215
p = 2, K = 35	0.084	0.052	0.134	0.046	0.038	0.117	0.044	0.043	0.118
p = 2, K = 50	0.115	0.050	0.227	0.252	0.025	0.202	0.017	0.016	0.158
p = 3, K = 35	0.064	0.049	0.108	0.033	0.033	0.072	0.043	0.042	0.071
p = 3, K = 50	0.117	0.048	0.190	0.309	0.023	0.170	0.018	0.015	0.129
DGP F.2 DGP S.2.1									
p = 1, K = 35	0.044	0.020	2.819	0.018	0.009	2.677	0.009	0.006	2.969
p = 1, K = 50	0.056	0.022	0.326	0.023	0.010	3.474	0.011	0.007	4.321
p = 2, K = 35	0.063	0.019	2.911	0.018	0.009	2.396	0.009	0.006	2.694
p = 2, K = 50	0.090	0.022	3.119	0.026	0.010	3.275	0.012	0.006	3.792
p = 3, K = 35	0.268	0.019	2.664	0.018	0.009	2.220	0.009	0.006	2.431
p = 3, K = 50	0.012	0.006	3.544	0.024	0.010	3.005	0.012	0.006	3.544
DGP F.2 DGP S.2.2									
p = 1, K = 35	0.072	0.072	2.453	0.076	0.025	1.984	0.036	0.023	1.261
p = 1, K = 50	0.089	0.082	2.662	0.033	0.093	2.545	0.028	0.041	2.053
p = 2, K = 35	0.128	0.061	2.325	0.076	0.025	1.889	0.032	0.024	1.205
p = 2, K = 50	2.598	0.089	2.615	0.033	0.082	2.323	0.026	0.041	1.733
p = 3, K = 35	0.071	0.073	2.301	0.026	0.075	1.683	0.024	0.036	1.083
p = 3, K = 50	0.269	0.081	2.684	0.067	0.084	2.343	0.026	0.039	1.680

Notes: DGPs of state variables are described in DGP S.2.1 and DGP S.2.2. The unknown function satisfies DGP F.1 and DGP F.2. K is the number of knots and these knots are equally spaced on the entire range of the state variable. p is the degree of B-splines. Specifically, $p = 1$ is piecewise linear B-spline, and $p = 2$ is piecewise quadratic B-splines and $p = 3$ is a cubic B-spline. We calculate the IMSE for each Monte Carlo study by implementing the estimation without and with penalty 100 times and λ^* is the average optimal penalty for the penalized case.

(1999) as follows:

$$\max_{\{C_t\}} E \sum_{t=0}^{\infty} \beta^t \frac{(C_t - H_t)^{1-\gamma} - 1}{1-\gamma},$$

where H_t is the level of habit. The consumption habit, $S_t = (C_t - H_t)/C_t$, is assumed to be exogenous and determined by the history of aggregate consumption rather than the history of individual consumption. The log surplus consumption ratio $s_t \equiv lnS_t$ evolves as a heteroskedastic AR(1) process as follows:

$$s_{t+1} = (1-\Gamma)\bar{s} + \Gamma s_t + l(s_t)(c_{t+1} - c_t - g_c),$$

where $\bar{s}$ is the steady state level and $l(s_t)$ is called the sensitivity function, specified as

$$l(s_t) = \begin{cases} \frac{1}{\bar{S}}\sqrt{1-2(s_t-\bar{s})} - 1, & \text{if } s_t \le s_{max} = ln(\bar{S}) + \frac{1}{2}(1-\bar{S})^2, \\ 0, & \text{if } s_t \ge s_{max}. \end{cases}$$

The price–dividend ratio f_t is embedded in the Euler equation[9]

$$f(X_t) = E\{\beta e^{-\gamma g_c} e^{-\gamma((\Gamma-1)(X_t-\bar{s})+(1-l(X_t))\Delta c_{t+1})} e^{\Delta d_{t+1}}[1 + f(X_{t+1})]|X_t\},$$

where $c_t = log(C_t)$, $d_t = log(D_t)$, $\Delta c_t = c_t - c_{t-1}$, $\Delta d_t = d_t - d_{t-1}$, $g_c = E(\Delta c_t)$ and $X_t = s_t$ is the state variable. In Campbell and Cochrane (1999), the model-implied P/D ratio also directly depends on the distributional assumption with respect to the dynamics of consumption and dividend growth. The authors first assume that the consumption and dividend processes are identical and also investigate the scenario with imperfectly correlated consumption and dividend processes, that $(\Delta c_t, \Delta d_t) \sim i.i.d.N(g_c, \sigma_c^2, g_d, \sigma_d^2, \rho)$. However, in this section, we remove such assumptions and study how the true dynamics of consumption and dividend shocks affect equity returns.

9 More details on the derivation of the Euler equation with consumption habit could be referred at Campbell and Cochrane (1999).

Table 3
Two-stage B-splines solution accuracy under nonstationary state variable.

	Sample size = 100			Sample size = 250			Sample size = 500		
	IMSE		λ^*	*IMSE*		λ^*	*IMSE*		λ^*
	Unpenalized	Penalized	Penalty	Unpenalized	Penalized	Penalty	Unpenalized	Penalized	Penalty
DGP F.1 DGP S.3									
p = 1, K = 35	0.299	0.107	1.345	0.086	0.062	1.082	0.046	0.046	0.689
p = 1, K = 50	0.271	0.125	1.722	0.092	0.071	1.635	0.059	0.047	1.171
p = 2, K = 35	0.206	0.106	1.206	0.175	0.062	0.917	0.101	0.041	0.603
p = 2, K = 50	0.586	0.126	1.600	0.173	0.067	1.463	0.494	0.045	1.047
p = 3, K = 35	0.196	0.107	1.163	0.394	0.059	0.877	0.045	0.040	0.517
p = 3, K = 50	0.529	0.127	1.498	0.124	0.070	1.408	0.046	0.045	0.964
DGP F.2 DGP S.3									
p = 1, K = 35	3.233	0.042	15.584	0.095	0.090	15.866	0.085	0.070	19.971
p = 1, K = 50	0.119	0.049	17.677	0.095	0.073	20.330	0.553	0.129	30.878
p = 2, K = 35	0.139	0.043	15.488	0.125	0.091	16.154	2.633	0.106	20.858
p = 2, K = 50	0.714	0.051	19.165	0.104	0.088	21.410	0.622	0.201	31.926
p = 3, K = 35	0.106	0.044	16.151	0.208	0.085	15.567	0.292	0.095	21.010
p = 3, K = 50	0.355	0.052	18.822	0.151	0.090	22.182	0.088	0.076	30.227
DGP F.1 DGP S.4									
p = 1, K = 35	0.273	0.151	0.991	0.126	0.077	1.527	0.054	0.044	2.189
p = 1, K = 50	0.318	0.166	1.234	0.179	0.080	2.086	0.064	0.046	3.268
p = 2, K = 35	1.166	0.151	0.904	0.183	0.073	1.327	0.061	0.042	1.994
p = 2, K = 50	0.445	0.168	1.102	0.268	0.079	1.941	0.108	0.045	3.029
p = 3, K = 35	0.370	0.147	0.824	0.836	0.073	1.249	0.061	0.040	1.813
p = 3, K = 50	1.045	0.168	1.075	0.228	0.078	1.851	0.090	0.044	2.891
DGP F.2 DGP S.4									
p = 1, K = 35	0.225	0.052	6.537	0.070	0.021	14.169	0.030	0.011	30.840
p = 1, K = 50	0.243	0.063	5.760	0.120	0.025	15.286	0.036	0.012	36.475
p = 2, K = 35	0.584	0.054	6.635	0.899	0.021	14.225	0.034	0.012	30.830
p = 2, K = 50	0.235	0.069	5.513	0.710	0.025	16.346	0.251	0.012	38.260
p = 3, K = 35	0.314	0.053	6.790	0.796	0.022	14.946	0.032	0.012	31.267
p = 3, K = 50	0.977	0.068	6.679	0.170	0.025	17.129	0.057	0.012	39.239

Notes: DGPs of state variables are described in DGP S.3 and DGP S.4. The unknown function satisfies DGP F.1 and DGP F.2. K is the number of knots and these knots are equally spaced on the entire range of the state variable. p is the degree of B-splines. Specifically, $p = 1$ is piecewise linear B-spline, and $p = 2$ is piecewise quadratic B-splines and $p = 3$ is a cubic B-spline. We calculate the IMSE for each Monte Carlo study by implementing the estimation without and with penalty 100 times and λ^* is the average optimal penalty for the penalized case.

Table 4
Two-stage B-splines solution accuracy under multivariate state variables.

	Sample size = 500				Sample size = 1000			
	IMSE		λ^*		*IMSE*		λ^*	
	Unpenalized	Penalized	Penalty1	Penalty2	Unpenalized	Penalized	Penalty1	Penalty2
DGP F.3 DGP S.5								
p = 1, K = 10	0.243	0.038	0.359	9.109	0.243	0.017	0.407	11.484
p = 1, K = 15	0.641	0.061	0.406	12.676	0.690	0.026	0.496	16.449
p = 2, K = 10	7.119	0.043	0.348	6.833	1.914	0.018	0.374	9.245
p = 2, K = 15	9.576	0.069	0.387	10.399	2.187	0.028	0.470	13.842
p = 3, K = 10	5.583	0.046	0.329	5.679	5.044	0.018	0.364	7.501
p = 3, K = 15	8.347	0.074	0.374	9.119	6.897	0.029	0.437	11.921
DGP F.4 DGP S.5								
p = 1, K = 10	0.271	0.040	0.425	2.332	0.247	0.017	0.482	3.065
p = 1, K = 15	0.607	0.073	0.510	3.423	0.710	0.029	0.622	4.422
p = 2, K = 10	7.031	0.044	0.397	1.288	1.936	0.018	0.463	2.497
p = 2, K = 15	9.675	0.076	0.479	2.664	4.947	0.031	0.567	3.717
p = 3, K = 10	6.023	0.046	0.380	1.529	2.419	0.018	0.451	1.905
p = 3, K = 15	8.453	0.080	0.446	2.385	6.712	0.032	0.541	3.125

Notes DGPs of state variables are described in DGP S.5. The unknown function satisfies DGP F.4 and DGP F.5. K is the number of knots and these knots are equally spaced on the entire range of the state variable. p is the degree of B-splines. Specifically, $p = 1$ is piecewise linear B-spline, and $p = 2$ is piecewise quadratic B-splines and $p = 3$ is a cubic B-spline. We calculate the IMSE for each Monte Carlo study by implementing the estimation without and with penalty 100 times and λ^* is the average optimal penalty for the penalized case.

