Solving Asset Pricing Models Via Nonparametric Two-Stage Penalized B-spline Regression*

Liyuan Cui[1] Yongmiao Hong[2] Yingxing Li[3]

Abstract

We present a nonparametric two stage Penalized B-spline regression method in solving consumption-based asset pricing models, which allows the true dynamics of state variables to determine asset prices. Unlike current numerical solution methods, this new method is a one-step procedure with a closed-form solution. It does not require imposing auxiliary assumptions on the conditional distributions of state variables. We establish the asymptotics of the estimation for a broad class of stationary Markov state Variables. Our estimator overcomes the ill-posed inverse problem which usually might be caused in nonparametric instrumental variable regression and achieves the optimal convergence rate. The approach is robust to the choice of the spline basis. We also design a fast generalized cross-validation procedure to well tune the penalty parameter for practical use. As an application, we propose a nonparametric decomposition of observed dividend yields and investigate its predictability ability on excess future returns.

Keywords: Nonparametri estimation, series estimation, integral equation of the second kind, penalized B-spline estimation, endogeneity, instrumental variables, rational price.
JEL classification: C1, C3, C4, C5, E1, G12.

---

*We thank Haiqiang Chen, Tim Christensen, Jiti Gao, Burce Hansen, Tao Li, Taehwy Lee, Yue Ma, Whitney Newey, Kristoffer Nimark, Jeffrey Rachine, Junbo 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 and the City University of Hong Kong. Any errors are solely ours.

[1]City University of Hong Kong, Department of Economics and Finance, Tat Chee Avenue, Kowloon Tong, Kowloon, Hong Kong, phone:+852-3442-8450, fax: +852-3442-0284,e-mail: liyuan.cui@cityu.edu.hk

[2]Cornell University, Department of Economics and Department of Statistical Sciences, 404 Uris Hall Cornell University, Ithaca, N.Y. 14853, phone: 607-255-5130, fax:607-255-2818,e-mail: yh20@cornell.edu

[3]Xiamen University, The Wang Yanan Institute of Studies in Economics, e-mail: amyli999@hotmail.com

Corresponding author. Yongmiao Hong, Department of Economics and Department of Statistical Sciences, 424 Uris Hall, Cornell University, Ithaca, NY 14850. e-mail: yh20@cornell.edu.

1. Introduction

Considerable attempts to enrich the explanatory powers of economic models have been witnessed in recent years. Economists load additional factors into canonical models to enhance the understanding of well-documented economic anomalies, thereby increasing model complexity. In doing so, analytic or closed-form solutions usually become extremely difficult, if not impossible, as is the case for dynamic stochastic general equilibrium (DSGE) models (Fernández-Villaverde, Ramírez, and Schorfheide, 2016). In macroeconomics and finance, Euler equations are often employed as a pivotal tool to understand the connection among agent preferences, asset prices and economic fundamentals (Campbell, 2003; Christensen, 2017). In modeling dynamic asset prices, the price-dividend (P/D) ratio function, which is recursively specified in Euler equations as a function of state variables, is one of the central quantities that must be solved for. Mehra and Prescott (1985) show that equilibrium asset returns can be formulated by P/D ratios. An accurate solution for the P/D ratio function helps understand market volatility puzzle and provides us a reliable channel to test the existence of bubbles or long-run risks (Campbell and Cochrane, 1999; Cochrane, 1992; Jagannathan and Marakani, 2015; Bansal and Yaron, 2004).

Numerical solution methods have been widely used in solving the P/D ratio function, as most most dynamic general equilibrium models do not offer analytical solutions (Pohl, Schmedders, and Wilms, 2014; Fernández-Villaverde and Rubio-Ramírez, 2007). Judd, Maliar, and Maliar (2011) divide most numerical methods into three broad classes: projection methods[1], perturbation methods[2], and regression-based methods[3]. Unfortunately, all existing numerical approaches that are popularly employed all suffer from approximation errors, which do not disappear even when the sample size goes to infinity. These proce-

[1]projection methods approximate the unknown functions through series expansions with a pre-specified order for the expansion on a pre-specified domain using deterministic integration.

[2]Perturbation methods seek local approximations via Taylor expansions around some pre-determined steady states.

[3]regression-based methods obtain approximations on a set of simulated points using Monte Carlo integration via a large number of iterations.

dures must pre-assume auxiliary fully specified data generating process for state variables or parametric functional form for unknown functions via iterative operations, which may lead to misspecification errors compared to their true underlying dynamics. Blanchard (2016) documents that misspecification will severely affect the estimation and model conclusions. Therefore, cautions must be practised when interpreting conclusions from dynamic general equilibrium models that are built on poorly approximated functions or spurious DGP of state variables.

Despite its important roles, there are not well-developed estimations and inferences of P/D ratios. To push beyond current methodological constraints, our paper will provide a novel nonparametric estimation for the P/D ratio function in Euler equations. Our approach does not require modelling the conditional distributions of state variables over time. Additionally, in a sharp contrast with current regression-based algorithms, our method does not assume parametric assumptions on unknown functions or conditional expectations and will obtain a closed-form estimation through a one-step procedure without iterations. We establish the desired asymptotic properties of the proposed method and examine its finite sample performance in comparison with popular numerical approximation methods in the literature.

We represent the Euler equation equivalently as a nonlinear time series regression model, where the regression function contains the unknown price-dividend ratio function over two time periods. To overcome the endogeneity, we adopt a two-stage nonparametric instrumental variable procedure. Nonparametric estimations and identifications with endogeneity have also been carefully studied in Newey (1997), Newey and Powell (2003), and Andrews (1991), Chen, Chernozhukov, Lee, and Newey (2014), where most of these models are structured as Fredholm equations of the first kind (also called the Type I equations). In the context of our work, we propose to transform the Type I equation into a Type II. Through this, we avoid the ill-posed inverse problem and potentially unstable inferences.

Our paper further complements the fundamental nonparametric 2SLS (NP2SLS) work of Newey (1997) and Newey and Powell (2003) by showing how to optimize the order of series estimations not only in the second stage but also in the first. We propose to estimate the unknown P/D ratio function via regression splines and penalized splines respectively. When regression splines are used, the procedure is similar to other series approximation procedure, except that we employ B-splines that are more numerical stable compared to global approximation series. Similar to all series approximation methods, the choice of the series order, i.e. the number of spline basis K, is extremely crucial. How to choose a proper order in each stage remains an open question (Newey and Powell, 2003). To reduce computation burden, we recommend using the same K for each stage. In order to adopt different smoothing amounts for each stage, we consider using penalized splines, which has an additional smoothing parameter λ. We let K play the key role of smoothing in the first stage, while λ play the key role of smoothing in the second stage. Theoretical justifications show that when K exceeds some minimum bound, a well tuned λ will yield an estimator of the optimal nonparametric convergence rate.

The theoretical analysis in our paper also sheds lights on practical implementation. In our two-stage procedure, the minimization criterion is a quadratic function and the solution enjoys a closed form. As K is not crucial, we could choose a relatively large K and then use data driven method to select the appropriate λ. We further design a fast generalized cross-validation (GCV) algorithm, which will further reduce the computation from $O(K^3)$ to $O(K)$ in each iteration when searching the optimal λ. Our simulation studies demonstrate excellent performance of our estimate even when the sample size is relatively small.

In the empirical analysis, we make use of empirical observations of consumption dynamics to estimate P/D ratios. We interpret our estimate as the implied P/D ratio in the spirit of Shiller (2014). We construct a new measure of the present value of future dividends, which exhibits superior ability to match historical price movements from 1947 to 1970. Its inability to produce similar magnitudes and fluctuations of price movements in recent years provide additional evidence that efficiency theory is lacking.

Based on the estimated implied P/D ratios, we further propose a nonparametric decomposition of observed dividend yields into rational and beyond-rational components. We find significant time-varying and heterogeneous predictability ability that rational and beyond-rational dividend yields exhibit on future excess returns with horizon from 1 quarter to 10 years.

The paper is organized as below. Section 2 is about the methodology, where we establish the model and estimation method, and we provide the asymptotic results as well as data-driven implementation procedure. In Section 3 reports simulation results and Section 4 is the empirical study. Section 5 concludes , while all mathematical proofs and technical details are collected in Appendix.

2. Literature Review

In the asset pricing literature, since Hansen and Singleton (1982), Gallant and Tauchen (1989), Hansen and Scheinkman (2012) and Christensen (2017), it has been conventional to work with stationary Markov state variables. We shall allow serial dependence under Markov processes. In a general consumption based asset pricing model populated by a representative agent in such an environment, a recent flourishing of work mainly focuses on identifications and nonparametric estimations of SDF (Chen et al., 2014; Escanciano, Hoderlein, Lewbel, Linton, and Srisuma, 2015; Christensen, 2017). However, there are not well-developed estimations and inferences of P/D ratios given a known or estimated SDF. Our paper aims to fill this gap by proposing a two-stage penalized B-spline regression method.

Chen, Favilukis, and Ludvigson (2013), Escanciano et al. (2015) and Christensen (2017) also propose to incorporate empirical data into Euler equations when estimating SDF. However, their methods do not provide a direct link to either model implied stock returns or P/D ratios. Most importantly, they do not answer the question of how to choose a proper order in the first stage to balance the trade-off between unbiasedness and efficiency. Our paper provides an alternative solution method that addresses these practical requirements and theoretical constraints. In addition, we propose a different identification strategy, under which we achieve the optimal rate of convergence with a data-driven GCV method and establish its asymptotic normality.

In the literature, enormous efforts have been devoted to approximating nonlinear functions in Euler equations (Judd, 1992; Judd, 1998; Fernández-Villaverde and Rubio-Ramírez, 2006 and Pohl et al., 2014). Aruoba, Fernández-Villaverde, and Rubio-Ramirez (2006) and Fernández-Villaverde et al. (2016) provide a comprehensive survey of these widely used numerical solution methods. Our paper finds that there is a trade-off between solution accuracy and robustness- the more accurate methods are less robust to DGP misspecification. Due to computational convenience, including discretization, projection, perturbation and regression-based methods, most current numerical solution methods for the P/D ratio function described in Euler equations require auxiliary assumptions on the conditional distribution of state dynamics. Despite the substantial progress that has been made in the development of more realistic and reasonable DGPs, there is no assurance that those pre-specified distributional assumptions made on state variables can capture their true underlying dynamics. Meghir and Pistaferri (2004) show that model misspecification on the stochastic process of the state variable, income innovations, can lead to incorrect conclusions about the effect of individual behavior on consumption decisions. In this paper, we show that model implied P/D ratios are sensitive to distributional assumptions of state variables. In our simulation study, we find that the skewness and nonlinearity properties of state variables not only affect the solution accuracy of policy functions but also determine the degree to which the equity premium puzzle can be explained.

The discretization method is accomplished by exactly solving a finite number of points within a support and interpolating the areas between grid points. Although various interpolation methods (e.g., linear and cubic interpolations) have been introduced, the discretization method may still suffer from interpolation biases, which do not disappear when the sample size goes to infinity. The perturbation method is popular because of its wide applications and computational convenience. The essence of this method is Taylor's theory (Judd, 1998). A pre-specified functional form is obtained by expanding the P/D ratio function around certain steady states. However, there is still a heated debate around the judgement of steady points (Juillard, 2011). Furthermore, the perturbation method is challenged by approximation errors, regardless of the choice of steady states (Aruoba et al., 2006). Because it needs extra effort to compute partial derivatives of Euler equations up to a higher order p, a popular approach is to linearize the model around some steady states. While the linearization method is computationally fast and can obtain reasonable solutions for simple functions, the approximation errors become substantially large for complex models. First introduced by Judd (1992), the projection method is appealing due to its global approximation in the entire domain. It delivers an approximation without additional interpolation techniques. The issue is that an appropriate polynomial order p must be specified as a priori. Furthermore, the boundary regions of state variables may become too wide when the dynamics of state variables have high persistence in absolute value, which may result in a loss of accuracy (Culham, 2005). Furthermore, Santos (2000) shows how changes in the curvature of the utility function and the time discount rate can influence the size of Euler equation errors and therefore bound the approximation errors of numerical solution methods. In contrast, the novel estimation methodology proposed in this paper enables consistent estimation of the P/D ratio function for the entire support and whole distribution of state variables, which also avoids interpolation biases when the sample size grows. The newly proposed nonparametric penalized B-splines series regression method works with both continuous and discrete state variables and does not involve computations of partial derivatives. As the sample size grows, the newly proposed method is asymptotically free of approximation errors no matter how complex the model is.

The PEA method provides an alternative solution method for unknown P/D ratios. By using either simulated or empirical data, Den Haan and Marcet (1990) propose to approximate the unknown function by parameterizing the conditional expectation in Euler equations using series expansions. This method is particularly useful and regarded for strong capability when there are many state variables with unknown dynamics in general equilibrium models. As an improvement, Judd et al. (2011) suggest non-stochastic quadrature-based PEA, but this method also requires specifying the conditional distribution of state variables. To alleviate the unstable performance of the original PEA algorithm, Judd et al. (2011) proposes to incorporate an L_1 penalty on the coefficient vector of the parameterized expectation. Both of these regression-based algorithms can only obtain approximations through iterative ordinary least squares, leaving the optimal order for series expansions and best penalization levels in each iteration unsolved. Also, their convergence rates are not known. In addition, the penalization procedure and choices for the order of series expansion are treated separately and their connections are not clear.

The numerical solution methods discussed above are all extensively used in the literature because of their wide scope of application, weaker restrictions and ease of computation. However, deciding which one performs the best is difficult because pros and cons accompany all of them (Culham, 2005). Taylor and Uhlig (1990) show that even for the simple growth model, different numerical solution techniques may display various results for the model. Den Haan and Marcet (1994) reach an important conclusion that numerical solution methods cannot be used interchangeably in general. In addition, one of the most commonly used measures for goodness of approximation is the relative error, which is defined as the approximated Euler equation divided by the approximated P/D ratio function. However, Calin, Chen, Cosimano, and Himonas (2005) point out that relative errors do not necessarily reflect the accuracy of P/D ratios.

Based on Euler equations, our newly proposed penalized B-splines approach is also a regression-based method. However, our paper strongly differs from the PEA algorithms as follows. First, our reason for imposing penalizations and the method by which we do so are completely different from PEA algorithms. The ill-conditioned problem is often encountered in nonparametric series estimation analysis (Newey, 1997). To reduce approximation errors, a large number of basis functions have to be incorporated, but singularity issues may arise as a side effect (Newey and Powell, 2003). Therefore, directly imposed on the coefficient vector of series expansions, the penalization in PEA algorithms mainly aims to alleviate the ill-conditioned problem (Judd et al., 2011). In our paper, we instead impose an asymmetric penalization on the second order differences of coefficient vectors. The role played by our new penalization is three-fold. It not only helps address the common noninvertible problems in series estimations (Newey, 1997), but also prevents the overfitting problem. In addition, our penalized B-splines technique offers an alternative method to achieve identification. As pointed out by Newey (1997) and Newey and Powell (2003), spline regressions are particularly flexible and can also largely reduce the colinearity problem. For the first time in the literature, our penalized B-splines estimation method allows us to achieve the optimal rate of convergence through a newly proposed fast data-driven cross-validation method. Second, our method is a pure nonparametric method, which does not require parameterizing either unknown functions or conditional expectations. Third, unlike current PEA algorithms, our method is a one-step procedure, which does not involve any iterations and numerical integrations. Our examination shows that the small sample performance is satisfactorily stable. Our method is asymptotically unbiased when the sample size increases. Judd et al. (2011)'s method, which also involves series expansions, is sensitive to the choice of basis functions and the order of series expansions as well as the level of penalizations. Our method unifies these problems. We show that our method is much less invariant to the order of series expansions when it is large in both stages. In contrast, the level of penalization is proven to be the major quantity that controls convergence rates and consistency. For the first time in the literature, we propose a fast-generalized cross-validation (GCV) method in the nonparamteric 2SLS analysis to offer additional help in enhancing the robust performance of the regression-based method.

Our newly proposed functional estimation method pushes beyond current methodological constraints. In a different strand of the literature, Woodford (2002) shows that the use of the log-linear approximation of unknown functions such as equilibrium fluctuations in consumption, inflation and output will lead to spuriously higher expected utility under autarchy. Kim and Kim (2003) document a welfare reversal due to approximation errors. Schmitt-Grohe and Uribe (2004) further confirm that a correct second-order approximation of the equilibrium welfare function relies on the accuracy of a second-order approximation to the policy function.

From an econometric perspective, all existing popular numerical approximation approaches are equivalent to various parametric models for P/D ratios, where an approximating functional form is pre-specified a priori. There is no assurance that a parametric model which is chosen for analytic or computational convenience will contain the true P/D ratio function or even a good approximation of it. Therefore, these parametric approximations can cause misleading inferences about and judgements of model performance due to potential approximation errors. It is important to provide an uniformly accurate numerical solution for the P/D ratio function f_t under various empirically relevant setups.

3. Methodology

3.1. Identification

Our work starts from considering the Euler equation in an exchange economy populated by a representative agent:

$$f_t = E[m(X_t, X_{t+1})(f_{t+1} + 1)|I_t], \tag{1}$$

where X_t is a vector of state variables that summarizes the law of motions, $m(X_t, X_{t+1})$ is a known SDF based on the state variables, I_t denotes all information available at time t, and f_t is the unknown P/D ratio function. Without loss of generality, we assume that $E(\cdot|I_t)$ is the rational expectation, which coincides with the mathematical conditional expectation [4].

To consistently estimate the recursively specified unknown function f_t in the Euler equation (1), it is essential to establish the existence and uniqueness of the solution f_t^o. Assume that X_t follows a Markov process. Using the linearity property of expectations, Equation (1) can be equivalently expressed as

$$f(X_t) = \int m(X_{t+1}) f(X_{t+1}) g(X_{t+1}|X_t) dX_t + \pi_t,$$

where $\pi_t = E[m(X_{t+1})|X_t]$ and $g(X_{t+1}|X_t)$ is the conditional density of X_{t+1} given X_t.

Let $K(X_t, X_{t+1}) = m(X_{t+1})g(X_{t+1}|X_t)$ be the kernel function associated with the linear operator A on a normed space $\mathbf{X}$ as

$$(Af)(X_t) \equiv \int K(X_t, X_{t+1}) f(X_{t+1}) dX_{t+1}, \tag{2}$$

Then we could rewrite the Euler equation as an integral equation of the second kind, i.e.

$$f(X_t) - (Af)(X_t) = \pi_t, \tag{3}$$

Correspondingly, we could ensure the existence of a unique solution f given the following assumptions.

Assumption 3.1. *Assume the following:*

(A1) There exists nonzero f satisfying equation (3)

(A2) A is a compact operator.

(A3) Af is a positive transformation for nonzero f.

The compactness of A guarantees that f is locally identified. Together with the positive assumption, A is irreducible. By the second Riesz Theorem, the operator $L = I - A$

[4] Irrational expectations occur when subjective expectations differ from objective expectations. We can convert the subjective expectation back to the mathematical one by the Radyon-Nikodym theory.

has continuous (bounded) Moore-Penrose pseudoinverse, which frees us from the ill-posed problem.

As the gross return is positive, (A1) and (A3) hold in our case. There are several ways to guarantees (A2). For example, if our interest domain $G \in \mathbf{X}$ has finite dimensional range, then a bounded operator A could suffice that A is compact. Or alternatively, we could adopt regularization or penalization method for compactness. More details about how to estimate f using an two stage nonparametric approach is given below.

3.2. Estimation

Following equation (3), we consider the nonlinear time series regression:

$$y_{t+1} = f(X_t) - y_{t+1} f(X_{t+1}) + \varepsilon_{t+1}, \tag{4}$$

where $y_{t+1} = m(X_{t+1})$, and ε_{t+1} is an unobservable martingale difference sequence with respect to the information set I_t, namely $E(\varepsilon_{t+1}|I_t) = 0$. Note that $\{\varepsilon_{t+1}\}$ can be interpreted as a sequence of aggregate pricing shocks. The martingale difference sequence property of $\{\varepsilon_{t+1}\}$ is a sufficient and necessary condition which guarantees the equivalence between the nonlinear time series regression model (4) and the Euler equation.