We consider the returns on S&P500 and a three-month Treasury bill for market returns and risk-free returns. The consumption growth rate is obtained from the personal consumption expenditures on non-durable goods per capita from the US National Income and Product Accounts. Our data are in quarterly frequency and span from 1947 to 2017. All the empirical returns are deflated by a fixed base consumption price index. We fit Mehra and Prescott's (1985)

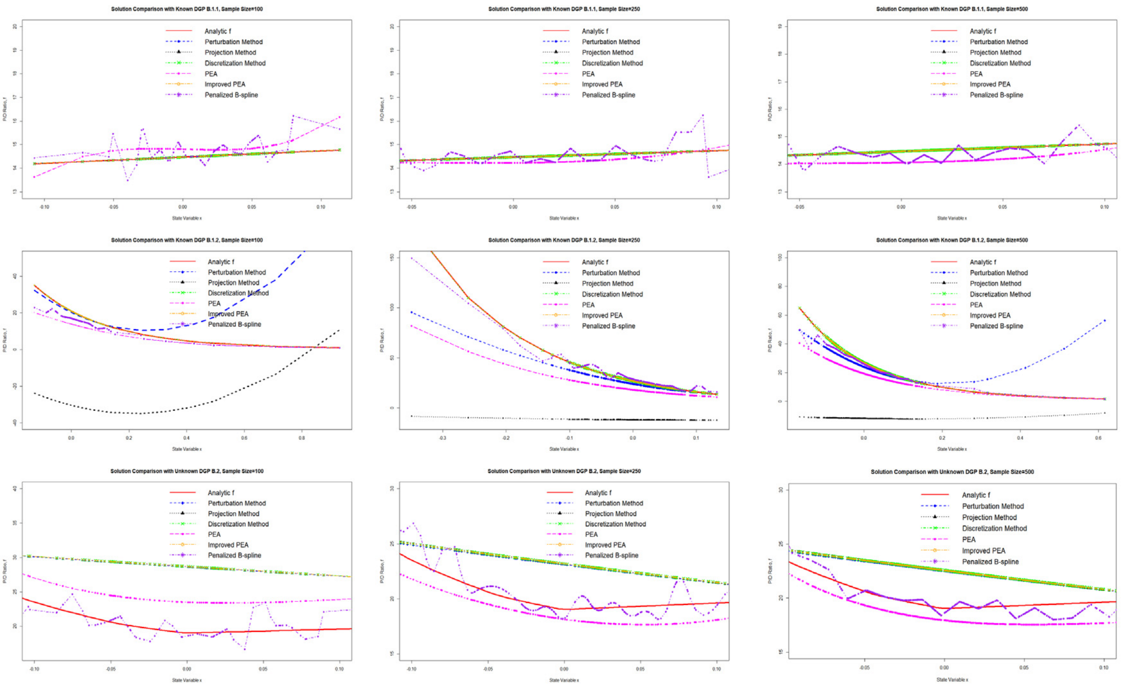

Fig. 5. P/D ratios solution comparison under known or unknown DGPs. This figure plots the P/D ratios solved from the Mehra and Prescott (1985)'s model from different solution methods given the state variable follows some known or unknown dynamics. The first two rows evaluate the solution accuracy under some known DGPs, where the analytic solution is constructed by Burnside (1998)'s method. Specifically, we assume the state variable follows DGP B.1.1 in the first row and DGP B.1.2 in the second. The third row compares the solution accuracy for the case with unknown DGP as specified in DGP B.2 and we proxy the analytic solution with the discretization method under the correct DGP and sufficiently fine grids. For our penalized two-stage B-splines approach, we report the results under $p = 1$ and $K = 35$ with equally spaced knots and the optimal penalty is determined via the proposed fast GCV algorithm. For the perturbation, projection, PEA and improved PEA methods, we report their associated results with order equal to three to encourage possible nonlinearity in unknown P/D ratios. For the PEA and Improved PEA algorithms, the depreciation parameter is set to be 0.5. For the analytic and numerical solution methods which require iterations, we set the tolerance level equal to $1e{-}7$. For each row, the sample size is 100 for the left column, 250 for the middle one and 500 for the right one.

and Campbell and Cochrane's (1999) models with the above empirical data. When solving the P/D ratio numerically using the perturbation, projection, discretization, and parametrized expected algorithm methods, a conventional protocol is to fit the empirical consumption growth data with an AR(1) process in addition to a parametric functional form of the P/D ratio. Therefore, we set the order of the Taylor expansion or series expansion in numerical solution methods to be 3 in order to incorporate possible non-linearities, given that there is no available data-driven order suggestion criterion for all these numerical solution methods. We then apply our newly proposed penalized two-stage nonparametric B-splines estimation method for the unknown P/D ratios for each model.

Table 6 reports the first two moments of the risk-free and risky assets using different solution methods for the P/D ratios in Mehra and Prescott's (1985) model. Except for the PEA method, because all the numerical solution methods listed above are subject to the same type of possible misspecification errors on the conditional distribution of state dynamics, the mean and variance of the risk-free asset from these methods are identical. In contrast to the results from the numerical solution methods above, when we incorporate the true dynamics of the consumption process using our penalized two-stage B-splines method, the model is able to generate a much lower correlation between the risk-free and risky assets.

For the Campbell and Cochrane's (1999) model, we first pin down the value of the persistence coefficient $\Gamma = 0.948$ such that it matches the annual serial correlation of the empirical $\log(P/D)$ ratios. We set the subjective discount factor $\beta = 0.9728$ to match the risk-free rate with the average annual nominal return on Treasury bills to be around 1.15%. g_c, δ_c, g_d and σ_d are the mean and standard deviation of log consumption and dividend growth from a normal distribution fitting of the consumption and dividend data, and $\rho = 0.147$ matches their correlation. We solve the model above with perfectly and imperfectly correlated consumption and dividend growth rates, respectively. For each numerical solution method, we first solve the model and obtain approximated P/D ratios based on the distributional assumptions stated above. We then obtain the model-implied P/D ratios for the empirical data using our newly proposed penalized two-stage B-splines regression without distributional assumptions on consumption and dividend processes, and without parametric form assumptions on the unknown P/D ratio function. We further compute the sample moments of equity returns based on the empirical data from 1947 to 2017. Table 7 compares the results from the discretization, perturbation, and

Table 5
Comparisons of solution accuracy with known and unknown DGP.

Solution method	Sample size = 100			Sample size = 250			Sample size = 500		
	Number of iterations	Computation time	IMSE	Number of iterations	Computation time	IMSE	Number of iterations	Computation time	IMSE
Panel A: with Known DGP									
DGP B.1.1									
Perturbation	1	0.022	<0.001	1	0.028	<0.001	1	0.002	<0.001
Projection	1	6.961	<0.001	1	6.912	<0.001	1	7.285	<0.001
Discretization	276	5.188	<0.001	276	5.297	<0.001	276	5.289	<0.001
PEA									
(order = 1)	203	0.112	0.097	197	0.113	0.038	195	0.131	0.002
(order = 2)	203	0.120	0.097	196	0.113	0.059	195	0.138	0.005
(order = 3)	204	0.116	0.125	197	0.116	0.055	195	0.136	0.006
Improved PEA									
(order = 1)	3169	2.654	<0.001	3169	6.930	<0.001	3169	18.600	<0.001
(order = 2)	3169	7.904	<0.001	3169	17.167	<0.001	3169	47.229	<0.001
(order = 3)	3169	13.077	<0.001	3169	24.335	<0.001	3169	80.182	<0.001
Penalized two-stage B-splines									
(p = 1)	1	0.410	0.251	1	0.410	0.109	1	0.470	0.025
(p = 2)	1	0.410	0.367	1	0.410	0.116	1	0.490	0.028
(p = 3)	1	0.410	0.216	1	0.533	0.106	1	0.490	0.027
DGP B.1.2									
Perturbation	1	0.043	45.056	1	0.041	51.657	1	0.044	22.507
Projection	1	7.358	1500	1	7.164	1810	1	7.411	1550
Discretization	477	9.815	<0.001	514	19.710	<0.001	481	8.777	<0.001
PEA									
(order = 1)	325	0.208	48.054	230	0.130	156.726	244	0.296	71.586
(order = 2)	337	0.178	45.269	230	0.135	155.739	244	0.159	70.081
(order = 3)	345	0.176	44.572	230	0.1249	153.630	244	0.155	70.058
Improved PEA									
(order = 1)	5527	5.174	6.604	5979	11.659	0.412	5563	22.129	5.574
(order = 2)	5946	15.201	0.064	5928	31.768	0.011	5936	52.081	0.035
(order = 3)	5876	23.972	<0.001	5939	47.175	<0.001	5886	82.132	<0.001
Penalized two-stage B-spline									
(p = 1)	1	0.420	22.831	1	0.420	10.857	1	0.450	4.746
(p = 2)	1	0.420	15.231	1	0.420	6.808	1	0.420	5.300
(p = 3)	1	0.410	20.149	1	0.411	8.756	1	0.420	6.432
Panel B: with unknown DGP									
DGP B.2									
Perturbation	1	0.043	75.839	1	0.025	9.939	1	0.004	7.155
Projection	1	0.793	76.665	1	1.052	10.273	1	0.785	7.395
Discretization	439	11.076	76.555	403	7.895	10.354	393	9.967	7.485
PEA									
(order = 1)	254	0.252	16.536	251	0.173	3.295	243	0.286	3.416
(order = 2)	255	0.223	16.416	251	0.186	1.846	244	0.220	2.319
(order = 3)	256	0.196	17.021	251	0.187	1.917	260	0.199	2.208
Improved PEA									
(order = 1)	4624	5.567	76.527	3609	10.340	10.346	3564	18.672	7.4775
(order = 2)	4624	15.138	76.528	3609	28.000	10.347	3564	43.758	7.480
(order = 3)	4624	24.276	76.528	3609	41.380	10.347	3564	64.608	7.479
Penalized two-stage B-splines									
(p = 1)	1	0.522	1.427	1	0.530	1.267	1	0.890	0.684
(p = 2)	1	0.610	1.859	1	0.530	1.057	1	0.510	0.718
(p = 3)	1	0.524	1.593	1	0.530	1.179	1	0.510	0.677

Notes: in Panel A, we evaluate the solution accuracy under some known DGPs as described in DGP B.1.1 and DGP B.1.2, where the analytic solution is constructed by Burnside (1998)'s method. In Panel B, we consider an unknown DGP as specified in DGP B.2 and proxy the analytic solution with the discretization method under the correct DGP and sufficiently fine grids. For our penalized two-stage B-splines approach, we report the results under $p = 1, 2$ and 3 with equally spaced knots $K = 35$. The optimal penalty is determined via the proposed fast GCV algorithm. For the perturbation and projection methods, we report the results with order equal to 3 and for the PEA and improved PEA methods, we report their associated results with order up to three and the involved depreciation parameter is set to 0.5. For the analytic and numerical solution methods which require iterations, we set the tolerance level equal to $1e{-}7$. For each case, we report the number of iterations, computation time (in seconds) and the IMSE for the sample size equal to 100, 250 and 500. Computation time is measured in seconds.

improved PEA and the penalized two-stage regression methods. We omit the results from the projection and the PEA methods because the former renders undesirable performance, given a near one Γ, while the latter yields a solution

Table 6
Asset returns comparisons under different solution methods based on Mehra and Prescott's (1985) model.

Solution method	The Mehra and Prescott (1985)'s model						
	Parametric assumptions on		The first two moments of asset returns				
	State dynamics	Functional form	μ_f	σ_f	μ_R	σ_R	$\rho_{f,R}$
Perturbation	Yes	Yes	5.410	0.790	5.380	1.414	0.567
Projection	Yes	Yes	5.410	0.790	5.407	1.732	0.527
Discretization	Yes	Yes	5.410	0.790	5.338	1.416	0.565
PEA	No	Yes	5.421	0.892	5.426	1.370	0.880
Improved PEA	Yes	Yes	5.410	0.790	5.380	1.417	0.566
Penalized two-stage B-splines							
(p = 1)	No	No	5.419	0.799	5.417	1.422	0.130
(p = 2)	No	No	5.419	0.799	5.417	1.422	0.125
(p = 3)	No	No	5.419	0.799	5.417	1.426	0.128

Notes: this table compares the first two moments of asset returns based on the P/D ratios solved from the Mehra and Prescott (1985)'s model from different solution methods. All returns are in annual percentage. μ_R and σ_R are the mean and standard deviation of the risky return; μ_f and σ_f are the mean and standard deviation of the risk-free return; $\rho_{f,R}$ is the correlation of the risk-free and risky returns. For numerical solution methods, which require a complete specification of the conditional distribution, we assume the state variable follows an AR(1) process with mean and variance estimated from the U.S. 1947–2017 quarterly data. For the perturbation, projection, PEA and improved PEA methods, we report the results with order equal to 3 (a complete table with numerical solution results based on high order expansions is available upon request). For the PEA and Improved PEA methods, the depreciation parameter is set to be 0.5. For the studied numerical solution methods which require iterations, we set the tolerance level equal to $1e{-}7$. For our penalized two-stage B-splines approach, we report the results under $p = 1, 2$ and 3 with equally spaced knots $K = 35$. The optimal penalty is determined via the proposed fast GCV algorithm.

with unconvergence because of the ill-posed problem.[10] First, after relaxing the normality assumption on consumption growth rates, we find that the data-driven model-implied risk-free rate has a negative average return. This may be because of our method's allowance of true asymmetric consumption shocks in the pricing procedure. Second, when pricing the imperfectly correlated consumption and dividend claims, we relax the bivariate normal assumption on consumption and dividend growth. By doing so, the true dynamics reveal a negative skewness for the equity premiums. Allowing the true data to determine equilibrium asset prices largely enhances the explanatory powers of Campbell and Cochrane's (1999) model.

5. Empirical applications

We now reinvestigate the present value of future dividends, equity returns, and stock predictability in a misspecification-free environment using our newly proposed penalized two-stage estimation method. In line with Ang and Bekaert (2006), consumptions and dividends are summed over the past one year to alleviate seasonality effects that are especially pronounced in the dividend payment process. We thus represent the annualized consumptions and dividends by $C_t^4 = C_t + C_{t-1} + C_{t-2} + C_{t-3}$ and $D_t^4 = D_t + D_{t-1} + D_{t-2} + D_{t-3}$.

5.1. Predictability of equity returns

The determination of the predictability of future stock returns is important. Unfortunately, no consensus has been reached in theory and empirical evidence (Campbell and Yogo, 2006; Phillips and Lee, 2013). The path-breaking theory developed by Campbell and Shiller (1988), which is based on a rational consumption-based asset pricing model and is under log linearization, suggests that high dividend yields predict high future excess returns, low dividend growth, or both. Ang and Bekaert (2006) relax log linearization treatment, and thus obtains an astonishingly different predictive relationship that states that dividend yield is not a natural predictor for returns, but is one for time-varying discount rates. Although both theories expect a negative predictive relationship between dividend yields and cash flows, scant empirical evidence has been found to confirm this expectation (Lettau and Ludvigson, 2005; Campbell and Yogo, 2006; Cujean and Hasler, 2017).

Built on equilibrium asset prices under full rationality in Euler equations, the implied dividend yields, $log(1/f_t)$, reflect consumption-based rational forecasts of future returns and dividend growth. In this section, we investigate how nonparametrically estimated implied dividend yields predict future excess returns and dividend growth.

We consider the following multivariate regression model to investigate the predictability relationship.

$$\tilde{y}_{t+j} = \beta_0 + \beta_{j,1} log(1/f_t) + \beta_{j,2} dy_t^4 + \beta_{j,3} r_{f,t} + \epsilon_t, \tag{9}$$

where $\tilde{y}_{t+j} = (4/j)[(r_{t+1} - r_{f,t}) + \cdots + (r_{t+j} - r_{f,t+j-1})]$ is the annualized j-period excess log return for the US aggregate market, $r_{f,t}$ is the log return on short interest rates and dy_t^4 is the annualized observed dividend yields from quarterly

[10] A full table is available upon request.

Table 7
Asset returns comparisons under different solution methods based on Campbell and Cochrane's (1999) model.

	Pricing the consumption claim						
	Discretization	Perturbation	Improved PEA	Penalized B-spline $p=1$	$p=2$	$p=3$	Postwar sample
$\sigma(R_f)$	0.000	0.000	0.000	17.707	17.719	17.869	0.844
$E(R-R_f)$	6.259	63.178	0.518	3.427	3.732	3.565	6.604
$\sigma(R-R_f)$	26.024	10.695	27.714	33.798	30.905	33.257	29.065
$skewness(R-R_f)$	0.552	1.694	0.397	1.087	1.523	1.139	−1.126
$Kurtosis(R-R_f)$	7.541	7.078	8.123	16.318	13.404	14.528	3.400
Distributional assumptions	Yes $\Delta c_t = \Delta d_t \sim i.i.d.N(g_c, \sigma_c^2)$	Yes	Yes	No	No	No	No
Functional form assumption	No	Yes	Yes	No	No	No	No
	Imperfectly correlated consumption and dividend claims						
	Discretization	Perturbation	Improved PEA	Penalized B-spline $p=1$	$p=2$	$p=3$	Postwar sample
$\sigma(R_f)$	0.000	0.000	0.000	17.439	17.719	17.869	0.844
$E(R-R_f)$	5.372	51.094	2.133	5.578	4.388	6.862	6.604
$\sigma(R-R_f)$	25.730	7.487	26.74	42.130	43.687	41.243	29.065
$skewness(R-R_f)$	0.552	1.724	0.522	−0.733	−1.879	−1.405	−1.126
$Kurtosis(R-R_f)$	7.830	6.691	8.369	7.185	7.573	5.871	3.400
Distributional assumptions	Yes $(\Delta c_t, \Delta d_t) \sim ii.d.N(g_c, g_d, \sigma_c^2, \sigma_d^2, \rho)$	Yes	Yes	No	No	No	No
Functional form assumption	No	Yes	Yes	No	No	No	No

Notes: this table compares the implications of asset returns based on the P/D ratios solved from the Campbell and Cochrane (1999)'s model from different solution methods. We set the subjective discount factor $\beta = 0.9728$ to match the risk-free rate with the average annual nominal return on Treasury bills to be around 1.15%. All returns are in annual percentage. R is the simple return of the risky asset. R_f is the simple return of the risk-free asset. For the perturbation and improved PEA methods, we only report the results with order equal to 2. For the improved PEA methods, the depreciation parameter is set to be 0.5. For the studied numerical solution methods which require iterations, we set the tolerance level equal to $1e{-7}$. We omit the results from the projection and PEA methods because the former renders undesirable performance given a near one Γ and the latter has a diverging solution due to the ill-posed problem (a complete table with numerical solution results based on higher order expansions is available upon request). For our penalized two-stage B-splines approach, we report the results under $p = 1, 2$ and 3 with equally spaced knots $K = 35$. The optimal penalty is determined via the proposed fast GCV algorithm.

S&P500 data. We also conduct two other sets of regression analysis, that the dependent variable represents dividend growth and short interest rates, respectively. To properly address the strong serial dependency in explanatory variables and the dependent variables, we follow the literature by reporting standard deviations with various treatments, including the Newey and West (1987) corrected standard deviation, the Hodrick-type standard errors and the Cochrane–Orcutt-type standard errors.

Table 8 displays the predictive power that short interest rates, observed dividend yield, and implied dividend yield exhibit on excess returns. First, our result confirms the short-horizon forecastability phenomenon when we use the Hodrick-type correct covariance estimation. We confirm that the implied dividend yield from a rational model is not a natural and significant predictor for excess returns. We find that the significant forecastability from observed dividend yields mainly arises from the component that is beyond rational models.

We also investigate forecasts of short-horizon dividend growth. Table 9 reports the forecasting results for the future dividend growth from the same exercise. The accounting identity of the P/D ratio indicates a negative relationship between dividend yields and future dividend growth (Lettau and Ludvigson, 2005). However, dividend yields have been documented to have little or mixed forecastability of dividend growth. By proposing a proxy through a counteraction linear regression model, Lettau and Ludvigson (2005) find that consumption-based dividend yields exhibit a reversed forecastability for dividend growth, that is, high dividend payments relative to prices predict higher dividend growth, not lower. In our empirical study, Table 9 shows that true implied dividend yields have negative predictability for dividend growth, but they are still a short-horizon phenomenon. When implied dividend yields are controlled, observed dividend yields exhibit significant positive predictability on dividend growth. Therefore, we provide an alternative explanation for this positive forecastability, namely, it is the component that is orthogonal to the rational dividend yields that leads to the positive forecastability in dividend growth. Our results provide clear evidence for time-varying facts of dividend growth, which has long been treated in the literature as a process with a constant growth rate.

Our study also investigates the forecastability relationship between dividend yields and short interest rates. Table 10 reports predictive regression coefficients and their t-statistics for future annualized short interest rates on implied dividend yields, lagged short interest rates and observed dividend yields. First, we find that short interest is also a highly persistent process. Therefore, as suggested by Ang and Bekaert (2006), to address highly correlated regression residuals in

ARTICLE IN PRESS

Table 8
Predictability on excess returns.

	j-period regression: $\tilde{y}_{t+j}=\beta_0+\beta_j' z_t+\epsilon_{t,t+h}$															
	f_t *estimated from the Mehra and Prescott (1985)'s model*															
	j = 1				j = 2				j = 3				j = 4			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
dy_t^4	0.079*	0.097**	0.100**		0.150	0.180	0.188*		0.211	0.253	0.269		0.267	0.320	0.344	
	(0.039)	(0.041)	(0.041)		(0.058)	(0.061)	(0.061)		(0.072)	(0.074)	(0.074)		(0.083)	(0.086)	(0.085)	
	[0.043]	[0.042]	[0.040]		[0.082]	[0.081]	[0.080]		[0.119]	[0.118]	[0.118]		[0.154]	[0.155]	[0.155]	
	{0.040}	{0.041}	{0.039}		{0.119}	{0.118}	{0.114}		{0.224}	{0.223}	{0.219}		{0.347}	{0.348}	{0.343}	
$r_{f,t}$		−3.500	−3.271	−1.894		−5.887	−5.239	−2.661		−8.144	−6.874	−3.222		−10.348	−8.403	−3.764
		(2.132)	(2.152)	(2.096)		(3.186)	(3.205)	(3.144)		(3.915)	(3.911)	(3.863)		(4.525)	(4.478)	(4.453)
		[2.403]	[2.461]	[2.440]		[4.427]	[4.522]	[4.453]		[5.841]	[5.901]	[5.751]		[6.950]	[6.930]	[6.625]
		{2.283}	{2.331}	{2.284}		{6.501}	{6.643}	{6.585}		{11.594}	{11.684}	{11.487}		{17.511}	{17.222}	{16.778}
$\log(1/f_t)$			−0.525	−0.390			−1.495	−1.240			−2.928	−2.565			−4.500	−4.037
			(0.642)	(0.645)			(0.955)	(0.966)			(1.164)	(1.185)			(1.330)	(1.363)
			[1.093]	[1.151]			[1.722]	[1.832]			[2.005]	[2.151]			[2.298]	[2.472]
			{1.001}	{1.034}			{2.604}	{2.696}			{3.972}	{4.117}			{5.375}	{5.579}
Constant	0.337**	0.438***	−1.136	−1.085	0.643	0.814**	−3.669	−3.575	0.918	1.154	−7.629	−7.496	1.177	1.474	−12.022	−11.858
	(0.135)	(0.148)	(1.931)	(1.948)	(0.202)	(0.221)	(2.872)	(2.917)	(0.249)	(0.272)	(3.500)	(3.577)	(0.288)	(0.313)	(4.001)	(4.114)
	[0.143]	[0.146]	[3.331]	[3.463]	[0.275]	[0.282]	[5.244]	[5.510]	[0.398]	[0.415]	[6.106]	[6.466]	[0.513]	[0.547]	[6.983]	[7.433]
	{0.136}	{0.143}	{3.046}	{3.113}	{0.398}	{0.406}	{7.938}	{8.112}	{0.751}	{0.776}	{12.112}	{12.399}	{1.165}	{1.218}	{16.316}	{16.831}
T	279	279	279	279	278	278	278	278	277	277	277	277	276	276	276	276
R^2	0.014	0.024	0.026	0.005	0.023	0.035	0.044	0.010	0.030	0.045	0.067	0.022	0.036	0.054	0.093	0.037
Adjusted R^2	0.011	0.017	0.016	−0.002	0.020	0.028	0.033	0.003	0.027	0.038	0.057	0.015	0.033	0.047	0.083	0.030
	f_t *estimated from the Campbell and Cochrane (1999)'s model*															
	j = 1				j = 2				j = 3				j = 4			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
dy_t^4	0.079*	0.097**	0.098**		0.150	0.180	0.180		0.211	0.253	0.253		0.267	0.320	0.320	
	(0.039)	(0.041)	(0.040)		(0.058)	(0.061)	(0.060)		(0.072)	(0.074)	(0.074)		(0.083)	(0.086)	(0.086)	
	[0.043]	[0.042]	[0.042]		[0.082]	[0.081]	[0.081]		[0.119]	[0.118]	[0.119]		[0.154]	[0.155]	[0.156]	
	{0.040}	{0.041}	{0.041}		{0.119}	{0.118}	{0.118}		{0.224}	{0.223}	{0.224}		{0.347}	{0.348}	{0.348}	
$r_{f,t}$		−3.500	−3.504	−2.089		−5.887	−5.944	−3.347		−8.144	−8.198	−4.582		−10.348	−10.410	−5.875
		(2.132)	(2.128)	(2.063)		(3.186)	(3.180)	(3.102)		(3.915)	(3.913)	(3.839)		(4.525)	(4.521)	(4.458)
		[2.403]	[2.412]	[2.432]		[4.427]	[4.425]	[4.416]		[5.841]	[5.829]	[5.735]		[6.950]	[6.936]	[6.675]
		{2.283}	{2.287}	{2.267}		{6.501}	{6.532}	{6.551}		{11.594}	{11.643}	{11.501}		{17.511}	{17.559}	{17.087}
$\log(1/f_t)$			0.060	0.059			0.093	0.092			0.091	0.089			0.109	0.108
			(0.041)	(0.042)			(0.063)	(0.064)			(0.077)	(0.079)			(0.089)	(0.091)
			[0.037]	[0.038]			[0.066]	[0.067]			[0.079]	[0.079]			[0.094]	[0.093]
			{0.037}	{0.036}			{0.088}	{0.088}			{0.178}	{0.180}			{0.248}	{0.250}
Constant	0.337**	0.438***	0.556***	0.203***	0.643	0.814**	0.994**	0.345*	0.918	1.154	1.329	0.419	1.177	1.474	1.685	0.534
	(0.135)	(0.148)	(0.168)	(0.084)	(0.202)	(0.221)	(0.252)	(0.130)	(0.249)	(0.272)	(0.310)	(0.161)	(0.288)	(0.313)	(0.358)	(0.186)
	[0.143]	[0.146]	[0.169]	[0.077]	[0.275]	[0.282]	[0.317]	[0.140]	[0.398]	[0.415]	[0.450]	[0.166]	[0.513]	[0.547]	[0.597]	[0.192]
	{0.13}	{0.143}	{0.164}	{0.073}	{0.398}	{0.406}	{0.456}	{0.182}	{0.751}	{0.776}	{0.894}	{0.363}	{1.165}	{1.218}	{1.355}	{0.506}