Compared to the Euler equation, the nonlinear time series regression model does not require specifying the conditional distribution, and thus could accommodate flexible dependence structure of X_t. However, the transformation from the Euler equation into the nonlinear time series regression model will cause endogeneity problem, as a result of recursive occurrences of the unknown P/D ratio function $f(x)$ over two time periods. To be specific, we could choose a set of basis functions $\{\varphi_1, \cdots, \varphi_q\}$ and approximate the unknown function as $f(x) = \sum_{j=1}^{q} \alpha_j \varphi_j(x)$. Define the control variable $\psi_{j,t} = \varphi_j(X_t) - y_{t+1}\psi_j(X_{t+1})$. Since it contains an ingredient y_{t+1} which leads to correlation between the control variable and the true regression error, namely $E(\psi_{j,t}\varepsilon_{t+1}) \neq 0$. As a consequence, the OLS series estimation will not be consistent for Equation (4). To eliminate the endogeneity biases, we need instrumental variables (IV) to conduct the 2SLS, where in the first stage, we regress the control variables $\psi_{j,t}$ on the instrument variables to obtain the fitted values $\hat{\psi}_{j,t}$; and in the second stage, we regress y_t on $\hat{\psi}_{j,t}$ to obtain the estimate of the P/D ratio.

The above estimation procedure could be summarized using the matrix form. Let $(y = y_1, \cdots, y_T)'$, $(\varepsilon = \varepsilon_1, \cdots, \varepsilon_T)'$ and $\alpha = (\alpha_1, \cdots, \alpha_q)$. Define the control variable matrix $\Psi = (\Psi_1, \cdots, \Psi_T)'$, where $\Psi_t = (\varphi_{1,t}, \cdots, \varphi_{q,T})'$. Denote $\phi_{j',t} = \phi_{j'}(X_t)$ for some basis functions $\{\phi_{j'}(z)\}_{j'=1}^{q'}$ which may or may not be the same as $\{\varphi_j(x)\}_{j=1}^{q}$. Note that $E(\varepsilon_{t+1}|X_t) = 0$ implies $E(\varepsilon_{t+1}|h(X_t)) = \mathbf{0}$ for any transformation function h. Therefore, we could define $\Phi = (\Phi_1, \cdots, \Phi_T)'$, where $\Phi_t = (\phi_{1,t}, \cdots, \phi_{q',t})'$, as the instrumental variable matrix. Then the auxiliary regression in the first stage yields the fitted values $\hat{\Psi} = \Phi(\Phi'\Phi)^{-1}\Phi'\Psi$, while the second stage yields that

$$\hat{\alpha} = (\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'Y.$$

Then the estimated P/D ratio is $f(x) = \sum_{j=1}^{q} \hat{\alpha}_j \varphi_j(x)$.

One appealing feature of this 2SLS procedure is its easy implementation. It always has a data-based closed-form solution. As a nonparametric approach, we do not have to specify the DGP for state variables $\{X_t\}$, nor approximating Euler equations by some parametric functional form, thus avoiding potential spurious conclusions due to misspecifications. In general, any series could be used, but the choice might affect the performance of the estimate. For example, using local series might yield more numerical stable estimate compared to using global series. Therefore, we use B-splines as our basis functions. Consider real values c_i, also called knots, such that $c_0 \leq c_1 \leq c_{K+p}$. The pth degree B-splines are the pth degree piecewise polynomials defined as in Eilers and Marks (1996). The spline basis could well approximate most smooth function given that $max(c_i - c_{i-1}) \to 0$. When the knots are equidistance, it requires that the number of total knots $K + p + 1$, or equivalently, K grows as the sample size T does. On the other hand, a relatively large K might result in a complicated model with lots of variability. Therefore, it requires careful tuning in order to well balance between model complexity and flexibility.

For simplicity, we first consider using the same B-spline basis in both stage, so the total numbers of spline basis satisfy $q = q' = K + p$. Theoretical investigations on the rate of K are given in subsection 3.3. However, such an approach might be inferior as it conducts the same amount of smoothing in both stages. A better procedure is to set $q' = o(q)$ so that the bias in the first stage become negligible and will not take in effects in the second stage. Yet this is not the complete story as we also require a practical solution to appropriately select the smoothing parameter. The computation expense of a data-driven method is worth consideration if we need to determine q and q' simultaneously.

This motivates us to consider an approach that could conduct different amounts of smoothing in the 2 stage procedure, while maintaining small computation costs via using the same number of spline basis. Our idea comes from introducing a new smoothing parameter that could replace the number of spline basis to control the roughness of the estimate. Let Δ be the difference operator that $\Delta\alpha_j = \alpha_j - \alpha_{j-1}$ and $\Delta^m = \Delta(\Delta^{m-1})$ for any positive integer m. We propose to estimate α by minimizing

$$\sum_{t=1}^{T}\{y_t - \sum_{j=1}^{K+p}\alpha_j\hat{\psi}_{j,tt}\}^2 + \lambda^* \sum_{k=m+1}^{K+p}\{\Delta^m(\alpha_k)\}^2,\ \lambda^* \geq 0, \tag{5}$$

where λ^* is the penalty parameter that controls the roughness measure that is similar as the total variation of the mth derivative of $f(x)$. In this procedure, K is the smoother in the first stage, while K and λ together are the smother in the second stage. We recommend to choose a relatively large K and let λ play the key role of smoothing. Note that there is no penalty when calculating $\hat{\psi}_{j,t}$, so the regression in the first stage is undersmoothing, while the rate of λ in the second stage will ultimately determine the convergence rate of the estimated function. More justifications are given below.

3.3. Asymptotics

In this section, we discuss the asymptotics for both the unpenalized and the penalized estimators. We have the following assumptions. Suppose f^o is differentiable up to order $d \geq 0$. Denote a truncated pth degree B-splines series $f_a(x) = \sum_{j=1}^{K+p} \alpha_j \varphi_j(x)$, where K denotes the number of knots. The first two assumptions are about K and T.

Assumption 3.2. *Assume that (i)* $K \to \infty$, *(ii)* $\frac{T}{K} \to \infty$.

Assumption 3.3. *Assume that* $T = o(K^{3+2*min(d,p)})$.

Remark: Assumption (3.2) are regularity conditions for consistent estimate of f_a. Note that $f^o - f_a = O(K^{-(1+min(d,p))})$. Assumption (3.3) simply guarantees that the approximated bias is negligible compared to the asymptotic variance which is of the order K/T.

Next we impose some mild conditions on state variables X_t and the unobservable aggregate pricing shock ε_{t+1}.

Assumption 3.4. *The state variable* X_t *follows a Markov process and has a positive density function that is continuous and bounded away from 0 and* ∞ *on a bounded support* $\mathbf{X}$.

Assumption 3.5. *For all* t *and* j, *there exists some* $\delta > 0$ *and* $0 < \Delta < \infty$ *such that (i)* $\{X_t, \varepsilon_{t+1}\}$ *is an* $\alpha-$*mixing sequence with mixing coefficients* $\alpha(j)$ *so that* $\sum_{j=1}^{\infty} \alpha(j)^{\frac{\delta}{4+\delta}} < \Delta$; *(ii)* $KE|\phi_{j,t}|^{4+\delta} < \Delta$, *and* $KE|\varepsilon_{t+1}|^{4+\delta} < \Delta$; *(iii)* $KE|\varphi_{j,t}|^{8+\delta} < \Delta$, *and* $KE|m(X_{t+1})|^{8+\delta} < \Delta$.

Denote $M = T/K$. Let $\lambda_{\min}[E(\frac{\Phi'\Phi}{M})]$ and $\lambda_{\max}[E(\frac{\Phi'\Phi}{M})]$ denote the minimum and maximum eigenvalues of the a $(K+p) \times (K+p)$ matrix $E(\frac{\Phi'\Phi}{M})$ respectively, where $K \to \infty$ as $T \to \infty$. We impose some mild conditions on $\lambda_{\min}[E(\frac{\Phi'\Phi}{M})]$ and $\lambda_{\max}[E(\frac{\Phi'\Phi}{M})]$ so that consistent estimation of the parameters α in the 2SLS series regression can be obtained.

Assumption 3.6. *For all* $M \geq 1$ *(i)* $\lambda_{\min}[E(\frac{\Phi'\Phi}{M})] > 0$; *(ii)* $\lambda_{\max}[E(\frac{\Phi'\Phi}{M})] < \infty$; *(iii)* $\lambda_{\max}[E(\frac{\Phi'\varepsilon\varepsilon'\Phi}{M})] < \infty$.

When the number of regressors is fixed, the well-known necessary and sufficient condition for consistent estimation of parameters in a linear regression model is $\lambda_{\min}[E(\frac{\Phi'\Phi}{T})] > c > 0$. However, the spline basis do not satisfy this assumption. Suppose the density dF_t/dX_t of state variables X_t with a compact support is bounded away from below from zero. Then it is easy to prove that $\lambda_{\min}E(\frac{\Phi'\Phi}{M})$ is bounded away from below from zero uniformly.

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

$$\int [\hat{f}_a(x) - f^o(x)]^2 dF(x) = O(K/T).$$

Theorem 3.1 is a global consistency result. Theorem 3.1 implies that our procedure is always free of misspecification for the price-dividend ratio function when the sample size $T \to \infty$,and we do not have to specify the DGP for state variables. This appealing property is not attainable by existing numerical solution methods in the literature that have to specify a model for the DGP of state variables, which therefore may suffer from model misspecification.

Theorem 3.1 provides a range of admissible rates for K. In practice, one may like to choose K via data-driven methods. However, each iteration requires inversion of a $(K+p)\times(K+p)$ matrix, i.e. a calculation of order $O(K^3)$. Moreover, the placement of the knots might also affect the performance of estimator under finite sample case. To make rigorous statistical inference such as confidence interval estimation and hypothesis testing, we shall derive the asymptotic distribution of the series estimator $\hat{f}_a(x)$. Put $S_T = E(\frac{\Phi'\varepsilon\varepsilon'\Phi}{M})$, $Q_T = E(\frac{\Phi'\Psi}{M})$ and $P_T = E(\frac{\Phi'\Phi}{M})$. Define $V_T \equiv \varphi(x)'E(\frac{\Phi\varepsilon\varepsilon'\Phi'}{M})\varphi(x) = \varphi(x)'S_T\varphi(x)$. Then the variance of the series estimator $\hat{f}_a(x)$ is

$$D_T(x) = \varphi(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}S_TP_T^{-1}Q_T(Q_T'P_T^{-1}Q_T)^{-1}\varphi(x). \tag{6}$$

If there exists conditional homoskedasticity, i.e., $E(\varepsilon_{t+1}^2|\Phi_{q,t}) = \sigma^2$ for all t, then we have

$$D_T(x) = \sigma^2 \varphi(x)'(Q_T' P_T^{-1} Q_T)^{-1} \varphi(x). \tag{7}$$

Theorem 3.2 (Asymptotic Normality)**.** *Suppose Assumptions 3.1-3.6 hold. Then for any given $x \in \mathbf{X}$, as $T \to \infty$,*

$$\frac{[\hat{f}_a(x) - f^o(x)]}{\sqrt{D_T(x)}} \xrightarrow{d} N(0,1). \tag{8}$$

Theorem 3.2 imply that the bias of the series estimator $\hat{f}_a(x)$ vanishes to zero sufficiently fast so that it does not affect the asymptotic normal distribution of $\hat{f}_a(x)$. However, this is not the optimal convergence rate as the bias is negligible compared to the standard deviation.

Our method is also applicable to hidden Markov processes. Suppose state variables X_t is not directly observable, but can be estimated via such methods as Kalman filters. Intuitively, the estimated state variables $\hat{x}$ converges in probability to the point x at a parametric rate $T^{-\frac{1}{2}}$, 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 impact on the asymptotic distribution of $\hat{f}_a(\hat{x})$.

Finally, we consider the penalized splines case. The following condition is about the penalty parameter λ.

Assumption 3.7. *Define $h = \lambda^{-1/(2d)}/K$. Assume that $h \to 0$ and $Th \to \infty$.*

Remark: h defined in Assumption (3.7) serves as the equivalent bandwidth in nonparametric smoothing procedure. Assumption (3.7) is used to prove the consistency of the estimate as it guarantees that the bias and the variance of the estimator shrink to 0. We could obtain the optimal rate of λ by selecting the optimal rate of h, which is $T^{-1/(4d+1)}$ when $f(x) \in W^{(2d)}$.

Theorem 3.3 (Consistency)**.** *Suppose Assumptions 3.1-3.7 hold. Then there exists a unique solution* $f^o(x)$ *to Equation (1), and the nonparametric 2SLS series estimator* $\hat{f}_a(x)$ *satisfies:*

$$\int [\hat{f}_a(x) - f^o(x)]^2 dF(x) = O[\frac{1}{Th} + h^{4d}]$$

Remark: When K grows faster than $T^{1/(3+2min(d,p))}$ and λ grows exactly at the rate of $K^{2d}(T^{-2d/(4d+1)})$, $\hat{f}(x)$ could reach the optimal convergence rate of $T^{(2d)/(4d+1)}$.

Similar as Theorem (3.2), we have the following results.

Theorem 3.4 (Asymptotic Normality)**.** *Suppose Assumptions 3.1-3.7 hold. Then for any given* $x \in \mathbf{X}$*, as* $T \to \infty$*,*

$$\frac{[\hat{f}_a(x) - f(x)]}{\sqrt{var(\hat{f}_a(x))}} \xrightarrow{d} N(0,1); \tag{9}$$

3.4. Data Driven Implementation

Subsection 3.3 provides the theoretical recommendation on how to choose 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 GCV method to select λ in order to prevent overfitting. Since K is not the crucial smoothing parameter, the choice of the degree of the B-spline basis, as well as the placement of the knots are not important. Unlike regression spline where no penalty is imposed, we do not need to worry about how to place our knots. We recommend select p from $0, 1, 2, 3$, choose a relatively K and use equi-distance knots.

The penalty order m as well as the penalty parameter are important. Common choices of m are 1 or 2, though one may further increase m if they expect the estimated function has higher order derivatives. To reduce computation burden, we propose to use the GCV approach to determine λ. We have the following results.

Proposition 1. *Let Y, Φ, Ψ and $\hat{\Psi}$ be defined as in subsection 3.2. Define the penalty matrix $P = D_m' D_m$, where D_m is the $(K+p-m) \times (K+p)$ difference matrix. Define $H = \Phi(\hat{\Psi}^T\hat{\Psi} + \lambda P)^{-1}\hat{\Psi}$ and $\Sigma = \Psi'\Phi(\Phi'\Phi)^{-1}\Phi'\Psi$. Denote r_j as the jth eigenvalues of $\Sigma^{-\frac{1}{2}}P\Sigma^{-\frac{1}{2}}$. Then the GCV value equals*

$$GCV = \frac{||Y - \hat{Y}||^2}{(trace(I-H))^2} = \frac{\sum_{i=1}^{T} y_i^2 - 2\sum_{i=1}^{K+p} \frac{z_{1,i}z_{2,i}}{1+\lambda r_i} + \sum_{i=1}^{K+p} \frac{z_{3,i}^2}{1+\lambda \tilde{r}_i}}{(T - tr_\lambda)^2}, \tag{10}$$

where $tr_\lambda = \sum_{j=1}^{K+p} \frac{1}{1+\lambda r_j}$, $z_{1,i}, z_{2,i}$ are provided under equation (22), and $\tilde{r}_i$ and $z_{3,i}$ are provided under equation (23).

Remark: Note that we could precalculate $\sum_{i=1}^{T} y_i^2$, all $z_{1,i}, z_{2,i}, z_{3,i}, r_i$ and $\tilde{r}_i$. Then in each evaluation, we could calculate $GCV(\lambda)$ by equation (10). Since we need not calculate the inverse of a matrix of order $(K+p) \times (K+p)$, we reduce the computation from $O(K^3)$ to $O(K)$.

4. Monte Carlo and Simulation Studies

4.1. Monte Carlo Simulations

We now examine the finite sample performance of the proposed B-spline estimation method in estimating functions with different smoothness under different conditional distributions of the state variable. For each Monte Carlo study, we generate samples of sizes of 250 and 500 and implement our new 2SLS B-spline regression with and without penalization. In each study, we also examine the estimation performance under 35 knots and 50 knots using piecewise-linear and piecewise-quadratic B-splines.

DGP F.1 $f(X_t) = 3 + 0.5sin(20X_t + 1.2) + cos(10x + 2)$.

DGP F.2 $f(X_t) = exp(x)$.

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.

To ensure our method works for a wide range of empirically relevant setups, we investigate the estimation results for the above two classes of functions under the following assumed data generating processes. Among others, autoregressive processes have been widely employed in modelling the dynamics of state variables, therefore we first explore the finite sample performance under a popularly used AR(1) process:

DGP S.1

$$X_{t+1} = \Gamma X_t + \epsilon_{t+1}, \tag{11}$$

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

Specifically, we further investigate two extreme situations for the above general AR(1) process. In DGP S.1.1, we assume $\Gamma = 0.1$ with $\sigma_s = 0.1$, where the serial dependency fades out exponentially fast as the time distance between state variables increases. Also, as the autocorrelation of state variables among two consecutive periods is small, the correlation between instruments and endogenous variables is also small in this scenario. In DGP S.1.2, we assume $\Gamma = 0.8$ with $\sigma_s = 0.5$. This is a situation where numerical solution methods face challenges, especially for the projection method. Calin et al. (2005) points out that the pre-specified order for series expansion must be large enough to ensure reasonable approximation of the projection method. 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 the second set of simulation studies, we examine the finite sample performance under a general autoregressive conditional heteroskedasticity (ARCH) process. Given their ability to capture volatility clustering and leverage effects, ARCH processes and its extensions have been proven to be popular. In addition, by studying the performance of our new estimation method under the ARCH process, we can facilitate future studies that aim to investigate the role played by time-varying volatility in modelling asset prices. We consider a general

ARCH(1) process for the DGP of state variables as follows.

DGP S.2

$$\begin{cases} X_t &= \epsilon_t \\ \epsilon_t &= z_t\sqrt{h_t}, \\ h_t &= \omega_0 + \omega_1 X_{t-1}^2, \\ z_t &\sim i.i.d.(0,1) \end{cases} \tag{12}$$

We investigate the finite sample performance under the ARCH DGP with high and low autocorrelations, namely DGP S.2.1: $\omega_0 = 0.1$ and $\omega_1 = 0.2$ and DGP S.2.2: $\omega_0 = 0.01$ and $\omega_1 = 0.6$. Using these two Monte Carlo studies, we can evaluate our newly proposed estimation method for state variables with large and small volatility clustering effects.

Tables 1 and 2 report the Monte Carlo results for the designed functions DGP F.1 and DGP F.2 for state variables with the above assumed DGP S.1.1, DGP S.1.2, DGP S.2.1 and DGP S.2.2. For each pair of designed function and DGP, we report estimation results with different degrees of B-splines and numbers of knots. Integrated mean squared error (IMSE) has been widely used in nonparametric series estimations to evaluate finite sample performance (Hansen, 2015). To evaluate the goodness of fit, we calculate IMSE for each Monte Carlo study by implementing the estimation method 400 times.

The IMSE of the penalized two-stage B-splines estimator $\hat{f}(x)$ is

$$IMSE = \int E[\hat{f}(x) - f(x)]^2 dF(x), \tag{13}$$

where $F(x)$ is the cumulative distribution function for the state variable.