(*continued on next page*)

Table 8 (*continued*).

	f_t estimated from the Campbell and Cochrane (1999)'s model															
	j = 1				j = 2				j = 3				j = 4			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
T	279	279	279	279	278	278	278	278	277	277	277	277	276	276	276	276
R^2	0.014	0.024	0.031	0.011	0.023	0.035	0.043	0.012	0.030	0.045	0.050	0.010	0.036	0.054	0.060	0.011
Adjusted R^2	0.011	0.017	0.021	0.004	0.020	0.028	0.032	0.004	0.027	0.038	0.040	0.002	0.033	0.047	0.049	0.004

Notes: for all forecasting horizon j, we have $z_t = (1, dy_t^4)'$ in (1), $z_t = (1, dy_t^4, r_{f,t})'$ in (2) , $z_t = (1, dy_t^4, r_{f,t}, \log(1/f_t))'$ in (3) and $z_t = (1, r_{f,t}, \log(1/f_t))'$ in (4). $\tilde{y}_{t+j} = (4/j)\sum_{i=1}^{j}(r_{t+i} - r_{f,t+i-1})$ is the annualized cumulative excess return, dy_t^4 is the observed dividend yields, $r_{f,t}$ is the log return on short interest rates and f_t is the estimated P/D ratios. For each regression, the table reports OLS estimates and the associated standard deviation in the parenthesis, Newey and West (1987) corrected standard deviation in the brackets and the Hodrick-type standard errors in the curly brackets. The significance level marked by $*$ is based on the Hodrick-type t-statistics.

*$p < 0.1$.

**$p < 0.05$.

***$p < 0.01$.

Table 9
Predictability on dividend growth.

	j-period regression: $d_{t+j} - d_t = \beta_0 + \beta_j' z_t + \epsilon_{t,t+h}$															
	f_t *estimated from the Mehra and Prescott (1985)'s model*															
	j = 1				j = 2				j = 3				j = 4			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
dy_t^4	0.008***	0.008***	0.008***		0.015	0.015	0.016		0.022	0.022	0.023		0.028	0.029	0.030	
	(0.002)	(0.002)	(0.002)		(0.004)	(0.004)	(0.004)		(0.006)	(0.006)	(0.006)		(0.008)	(0.008)	(0.008)	
	[0.004]	[0.004]	[0.004]		[0.009]	[0.010]	[0.010]		[0.015]	[0.016]	[0.016]		[0.020]	[0.022]	[0.022]	
	{0.002}	{0.003}	{0.003}		{0.010}	{0.011}	{0.01}		{0.022}	{0.024}	{0.023}		{0.038}	{0.041}	{0.040}	
$r_{f,t}$		−0.045	−0.010	0.107		−0.103	−0.046	0.175		−0.178	−0.119	0.197		−0.279	−0.238	0.163
		(0.118)	(0.118)	(0.116)		(0.232)	(0.233)	(0.230)		(0.342)	(0.344)	(0.339)		(0.444)	(0.448)	(0.442)
		[0.207]	[0.196]	[0.180]		[0.461]	[0.437]	[0.400]		[0.740]	[0.700]	[0.642]		[1.030]	[0.972]	[0.896]
		{0.126}	{0.119}	{0.109}		{0.497}	{0.468}	{0.421}		{1.086}	{1.019}	{0.915}		{1.863}	{1.734}	{1.559}
$\log(1/f_t)$			−0.080**	−0.069**			−0.130	−0.108			−0.137	−0.105			−0.094	−0.054
			(0.035)	(0.036)			(0.070)	(0.071)			(0.102)	(0.104)			(0.133)	(0.135)
			[0.050]	[0.051]			[0.113]	[0.116]			[0.184]	[0.192]			[0.263]	[0.277]
			{0.033}	{0.033}			{0.133}	{0.137}			{0.318}	{0.333}			{0.611}	{0.639}
Constant	0.042***	0.043***	−0.197**	−0.193*	0.081**	0.084**	−0.306	−0.298	0.118	0.123	−0.287	−0.276	0.154	0.162	−0.121	−0.106
	(0.007)	(0.008)	(0.106)	(0.108)	(0.015)	(0.016)	(0.209)	(0.214)	(0.022)	(0.024)	(0.308)	(0.314)	(0.028)	(0.031)	(0.400)	(0.408)
	[0.014]	[0.017]	[0.151]	[0.151]	[0.032]	[0.038]	[0.341]	[0.345]	[0.051]	[0.060]	[0.559]	[0.572]	[0.070]	[0.083]	[0.798]	[0.826]
	{0.009}	{0.010}	{0.098}	{0.099}	{0.034}	{0.041}	{0.403}	{0.411}	{0.076}	{0.090}	{0.967}	{0.995}	{0.132}	{0.154}	{1.852}	{1.910}
T	279	279	279	279	278	278	278	278	277	277	277	277	276	276	276	276
R^2	0.045	0.046	0.063	0.014	0.043	0.043	0.056	0.009	0.042	0.043	0.049	0.004	0.041	0.043	0.045	0.001
Adjusted R^2	0.042	0.039	0.053	0.007	0.039	0.037	0.045	0.002	0.038	0.036	0.038	−0.003	0.038	0.036	0.034	−0.006
	impdy estimated from the Campbell and Cochrane (1999)'s model															
	j = 1				j = 2				j = 3				j = 4			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
dy_t^4	0.0002	0.0004	0.0003		0.001	0.002	0.002		0.001	0.003	0.003		0.001	0.005	0.004	
	(0.003)	(0.003)	(0.003)		(0.006)	(0.006)	(0.006)		(0.008)	(0.009)	(0.009)		(0.011)	(0.011)	(0.011)	
	[0.005]	[0.005]	[0.005]		[0.011]	[0.011]	[0.011]		[0.018]	[0.017]	[0.017]		[0.024]	[0.023]	[0.023]	
	{0.003}	{0.003}	{0.003}		{0.013}	{0.013}	{0.013}		{0.029}	{0.028}	{0.028}		{0.049}	{0.048}	{0.048}	
$r_{f,t}$		−0.026	−0.026	−0.021		−0.162	−0.156	−0.130		−0.385	−0.377	−0.330		−0.669	−0.659	−0.595
		(0.168)	(0.168)	(0.161)		(0.316)	(0.316)	(0.303)		(0.456)	(0.455)	(0.437)		(0.580)	(0.579)	(0.557)
		[0.253]	[0.251]	[0.264]		[0.529]	[0.525]	[0.559]		[0.791]	[0.784]	[0.846]		[1.016]	[1.007]	[1.099]
		{0.167}	{0.166}	{0.171}		{0.603}	{0.599}	{0.625}		{1.228}	{1.218}	{1.278}		{1.925}	{1.906}	{1.995}
$\log(1/f_t)$			−0.005**	−0.005**			−0.009	−0.009			−0.014	−0.014			−0.017	−0.017
			(0.003)	(0.003)			(0.006)	(0.006)			(0.009)	(0.009)			(0.011)	(0.011)
			[0.003]	[0.003]			[0.006]	[0.006]			[0.009]	[0.009]			[0.011]	[0.011]
			{0.002}	{0.002}			{0.007}	{0.007}			{0.013}	{0.014}			{0.020}	{0.020}
Constant	0.011	0.012	0.002	0.001	0.024	0.029	0.011	0.004	0.036	0.047	0.021	0.009	0.046	0.065	0.033	0.017
	(0.011)	(0.012)	(0.013)	(0.007)	(0.020)	(0.022)	(0.025)	(0.013)	(0.029)	(0.032)	(0.036)	(0.018)	(0.037)	(0.040)	(0.046)	(0.023)
	[0.017]	[0.017]	[0.018]	[0.007]	[0.037]	[0.036]	[0.037]	[0.015]	[0.057]	[0.055]	[0.057]	[0.022]	[0.077]	[0.074]	[0.078]	[0.029]
	{0.011}	{0.012}	{0.012}	{0.005}	{0.043}	{0.044}	{0.045}	{0.017}	{0.093}	{0.094}	{0.097}	{0.033}	{0.157}	{0.160}	{0.167}	{0.050}

(continued on next page)

Table 9 (*continued*).

	impdy estimated from the Campbell and Cochrane (1999)'s model															
	j = 1				j = 2				j = 3				j = 4			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
T	279	279	279	279	278	278	278	278	277	277	277	277	276	276	276	276
R^2	0.00002	0.0001	0.008	0.008	0.0001	0.001	0.009	0.009	0.0001	0.003	0.011	0.010	0.00004	0.005	0.013	0.012
Adjusted R^2	−0.004	−0.007	−0.003	0.001	−0.004	−0.006	−0.002	0.002	−0.004	−0.005	0.0001	0.003	−0.004	−0.002	0.002	0.005

Notes: for all forecasting horizon j, we have $z_t = (1, dy_t^4)'$ in (1), $z_t = (1, dy_t^4, r_{f,t})'$ in (2) , $z_t = (1, dy_t^4, r_{f,t}, \log(1/f_t))'$ in (3) and $z_t = (1, r_{f,t}, \log(1/f_t))'$ in (4). dy_t^4 is the observed dividend yields, $r_{f,t}$ is the log return on short interest rates and f_t is the estimated P/D ratios. For each regression, the table reports OLS estimates and the associated standard deviation in the parenthesis, Newey and West (1987) corrected standard deviation in the brackets and the Hodrick-type standard errors in the curly brackets. The significance level marked by $*$ is based on the Hodrick-type t-statistics.

$^*p < 0.1$.

$^{**}p < 0.05$.

$^{***}p < 0.01$.

Table 10
Predictability on short interest rates.

	j-period regression: $\sum_{i=1}^{j} r_{f,t+i} = \beta_0 + \beta_j' z_t + \epsilon_{t,t+h}$															
	f_t estimated from the Mehra and Prescott (1985)'s model															
	j = 1				j = 2				j = 3				j = 4			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
dy_t^4	−0.0002	0.0005	0.0004		−0.003	0.002**	0.002*		−0.002	−0.002	−0.003		0.002	0.002	0.001	
	(0.001)	(0.0004)	(0.0004)		(0.002)	(0.001)	(0.001)		(0.003)	(0.001)	(0.001)		(0.004)	(0.002)	(0.002)	
	[0.002]	[0.0003]	[0.0003]		[0.005]	[0.001]	[0.001]		[0.008]	[0.002]	[0.002]		[0.012]	[0.003]	[0.003]	
	{0.001}	{.0003}	{0.0003}		{0.005}	{0.001}	{0.001}		{0.011}	{0.003}	{0.003}		{0.019}	{0.006}	{0.006}	
	⟨0.003⟩	⟨0.0003⟩	⟨0.0003⟩		⟨0.003⟩	⟨0.001⟩	⟨0.001⟩		⟨0.003⟩	⟨0.003⟩	⟨0.003⟩		⟨0.004⟩	⟨0.004⟩	⟨0.004⟩	
$r_{f,t}$		0.955***	0.948***	0.953***		1.704***	1.713***	1.741***		−0.240***	−0.226***	−0.225***		−0.225***	−0.203**	−0.203**
		(0.023)	(0.023)	(0.022)		(0.044)	(0.043)	(0.042)		(0.066)	(0.063)	(0.062)		(0.092)	(0.088)	(0.086)
		[0.033]	[0.032]	[0.032]		[0.070]	[0.067]	[0.068]		[0.111]	[0.109]	[0.113]		[0.166]	[0.162]	[0.171]
		{0.051}	{0.053}	{0.055}		{0.162}	{0.171}	{0.180}		{0.302}	{0.318}	{0.339}		{0.541}	{0.558}	{0.593}
		⟨0.017⟩	⟨0.017⟩	⟨0.016⟩		⟨0.053⟩	⟨0.050⟩	⟨0.048⟩		⟨0.074⟩	⟨0.074⟩	⟨0.074⟩		⟨0.086⟩	⟨0.086⟩	⟨0.086⟩
$\log(1/f_t)$			0.018***	0.019***			0.060***	0.063***			0.025	0.024			0.041**	0.042**
			(0.007)	(0.007)			(0.013)	(0.013)			(0.019)	(0.019)			(0.026)	(0.026)
			[0.006]	[0.006]			[0.014]	[0.014]			[0.025]	[0.025]			[0.039]	[0.039]
			{0.010}	{0.010}			{0.035}	{0.034}			{0.066}	{0.065}			{0.115}	{0.113}
			⟨0.005⟩	⟨0.005⟩			⟨0.015⟩	⟨0.015⟩			⟨0.018⟩	⟨0.018⟩			⟨0.021⟩	⟨0.020⟩
Constant	0.011	0.002*	0.057***	0.057***	0.011	0.011***	0.191***	0.191***	0.029	0.030	0.103*	0.108*	0.052*	0.053	0.174**	0.173**
	(0.004)	(0.002)	(0.021)	(0.021)	(0.007)	(0.003)	(0.039)	(0.039)	(0.011)	(0.005)	(0.057)	(0.057)	(0.015)	(0.006)	(0.078)	(0.079)
	[0.008]	[0.001]	[0.018]	[0.018]	[0.018]	[0.003]	[0.041]	[0.041]	[0.030]	[0.006]	[0.074]	[0.075]	[0.043]	[0.010]	[0.116]	[0.118]
	{0.005}	{0.001}	{0.029}	{0.029}	{0.018}	{0.004}	{0.103}	{0.104}	{0.040}	{0.010}	{0.198}	{0.199}	{0.071}	{0.019}	{0.342}	{0.344}
	⟨0.009⟩	⟨0.001⟩	⟨0.015⟩	⟨0.015⟩	⟨0.015⟩	⟨0.004⟩	⟨0.044⟩	⟨0.044⟩	⟨0.022⟩	⟨0.026⟩	⟨0.058⟩	⟨0.057⟩	⟨0.031⟩	⟨0.034⟩	⟨0.069⟩	⟨0.069⟩
T	279	279	279	279	278	278	278	278	277	277	277	277	276	276	276	276
R^2	0.082	0.863	0.868	0.867	0.088	0.871	0.880	0.878	0.093	0.871	0.883	0.881	0.098	0.857	0.873	0.870
Adjusted R^2	0.079	0.862	0.866	0.866	0.085	0.870	0.878	0.878	0.090	0.871	0.882	0.880	0.095	0.856	0.871	0.869
	f_t estimated from the Campbell and Cochrane (1999)'s model															
	j = 1				j = 2				j = 3				j = 4			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
dy_t^4	−0.0002	0.0005	0.0005		−0.003	0.002**	0.002**		−0.002	−0.002	−0.002		0.002	0.002	0.002	
	(0.001)	(0.0004)	(0.0004)		(0.002)	(0.001)	(0.001)		(0.003)	(0.001)	(0.001)		(0.004)	(0.002)	(0.002)	
	[0.002]	[0.0003]	[0.0003]		[0.005]	[0.001]	[0.001]		[0.008]	[0.002]	[0.002]		[0.012]	[0.003]	[0.003]	
	{ 0.001}	{0.0003}	{0.0003}		{0.005}	{0.001}	{0.001}		{0.011}	{0.003}	{0.003}		{0.019}	{0.006}	{0.006}	
	⟨0.003⟩	⟨0.0003⟩	⟨0.0003⟩		⟨0.003⟩	⟨0.001⟩	⟨0.001⟩		⟨0.003⟩	⟨0.003⟩	⟨0.003⟩		⟨0.004⟩	⟨0.004⟩	⟨0.004⟩	
$r_{f,t}$		0.955***	0.955***	0.962***		1.704***	1.707***	1.749***		−0.240***	−0.229***	−0.227***		−0.225***	−0.233***	−0.234***
		(0.023)	(0.023)	(0.022)		(0.044)	(0.044)	(0.043)		(0.066)	(0.066)	(0.064)		(0.092)	(0.092)	(0.090)
		[0.033]	[0.033]	[0.032]		[0.070]	[0.070]	[0.069]		[0.111]	[0.111]	[0.110]		[0.166]	[0.166]	[0.166]
		{0.051}	{0.051}	{0.052}		{0.162}	{0.163}	{0.169}		{0.302}	{0.302}	{0.318}		{0.541}	{0.542}	{0.570}
		⟨0.017⟩	⟨0.017⟩	⟨0.016⟩		⟨0.053⟩	⟨0.053⟩	⟨0.051⟩		⟨0.074⟩	⟨0.074⟩	⟨0.073⟩		⟨0.086⟩	⟨0.086⟩	⟨0.086⟩
$\log(1/f_t)$			−0.0003	−0.0003			−0.0003	−0.0003			0.001**	0.001**			−0.001	−0.001
			(0.0004)	(0.0005)			(0.001)	(0.001)			(0.001)	(0.001)			(0.002)	(0.002)
			[0.001]	[0.001]			[0.001]	[0.001]			[0.003]	[0.003]			[0.003]	[0.003]
			{0.001}	{0.001}			{0.002}	{0.002}			{0.003}	{0.003}			{0.005}	{0.005}
			⟨0.0004⟩	⟨0.0004⟩			⟨0.001⟩	⟨0.001⟩			⟨0.001⟩	⟨0.001⟩			⟨0.001⟩	⟨0.001⟩
Constant	0.011	0.002*	0.002	−0.0001	0.011	0.011***	0.010**	0.002	0.029	0.030	0.031	0.039*	0.052*	0.053	0.052	0.046
	(0.004)	(0.002)	(0.002)	(0.001)	(0.007)	(0.003)	(0.004)	(0.002)	(0.011)	(0.005)	(0.005)	(0.003)	(0.015)	(0.006)	(0.007)	(0.004)
	[0.008]	[0.001]	[0.002]	[0.001]	[0.018]	[0.003]	[0.004]	[0.001]	[0.030]	[0.006]	[0.008]	[0.002]	[0.043]	[0.010]	[0.012]	[0.003]
	{0.005}	{0.001}	{0.002}	{0.001}	{0.018}	{0.004}	{0.006}	{0.004}	{0.040}	{0.010}	{0.012}	{0.007}	{0.071}	{0.019}	{0.022}	{0.010}
	⟨0.009⟩	⟨0.001⟩	⟨0.001⟩	⟨0.001⟩	⟨0.015⟩	⟨0.004⟩	⟨0.004⟩	⟨0.002⟩	⟨0.022⟩	⟨0.026⟩	⟨0.026⟩	⟨0.021⟩	⟨0.031⟩	⟨0.034⟩	⟨0.034⟩	⟨0.032⟩

(*continued on next page*)

ARTICLE IN PRESS

ARTICLE IN PRESS

Table 10 (*continued*).

	f_t *estimated from the Campbell and Cochrane (1999)'s model*															
	j = 1				j = 2				j = 3				j = 4			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
T	279	279	279	279	278	278	278	278	277	277	277	277	276	276	276	276
R^2	0.082	0.863	0.863	0.862	0.088	0.871	0.872	0.870	0.093	0.871	0.871	0.868	0.098	0.857	0.857	0.852
Adjusted R^2	0.079	0.862	0.862	0.861	0.085	0.870	0.870	0.869	0.090	0.871	0.870	0.867	0.095	0.856	0.855	0.851

Notes: for all forecasting horizon j, we have $z_t = (1, dy_t^4)'$ in (1), $z_t = (1, dy_t^4, r_{f,t})'$ in (2) , $z_t = (1, dy_t^4, r_{f,t}, \log(1/f_t))'$ in (3) and $z_t = (1, r_{f,t}, \log(1/f_t))'$ in (4). dy_t^4 is the observed dividend yields, $r_{f,t}$ is the log return on short interest rates and f_t is the estimated P/D ratios. For each regression, the table reports OLS estimates and the associated standard deviation in the parenthesis, Newey and West (1987) corrected standard deviation in the brackets, the Hodrick-type standard deviation in the curly brackets. The significance level is based on the Cochrane–Orcutt-type standard deviation marked in $\langle\cdot\rangle$.

$^{*}p < 0.1$.

$^{**}p < 0.05$.

$^{***}p < 0.01$.

short-run predictions, as a robustness check, we also report the t-statistics based on Cochrane–Orcutt-type standard errors. Consistent with the results obtained in Ang and Bekaert (2006), we find no evidence for the predictability of observed dividend yields for future interest rates. However, when we include the implied dividend yields as an additional predictor, we find that those from a rational model have significant positive forecastability on future short interest rates. This result strongly supports the theory that discount rates play a critical role in determining dividend yields.

6. Conclusion

We have proposed a general framework to solve the Euler Equation via a penalized B-splines procedure. When solving Euler equations, compared to existing numerical solution methods that often impose auxiliary assumptions on the state dynamics and functional form of unknown functions for ease of implementation, our method avoids potential misspecification while inheriting a closed-form solution. By transforming the Euler equation into a regression model (with endogeneity), our estimate enjoys computational advantages, as well as the optimal theoretical convergence rate and robust finite sample performance. Our newly designed penalized splines regression also distinguishes itself in the nonparametric literature by weakening the impact of the spline setting and instead letting the penalty play the key role in smoothing. Through empirical analysis on return predictability, we find that higher implied dividend yields predict lower future cash flows, but predict higher future interest rates at short horizons. Moreover, the implied and the observed dividend yields have opposite impacts on cash flow predictions, which may indicate that they represent different sets of information.