Overall, our newly proposed penalized two-stage B-spline estimator works uniformly well for both AR and ARCH processes with different levels of serial dependencies. From Figures 1-8, comparing true P/D ratio functions, we can visualize the estimation results with and without penalties. The penalized B-splines will generally enhance the estimation results, especially when estimating smoother functions. The outstanding accuracy for small sample studies of size 250 demonstrates how our method offers significant advantages for current asset pricing and macroeconomic general equilibrium modelling, where the state variables are mostly quarterly with an effective sample size around 300 for the postwar period. The sample size of 500, which amounts to 42 years of monthly observations, will be sufficiently large to ensure more accurate estimation results using our method.

In addition, we show that the number of knots K and degrees of B-splines are not crucial in estimation. The mechanism can be clearly seen from Tables 1 and 2. When a larger number of knots is used, our fast GCV algorithm will automatically generate a larger penalty to correct for the potential overfitting problem. This further strengthens the superiority of our new method, which successfully optimizes the function to be estimated via a single data-driven penalty term λ^*. As the question of how to choose optimal orders for series expansions in both stages remains an open one, our paper contributes to the literature by providing a convenient solution via the fast GCV algorithm.

4.2. Comparison with Numerical Solutions

We confine the scope of this section to situations where analytic solutions of the P/D ratios exist (Burnside, 1998). We provide a detailed comparison of our nonparametric penalized B-spline series regression method with some representative and popularly used numerical solution methods. For the numerical solution method, we consider perturbation, projection and discretization methods as well as the PEA and 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 due to limited skill, time, or noisy observations. In practice, a proxy for the dynamics of state variables can be obtained via various techniques. For example, using simple rules of thumb, investors may obtain an estimated DGP which actually deviates from the true one in many dimensions (Cecchetti, Lam, and Mark, 2000). Cecchetti et al. (2000) point out that this discrepancy between the true and subjective beliefs in the DGP of state variables is a key ingredient in addressing the equity premium puzzle. Even though it is common to encounter misspecified DGPs, very little attention has been paid to examining how asset pricing models can be affected when DGPs of state variables are misspecified.

Built upon the Mehra and Prescott (1985) model, analytic solutions are obtainable under special circumstances (Burnside, 1998). In this exchange economy, there is an infinitely-lived representative agent who wishes to maximize her expected lifetime utility at time zero:

$$\begin{aligned} &\max_{\{C_t\}} E\sum_{t=0}^{\infty}\beta^{t-1}\frac{C_t^{1-\gamma}}{1-\gamma} \\ &\text{s.t.}\, C_t + P_{t+1}\theta_{t+1} + Q_t b_{t+1} = b_t + (D_t + P_t)\theta_t, \end{aligned} \tag{14}$$

where $X_t = ln(C_t/C_{t-1})$, $X_{t+1} - \mu = \Gamma(X_t - \mu) + u_{t+1}$, and $u_{t+1} \sim i.i.d.N(0, \sigma_u^2)$. 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:

$$f_t = \beta E[e^{(1-\gamma)X_{t+1}}(f_{t+1}+1)|X_t]. \tag{15}$$

We consider two numerical studies to examine the performance of our newly proposed estimation method with finite samples 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 without knowing their conditional distributions. We first compare solution performance in scenario one.

DGP B.1: We consider the true DGP of state variable is fully acknowledged by public and is assumed to follow the following AR(1) process:

$$X_{t+1} - \mu = \Gamma(X_t - \mu) + \epsilon_{t+1}, \tag{16}$$

where $\epsilon_t \sim IIDN(0, \sigma^2)$. In this section, for ease of comparisons, we set $\beta = 0.96$, $\gamma = 2.5$, $E(X_t) = 0.0179$ and $\sigma = 0.0379$. Similar to the Monte Carlo simulation studies, we consider two sub-scenarios, namely DGP B.1.1: $\Gamma = -0.139$ and $\Gamma = 0.8$.

In scenario two, we further investigate possible consequences when the dynamics of state variables are misspecified using Mehra and Prescott's (1985) model.

DGP B.2: We have a threshold model for the true DGP, which is a nonlinear stationary process:

$$X_{t+1} = \begin{cases} \mu + \Gamma_1 X_t + u^*_{1,t+1},\ u^*_{1,t} \sim IIDN(0, \sigma_1^2) & \text{if } X_t > 0, \\ \mu + \Gamma_2 X_t + u^*_{2,t+1},\ u^*_{2,t} \sim IIDN(0, \sigma_2^2) & \text{if } X_t \leq 0 \end{cases} \tag{17}$$

where $\sigma_1 = 3.48\%$, $\sigma_2 = 2\sigma_1$, $\beta = 0.96$, $\gamma = 2.5$, $\Gamma_1 = 0.8$, and $\Gamma_2 = -0.139$. A misspecified DGP for such a process is as follows:

$$\tilde{X}_{t+1} - \mu = \bar{\Gamma}(\tilde{X}_t - \mu) + v_{t+1}, \text{ and } v_{t+1} \sim IIDN(0, \sigma_v^2), \tag{18}$$

where $\bar{\Gamma}$ and σ_v^2 are chosen so that it can match the autocorrelation with the true DGP. DGP B.2 explores a threshold structure, whose imporance has been widely acknowledged in many economic studies (e.g., Hong, Li, and Zhao, 2012). In the true DGP, the state variable X_t is assumed to enjoy higher persistency level in mean and lower volatility when the consumption growth rate is positive, and will exhibit a mean-reverting pattern when the consumption growth rate is negative.

Figures 9-11 compare the approximated P/D ratio function from different solution methods under DGP B.1 and DGP B.2 with sample sizes of 250 and 1000 respectively. Specifically, DGP B.1.1 has a small serial correlation of state variables in absolute values and the true

P/D ratio behaves as a linear function of it. Therefore, as shown in Figure 9, all studied numerical methods can provide accurate approximations. The performance of our regression-based nonparametric penalized B-spline 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-spline method always exhibits superior performance compared to the PEA method. It is worth mentioning that the discretization and improved PEA methods work well when the DGP of state variables is correctly specified.

Therefore, in DGP B.2, where the analytic solution does not exist (Burnside, 1998), we first assume 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 comparison purposes. As can be seen from Figure 11, 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. Using our penalized B-splines regression, we can obtain consistent, unbiased and efficient estimation of the unknown P/D ratios in the presence of an unknown DGP.

Tables 3 and 4 further report and compare the number of iterations, real computational time and mean squared errors for the above simulation studies. Our method is a speedy one-step procedure with an endogenously 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 and the orders in both stages are optimized through a penalization factor, which is itself optimized in the GCV procedure.

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 2SLS series regression method will become an indispensable approach for obtaining a consistent estimate of the P/D ratio function, and constructing the most reliable and accurate model implications.

5. Empirical Applications

5.1. *Present Value of Future Dividends*

We now compute the present value of future dividend payments using our nonparametric 2SLS penalized B-splines regression procedure for the US from 1947 Q1-2016 Q1. We intend to provide an alternative way to allow time-varying stochastic discount factors to determine equilibrium asset prices that reflect full rationality and future dividend payments. Following the modelling setup in Shiller (2014), under the fundamental work of Mehra and Prescott's (1985) CAPM model, we use the marginal rate of substitution between consumptions in consecutive periods as the discount factor.

Traditionally, the present value of future dividends is constructed as

$$P_t = E_t \sum_{t=0}^{\infty} \Pi_{j=0}^{k} M_{t+j} D_{t+k}, \tag{19}$$

where $M_{t+j} = \beta(\frac{C_{t+1}}{C_t})^{-\gamma}$, C_t is the real per capita consumption at time t, β is the constant time discount factor and γ is the constant relative risk aversion level.

One practical issue of Equation (19) is the unknown consumption growth rates after the termination date. The common strategy is to assume a constant growth rate and take it as the geometric average growth over the last 30 or 10 years. However, as Shiller (2014) emphasizes, there is no objective way to forecast dividends out for decades. Fortunately, using our new penalized B-splines regression, we can obtain the present value of future dividends P_t without this problem.

In estimating the present value of future dividends through f_t in the Euler equation (15), we use the US quarterly real consumption per capita data from 1947 Q1- 2016 Q1, which amounts to 277 observations [5]. By feeding Euler equation (15) with empirical observations, we do not need to estimate or assume conditional distributions for X_t.

For simplicity, we estimate f using $K(T) = 50$ equidistant knots with piecewise-linear ($w = 1$) B-splines. The optimal penalty which ensures efficient estimator will be determined by the fast GCV procedure. Therefore, we can present the present value P_t alternatively using $P_t^* = f_t D_t$.

We plot the real Standard and Poor's composite stock price index along with three present values of subsequent real dividends as used in Shiller (2014) together with our new measure. In Figure 9, except for the consumption discounted dividends, shown in Equation (19), we also produce the present price (in an orange dotted line) with dividends discounted by a constant interest rate, and the present price (in a black dotted line) with dividends discounted by actual future interest rates.

Compared to the other three present value measures, our new consumption-based present value P_t^* is able to reproduce most of the magnitudes and fluctuations from 1947 to the middle of 1970. In recent years, as consumption growth has been very smooth, it is not surprising to see P_t^*'s lack of ability to mimic both the magnitude and volatility of the observed price. However, the present value using our method still reflects the fact that economic fundamentals, such as the consumption/dividend growth, can explain certain price changes to some extent.

[5]The US quarterly real consumption per capita data can be downloaded from the Fred St. Louis

5.2. A Nonparametric Decomposition of Dividend Yield and Predictability of Equity Returns

we consider the returns on $S\&P500$ and 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 U.S. National Income and Product Accounts (NIPA). Our data are in quarterly frequency and spans from 1962 to 2015. The inflation series is constructed using the consumption price deflater. In constructing observed dividend yields from empirical dataset, dividends are summed over the past one year to alleviate seasonality effects which are especially pronounced in the dividend payment process(Ang and Bekaert, 2006).

Built on equilibrium asset prices under full rationality, sharing the same spirit of the implied volatility (Shiller, 2014), the implied dividend yields, $log(1/f_t)$, reflect consumption-based rational forecasts of future returns and dividend growth. In this section, we will investigate how this non-parametrically estimated implied dividend yields predict future excess returns and dividend growth.

We further propose a nonparametric decomposition of stock dividend yields into two orthogonal components, namely the data-driven nonparametrically estimated implied dividend yield under full rationality and the residuals. To achieve such a decomposition, we propose the following nonparametric series regression model:

$$dy_t^4 = F(1/f_t) + \epsilon_t, \tag{20}$$

where f_t is the implied P/D ratios estimated from the Euler equation (1), $F(1/f_t)$ is the component that reflects full rationality and risk averse attitude which will be estimated nonparametrically via a penalized B-spline regression. $dy_t^4 = D_t^4/P_t$ represents the observed dividend yields with dividends $D_t^4 = \sum_{j=0}^{3} D_{t-j}$ summed over the past year.

Let $diff_t = dy_t^4 - F(f_t)$ denote the difference between the observed and the implied dividend yield component. It reflects non-equilibrium or irrational information that might be possibly caused by bubbles, heterogeneous time effects, etc.. We create a dummy variable, $asym_t = I_t\{diff_t < 0\}$, to investigate the existence of nonlinear predictive ability that such residual part may exhibit on future excess returns and dividend growth.

Instead of predicting future cumulative excess returns using observed log dividend yields alone as the single univariate regressor, Ang and Bekaert (2006) documents the superior and enhanced predictability ability that dividend yields exhibit on excess returns when an additional regressor short interest rate is included. Therefore, we consider the main regression model as

$$\tilde{y}_{t+j} = \alpha_j + \beta_{j,1} log(1/f_t) + \beta_{j,2} dy_t^4 + \beta_{j,3} r_t + \epsilon_t, \tag{21}$$

where $\tilde{y}_{t+j} = (4/j)[(y_{t+1} - r_t) + (y_t - r_{t-1}) + \cdots + (y_{t+j} - r_{t+j-1})]$ is the annualized $j-$period excess return for the U.S. aggregate market, $y_{t+1} = log(\frac{P_{t+1}+D_{t+1}}{P_t})$ is the log return on equity and r_t is the log return on short interest rate.

Starting with a univariate regressor, observed dividend yields, Table 5 reports the predictability regression of future excess returns by short interest rate, implied dividend yields, the observed dividend yields and asymmetric dummy variable for horizon of 1, 4, 12 and 20 quarters. Compared with the regression on the observed dividend yields only, the new regression maintains better predictability by achieving higher R^2. First, the model convinces the positive predictability that dividend yields exhibit on excess returns, where the predictability ability increases significantly when forecasting horizon increases. Second, higher prices implies lower future excess returns. Additionally, an increase in implied dividend yields predicts an increase in expected future excess returns. Therefore, implied dividend yields could serve as an additional predictor when predicting future stock performance.

Figures 13-14 report the predictability ability that nominal short interest rates, the rational and beyond-rational components of dividend yields exhibit on future excess returns for horizons of 1 quarter to 10 years. We compute t-statistics based on standard errors and Newey-West ones. Over 1962 to 1990, real short interest rate has decreasing predictive power on future excess returns. Higher observed and implied dividend yields both predict higher future excess returns. However, during such a period, the predictive marginal effect of observed dividend yields is significant only during the short and long horizons. We find significant nonlinear predictability that observed dividend yields exhibit during short and medium horizons. During 1990 to 2015, we find little forecastablity in future excess returns using implied dividend yields. But the asymmetric effect of dividend yields is still proven to be a strong predictor [6].

We also investigate forecasts of long-horizon dividend growth. Table 6 reports forecasting results for the same exercise for future dividend growth. Dividend yields have been documented to have little forecastability in dividend yields (Lettau and Ludvigson, 2005). To capture the consumption-based present-value of future returns and dividends, Lettau and Ludvigson (2005) propose a proxy through a cointegration linear regression model. They found reversed forecastability that consumption-based dividend yields on dividend growth: they find that high dividend payments relative to prices predicts higher dividend growth, not lower. In our empirical study, we find that observed dividend yields have little predictability in dividend growth. However, implied dividend yields have strong but time-varying forecastability in dividend yields. During 1952-1990, we find that higher dividend payments predict lower dividend growth in the future, which is consistent with the theory (Campbell and Shiller, 1988). However, this forecastability is found to be time-varying. During 1990-2015, we find the opposite effect as that in Lettau and Ludvigson (2005). The asymmetric effect of dividend yields that does not depend on long-run equilibrium has strong time-varying forecastability in dividend yields. The asymmetric effect indicator variable exhibits time-varying and significant predictability on future dividend yields.

[6] We also conduct a regression analysis by pooling all data from 1952 to 2015 together, which results in clear differences in the predictability ability for both rational and beyond-ration components. Such a difference is mainly due to inconsistent estimations of the regression coefficients provided the predictive marginal effects are time-varying during the two samples. Therefore, it provides further evidence on time-varying predictive relationship that short interest rate and dividend yields have on future excess returns.

6. Conclusion

Considerable attempts to enrich the explanatory powers of economic models have been witnessed in recent years. In modelling dynamic asset prices and enhancing the understanding of well-documented financial anomalies, P/D ratios are one of the central quantities that must be solved for. Unlike current numerical solution methods, we do not assume any distributional assumptions on the dynamics of state variables, and we propose a 2SLS penalized B-splines regression method, which has a convenient data-based closed-form solution regardless of model complexity, thus making the implementation particularly easy in practice. Our method is free of endogeneity biases and functional form misspecification when the sample size increases.

For the first time in the literature, we complement the work of Newey and Powell (2003) by showing how to optimize the order of series expansion in the nonparametric 2SLS analysis. Using spline techniques, we show that it is equivalent to set the number of knots large in both stages and let our newly proposed fast GCV algorithm determine the optimal penalty, which prevents the overfitting problem. Our method is data-driven and easy to implement. It does not involve numerical integrations or optimizations.

We also establish new local identification strategies for the integral equation of the first kind, which helps prevent the ill-posed inverse problem in the nonparametric estimation literature. Under our identification strategy, for the first time in the literature, we achieve the optimal rate of convergence using our newly proposed penalized B-splines regression. It will become an important tool to construct reliable and correct conclusions about model implications and to evaluate general equilibrium models.

In the empirical application, we apply our method to construct a new measure for the present value that reflects all future dividend payments. Our approach can be generalized in several directions. Meghir and Pistaferri (2004) model the conditional variance of the income shocks as a parsimonious ARCH process. It helps them achieve significant improvement in understanding household counterfactual consumptions by capturing education- and time-specific differences in the stochastic process for earnings and for measurement error. By applying this newly proposed method, we can learn the extent to which income risks affect equity prices without modelling the stochastic process of income risks. Also, by incorporating empirical observations of state variables and avoiding model misspecification, we can better understand how monetary and fiscal policies will actually function in the real economy. In addition, this method can be extended to DSGE models in the production economy, where the log linearization method is widely used (Zietz, 2006). Lastly, under a system of multiple Euler equations, we must solve multiple unknown functions rather than just the P/D ratio function (Epstein and Zin, 1989). It is important to solve all unknown functions accurately because possible functional form misspecification from one solution may be amplified and adversely affect the others, eventually seriously discrediting model implications. Our 2SLS series regression approach can be extended to this more general and complex setup, eliminating all possible functional form misspecification in large samples. This newly proposed functional estimation method will facilitate a more reliable understanding of existing DSGE models that are now widely used in both macroeconomics and finance.

One limitation of the current research is that we have not obtained explicit formula for the asymptoic bias and variance of the penalized spline estimate. For this direction, a potential try is to consider more complicated weighted (possibly asymmetric) penalty. Besides, our approach could also be extended to multivariate D-dimensional state variable X_t, together with the use of tensor product and d penalty parameters.

References

Andrews, D. W., 1991. Asymptotic normality of series estimators for nonparametric and semiparametric regression models. Econometrica: Journal of the Econometric Society pp. 307–345.

Ang, A., Bekaert, G., 2006. Stock return predictability: Is it there? The Review of Financial Studies 20, 651–707.

Aruoba, S. B., Fernández-Villaverde, J., Rubio-Ramirez, J. F., 2006. Comparing solution methods for dynamic equilibrium economies. Journal of Economic dynamics and Control 30, 2477–2508.

Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. The Journal of Finance 59, 1481–1509.

Blanchard, O., 2016. Do dsge models have a future? Tech. rep.

Brown, B. M., 1971. Martingale central limit theorems. The Annals of Mathematical Statistics 42, 59–66.

Burnside, C., 1998. Solving asset pricing models with gaussian shocks. Journal of Economic Dynamics and Control 22, 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, 961–982.

Campbell, J. Y., 2003. Consumption-based asset pricing. Handbook of the Economics of Finance 1, 803–887.

Campbell, J. Y., Cochrane, J. H., 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. Journal of political Economy 107, 205–251.

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

Cecchetti, S. G., Lam, P.-S., Mark, N. C., 2000. Asset pricing with distorted beliefs: Are equity returns too good to be true? The American Economic Review .

Chen, X., Chernozhukov, V., Lee, S., Newey, W. K., 2014. Local identification of nonparametric and semiparametric models. Econometrica 82, 785–809.

Chen, X., Favilukis, J., Ludvigson, S. C., 2013. An estimation of economic models with recursive preferences. Quantitative Economics 4, 39–83.

Christensen, T. M., 2017. Nonparametric stochastic discount factor decomposition. Econometrica 85, 1501–1536.

Cochrane, J. H., 1992. Explaining the variance of price–dividend ratios. Review of Financial Studies 5, 243–280.

Culham, A. J., 2005. Computational techniques for solving asset pricing models. Tech. rep., Working Papers 1, Florida State University, Department of Mathematics.

Den Haan, W. J., Marcet, A., 1990. Solving the stochastic growth model by parameterizing expectations. Journal of Business & Economic Statistics 8, 31–34.

Den Haan, W. J., Marcet, A., 1994. Accuracy in simulations. The Review of Economic Studies 61, 3–17.

Epstein, L. G., Zin, S. E., 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. Econometrica: Journal of the Econometric Society pp. 937–969.