Our work opens several avenues for future research. Although our method is designed to solve asset pricing models with time-separable preferences, it may be extended to recursive preference models. With the use of iteration, we may enable estimation of multiple unknown policy functions without solution misspecification and accumulative approximation errors. However, it may exclude the availability of a closed-form solution. Moreover, we point out that our new solution method is built upon a well-posed linear Type-II integral equation framework and it is interesting to consider extending it for solving the general Type I equation.

Acknowledgments

We gratefully thank Yacine Aït-Sahalia (Co-Editor), the Associate Editor and three anonymous referees for their insightful suggestions and comments. We also thank Haiqiang Chen, Tim Christensen, Zhonghao Fu, Burce Hansen, Tao Li, Taehwy Lee, Yue Ma, Whitney Newey, Kristoffer Nimark, Jeffrey Rachine, Junbo Wang, Junhui Wang, Xueping Wu and Aman Ullah for all their useful comments. An early version of this paper was presented at the NY Camp Econometrics X, International Symposium on Recent Developments in Econometric Theory with Applications in Honor of Professor Takeshi Amemiya, China Youth Economists Forum 2015, the State University of New York at Buffalo, and the City University of Hong Kong. We thank Zengguang Zhong and Jiajun Yuan for providing Lab supports. This work is part of Cui's doctoral dissertation. Hong's research is supported by National Science Foundation of China (NSFC) (No. 71988101), which is the Basic Scientific Center Project entitled as Econometric Modelling and Economic Policy Studies. Cui's research is supported by the Research Grants Council of Hong Kong, China (No. 11500119 and 21504818) and NSFC (No. 71803166). Li's research is supported by NSFC (No. 71571154 and No. 71631004).

Appendix A. Proofs of theorems

A.1. Proof of Theorem 3.1

Recall that the unpenalized estimator is expressed as $\hat{f}_a(x) = \phi(x)'\tilde{b}$, where $\tilde{b} = (\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}Y$. Note that

$$\tilde{b} - b = (\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'\varepsilon + (\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'(Y-\varepsilon) - b. \tag{A.1}$$

First consider the variance part, which is associated with the first term in Eq. (A.1). Because of Lemmas B.1 and B.2, we have

$$\begin{aligned} var(\hat{f}_a(x)|\mathcal{X}) &= \frac{1}{M}\phi(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{M})^{-1}\frac{\hat{\Psi}'E(\varepsilon\varepsilon')\hat{\Psi}}{M}(\frac{\hat{\Psi}'\hat{\Psi}}{M})^{-1}\phi(x) \\ &\leq d_{\min}^{-4}(\frac{\hat{\Psi}'\hat{\Psi}}{M})E_{\mathcal{X}}[(\frac{\phi(x)'\hat{\Psi}'\varepsilon}{M})^2] = O_p(\frac{1}{M}), \end{aligned}$$

For the bias term, note that $\hat{\Psi}'\hat{\Psi} = \Psi'\Phi(\Phi'\Phi)^{-1}\Phi'\Psi = \hat{\Psi}'\Psi$. As $\max_x|f(x) - f_a(x)| = O(K^{-\min(1+p,2\nu)})$ and $y_t = O_p(1)$, we have

$$y_t - \varepsilon_t - \psi'(X_t)b = f(X_t) - \phi'(X_t)b + y_{t+1}(f(X_t) - \phi'(X_t)b) = O_p(K^{-\min(1+p,2\nu)}).$$

Since the B-splines basis $\phi(x)$ has no more than $p+1$ nonzero elements, $\phi(x)'\hat{\Psi}'(Y-\varepsilon-\Psi b)/M = O_p(K^{-\min(1+p,2\nu)})$. Therefore,

$$\phi(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{M})^{-1}\frac{\hat{\Psi}'(Y-\varepsilon)}{M} - \phi(x)'b = \phi(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{M})^{-1}\frac{\hat{\Psi}'(Y-\varepsilon-\Psi b)}{M} = O_p(K^{-\min(1+p,2\nu)}).$$

By Assumption 3.5, the square of the bias term is negligible compared with the variance term. Therefore we conclude that

$$E_{\mathcal{X}}[(\hat{f}_a(x)) - f^o(x)]^2 = O_p(K^{-2\min(1+p,2\nu)}) + O_p(\frac{1}{M}) = O_p(\frac{1}{M}) = O_p(\frac{K}{T}). \quad \square$$

A.2. Proof of Theorem 3.2

Note that $var[\hat{f}(x)|\mathcal{X}] = O_p(1/M)$. By Assumption 3.5, $E_{\mathcal{X}}[\hat{f}_a(x) - f^o(x)] = o_p\{\sqrt{var[\hat{f}_a(x)|\mathcal{X}]}\}$. Thus it suffices to show that condition on $\mathcal{X}$, $\frac{\hat{f}_a(x)-E_{\mathcal{X}}[\hat{f}_a(x)]}{\sqrt{var[\hat{f}_a(x)|\mathcal{X}]}} \xrightarrow{d} N(0,1)$. Note that $\hat{f}_a(x) - E_{\mathcal{X}}[\hat{f}_a(x)] = \sum_{i=1}^T a_i\varepsilon_i$, where $a_i = \phi(x)'(\hat{\Psi}'\hat{\Psi})^{-1}\Psi'\Phi(\Phi'\Phi)^{-1}\phi(X_i)$. Following Fan (1992) and Huang (2003), it suffices to verify the Lindeberg condition such that

$$\max_{1\le i\le T} a_i^2/var[\hat{f}_a(x)|\mathcal{X}] = \delta_T(1+o_p(1)),$$

where $\delta_T \to 0$ as $T\to\infty$.

Note that the B-splines basis satisfies $\phi_i(x)\ge 0$ and $\sum_{i=1}^q \phi_i(x) = 1$. Hence $0 < \phi(x)'\phi(x) \le 1$ for any x. Therefore,

$$\begin{aligned} a_i^2M^2 &= \phi(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{M})^{-1}\frac{\Psi'\Phi}{M}(\frac{\Phi'\Phi}{M})^{-1}\phi(X_i)\phi'(X_i)(\frac{\Phi'\Phi}{M})^{-1}\frac{\Phi'\Psi}{M}(\frac{\hat{\Psi}'\hat{\Psi}}{M})^{-1}\phi(x) \\ &\le \phi(x)'\phi(x)\phi'(X_i)\phi(X_i)d_{\min}^{-4}(\hat{\Psi}'\hat{\Psi}/M)d_{\max}^{2}(\hat{\Psi}'\hat{\Psi}/M)d_{\min}^{-2}(\Phi'\Phi/M). \end{aligned}$$

Together with Assumption 3.6 and Lemma B.1, a_i^2 is of order $1/M^2$. Since $var[\hat{f}_a(x)|\mathcal{X}]$ is of order $1/M$, we conclude there exists some constant c_6 such that $\max_{1\le i\le T} a_i^2/var[\hat{f}_a(x)|\mathcal{X}] \le \frac{c_6}{M}(1+o_p(1))$. Therefore, Theorem 3.2 holds. $\square$

A.3. Proof of Theorem 3.3

Recall that the penalized estimator is expressed as $\hat{f}_\lambda(x) = \phi(x)'\hat{b}$, where

$$\hat{b} = (\hat{\Psi}'\hat{\Psi} + \lambda^* D_\nu' D_\nu)^{-1}\hat{\Psi}Y = (\frac{\hat{\Psi}'\hat{\Psi}}{M} + \lambda D_\nu' D_\nu)^{-1}\frac{\hat{\Psi}'Y}{M}.$$

$$\hat{b} - b = (\hat{\Psi}'\hat{\Psi} + \lambda^* D_\nu' D_\nu)^{-1}\hat{\Psi}'\varepsilon + (\hat{\Psi}'\hat{\Psi} + \lambda^* D_\nu' D_\nu)^{-1}\hat{\Psi}'(Y-\varepsilon) - b.$$

First consider the variance part. Note that

$$var(\hat{f}_\lambda(x)|\mathcal{X}) = \phi(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{M} + \lambda D_\nu' D_\nu)^{-1}\frac{\hat{\Psi}'E(\varepsilon\varepsilon')\hat{\Psi}}{M^2}(\frac{\hat{\Psi}'\hat{\Psi}}{M} + \lambda D_\nu' D_\nu)^{-1}\phi(x).$$

As in Lemma B.2, $d_{\max}^2\{\frac{\hat{\Psi}'E(\varepsilon\varepsilon')\hat{\Psi}}{M^2}\} = O_p(1/M)$. Since $T = MK$, it suffices to show that

$$\phi(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{M} + \lambda D_\nu' D_\nu)^{-1}(\frac{\hat{\Psi}'\hat{\Psi}}{M} + \lambda D_\nu' D_\nu)^{-1}\phi(x) = O_p(\frac{1}{Kh}). \tag{A.2}$$

Denote $V(c) = \phi(x)'(cI_q + \lambda D_\nu' D_\nu)^{-1}(cI_q + \lambda D_\nu' D_\nu)^{-1}\phi(x)$, and let a_{ij} be the (i,j)th element of $(I_q + \lambda D_\nu' D_\nu)^{-1}$. When $\lambda \sim (Kh)^{2\nu}$ and i satisfies $i/K \to x$, Li and Ruppert (2008) and Xiao (2019) show that $a_{ij}, j = 1,\ldots,q$ are asymptotically equivalent to K Nadaraya–Watson kernel weights yielded by the 2νth order kernel function $H_\nu(z)$[11] using the equivalent bandwidth h. Hence for any constant $0 < c < \infty$, $V(c) = O(\sum_{i=1}^q a_{ij}^2) = O(\frac{1}{Kh})$.

By Lemma B.1, we have

$$V(\bar{c}_2)(1+o_p(1)) \le \phi(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{M} + \lambda D_\nu' D_\nu)^{-1}(\frac{\hat{\Psi}'\hat{\Psi}}{M} + \lambda D_\nu' D_\nu)^{-1}\phi(x) \le V(c_2)(1+o_p(1)). \tag{A.3}$$

Therefore, Eq. (A.2) holds and hence we prove $var(\hat{f}_\lambda(x)|\mathcal{X}) = O_p(\frac{1}{Kh})O_p(\frac{1}{M}) = O_p(\frac{1}{Th})$.

[11] The kernel function is defined as $H_\nu(z) = \frac{1}{2\nu}\sum_{l=1}^{\nu}\exp(-\zeta_l|z|)$, and $\zeta_1,\ldots,\zeta_\nu$ are the ν complex roots of $z^{2\nu} + (-1)^\nu = 0$ such that all ζ_l, $1\le l\le \nu$ have positive real parts.

For the bias term, using the fact that $\hat{\Psi}'\hat{\Psi} = \hat{\Psi}'\Psi$, we have

$$\begin{aligned} &\phi(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{M} + \lambda D_\nu' D_\nu)^{-1}\frac{\hat{\Psi}'(Y-\varepsilon)}{M} - \phi(x)'b \\ &= \phi(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{M} + \lambda D_\nu' D_\nu)^{-1}\frac{\hat{\Psi}'(Y-\varepsilon-\Psi b)}{M} - \phi(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{M} + \lambda D_\nu' D_\nu)^{-1}\lambda D_\nu' D_\nu b. \end{aligned} \tag{A.4}$$

Similar as in the proof of Theorem 3.1, we have $\phi(x)'\hat{\Psi}'(Y-\varepsilon-\Psi b)/M = O_p(K^{-\min(1+p,2\nu)})$. Note that $d^2_{\max}(\frac{\hat{\Psi}'\hat{\Psi}}{M} + \lambda D_\nu' D_\nu)^{-1} \le d^2_{\min}(\frac{\hat{\Psi}'\hat{\Psi}}{M})^{-1} \le c_2^{-1}(1+o_p(1))$. Hence the first term in Eq. (A.4) is $O_p(K^{-\min(1+p,2\nu)})$. Note that $\phi(x)'D_\nu' D_\nu b = O(K^{-2\nu})$. Together with Assumption 3.7,

$$\phi(x)'(\frac{\hat{\Psi}'\hat{\Psi}}{M} + \lambda D_\nu' D_\nu)^{-1}\lambda D_\nu' D_\nu b \le \lambda d_{\min}(\frac{\hat{\Psi}'\hat{\Psi}}{M})^{-1}\phi(x)'D_\nu' D_\nu b = (Kh)^{2\nu}O_p(K^{-2\nu}) = O_p(h^{2\nu}).$$

Since Assumption 3.7 also implies that $h^{2\nu}$ dominates the spline approximation bias $K^{-\min(1+p,2\nu)}$, we conclude that

$$E_{\mathcal{X}}[(\hat{f}_\lambda(x)) - f^o(x)]^2 = O_p[(\frac{1}{Th} + h^{4\nu})].$$

Correspondingly, the optimal convergence rate will be achieved when the rate of the square of the bias, $h^{4\nu}$, grows at the same rate as the variance, $\frac{1}{Th}$. Suppose $h_{opt} \sim cT^{-1/(4\nu+1)}$ for some positive constant c. Then the optimal rate of $\lambda = (Kh_{opt})^{2\nu} \sim c^{2\nu}K^{2\nu}T^{-2\nu/(4\nu+1)}$. □

A.4. Proof of Theorem 3.4

First, we show that Lemma B.1 still holds if Assumption 3.1 is replaced by Assumption 3.8.

According to de Boor (1978), we have

$$c_{0U}(1+o(1))(\sum_{i=1}^q a_i^2/K) \le \int s^2(x)\pi_U(x)dx \le \bar{c}_{0U}(1+o(1))(\sum_{i=1}^q a_i^2/K), \tag{A.5}$$

$$c_{0L}(1+o(1))(\sum_{i=1}^q a_i^2/K) \le \int s^2(x)\pi_L(x)dx \le \bar{c}_{0L}(1+o(1))(\sum_{i=1}^q a_i^2/K). \tag{A.6}$$

By Assumption 3.8, we have

$$c_{0L}(1+o_p(1))(\sum_{i=1}^q a_i^2/K) \le \frac{1}{T}\sum_{i=1}^T s^2(X_i) \le \bar{c}_{0U}(1+o_p(1))(\sum_{i=1}^q a_i^2/K). \tag{A.7}$$

Note that the relationship $\frac{1}{T}\sum_{i=1}^T s^2(X_i) = a'\Phi'\Phi a/M$ holds for any $a = (a_1, \dots, a_q)$ with norm 1, if we define $s(x)$ such that its ith coefficient associated with $\phi_i(x)$ is $\sqrt{K}a_i$.

Hence Eq. (B.1) still holds. Similarly, we could also conclude that Eq. (B.2) holds.

Then using the same arguments as in Theorems 3.1 and 3.3, we could conclude that Theorem 3.4 holds. □

A.5. Proof of Theorem 3.5

Recall that $\hat{\Psi} = \Phi(\Phi'\Phi)^{-1}\Phi'\Psi$. Hence $\hat{\Psi}'\hat{\Psi} = \hat{\Psi}'\Psi = \Psi'\hat{\Psi} = \Sigma$. Let I_q be the $q \times q$ identity matrix with $q = K + p$. Let $U\Gamma U'$ be the eigendecomposition of $\Sigma^{-1/2}D_\nu' D_\nu \Sigma^{-1/2}$. Note that r_i is the ith diagonal element of Γ. Define $H = \Psi(\hat{\Psi}^T\hat{\Psi} + \lambda^* D_\nu' D_\nu)^{-1}\hat{\Psi}'$. Then

$$\begin{aligned} trace(H) &= trace[\hat{\Psi}'\Psi(\hat{\Psi}^T\hat{\Psi} + \lambda^* D_\nu' D_\nu)^{-1}] \\ &= trace[(I + \lambda^*\Sigma^{-1/2}D_\nu' D_\nu\Sigma^{-1/2})^{-1}] \\ &= \sum_{j=1}^{K+p}\frac{1}{1+\lambda^* r_j}, \end{aligned}$$

and it follows immediately that

$$\begin{aligned} (\hat{\Psi}'\hat{\Psi} + \lambda^* D_\nu' D_\nu)^{-1} &= \Sigma^{-1/2}(I_q + \lambda^*\Sigma^{-1/2}D_\nu' D_\nu\Sigma^{-1/2})^{-1}\Sigma^{-1/2} \\ &= \Sigma^{-1/2}U(I_q + \lambda^*\Gamma)^{-1}U'\Sigma^{-1/2}. \end{aligned}$$

Define $Z_1 = U'\Sigma^{-1/2}\Psi'Y$ and $Z_2 = U'\Sigma^{-1/2}\hat{\Psi}'Y$. Then

$$\begin{aligned} Y'\hat{Y} &= Y'HY \\ &= Y'\Psi\Sigma^{-1/2}(I_{K+p} + \lambda^*\Sigma^{-1/2}D_\nu'D_\nu\Sigma^{-1/2})^{-1}\Sigma^{-1/2}\hat{\Psi}'Y \\ &= \sum_{i=1}^{K+p}\frac{1}{1+\lambda^* r_i}z_{1,i}z_{2,i}. \end{aligned} \tag{A.8}$$

where $z_{1,i}$ and $z_{2,i}$ are the ith element of Z_1 and Z_2 respectively. Moreover,

$$\begin{aligned} \|\hat{Y}\|^2 &= \hat{Y}'\hat{Y} \\ &= Y'\hat{\Psi}(\hat{\Psi}'\hat{\Psi} + \lambda^* D_\nu' D_\nu)^{-1}\Psi'\Psi(\hat{\Psi}'\hat{\Psi} + \lambda^* D_\nu' D_\nu)^{-1}\hat{\Psi}'Y \\ &= Y'\hat{\Psi}\Sigma^{-1/2}U(I_q + \lambda^*\Gamma)^{-1}U'\Sigma^{-1/2}\Psi'\Psi\Sigma^{-1/2}U(I_q + \lambda^*\Gamma)^{-1}U'\Sigma^{-1/2}\hat{\Psi}'Y \\ &= Z_2'(I_q + \lambda^*\Gamma)^{-1}\tilde{U}\tilde{\Gamma}\tilde{U}'(I_q + \lambda^*\Gamma)^{-1}Z_2 \\ &= \|\sqrt{\tilde{\Gamma}}\tilde{U}'(I_q + \lambda^*\Gamma)^{-1}Z_2\|^2 \end{aligned}$$

where $\tilde{U}\tilde{\Gamma}\tilde{U}'$ is the eigendecomposition of the matrix $U'\Sigma^{-1/2}\Psi'\Psi\Sigma^{-1/2}U$.

Note that the (i,j)th element of $\sqrt{\tilde{\Gamma}}\tilde{U}'(I_q + \lambda^*\Gamma)^{-1}$ is $w_{i,j}(\lambda^*)\tilde{U}_{j,i}$, where $w_{i,j}(\lambda^*) = \frac{\sqrt{\tilde{\Gamma}_{i,i}}}{1+\lambda^*\Gamma_{j,j}}$. Define $W(\lambda^*)$ be the matrix whose (i,j)th element is $w_{i,j}(\lambda^*)$. Then $\sqrt{\tilde{\Gamma}}\tilde{U}'(I_q+\lambda^*\Gamma)^{-1} = W(\lambda^*)\odot\tilde{U}'$, where $\odot$ is the Hadamard product which calculate elementwise product. Define $Z_3(\lambda)$ such that

$$Z_3(\lambda^*) = (W(\lambda^*)\odot\tilde{U}')Z_2. \tag{A.9}$$

Let $z_{3,i}(\lambda^*)$ be the ith element of $Z_3(\lambda^*)$. Then we have $\|\hat{Y}\|^2 = \|Z_3(\lambda^*)\|^2 = \sum_{i=1}^{K+p} z_{3,i}^2(\lambda^*)$.

Together with Eq. (A.8), we prove that Theorem 3.5 holds. □

Appendix B. Lemmas with proofs

We state some technical lemmas. The first lemma is about the design matrix $\Phi'\Phi$, $\Psi'\Psi$, and $\hat{\Psi}'\hat{\Psi}$.

Lemma B.1. *Let $M = T/K$. Suppose Assumptions 2.1–3.3 hold. Then there exist positive constants c_1, $\bar{c}_1$, c_2, $\bar{c}_2$ such that*

(a)

$$c_1(1+o_p(1)) \le d_{\min}^2[\frac{\Phi'\Phi}{M}] \le d_{\max}^2[\frac{\Phi'\Phi}{M}] \le \bar{c}_1(1+o_p(1)) \tag{B.1}$$

(b)

$$c_2(1+o_p(1)) \le d_{\min}^2[\frac{\hat{\Psi}'\hat{\Psi}}{M}] \le d_{\max}^2[\frac{\hat{\Psi}'\hat{\Psi}}{M}] \le \bar{c}_2(1+o_p(1)), \tag{B.2}$$

where $d_{\min}^2(A)$ and $d_{\max}^2(A)$ denote the minimum and maximum singular values of A respectively.

Proof of Lemma B.1. Follow the proof of Zhou et al. (1998), we first establish the following results: There exist constants $0 < c_0 < \bar{c}_0 < \infty$ such that, for any $s(x) = \sum_{i=1}^q a_i\phi_i(x)$,

$$c_0(1+o_p(1)) \le \frac{1}{T}\sum_{i=1}^T s^2(X_i) \le \bar{c}_0(1+o_p(1)), \tag{B.3}$$

According to de Boor (1978), there exist constants $\tilde{c}_L$ and $\tilde{c}_U$ such that

$$\tilde{c}_L\sum_{i=1}^q a_i^2/K \le \int s^2(x)dx \le \tilde{c}_U\sum_{i=1}^q a_i^2/K. \tag{B.4}$$

Denote $Q_T(x)$ as the empirical density of X. According to Glivenko–Cantelli theorem, $\max_x|Q_T(x) - Q(x)| = O_p(T^{-1/2})$ and hence

$$\frac{1}{T}\sum_{i=1}^T s^2(X_i) = \int s^2(x)dQ_T(x) = \int s^2(x)dQ(x) + o_p(1)(\sum_{i=1}^q a_i^2/K).$$

Together with Eq. (B.4), we conclude Eq. (B.3) holds because

$$(\min_x \pi(x)\tilde{c}_L + o_p(1))(\sum_{i=1}^q a_i^2/K) \le \frac{1}{T}\sum_{i=1}^T s^2(X_i) \le (\max_x \pi(x)\tilde{c}_U + o_p(1))(\sum_{i=1}^q a_i^2/K).$$

Note that $d^2_{\min}\frac{\Phi'\Phi}{M} = \min_{\|a\|=1} a'\Phi'\Phi a/M$, $d^2_{\max}\frac{\Phi'\Phi}{M} = \max_{\|a\|=1} a'\Phi'\Phi a/M$. For any $a = (a_1, \ldots, a_q)$ with norm 1, we could define a specific function $s(x)$ such that its ith coefficient associated with $\phi_i(x)$ is $\sqrt{K}a_i$. Then $a'\Phi'\Phi a/M = \frac{1}{T}\sum_{i=1}^T s^2(X_i)$. Hence Eq. (B.1) holds.