Escanciano, J. C., Hoderlein, S., Lewbel, A., Linton, O., Srisuma, S., 2015. Nonparametric euler equation identification and estimation. Tech. rep., cemmap working paper, Centre for Microdata Methods and Practice.

Fernández-Villaverde, J., Ramírez, J. F. R., Schorfheide, F., 2016. Solution and estimation methods for dsge models. Tech. rep., National Bureau of Economic Research.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., 2006. Solving dsge models with perturbation methods and a change of variables. Journal of Economic Dynamics and Control 30, 2509–2531.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., 2007. Estimating macroeconomic models: A likelihood approach. The Review of Economic Studies 74, 1059–1087.

Gallant, A. R., Tauchen, G., 1989. Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. Econometrica: Journal of the Econometric Society pp. 1091–1120.

Hansen, B. E., 2015. The integrated mean squared error of series regression and a rosenthal hilbert-space inequality. Econometric Theory 31, 337–361.

Hansen, L. P., Scheinkman, J. A., 2012. Recursive utility in a markov environment with stochastic growth. Proceedings of the National Academy of Sciences 109, 11967–11972.

Hansen, L. P., Singleton, K. J., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica: Journal of the Econometric Society pp. 1269–1286.

Hong, Y., Li, H., Zhao, F., 2012. Out-of-sample performance of discrete-time spot interest rate models. Journal of Business & Economic Statistics .

Jagannathan, R., Marakani, S., 2015. Price-dividend ratio factor proxies for long-run risks. Review of Asset Pricing Studies 5, 1–47.

Judd, K. L., 1992. Projection methods for solving aggregate growth models. Journal of Economic Theory 58, 410–452.

Judd, K. L., 1998. Numerical methods in economics. MIT press.

Judd, K. L., Maliar, L., Maliar, S., 2011. Numerically stable and accurate stochastic simulation approaches for solving dynamic economic models. Quantitative Economics 2, 173–210.

Juillard, M., 2011. Local approximation of dsge models around the risky steady state. Tech. rep., Mimeo October.

Kim, J., Kim, S. H., 2003. Spurious welfare reversals in international business cycle models. journal of International Economics 60, 471–500.

Lettau, M., Ludvigson, S. C., 2005. Expected returns and expected dividend growth. Journal of Financial Economics 76, 583–626.

Meghir, C., Pistaferri, L., 2004. Income variance dynamics and heterogeneity. Econometrica 72, 1–32.

Mehra, R., Prescott, E. C., 1985. The equity premium: A puzzle. Journal of monetary Economics 15, 145–161.

Newey, W. K., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147–168.

Newey, W. K., Powell, J. L., 2003. Instrumental variable estimation of nonparametric models. Econometrica 71, 1565–1578.

Pohl, W., Schmedders, K., Wilms, O., 2014. Solving asset-pricing models with recursive preferences .

Santos, M. S., 2000. Accuracy of numerical solutions using the euler equation residuals. Econometrica 68, 1377–1402.

Schmitt-Grohe, S., Uribe, M., 2004. Solving dynamic general equilibrium models using a second-order approximation to the policy function. Journal of economic dynamics and control 28, 755–775.

Shiller, R. J., 2014. Speculative asset prices. American Economic Review 104, 1486–1517.

Taylor, J. B., Uhlig, H., 1990. Solving nonlinear stochastic growth models: A comparison of alternative solution methods. Journal of Business & Economic Statistics 8, 1–17.

Woodford, M., 2002. Inflation stabilization and welfare. Contributions in Macroeconomics 2.

Zietz, J., 2006. Log-linearizing around the steady state: A guide with examples. Available at SSRN 951753 .

Table 1: IMSE of Penalized B-spline under AR State Variables

		Sample Size=250			Sample Size=500		
		$IMSE$	$IMSE$	$\bar{\lambda}$	$IMSE$	$IMSE$	$\bar{\lambda}$
		B-spline without penalty	Penalized B-spline	Optimal Penalization	B-spline without penalty	Penalized B-spline	Optimal Penalization
DGP F.1 DGP S.1.1	w=1, K=35	0.01826872	0.0131101	2.532474	0.009258146	0.006898432	2.876585
	w=1, K=50	0.02347168	0.01470143	3.527953	0.01183204	0.007528154	4.268646
	w=2, K=35	0.0221372	0.01282211	2.199518	0.009463132	0.006621339	2.492744
	w=2, K=50	0.03377577	0.01468604	3.251942	0.01202409	0.007283417	3.841638
DGP F.1 DGP S.1.2	w=1, K=35	0.06700802	0.08259431	0.2080672	0.0672485	0.08046206	0.3690953
	w=1, K=50	0.06920832	0.06256849	0.2743021	0.02954389	0.039219	0.336305
	w=2, K=35	0.08284512	0.06503849	0.09964043	0.07453333	0.06287254	0.1606192
	w=2, K=50	0.05617438	0.05465477	0.1984541	0.02482012	0.03087558	0.2225816
DGP F.2 DGP S.1.1	w=1, K=35	0.01334645	0.004341459	70.34631	0.006803884	0.002328368	128.948
	w=1, K=50	0.01796366	0.005228453	85.77986	0.009216935	0.00259976	161.8562
	w=2, K=35	0.01505069	0.004494111	76.53147	0.007016259	0.002367346	133.0716
	w=2, K=50	0.01981698	0.005336624	88.68647	0.009392146	0.002611739	173.6824
DGP F.2 DGP S.1.2	w=1, K=35	0.03262231	0.07321093	2.238989	0.01536193	0.05109586	2.37595
	w=1, K=50	0.04434008	0.08083232	3.07003	0.01941035	0.05765428	3.514282
	w=2, K=35	0.03506841	0.07144779	2.1535	0.03024117	0.05027726	2.309324
	w=2, K=50	0.04347047	0.07851707	2.924002	0.02021131	0.05657635	3.348899

Notes: State variable X_t is assumed to follow AR(1) processes in DGP.S.1.1 and DGP.S.1.2. W is the degree of B-splines and K is the number of knots. Knots are equally spaced on the entire range of the state variable X_t. $w = 1$ is piecewise linear B-spline and $p = 2$ is piecewise quadratic B-spline.

Table 2: IMSE of Penalized B-spline under ARCH State Variables

		Sample Size=250			Sample Size=500		
		$IMSE$	$IMSE$	$\bar{\lambda}$	$IMSE$	$IMSE$	$\bar{\lambda}$
		B-spline without penalty	Penalized B-spline	Optimal Penalization	B-spline without penalty	Penalized B-spline	Optimal Penalization
DGP F.1 DGP S.2.1	w=1, K=35	0.02464076	0.029118	0.5715628	0.01405926	0.01541851	0.535519
	w=1, K=50	0.03249247	0.03360804	0.9272623	0.01582086	0.01751706	0.9680503
	w=2, K=35	0.02450054	0.02705671	0.4754202	0.01267992	0.01398112	0.449598
	w=2, K=50	0.03207846	0.0321732	0.816137	0.01559074	0.01660778	0.8643913
DGP F.1 DGP S.2.2	w=1, K=35	0.02143143	0.02016245	0.8702906	0.01191423	0.01163581	0.9533953
	w=1, K=50	0.02615308	0.02246086	1.342364	0.01357569	0.01227891	1.570881
	w=2, K=35	0.02156865	0.01912847	0.7659362	0.01258897	0.010924	0.8547724
	w=2, K=50	0.02655386	0.02150427	1.225838	0.01370247	0.01179226	1.42465
DGP F.2 DGP S.2.1	w=1, K=35	0.01790781	0.01011446	25.93743	0.009273527	0.008092244	47.14169
	w=1, K=50	0.02795115	0.01050107	30.21588	0.01227992	0.00802897	58.63276
	w=2, K=35	0.01841731	0.01026836	26.0214	0.009447733	0.008074038	46.83836
	w=2, K=50	0.02707516	0.01068826	30.54362	0.0123547	0.007841711	58.13063
DGP F.2 DGP S.2.2	w=1, K=35	0.0211788	0.006659848	29.98303	0.01400902	0.009780191	45.41252
	w=1, K=50	0.02536098	0.007435542	35.83577	0.01614226	0.009880039	53.82317
	w=2, K=35	0.02152584	0.006914626	30.02373	0.01442173	0.00971369	44.02125
	w=2, K=50	0.02586618	0.007442472	36.81254	0.01637856	0.009929373	55.61419

Notes: State variable X_t is assumed to follow ARCH(1) processes in DGP.S.2.1 and DGP.S.2.2. w is the degree of B-splines and K is the number of knots. Knots are equally spaced on the entire range of the state variable X_t. $w = 1$ is piecewise linear and $w = 2$ is piecewise quadratic B-spline.

Table 3: Comparisons of Different Solution Methods with Known DGP

		Sample Size=250			Sample Size=1000		
		Iterations	Computation Time	MSE	Iterations	Computation Time	MSE
DGP B.1.1	Perturbation	1	0.0278	$1.4485e-08$	1	0.0012	$1.2969e-08$
	Projection	1	6.9115	$1.0012e-09$	1	7.2853	$9.7285e-10$
	Discretization	276	4.7639	$1.9333e-14$	276	4.72553	$1.9331e-14$
	PEA	197	0.1124	0.0382	199	0.2375	0.0023
	Improved PEA	3169	5.7877	$1.9823e-07$	3169	18.5996	$1.9902e-07$
	Penalized B-splines	1	0.36	0.1270	1	0.61	0.0.005992
DGP B.1.2	Perturbation	1	0.0254	62.6267	1	0.0697	26.9855
	Projection	1	6.6360	$1.8408e+03$	1	6.9137	$1.4427e+03$
	Discretization	509	9.2022	$1.8174e-06$	481	9.1459	$8.8945e-06$
	PEA	264	0.2357	186.3197	188	0.100	144.7383
	Improved PEA	4160	7.380514	0.5174	5303	30.5248	11.5722
	Penalized B-splines	1	0.34	60.75646	1	0.45	9.7004

Notes: State variable X_t is assumed to follow AR(1) processes in DGP B.1. In this example, for simplicity, we set $w = 2$ is the degree of B-splines and $K = 50$ is the number of knots. Knots are equally spaced on the entire range of the state variable X_t. Throughout this simulation study, we set $\gamma = 2.5$D, $\beta = 0.96$, $\mu = 0.0179$, $\sigma = 0.0348$, with $\Gamma = -0.179$ and $\Gamma = 0.8$ in DGP B.1.1 and DGP B.1.2 respectively. Given the true analytic solution is simple, we report PEA and Improved PEA based on the following parametric assumptions $E(\cdot|x) = exp(a_1 + a_2x)$ and $f(x) = exp(a_1 + a_2x)$. By increasing the order for PEA and Improved PEA, the number of iterations and amount of computational time increase accordingly. The solution accuracy does not improve very much. The analytic solution is based on the Burnside (1998)'s algorithm. For analytic, discretion, PEA and Improved PEA methods, the tolerance level for convergence in each iteration is set to be $1e-7$. The order for perturbation and projection methods are set to be three. For the perturbation method, increasing the number of series expansions does not necessarily enhance the approximation results Calin et al. (2005). But the projection methods will enjoy better approximation with a larger order, especially when Γ is large. Given a misspecificed DGP, increasing the order for PEA and Improved PEA does not reduce approximation errors.

Table 4: Comparisons of Different Solution Methods with Unknown DGP

		Sample Size=250			Sample Size=5000		
		Iterations	Computation Time	MSE	Iterations	Computation Time	MSE
DGP B.2	Perturbation	1	0.0251	12.2352	1	0.0035	13.0844
	Projection	1	0.7567	11.9826	1	0.7850	12.8628
	Discretization	315	5.5417	11.9744	312	5.8937	12.8682
	PEA (order=1)	251	0.2132	3.2947	243	0.1859	4.8215
	PEA (order=2)	251	0.1618	1.8456	244	0.1790	3.1448
	PEA (order=3)	251	0.1709	1.9172	260	0.1756	3.0675
	Improved PEA (order=1)	3609	6.3269	11.9783	3564	20.5544	12.8721
	Improved PEA (order=2)	3609	16.7417	11.9783	3564	57.2894	12.8721
	Improved PEA (order=3)	3609	26.3455	11.9783	3564	98.4927	12.8727
	Penalized B-splines	1	0.49	1.7428	1	0.51	0.5564795

Notes: State variable X_t is assumed to follow AR(1) processes in DGP B.1. In this example, for simplicity, we set $w = 2$ is the degree of B-splines and $K = 50$ is the number of knots. Knots are equally spaced on the entire range of the state variable X_t. Throughout this simulation study, we set $\gamma = 2.5$D, $\beta = 0.96$, $\mu = 0.0179$, $\sigma = 0.0348$, with $\Gamma = -0.179$ and $\Gamma = 0.8$ in DGP B.1.1 and DGP B.1.2 respectively. By increasing the order for PEA and Improved PEA, the number of iterations and amount of computational time increase accordingly. The solution accuracy does not improve very much. The analytic solution is based on the Burnside (1998)'s algorithm. For analytic, discretion, PEA and Improved PEA methods, the tolerance level for convergence in each iteration is set to be $1e-7$. The order for perturbation and projection methods are set to be three. For the perturbation method, increasing the number of series expansions does not necessarily enhance the approximation results Calin et al. (2005). We conducted PEA and Improved PEA under different orders. The PEA algorithm with order=2 has the best approximation performance. Given misspecified DGP, increasing the order for Improved PEA does not reduce approximation errors. Therefore, we report PEA and Improved PEA based on the following parametric assumptions $E(\cdot|x) = exp(a_1 + a_2x + a_3x^2)$ and $f(x) = exp(a_1 + a_2x)$.

Table 5: Predictability on Excess Returns

	Dependent variable: Future Cumulative Excess Returns															
	j=1				j=4				j=12				j=20			
1952-1990	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Observed dy^4	0.311**	0.713***	0.704***	0.882***	0.216**	0.458***	0.460***	0.737***	0.106*	0.111**	0.111**	0.207***	0.135***	0.135***	0.135***	0.176***
	(0.141)	(0.171)	(0.167)	(0.259)	(0.099)	(0.133)	(0.132)	(0.185)	(0.061)	(0.052)	(0.053)	(0.056)	(0.043)	(0.048)	(0.048)	(0.041)
Short Interest Rate		−4.814***	−4.375***	−4.543***		−2.830***	−2.926***	−3.199***		−0.060	−0.031	−0.131		−0.002	−0.003	−0.044
		(1.197)	(1.219)	(1.159)		(0.981)	(0.993)	(0.930)		(0.484)	(0.585)	(0.549)		(0.405)	(0.485)	(0.462)
Implied Dividend Yield			0.788	1.007			−0.178	0.192			0.052	0.182			−0.0002	0.056
			(0.898)	(0.887)			(0.515)	(0.505)			(0.289)	(0.295)			(0.164)	(0.171)
Asymmetric Effect				−0.341				−0.531**				−0.189***				−0.084**
				(0.405)				(0.215)				(0.054)				(0.035)
Constant	1.034**	2.682***	12.020	15.263	0.725**	1.711***	−0.407	4.993	0.357*	0.378**	0.995	2.889	0.449***	0.450**	0.448	1.261
	(0.467)	(0.627)	(10.742)	(10.806)	(0.325)	(0.490)	(6.267)	(6.418)	(0.200)	(0.182)	(3.357)	(3.467)	(0.142)	(0.175)	(1.898)	(1.991)
Observations	115	115	115	115	112	112	112	112	104	104	104	104	96	96	96	96
R^2	0.044	0.139	0.144	0.149	0.091	0.228	0.229	0.284	0.124	0.124	0.125	0.162	0.346	0.346	0.346	0.358
Adjusted R^2	0.036	0.123	0.121	0.119	0.083	0.214	0.208	0.257	0.115	0.107	0.098	0.129	0.339	0.332	0.324	0.329
1990-2015																
Observed Dividend Yields	0.226**	0.224**	0.219**	0.378*	0.229***	0.223***	0.221***	0.374***	0.222***	0.211***	0.211***	0.301***	0.209***	0.204***	0.204***	0.230***
	(0.108)	(0.105)	(0.102)	(0.194)	(0.088)	(0.081)	(0.080)	(0.112)	(0.067)	(0.056)	(0.055)	(0.077)	(0.031)	(0.032)	(0.032)	(0.049)
Short Interest Rate		−0.317	0.081	0.451		−1.027	−0.896	−0.576		−1.772*	−1.893*	−1.787*		−0.759*	−0.695	−0.681
		(1.282)	(1.257)	(1.156)		(1.023)	(1.062)	(0.986)		(1.019)	(1.020)	(0.965)		(0.438)	(0.474)	(0.466)
Implied Dividend Yield			−1.060	−0.963			−0.344	−0.244			0.215	0.294			−0.080	−0.055
			(0.949)	(0.900)			(0.610)	(0.574)			(0.230)	(0.214)			(0.142)	(0.146)
Asymmetric Effect				−0.365				−0.355*				−0.210				−0.067
				(0.331)				(0.216)				(0.133)				(0.071)
Constant	0.945**	0.946**	−11.735	−9.926	0.956***	0.960***	−3.155	−1.329	0.924***	0.936***	3.511	4.824*	0.869***	0.876***	−0.087	0.327
	(0.427)	(0.422)	(11.270)	(10.748)	(0.341)	(0.314)	(7.284)	(6.748)	(0.268)	(0.213)	(2.789)	(2.627)	(0.121)	(0.119)	(1.684)	(1.845)
Observations	99	99	99	99	96	96	96	96	88	88	88	88	80	80	80	80
R^2	0.045	0.046	0.061	0.074	0.146	0.162	0.167	0.205	0.368	0.476	0.480	0.518	0.663	0.694	0.695	0.703
Adjusted R^2	0.036	0.026	0.031	0.034	0.137	0.144	0.140	0.170	0.361	0.463	0.462	0.494	0.658	0.686	0.683	0.688

Note: *p<0.1; **p<0.05; ***p<0.01

Note: This table provides evidence on the predictability of future excess returns by rationalized implied dividend yields and the difference between the observed dividend yields and implied one. Model 1 in this Table corresponds to regressing $\tilde{y}_{t+j} = \beta_0 + \beta_1 F_t + \beta_2 diff_t + \beta_3 r_t + \epsilon_t$. Model 2 corresponds to regressing $\tilde{y}_{t+j} = \beta_0 + \beta_1 dy_t^4 + \beta_2 r_t + \epsilon_t$.Model 3 corresponds to regression $\tilde{y}_{t+j} = \beta_0 + \beta_1 dy_t^4 + \epsilon_t$. $\tilde{y}_{t+j}$ is the annualized excess return. dy^4 is the observed dividend yields from quarterly $S\&P500$ data. f_t is the implied P/D ratios derived from Equation (19), where $\beta = 0.96$ and $\gamma = 2.5$. F is a nonparametric estimation based on implied dividend yields $1/f_t$ and $diff_t$ is the difference between the observed yield and F_t. F_t and $diff_t$ are a pair of orthogonal decomposition of the observed dividend yields dy_t^4. The risk-free rate is based on the 3-month U.S. treasury bond. The value in the parenthesis is the Newey-West t-value with $j+1$ lags. The entries for the model are based on the U.S. Quarterly data from 1962 to 2015.