Recall that $\hat{\Psi}'\hat{\Psi} = \Psi'\Phi(\Phi'\Phi)^{-1}\Phi'\Psi$. Hence $d^2_{\min}\frac{\hat{\Psi}'\hat{\Psi}}{M} \ge d^2_{\min}\frac{\Psi'\Phi}{M}(d^2_{\max}\frac{\Phi'\Phi}{M})^{-1}$. Together with Assumption 3.6 and Eq. (B.1), the left hand side of Eq. (B.2) holds. To prove the right hand side of Eq. (B.2), note that the maximum eigenvalue of $\Phi(\Phi'\Phi)^{-1}\Phi'$ is 1. It suffices to show that $d^2_{\max}\frac{\Psi'\Psi}{M} \le c_2(1 + o_p(1))$ for some positive constant c_2.

For any $s(x) = \sum_{i=1}^q a_i\phi_i(x)$, we have $(s(X_i) - y_{i+1}s(X_{i+1}))^2 \le 2s^2(X_i) + y^4_{i+1} + s^4(X_i)$. Similar as Eq. (B.3), there exists some constant c_3 such that $\frac{1}{T}\sum_{i=1}^T s^4(X_i) \le c_3(1 + o_p(1))$. By Law of large numbers, $\frac{1}{T}\sum_{i=1}^T y^4_{i+1} \le \bar{c}_3(1 + o_p(1))$ for some constant $\bar{c}_3$. Hence

$$\frac{1}{T}\sum_{i=1}^T (s(X_i) - y_{i+1}s(X_{i+1}))^2 \le (2\bar{c}_1 + c_3 + \bar{c}_3)(1 + o_p(1)), \tag{B.5}$$

Therefore, $d^2_{\max}\frac{\Psi'\Psi}{M} \le (2\bar{c}_1 + c_3 + c_4)(1 + o_p(1))$, and the right hand side of Eq. (B.2) also holds. Therefore, Lemma B.1 is proved. □

Our second lemma facilitates us to control the bound for the variance similarly as the dependence among the errors is not very strong.

Lemma B.2. *Assume Assumptions 3.2 and 3.4. Then*

(a) There exists a finite positive constant c_4 such that $E(\varepsilon\varepsilon') \le c_4 I_T$;
(b) $E_{\mathcal{X}}[(\phi(x)'\hat{\Psi}'\varepsilon/M)^2] = O_p(\frac{1}{M})$.

Proof of Lemma B.2. First, we prove part (a). Suppose an arbitrary vector $\bar{b} = (\bar{b}_1, \ldots, \bar{b}_T)$ and a finite number $c_4 > 0$ so that $c_4 \ge c_5$, where $c_5 = \tilde{\Delta}^{\frac{2}{4+\delta}}\sum_{\tau=0}^\infty \frac{2^{2-2/(4+\delta)}(4+\delta)}{2+\delta}\alpha(\tau)^{1-\frac{2}{4+\delta}}$ for some $\delta > 0$, and $\alpha(\tau)$ is the mixing coefficients. Then we have

$$\begin{aligned}
\bar{b}'(c_4 I_T - E\varepsilon\varepsilon')\bar{b} &= c_4\sum_{t=2}^T \bar{b}_t^2 - \sum_{t=2}^T\sum_{s=2}^T \bar{b}_t\bar{b}_s E(\varepsilon_t\varepsilon_s) \\
&\ge c_4\sum_{t=2}^T \bar{b}_t^2 - \frac{1}{2}\sum_{t=2}^T\sum_{s=2}^T(\bar{b}_t^2 + \bar{b}_s^2)E|\varepsilon_t\varepsilon_s| \\
&= c_4\sum_{t=2}^T \bar{b}_t^2 - \sum_{t=2}^T \bar{b}_t^2\sum_{s=2}^T E|\varepsilon_t\varepsilon_s| \\
&\ge c_4\sum_{t=2}^T \bar{b}_t^2 - \sum_{\tau=0}^\infty \frac{2^{2-2/(4+\delta)}(4+\delta)}{2+\delta}\alpha(\tau)^{1-\frac{2}{4+\delta}}(E|\varepsilon_t|^{4+\delta})^{\frac{2}{4+\delta}}\sum_{t=2}^T \bar{b}_t^2 \\
&\ge (c_4 - c_5)\sum_{t=2}^T \bar{b}_t^2.
\end{aligned}$$

Hence, $c_4 I_T - E(\varepsilon\varepsilon')$ is positive semi-definite. Recall that the B-splines basis $\phi(x)$ has no more than $p+1$ nonzero elements and hence $\phi(x)'\phi(x) = O(1)$. Together with Lemma B.1, we have

$$E_{\mathcal{X}}\{\left(\frac{\phi(x)'\hat{\Psi}'\varepsilon}{M}\right)^2\} = \frac{1}{M}[\phi(x)'(\frac{\hat{\Psi}'E(\varepsilon\varepsilon')\hat{\Psi}}{M})\phi(x)] \le \frac{\phi(x)'\phi(x)}{M}d^2_{\max}(\frac{\hat{\Psi}'\hat{\Psi}}{M})d^2_{\max}(E\varepsilon\varepsilon') = O_p(\frac{1}{M}). \quad \square$$

References

Ai, C., Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica 71 (6), 1795–1843.
Aït-Sahalia, Y., 2002. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Econometrica 70 (1), 223–262.
Andrei, D., Hasler, M., 2014. Investor attention and stock market volatility. Rev. Financ. Stud. 28 (1), 33–72.

Ang, A., Bekaert, G., 2006. Stock return predictability: Is it there? Rev. Financ. Stud. 20 (3), 651–707.
Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. J. Finance 59 (4), 1481–1509.
Borovička, J., Hansen, L.P., Scheinkman, J.A., 2016. Misspecified recovery. J. Finance 71 (6), 2493–2544.
Burnside, C., 1998. Solving asset pricing models with Gaussian shocks. J. Econom. Dynam. Control 22 (3), 329–340.
Calin, O.L., Chen, Y., Cosimano, T.F., Himonas, A.A., 2005. Solving asset pricing models when the price–dividend function is analytic. Econometrica 73 (3), 961–982.
Campbell, J.Y., Cochrane, J.H., 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. J. Political Econ. 107 (2), 205–251.
Campbell, J.Y., Cochrane, J.H., 2000. Explaining the poor performance of consumption-based asset pricing models. J. Finance 55 (6), 2863–2878.
Campbell, J.Y., Shiller, R.J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. Rev. Financ. Stud. 1 (3), 195–228.
Campbell, J.Y., Yogo, M., 2006. Efficient tests of stock return predictability. J. Financ. Econ. 81 (1), 27–60.
Canay, I.A., Santos, A., Shaikh, A.M., 2013. On the testability of identification in some nonparametric models with endogeneity. Econometrica 81 (6), 2535–2559.
Carrasco, M., Florens, J.-P., Renault, E., 2007. Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. In: Handbook of Econometrics, Vol. 6. Elsevier, pp. 5633–5751.
Cecchetti, S.G., Lam, P.-s., Mark, N.C., 2000. Asset pricing with distorted beliefs: Are equity returns too good to be true? Amer. Econ. Rev. 90 (4), 787–805.
Chen, H., Fang, Y., Li, Y., 2015. Estimation and inference for varying-coefficient models with nonstationary regressors using penalized splines. Econometric Theory 31 (4), 753–777.
Chen, X., Pouzo, D., 2015. Sieve wald and QLR inferences on semi/nonparametric conditional moment models. Econometrica 83 (3), 1013–1079.
Christensen, T.M., 2017. Nonparametric stochastic discount factor decomposition. Econometrica 85 (5), 1501–1536.
Claeskens, G., Krivobokova, T., Opsomer, J.D., 2009. Asymptotic properties of penalized spline estimators. Biometrika 96 (3), 529–544.
Cochrane, J.H., 2009. Asset Pricing:(Revised Edition). Princeton university press.
Cujean, J., Hasler, M., 2017. Why does return predictability concentrate in bad times? J. Finance 72 (6), 2717–2758.
Darolles, S., Fan, Y., Florens, J.-P., Renault, E., 2011. Nonparametric instrumental regression. Econometrica 79 (5), 1541–1565.
de Boor, C., 1978. A Practical Guide to Splines, Vol. 27. Springer-Verlag New York.
Den Haan, W.J., Marcet, A., 1990. Solving the stochastic growth model by parameterizing expectations. J. Bus. Econom. Statist. 8 (1), 31–34.
Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with *b*-splines and penalties. Statist. Sci. 11 (2), 89–102.
Elliott, G., Müller, U.K., Watson, M.W., 2015. Nearly optimal tests when a nuisance parameter is present under the null hypothesis. Econometrica 83 (2), 771–811.
Epstein, L.G., Zin, S.E., 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. Econometrica 57 (4), 937–969.
Escanciano, J.C., Hoderlein, S., Lewbel, A., Linton, O., Srisuma, S., 2020. Nonparametric Euler equation identification and estimation. Econometric Theory, forthcoming.
Fan, J., 1992. Design-adaptive nonparametric regression. J. Amer. Statist. Assoc. 87 (420), 998–1004.
Gagliardini, P., Gourieroux, C., Renault, E., 2011. Efficient derivative pricing by the extended method of moments. Econometrica 79 (4), 1181–1232.
Grossman, S.J., Shiller, R.J., 1981. The determinants of the variability of stock market prices. Amer. Econ. Rev. 71 (2), 222–227.
Hansen, B.E., 2001. The new econometrics of structural change: Dating breaks in U.S. labour productivity. J. Econ. Perspect. 15 (4), 117–128.
Hansen, B.E., 2015. The integrated mean squared error of series regression and a rosenthal Hilbert-space inequality. Econometric Theory 31 (2), 337–361.
Hansen, L.P., Singleton, K.J., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica 50 (5), 1269–1286.
Hong, Y., Li, H., 2005. Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. Rev. Financ. Stud. 18 (1), 37–84.
Hong, Y., White, H., 2005. Asymptotic distribution theory for nonparametric entropy measures of serial dependence. Econometrica 73 (3), 837–901.
Horowitz, J.L., 2011. Applied nonparametric instrumental variables estimation. Econometrica 79 (2), 347–394.
Huang, J.Z., 2003. Local asymptotics for polynomial spline regression. Ann. Statist. 31 (5), 1600–1635.
Juillard, M., 2011. Local approximation of DSGE models around the risky steady state. Wp. Comunite 87, 21–42.
Kress, R., 1989. Linear Integral Equations, Vol. 82. Springer.
Lettau, M., Ludvigson, S.C., 2005. Expected returns and expected dividend growth. J. Financ. Econ. 76 (3), 583–626.
Li, Y., Ruppert, D., 2008. On the asymptotics of penalized splines. Biometrika 95 (2), 415–436.
Lucas, R.E., 1978. Asset prices in an exchange economy. Econometrica 46 (6), 1429–1445.
Mehra, R., Prescott, E.C., 1985. The equity premium: A puzzle. J. Monetary Econ. 15 (2), 145–161.
Newey, W.K., Powell, J.L., 2003. Instrumental variable estimation of nonparametric models. Econometrica 71 (5), 1565–1578.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.
Phillips, P.C., Lee, J.H., 2013. Predictive regression under various degrees of persistence and robust long-horizon regression. J. Econometrics 177 (2), 250–264.
Pohl, W., Schmedders, K., Wilms, O., 2018. Higher order effects in asset pricing models with long-run risks. J. Finance 73 (3), 1061–1111.
Xiao, L., 2019. Asymptotic theory of penalized splines. Electron. J. Stat. 13 (1), 747–794.
Xiu, D., 2014. Hermite polynomial based expansion of european option prices. J. Econometrics 179 (2), 158–177.
Xu, K., 2018. Robust Inference of Long-Horizon Predictability. Tech. rep., Indianna University.
Zhou, S., Shen, X., Wolfe, D.A., 1998. Local asymptotics for regression splines and confidence regions. Ann. Statist. 26 (5), 1760–1782.