Table 6: Predictability on Future Dividend Growth

	Dependent variable: Future Dividend Growth															
	j=1				j=4				j=12				j=20			
1962-1990	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Observed Dividend Yields	0.018***	0.014	0.015*	−0.003	0.014*	0.010	0.011	−0.005	0.008	0.014	0.014	0.0005	0.010*	0.018**	0.019***	0.013*
	(0.006)	(0.009)	(0.009)	(0.010)	(0.007)	(0.010)	(0.010)	(0.010)	(0.007)	(0.009)	(0.009)	(0.009)	(0.006)	(0.007)	(0.007)	(0.007)
Short Interest Rate		0.042	0.020	0.035		0.040	0.017	0.032		−0.069	−0.080	−0.067		−0.087**	−0.110***	−0.104***
		(0.056)	(0.056)	(0.052)		(0.066)	(0.068)	(0.064)		(0.056)	(0.058)	(0.055)		(0.041)	(0.039)	(0.039)
Implied Dividend Yield			−0.040	−0.062**			−0.042	−0.063**			−0.019	−0.038***			−0.040***	−0.048***
			(0.029)	(0.031)			(0.026)	(0.025)			(0.013)	(0.013)			(0.012)	(0.012)
Asymmetric Effect				0.033***				0.030**				0.027**				0.011
				(0.012)				(0.012)				(0.013)				(0.010)
Constant	0.073***	0.058*	−0.418	−0.735**	0.060**	0.046	−0.448	−0.755***	0.040*	0.064*	−0.162	−0.432***	0.048**	0.080***	−0.399***	−0.512***
	(0.020)	(0.031)	(0.350)	(0.374)	(0.024)	(0.037)	(0.308)	(0.290)	(0.024)	(0.034)	(0.159)	(0.153)	(0.019)	(0.026)	(0.149)	(0.151)
Observations	114	114	114	114	111	111	111	111	103	103	103	103	95	95	95	95
R^2	0.157	0.166	0.182	0.238	0.108	0.116	0.136	0.187	0.056	0.095	0.102	0.167	0.133	0.216	0.259	0.274
Adjusted R^2	0.150	0.150	0.160	0.210	0.100	0.099	0.112	0.157	0.047	0.077	0.074	0.133	0.124	0.199	0.235	0.242
1990-2016																
Observed dy^4	0.0003	0.001	0.001	0.028**	−0.006	−0.006	−0.006	0.025*	−0.005	−0.006	−0.006	0.005	−0.0005	−0.002	−0.002	0.001
	(0.006)	(0.006)	(0.007)	(0.013)	(0.011)	(0.010)	(0.010)	(0.014)	(0.012)	(0.011)	(0.011)	(0.015)	(0.005)	(0.005)	(0.005)	(0.006)
Short Interest Rate		0.046	−0.008	0.056		0.009	−0.033	0.030		−0.122	−0.150	−0.137		−0.222**	−0.224**	−0.222**
		(0.169)	(0.146)	(0.139)		(0.216)	(0.202)	(0.190)		(0.177)	(0.173)	(0.160)		(0.109)	(0.111)	(0.106)
Implied Dividend Yield			0.142	0.158*			0.111	0.130			0.049	0.058			0.002	0.005
			(0.091)	(0.091)			(0.080)	(0.081)			(0.053)	(0.055)			(0.016)	(0.015)
Asymmetric Effect				−0.063***				−0.070**				−0.025				−0.008
				(0.023)				(0.027)				(0.017)				(0.007)
Constant	0.014	0.013	1.710	2.020*	−0.013	−0.013	1.310	1.670*	−0.009	−0.008	0.576	0.733	0.009	0.011	0.037	0.090
	(0.025)	(0.026)	(1.079)	(1.096)	(0.042)	(0.043)	(0.940)	(0.966)	(0.046)	(0.046)	(0.619)	(0.629)	(0.018)	(0.019)	(0.181)	(0.171)
Observations	99	99	99	99	96	96	96	96	88	88	88	88	80	80	80	80
R^2	0.00001	0.003	0.066	0.157	0.010	0.010	0.053	0.182	0.013	0.047	0.063	0.098	0.0004	0.306	0.306	0.322
Adjusted R^2	-0.010	-0.018	0.037	0.122	-0.001	-0.012	0.023	0.146	0.002	0.025	0.030	0.055	-0.012	0.288	0.279	0.286

Note: *p<0.1; **p<0.05; ***p<0.01

Note: This table provides evidence on the predictability of future excess returns by rationalized implied dividend yields and the difference between the observed dividend yields and implied one. Model 1 in this Table corresponds to regressing $d_{t+j} = \beta_0 + \beta_1 F_t + \beta_2 diff_t + \beta_3 r_t + \epsilon_t$. Model 2 corresponds to regressing $d_{t+j} = \beta_0 + \beta_1 dy_t^4 + \beta_2 r_t + \epsilon_t$.Model 3 corresponds to regression $dy_{t+j} = \beta_0 + \beta_1 dy_t^4 + \epsilon_t$. dy_{t+j} is the annualized dividend growth. dy^4 is the observed dividend yields from quarterly $S\&P500$ data. f_t is the implied P/D ratios derived from Equation (19), where $\beta = 0.96$ and $\gamma = 2.5$. F is a nonparametric estimation based on implied dividend yields $1/f_t$ and $diff_t$ is the difference between the observed yield and F_t. F_t and $diff_t$ are a pair of orthogonal decomposition of the observed dividend yields dy_t^4. The risk-free rate is based on the 3-month U.S. treasury bond. The value in the parenthesis is the Newey-West t-value with $j+1$ lags. The entries for the model are based on the U.S. Quarterly data from 1962 to 2015.

Fig. 1. Simulation Study of DGP S.1.1 and F.1, Sample Size=250 and 500, Knots=50

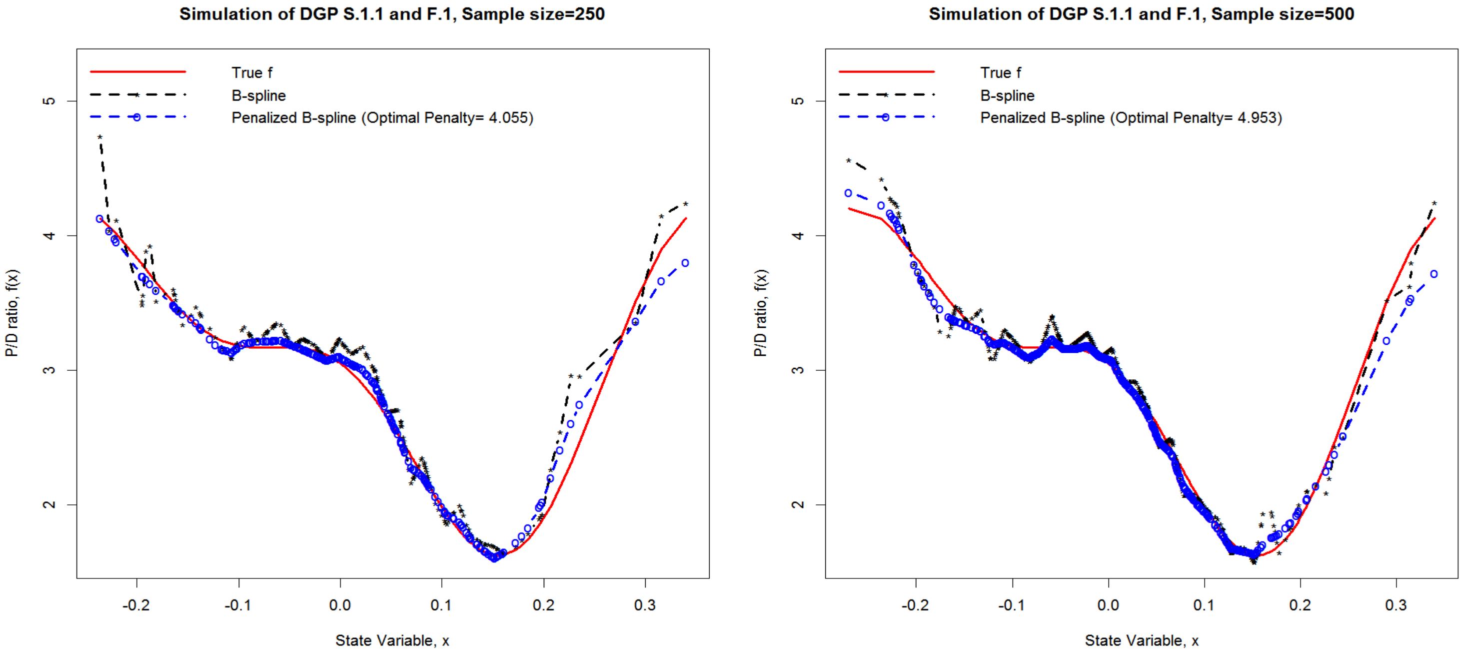


Fig. 2. Simulation Study of DGP S.1.2 and F.1, Sample Size=250 and 500, Knots=50

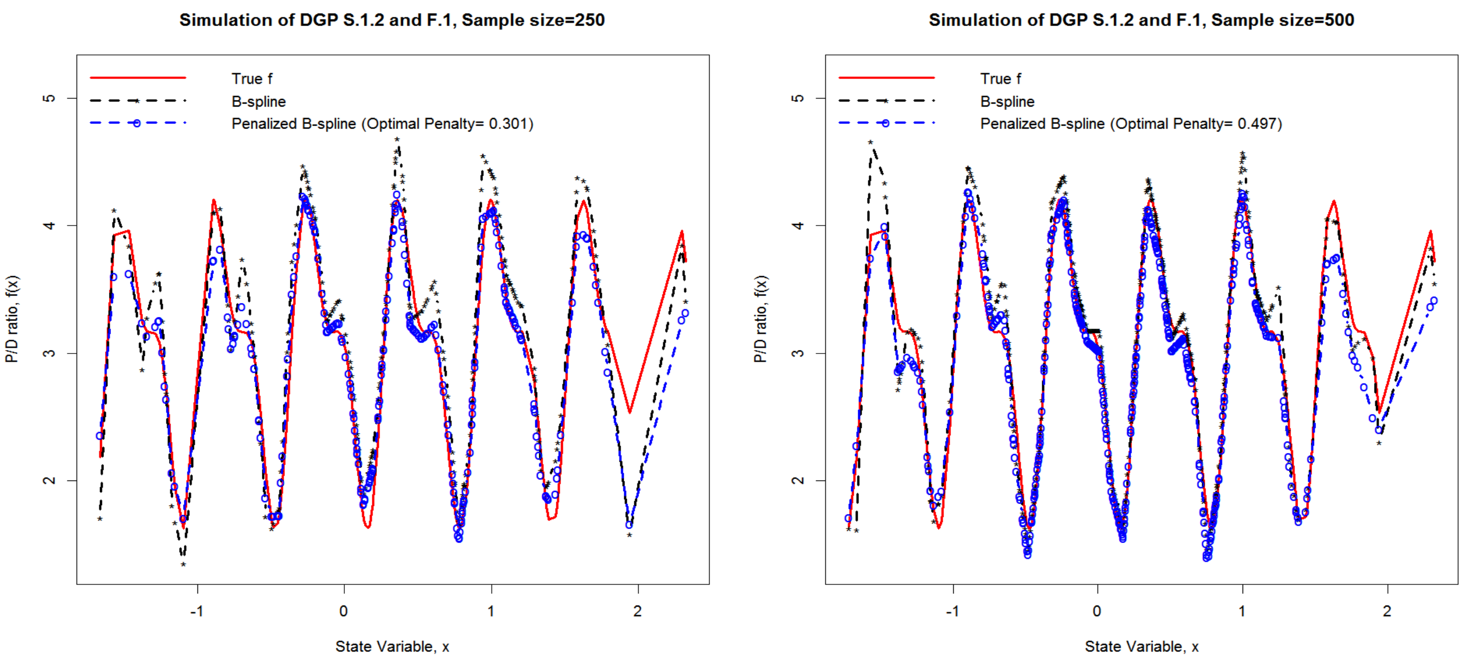


Fig. 3. Simulation Study of DGP S.1.1 and F.1, Sample Size=250 and 500, Knots=50

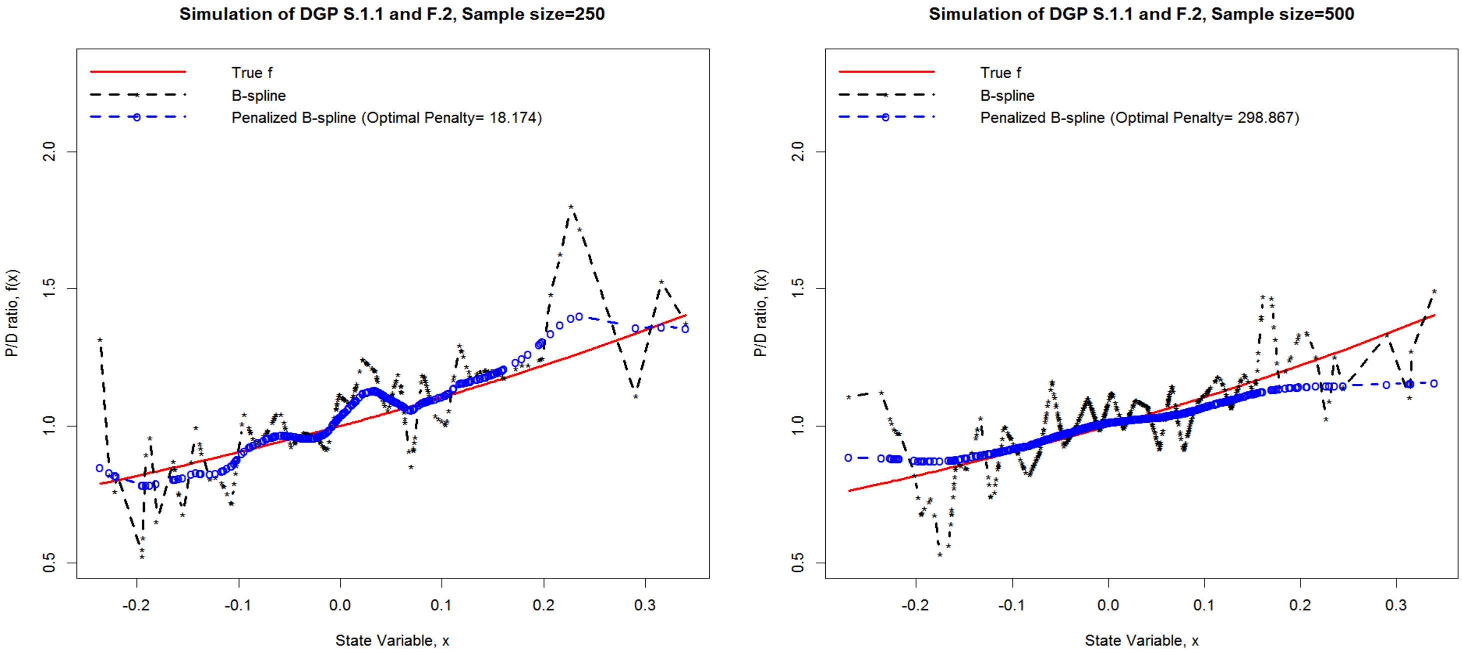


Fig. 4. Simulation Study of DGP S.1.2 and F.1, Sample Size=250 and 500, Knots=50

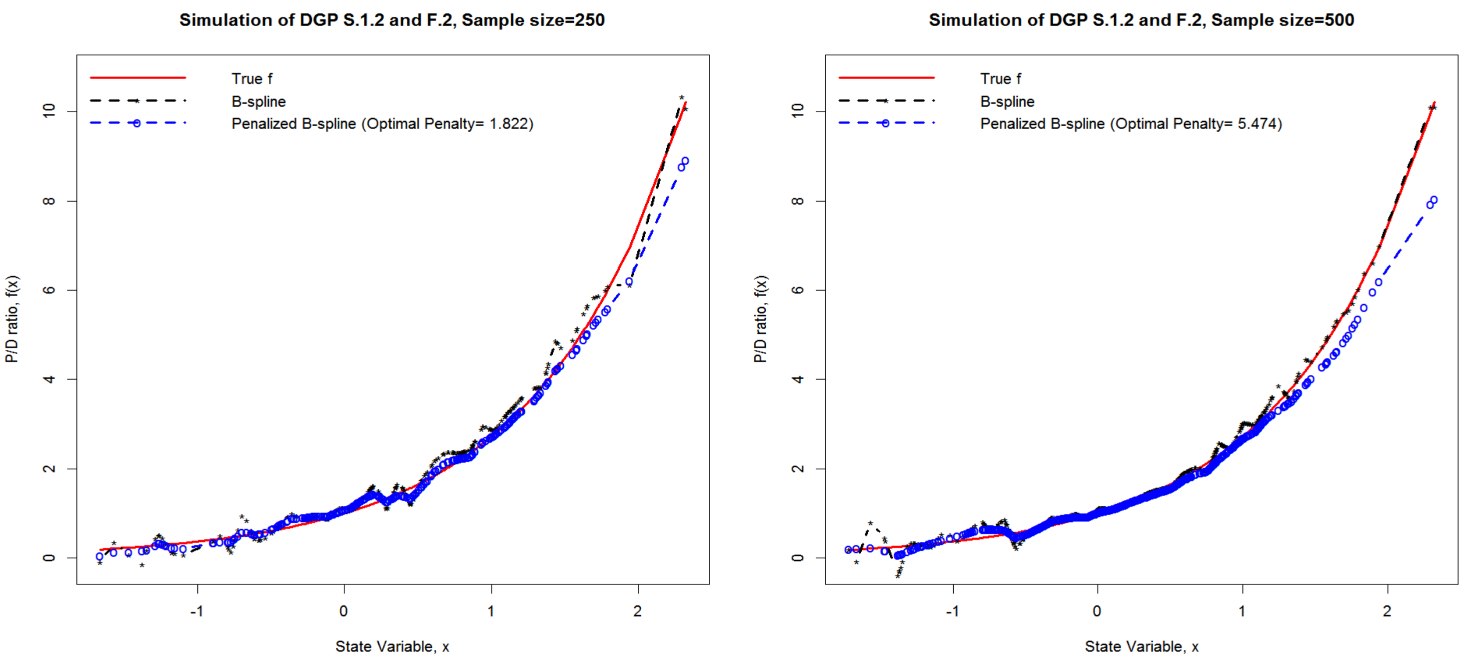


Fig. 5. Simulation Study of DGP S.2.1 and F.1, Sample Size=250 and 500, Knots=50

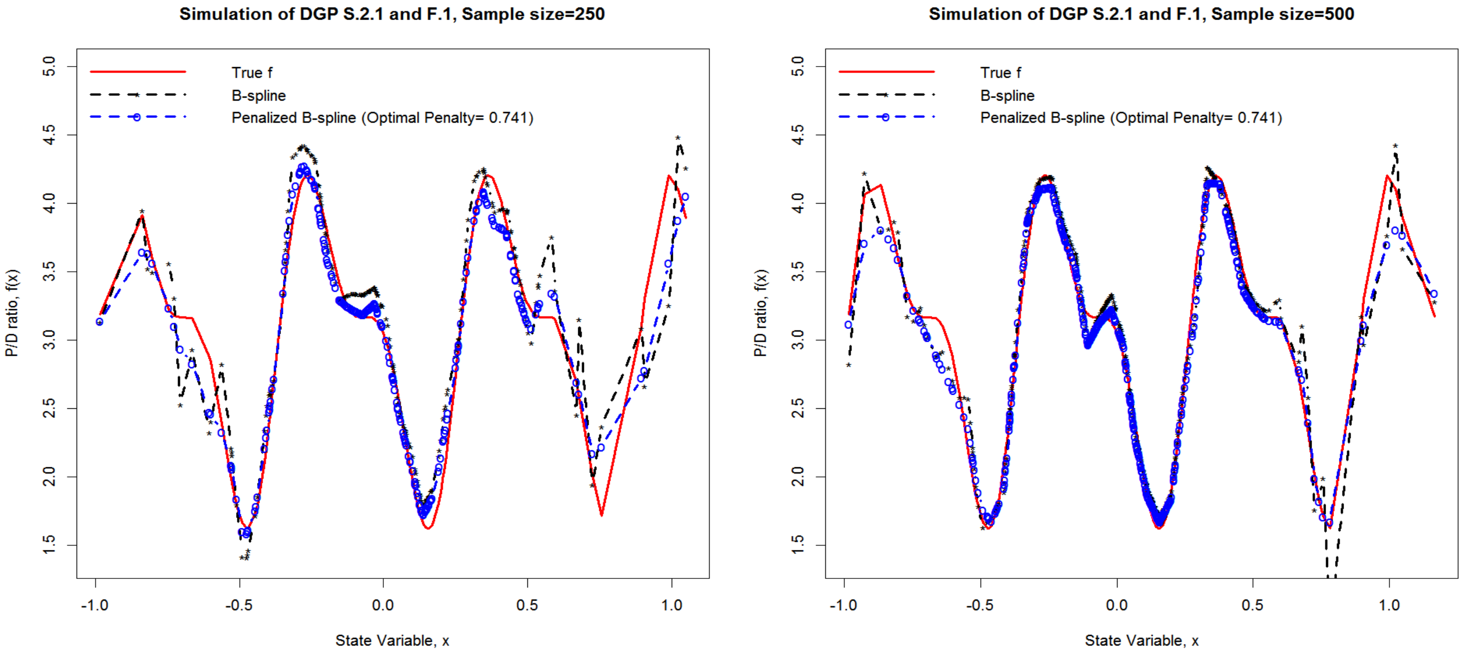


Fig. 6. Simulation Study of DGP S.2.2 and F.1, Sample Size=250 and 500, Knots=50

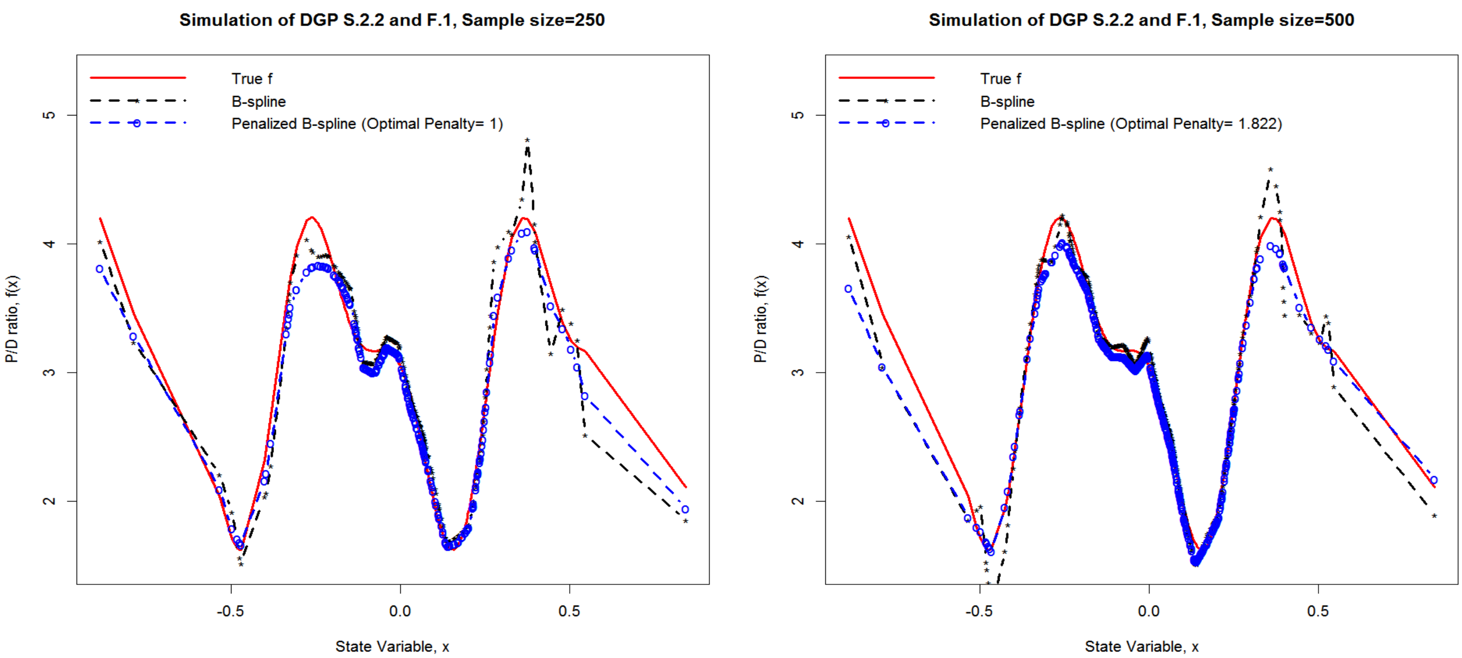


Fig. 7. Simulation Study of DGP S.2.1 and F.2, Sample Size=250 and 500, Knots=50

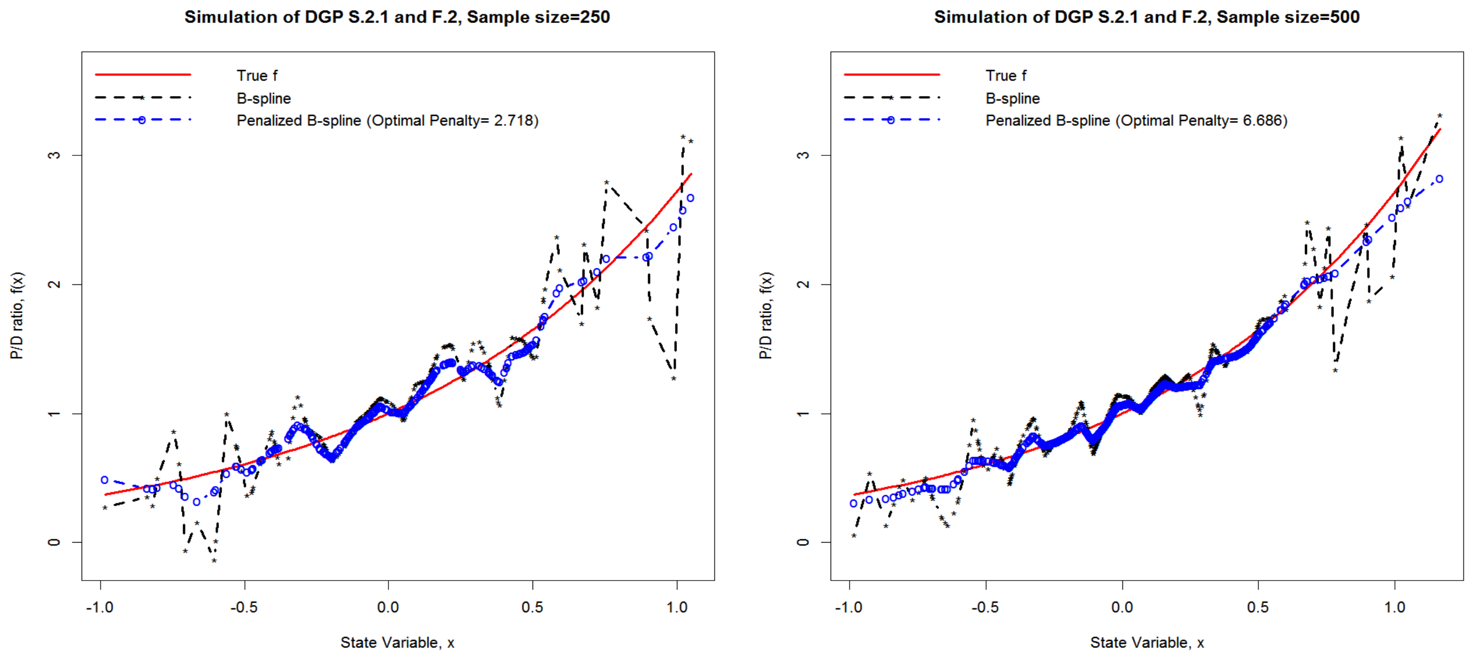


Fig. 8. Simulation Study of DGP S.2.2 and F.2, Size=250 and 500, Knots=50

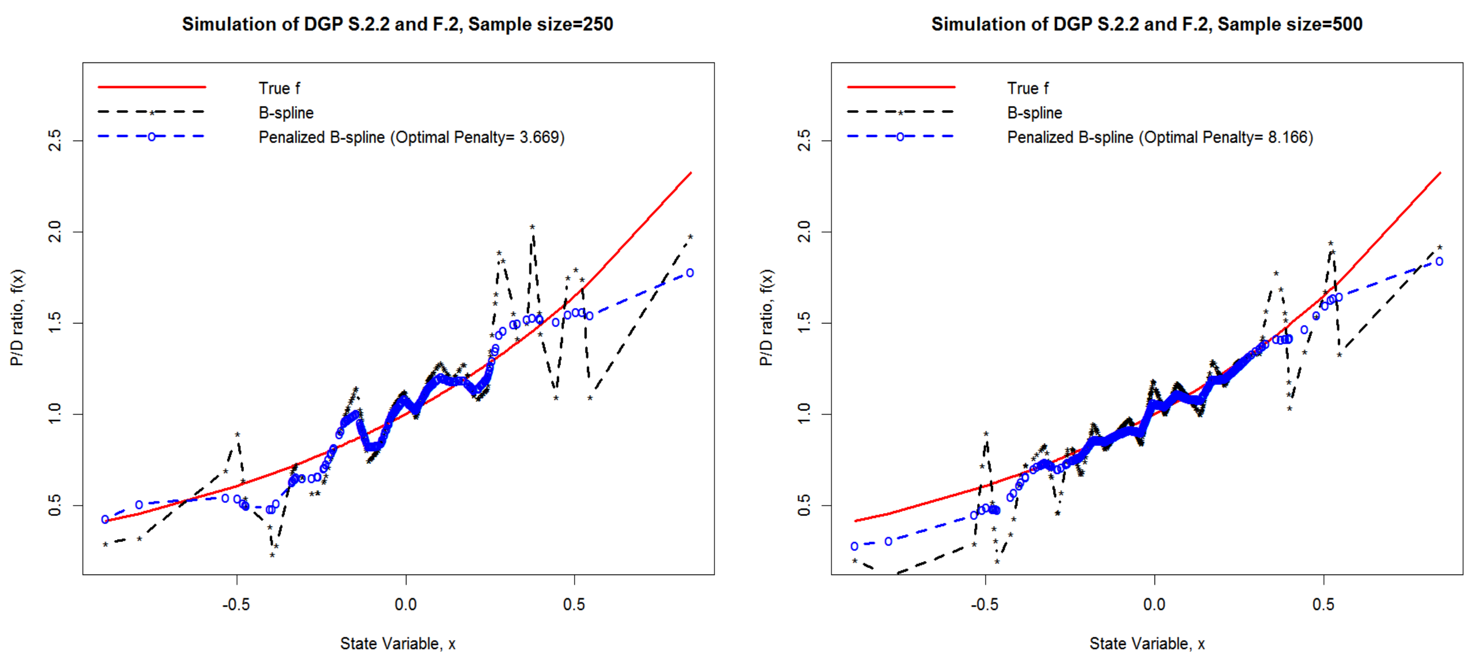


Fig. 9. Simulation Study with Known DGP B.1.1 ($\Gamma = -0.139$), Sample Size=250 (left) and 1000 (right)

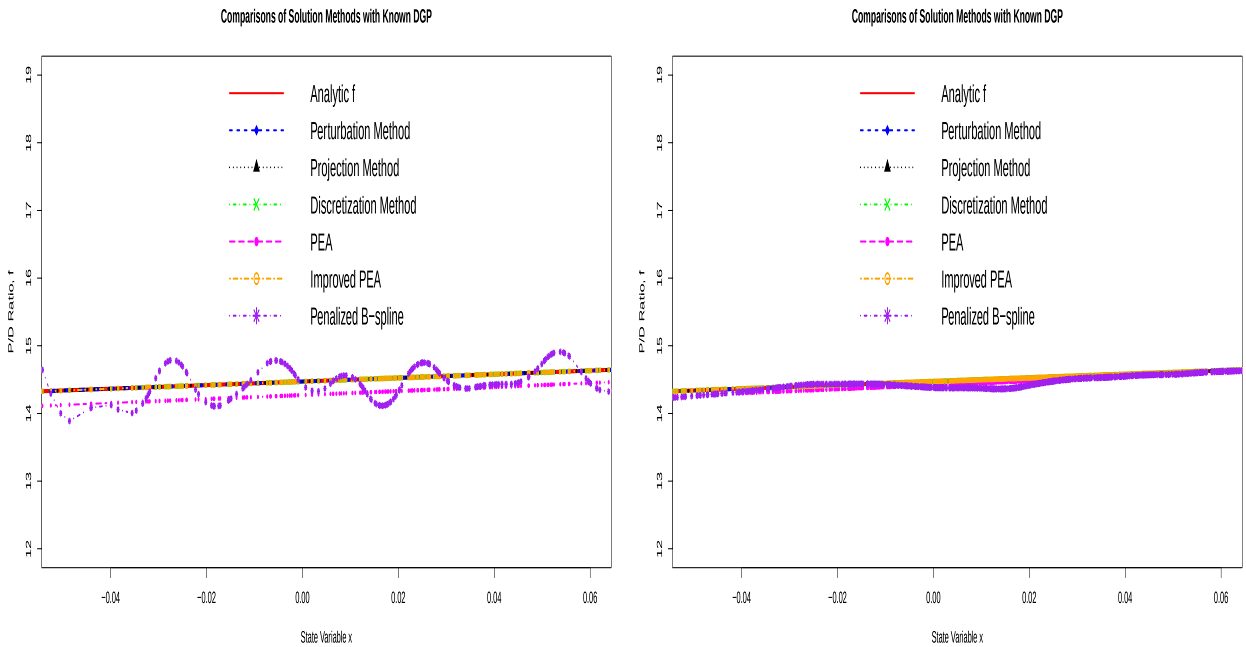


Fig. 10. Simulation Study with Known DGP B.1.2 ($\Gamma = 0.8$), Sample Size=250 (left) and 1000 (right)

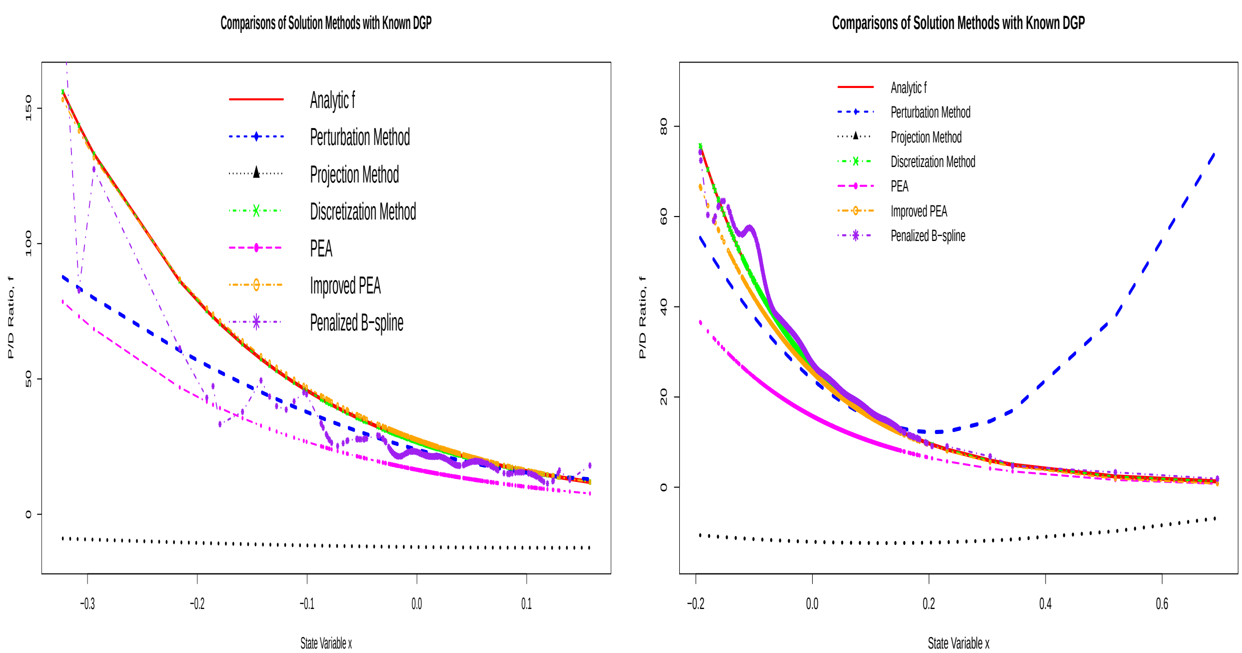


Fig. 11. Simulation Study with Misspecified DGP B.2, Sample Size=250 (left) and 1000 (right)

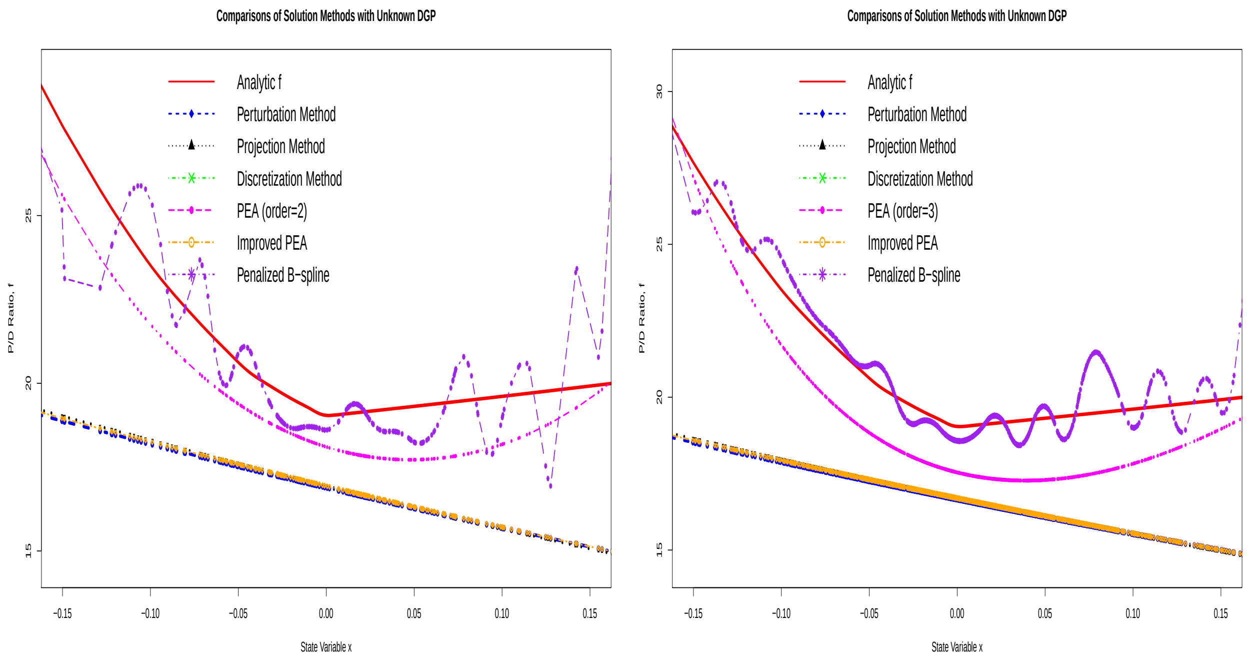


Fig. 12. Real $S\&P500$ Composite Stock Price Index and Measures of Present Value Prices

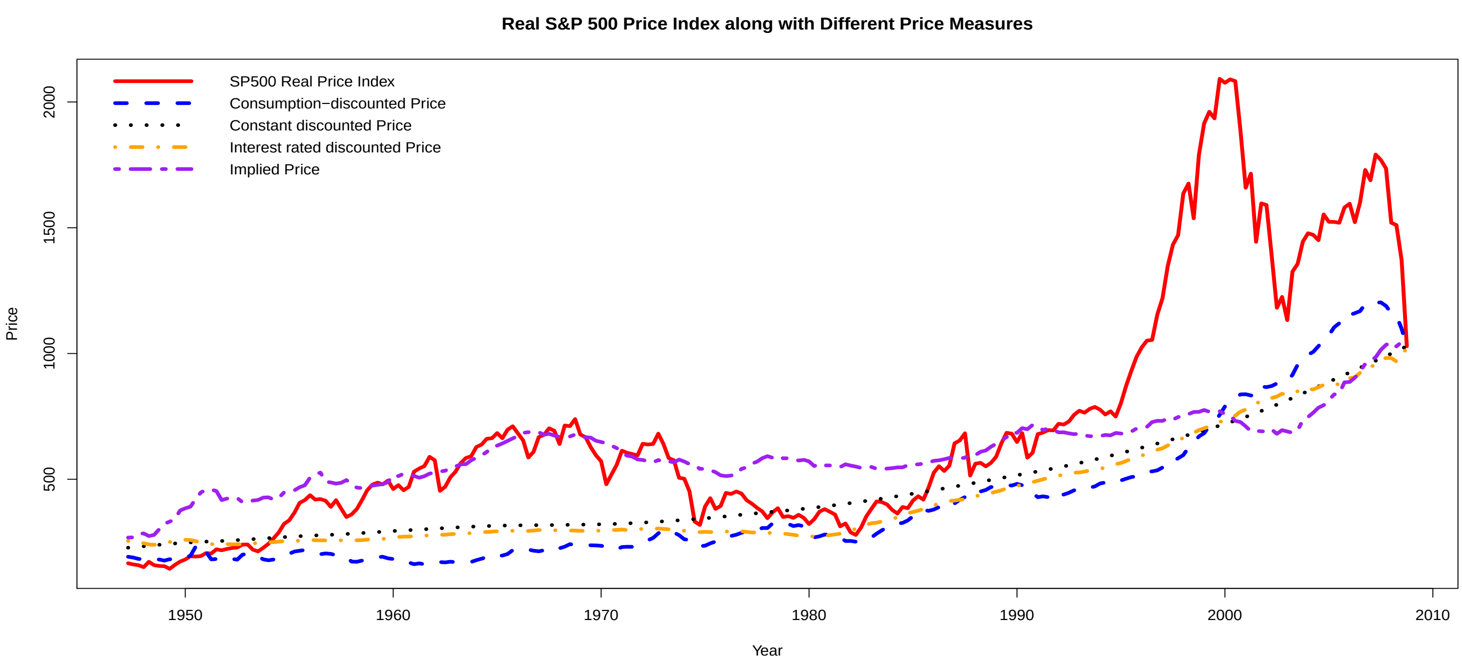


Fig. 13. Predictability Analysis: 1962-1990 Nominal Returns

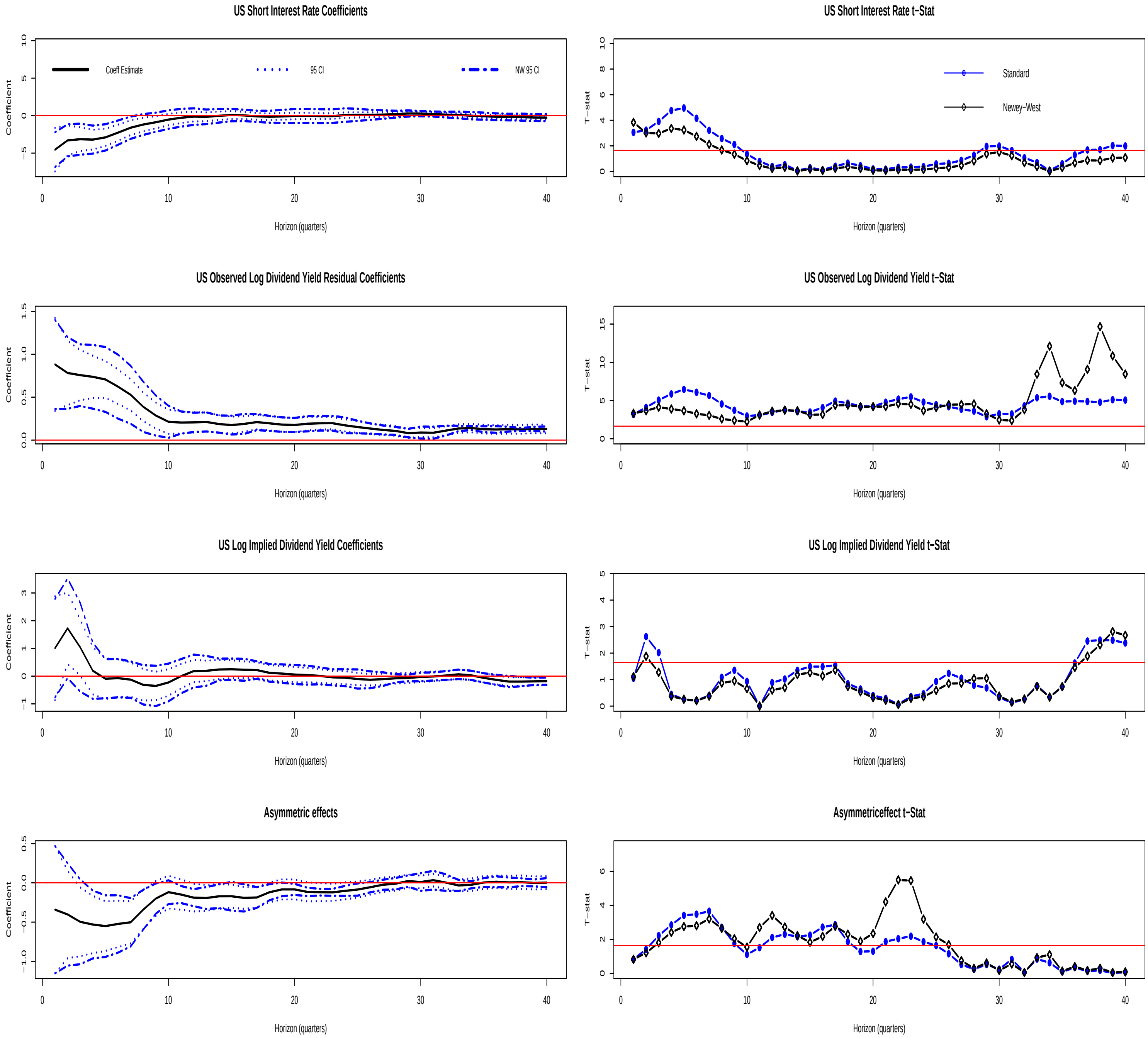


Fig. 14. Predictability Analysis: 1990-2015 Nominal Returns

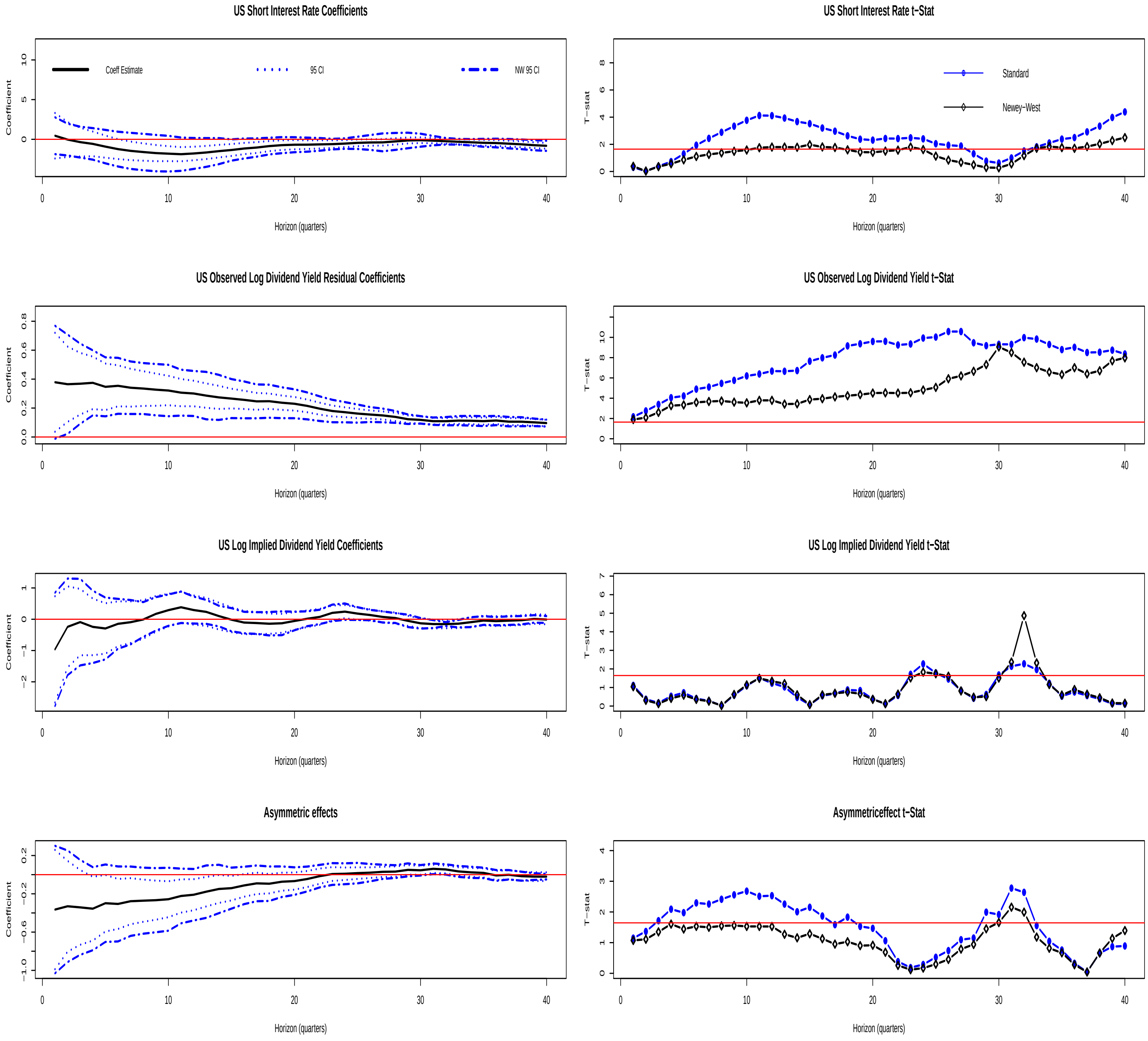


Fig. 15. Predictability Analysis: 1962-1990 Dividend Growth

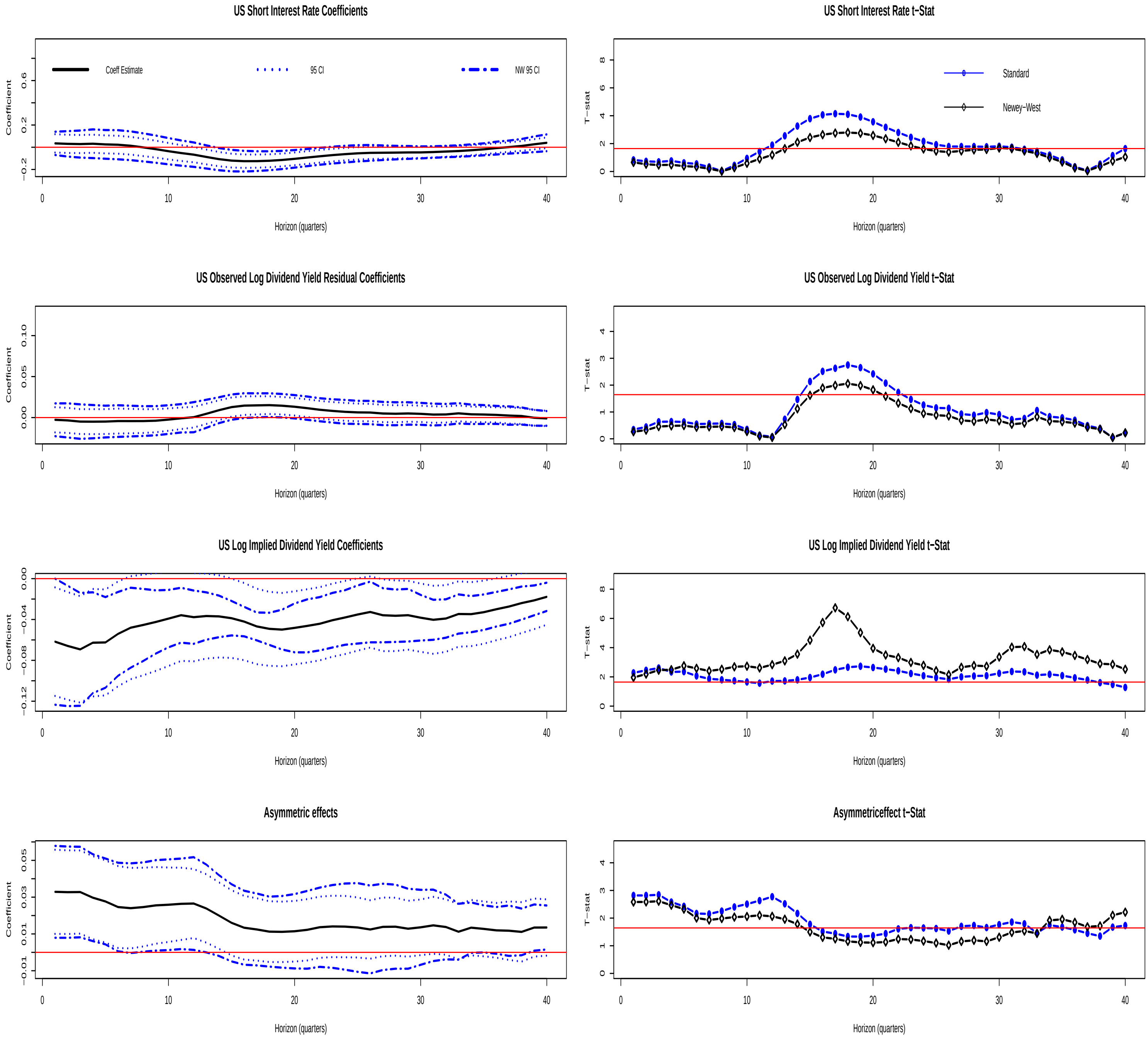
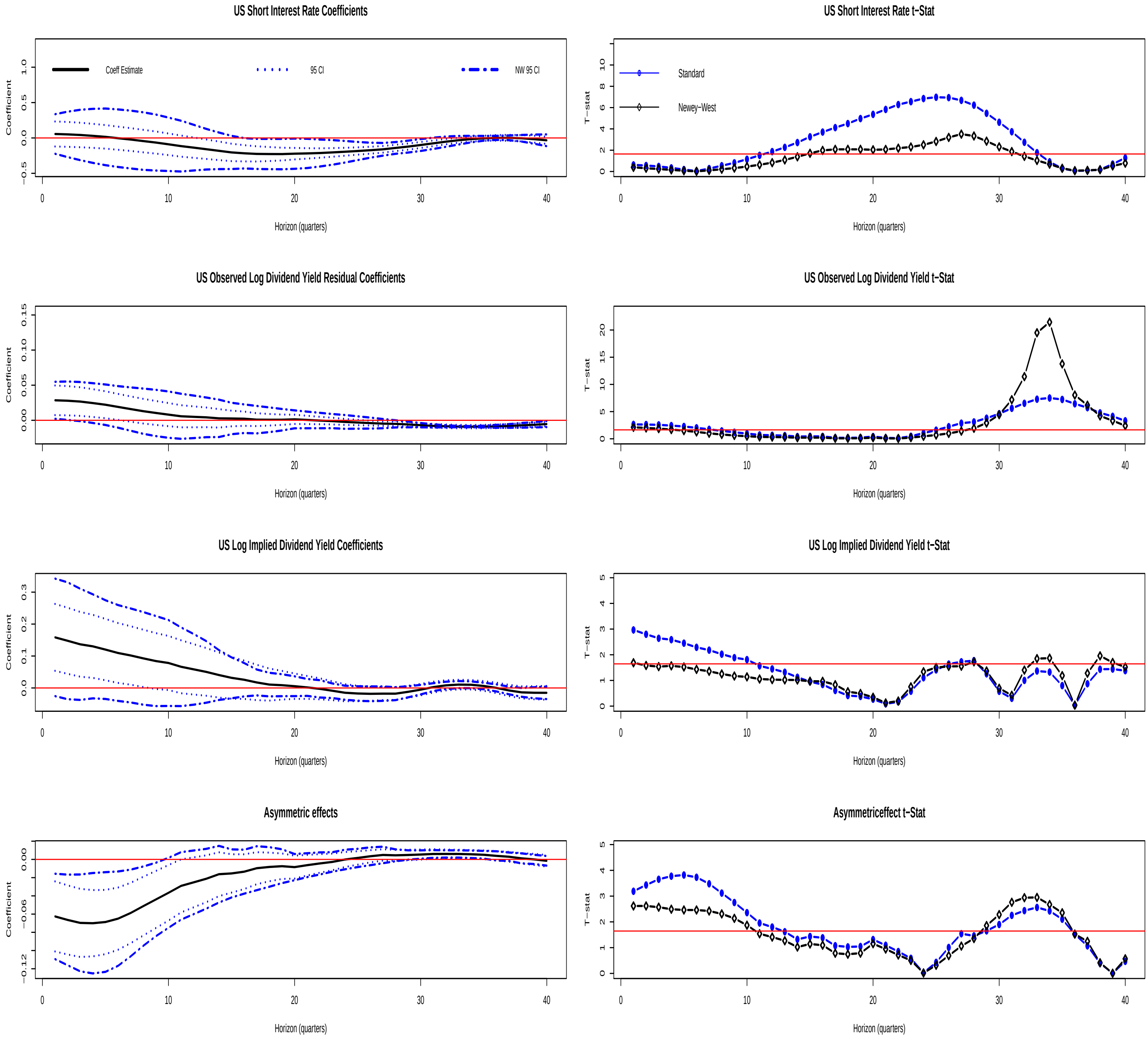


Fig. 16. Predictability Analysis: 1990-2015 Dividend Growth



Appendix

6.1. Appendix A: Mathematical Proofs

Lemma 6.1. *Since $f \in L^p$. Then there exists a continuous function $\bar{f}$ whose support lies in a bounded interval $[-A, A]$ so that*

$$||f - \bar{f}||_p < \epsilon.$$

Proof of Lemma 6.1 This is immediate from the Stone-Weierstrass theorem. $\square$

Lemma 6.2. *Suppose Assumption 3.4-3.6 hold. Then*

(a) $||\frac{\Phi'\Phi}{M} - E(\frac{\Phi'\Phi}{M})|| = O_p(M^{-1/2})$;

(b) $||\frac{\Psi'\Phi}{M} - E(\frac{\Psi'\Phi}{M})|| = O_p(M^{-1/2})$;

(c) $\lambda_{\min}[\frac{\Phi'\Phi}{M}] \to \lambda_{\min}[E(\frac{\Phi'\Phi}{M})]\, a.s.$;

(d) $\lambda_{\min}(\frac{\Phi'\Phi}{M}) > 0\, a.s.$;

(e) $\lambda_{\max}(\frac{\Phi'\Phi}{M}) < \infty\, a.s.$

Proof of Lemma 6.2 Since the proof of (b) is analogous to (a), we only prove (a) here. We have

$$E[||\frac{\Phi'\Phi}{M} - E(\frac{\Phi'\Phi}{M})||^2] = \sum_{j=1}^{K+p}\sum_{i=1}^{K+p} E\{\sum_{l=0}^{T-1} \phi_i(X_l)\phi_j(X_l)/M - E[\sum_{l=0}^{T-1} \phi_i(X_l)\phi_j(X_l)/M]\}^2$$

$$\leq \sum_{j=1}^{K+p}\sum_{i=1}^{K+p}\{\frac{T}{M^2}\sup_{X_t \in X} var[\phi_i(X_t)\phi_j(X_t)] + 2\sum_{0<k<m<T-1} cov[\phi_i(X_k)\phi_j(X_k), \phi_i(X_m)\phi_j(X_m)]/M^2\}$$

$$= A_1 + A_2, \text{ say.}$$

Because $KE||\phi_i(X_t)||_{4+\delta} < \Delta < \infty$ for some $\delta > 0$ by assumption, we have

$$A_1 \leq \frac{K}{M}\sum_{j=1}^{K+p}\sum_{i=1}^{K+p}\frac{E[K\phi_i(X_t)\phi_j(X_t)]^2}{K^2} = O(\frac{K}{M}\frac{K}{K^2}) = O(\frac{1}{M}).$$

Given that $\{X_t\}$ is α mixing with coefficients $\alpha(j)$, by using the Davydov inequality and the condition on $KE||\phi_i(X_t)||_{4+\delta} < \Delta < \infty$ for some $\delta > 0$, we have

$$
\begin{aligned}
A_2 &= \frac{1}{M^2}\sum_{j=1}^{K+p}\sum_{i=1}^{K+p}\sum_{\tau=1}^{T-1}(1-\frac{\tau}{T})cov[\phi_i(X_t)\phi_j(X_t), \phi_i(X_{t+\tau})\phi_j(X_{t+\tau})] \\
&\leq \frac{2^{(4+2\delta)/(4+\delta)}(4+\delta)/\delta}{T^2}\sum_{j=1}^{K+p}\sum_{i=1}^{K+p}\sum_{\tau=1}^{T-1}|1-\frac{\tau}{T}|\alpha(\tau)^{\frac{\delta}{4+\delta}}\{E|K\phi_i(X_t)\phi_j(X_t)|^{2+\delta/2}\}^{\frac{4}{4+\delta}} = O(\frac{1}{M}).
\end{aligned}
$$

It follows that,

$$
||\frac{\Phi'\Phi}{M} - E(\frac{\Phi'\Phi}{M})|| = O_p(\frac{1}{\sqrt{M}}) = o_p(1).
$$

To obtain an almost sure convergence result for (c), we first establish a similar result under the convergence in probability. Similar to part (a), using the Markov, Cauchy-Schwarz and Hölder's inequalities, we have

$$
\begin{aligned}
&P[|\lambda_{\min}(\frac{1}{M}\sum_{t=0}^{T-1}\Phi_{qt}\Phi'_{qt}) - \lambda_{\min}(\frac{1}{M}\sum_{t=0}^{T-1}E\Phi_{qt}\Phi'_{qt})| > \epsilon] \\
&\leq P\{\sum_{i=1}^{K+p}\sum_{j=1}^{K+p}|\frac{1}{M}\sum_{t=0}^{T-1}[\phi_i(X_t)\phi_j(X_t) - E\phi_i(X_t)\phi_j(X_t)]| > \epsilon\} \\
&\leq \frac{1}{\epsilon}\sum_{i=1}^{K+p}\sum_{j=1}^{K+p}E|\frac{1}{M}\sum_{t=0}^{T-1}[\phi_i(X_t)\phi_j(X_t) - E\phi_i(X_t)\phi_j(X_t)]| \\
&\leq \frac{1}{\epsilon}\sum_{i=1}^{K+p}\sum_{j=1}^{K+p}\{\frac{T}{M^2}\{\sup_{X\in X} var[\phi_i(X)\phi_j(X)] + 2\sum_{0<k<m<T-1} cov[\phi_i(X_k)\phi_j(X_k), \phi_i(X_m)\phi_j(X_m)]/M^2\}\}^{\frac{1}{2}} \\
&= o(1)
\end{aligned}
$$

Thus we have proved the convergence in probability for part (c). Conclusions under convergence in probability for (d) and (e) follow analogously. The proof of almost sure convergence follows Andrews (1991). □

Lemma 6.3. *For K and T satisfying Assumption 3.5,*

(a) $\{\phi_{i,t}\psi_{j,\tau}\}$, $\{\Phi_{p,t}\varepsilon_{t+1}\}$, $\{\Psi_{qt}\varepsilon_{t+1}\}$ and $\{\Phi_{p,t}\Psi_{p,t}\}$ are α- mixing sequences with coefficients

$\alpha(j)$;

(b) $KE|\phi_{i,t}\psi_{j,t}|^{2r'} < \Delta' < \infty$ for $r' = r + \delta > 1$,

Proof of Lemma 6.3 $\{\phi_{i,t}\}$ is a measurable function of X_t. Because $\{X_t\}$ is assumed to be an α-mixing process of size $r/(r-1)$, $\{\phi_{i,t}\}$ is also an α-mixing process of size $r/(r-1)$ using the Thereom 3.49 of White (1996). Similarly, $\{\psi_{i,t}\}$ is also an α-mixing process with size $r/(r-1)$. Immediately, from proposition 3.50 of White (1996), $\{\phi_{i,t}\phi_{j,\tau}\}$, $\{\phi_{i,t}\psi_{j,\tau}\}$, $\{\Phi_{p,t}\varepsilon_{t+1}\}$ and $\{\Psi_{q,t}\varepsilon_{t+1}\}$ are mixing sequences of size $r/(r-1)$.

For part (b), it immediately follows from the definition of $\psi_{i,t}$ and Minkowski's inequality that

$$\begin{aligned}
&KE|\phi_{i,t}\psi_{j,t}|^{2+\delta/2} = KE|\phi_{i,t}\varphi_{j,t} - \phi_{i,t}\varphi_{j,t+1}m(X_{t+1})|^{2+\delta/2} \\
&\leq K\{[E|\phi_{i,t}\varphi_{j,t}|^{2+\delta/2}]^{\frac{1}{2+\delta/2}} + [E|\phi_{i,t}\varphi_{j,t+1}m(X_{t+1})|^{2+\delta/2}]^{\frac{1}{2+\delta/2}}\}^{2+\delta/2} \\
&\leq K\{[E|\phi_{i,t}^{4+\delta}|E|\varphi_{i,t}^{4+\delta}|]^{\frac{1}{4+\delta}} + [E|\phi_{i,t}|^{4+\delta}E|\varphi_{j,t}m(X_{t+1})|^{4+\delta}]^{\frac{1}{4+\delta}}\}^{2+\delta/2} < \Delta < \infty.
\end{aligned}$$

□

Lemma 6.4. *Define $G = E(\frac{\Psi'\Phi}{M})[E(\frac{\Phi'\Phi}{M})]^{-1}E(\frac{\Phi\Psi'}{M})$ and $G_T = \frac{\hat{\Psi}'\hat{\Psi}}{M} = \frac{1}{M}\Psi'\Phi(\Phi'\Phi)^{-1}\Phi\Psi'$. Suppose Assumptions 3.5-3.7 hold. Then*

(a) $||G_T - G|| = O_p(\frac{1}{\sqrt{M}})$;

(b) $\lambda_{\max}(G_T) = \lambda_{\max}(G) + O_p(\frac{1}{\sqrt{M}})$;

(c) $\lambda_{\min}(G_T) \geq \frac{1}{2}\lambda_{\min}(G)$ with probability approaching 1 as $T \to \infty$.

Proof of Lemma 6.4 For part (a), using the triangular inequality, we have

$$\begin{aligned}
&||G_T - G_|| \leq ||(\frac{\Psi'\Phi}{M} - E\frac{\Psi'\Phi}{M})(\frac{\Phi'\Phi}{M})^{-1}\frac{\Phi\Psi'}{M}|| + ||E\frac{\Psi'\Phi}{M}[(\frac{\Phi'\Phi}{M})^{-1} - (E\frac{\Phi'\Phi}{M})^{-1}]\frac{\Phi'\Psi}{M}|| \\
&+ ||E\frac{\Psi'\Phi}{M}(E\frac{\Phi'\Phi}{M})^{-1}[\frac{\Phi'\Psi}{M} - E\frac{\Phi'\Psi}{M}]|| = A_3 + A_4 + A_5, \text{ say.}
\end{aligned}$$

Using the results from Lemma 6.3, we have

$$A_3 \leq [\lambda_{\min}(\frac{\Phi'\Phi}{M})]^{-1}\lambda_{\max}(\frac{\Psi'\Phi}{M})||\frac{\Psi'\Phi}{M} - E\frac{\Psi'\Phi}{M}|| = O_p(\frac{1}{\sqrt{M}}).$$

Then, we show that $||(\frac{\Phi'\Phi}{M})^{-1} - (E\frac{\Phi'\Phi}{M})^{-1}|| \leq [\lambda_{\min}(\frac{\Phi'\Phi}{M})]^{-1}[\lambda_{\min}(E\frac{\Phi'\Phi}{M})]^{-1}||\frac{\Phi'\Phi}{M} - E\frac{\Phi'\Phi}{M}|| = O_p(\frac{1}{\sqrt{M}})$. Thus

$$A_4 \leq \lambda_{\max}(E\frac{\Psi'\Phi}{M})\lambda_{\max}(\frac{\Psi'\Phi}{M})||(\frac{\Phi'\Phi}{M})^{-1} - (E\frac{\Phi'\Phi}{M})^{-1}|| = O_p(\frac{1}{\sqrt{M}}).$$

Therefore, the last term

$$A_5 \leq \lambda_{\max}(E\frac{\Psi'\Phi}{M})(\lambda_{\min}E\frac{\Phi'\Phi}{M})^{-1}||\frac{\Phi'\Psi}{M} - E\frac{\Phi'\Psi}{M}|| = O_p(\frac{1}{\sqrt{M}}).$$

It follows that $||G_T - G|| = O_p(\frac{1}{\sqrt{M}})$.

Now we prove part (b):

$$\lambda_{\max}(G_T) = \lambda_{\max}(G + G_T - G) = \lambda_{\max}(G) + ||G_T - G|| = \lambda_{\max}(G) + O_p(\frac{1}{\sqrt{M}}).$$

Next, we prove part (c). Similarly, we have

$$\lambda_{\min}(G_T) \geq \lambda_{\min}(G) - ||G_T - G|| \geq \lambda_{\min}(G) - O_p(\frac{1}{\sqrt{M}}) \geq \frac{1}{2}\lambda_{\min}(G).$$

□

Lemma 6.5. *Suppose Assumptions 3.4-3.6 hold. Then there exists $c_0 > 0$ so that*

(a) $\lambda_{\min}G \geq c_0 > 0$*;*

(b) $\lambda_{\min}G_T \geq \frac{c_0}{2} > 0\, a.s.$*;*

(c) $|||G_T^{-1}(\frac{\Psi'\Phi}{M})(\frac{\Phi'\Phi}{M})^{-1} - G^{-1}E\frac{\Psi'\Phi}{M}(E\frac{\Phi'\Phi}{M})^{-1}|| = O_p(\frac{1}{\sqrt{M}})$

Proof of Lemma 6.5 We first prove part (a). Denote a lead of matrix Φ as $\Phi_a = \sum_{t=1}^{T}\Phi_t\Phi_t'$ and a diagonal matrix $M = diag\{m_1, \cdots, m_T\}$. Recall the underlying structure of the asset pricing theory and our model construction. It is helpful to express $\Psi_t = \Phi_t - m_t\Phi_{t-1}$. Under Assumption 3.5, $E(y_t^2) < \infty$. Let $c, b \in R^p$ so that $c'(E\frac{\Psi'\Phi}{M})c = \lambda_{\min}(E\frac{\Psi'\Phi}{M})$, and

$b'(E\frac{\Psi'\Phi}{M})b = \lambda_{\max}(E\frac{\Psi'\Phi}{M})$. Applying the Cauchy-Schwarz inequality and Lemma 6.3, we have

$$\begin{aligned}
\lambda_{\max}E(\frac{\Psi'\Phi}{M}) &= b'E(\frac{\Phi'\Phi}{M})b - b'E(\sum_{t=0}^{T-1}\frac{\Phi_{t+1}\Phi_t' y_{t+1}}{M})b \le \lambda_{\max}E(\frac{\Phi'\Phi}{M}) + |b'E[\sup_{1\le t\le T}|y_t|\frac{\Phi_a'\Phi}{M}]b| \\
&\le \lambda_{\max}E(\frac{\Phi'\Phi}{M}) + (E\sup_{1\le t\le T} y_t^2)^{\frac{1}{2}} b'[\frac{1}{M^2}E(\Phi_a'\Phi\Phi'\Phi_a)]^{\frac{1}{2}} b \\
&= \lambda_{\max}E(\frac{\Phi'\Phi}{M}) + (E\sup_{1\le t\le T} m_t^2)^{\frac{1}{2}} b'[\frac{1}{M}E(\Phi_a'\Phi(\Phi'\Phi)^{-1}(\frac{\Phi'\Phi}{M})\Phi'\Phi_a)]^{\frac{1}{2}} b \\
&\le \lambda_{\max}E(\frac{\Phi'\Phi}{M}) + \sqrt{\lambda_{\max}(\frac{\Phi'\Phi}{M})}(E\sup_{1\le t\le T} y_t^2)^{\frac{1}{2}} b'[\frac{1}{M}E(\Phi_a'\Phi(\Phi'\Phi)^{-1}\Phi'\Phi_a)]^{\frac{1}{2}} b \\
&\le \lambda_{\max}E(\frac{\Phi'\Phi}{M}) + \sqrt{\lambda_{\max}(\frac{\Phi'\Phi}{M})}(E\sup_{1\le t\le T} y_t^2)^{\frac{1}{2}} \lambda_{\max}^{\frac{1}{2}} E(\frac{\Phi_a'\Phi_a}{M}) < \infty.
\end{aligned}$$

Therefore, $\lambda_{\max}E(\frac{\Psi'\Phi}{M}) = O_p(1)$. Because $\Phi(\Phi'\Phi)^{-1}\Phi'$ is an idempotent matrix, G is a square matrix and $\Psi'\Psi/M$ is invertible, we have $\lambda_{\min}G_T \ge c_0$ for some constant c_0. Using these facts, we can establish the following result that,

$$\lambda(p)\lambda_{\min}G \ge c_0/2.$$

The almost sure convergence theorem in part (b) follows immediately by combining Lemma 6.3 and 6.4 together with Assumption 3.1. Finally, we prove part (c). It is easy to show that

$$||G^{-1} - G_T^{-1}|| = ||G^{-1}(G - G_T)G_T^{-1}|| \le [\lambda_{\min}(G)]^{-1}[\lambda_{\min}(G_T)]^{-1}||G - G_T|| = O_p(\frac{1}{\sqrt{M}}).$$

Plugging this result into the following inequality, we have

$$\begin{aligned}
&||G_T^{-1}(\frac{\Psi'\Phi}{M})(\frac{\Phi'\Phi}{M})^{-1} - G^{-1}E\frac{\Psi'\Phi}{M}(E\frac{\Phi'\Phi}{M})^{-1}|| \\
&\le ||(G_T^{-1} - G^{-1})||\lambda_{\max}(\frac{\Psi'\Phi}{M})[\lambda_{\min}(\frac{\Phi'\Phi}{M})]^{-1} + ||\frac{\Psi'\Phi}{M} - (E\frac{\Psi'\Phi}{M})||[\lambda_{\min}(G)]^{-1}[\lambda_{\min}(E\frac{\Phi'\Phi}{M})]^{-1} \\
&+ ||(\frac{\Phi'\Phi}{M})^{-1} - (E\frac{\Phi'\Phi}{M})^{-1}||[\lambda_{\min}(G)]^{-1}\lambda_{\max}(E\frac{\Psi'\Phi}{M}) \\
&= O_p(\frac{1}{\sqrt{M}}).
\end{aligned}$$

$\square$

Lemma 6.6. *Suppose Assumptions 3.4-3.6 hold. Then*

(a) There exists a finite number $C > 0$ so that $E(\varepsilon\varepsilon') \leq CI_T$;

(b) $E[||\varphi(x)'\hat{\Psi}'\varepsilon/M||^2] = O_p(\frac{1}{M})$.

Proof of Lemma 6.6 First, we prove part (a). Suppose an arbitrary vector $b = (b_1, b_2, \cdots, b_T)$ and a finite number $C > 0$ so that $C \geq c_\tau$, where $c_\tau = \Delta^{\frac{2}{4+\delta}} \sum_{\tau=0}^{\infty} \frac{2^{2-2/(4+\delta)}(4+\delta)}{2+\delta} \alpha(\tau)^{1-\frac{2}{4+\delta}}$ for some $\delta > 0$, and $\alpha(\tau)$ is the mixing coefficients. Then we have

$$
\begin{aligned}
& b'(CI_T - E\varepsilon\varepsilon')b = C\sum_{t=1}^{T} b_t^2 - \sum_{t=1}^{T}\sum_{s=1}^{T} b_t b_s E(\varepsilon_t\varepsilon_s) \\
& \geq C\sum_{t=1}^{T} b_t^2 - \frac{1}{2}\sum_{t=1}^{T}\sum_{s=1}^{T}(b_t^2 + b_s^2)E|\varepsilon_t\varepsilon_s| = C\sum_{t=1}^{T} b_t^2 - \sum_{t=1}^{T} b_t^2 \sum_{s=1}^{T} E|\varepsilon_t\varepsilon_s| \\
& \geq C\sum_{t=1}^{T} b_t^2 - \sum_{\tau=0}^{\infty} \frac{2^{2-2/(4+\delta)}(4+\delta)}{2+\delta}\alpha(\tau)^{1-\frac{2}{4+\delta}}(E|\varepsilon_t|^{4+\delta})^{\frac{2}{4+\delta}} \sum_{t=1}^{T} b_t^2 \geq (C - c_\tau)\sum_{t=1}^{T} b_t^2.
\end{aligned}
$$

Hence, $CI_T - E(\varepsilon\varepsilon')$ is positive semidefinite. Using the result from part (a) and Lemma 6.4, we are can prove part (b) immediately:

$$
\begin{aligned}
& E(||\varphi(x)'\hat{\Psi}'\varepsilon/M||^2) = \frac{1}{M}E\{tr[(\frac{\hat{\Psi}'\hat{\Psi}}{M})\varphi(x)'\varphi(x)\varepsilon'\varepsilon]\} \leq \lambda_{\max}(\frac{\hat{\Psi}'\hat{\Psi}}{M})\frac{1}{M}\lambda_{\max}(E\varphi(x)'\varepsilon\varepsilon'\varphi(x)) \\
& = O_p(1)\frac{1}{M}\lambda_{\max}(CI_T) = O_p(\frac{1}{M}).
\end{aligned}
$$

$\square$

Proof of theorem 3.1. In our 2SLS series regression procedure, an estimator of f_p is expressed as $\hat{f}_p(x) = \varphi^p(x)'\hat{\alpha}^p$. By the Minkowski inequality,

$$
||\hat{f}_p - f|| = ||\hat{f}_p - f_p + f_p - f|| \leq ||\hat{f}_p - f_p|| + ||f_p - f|| \leq ||\hat{f}_p - f_p|| + O(K^{-1+min(d,p)})
$$

In the second step, we obtain $\hat{\alpha}^p = (\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}Y$. Consider the original time series nonlinear regression model, $\varepsilon_{t+1} = y_{t+1} - g_0(x_t, x_{t+1})$. Denote $G_0 = (g_0(x_0, x_1), \cdots, g_0(x_{T-1}, x_T))$.

Under Assumption 3.4 that the variance of ε_{t+1} is finite, thus $E(\varepsilon\varepsilon')-\overline{\sigma}^2 I$ is positive semidefinite. We modify the proof of Theorem 1 of Newey (1994), and use the triangular inequality that

$$\begin{aligned}
&\int[\hat{f}_p(x)-f(x)]^2 dF(x)=\int[\hat{f}_p-f_p+f_p-f]^2 dF(x)\\
&=\int[\varphi^p(x)'(\hat{\alpha}_p-\alpha_p)+\varphi^p(x)'\alpha_p-f]^2 dF(x)\le||\hat{\alpha}_p-\alpha_p||^2+O(K^{-2(1+min(d,p))}).
\end{aligned}$$

Thus, we can focus on relevant properties of $||\hat{\alpha}_p-\alpha_p||$. It is immediately follows that,

$$\begin{aligned}
&||(\hat{\alpha}^p-\alpha^p)||=||(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'Y-(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'\hat{\Psi}\alpha^p||\\
&\le||(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'Y-(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'G_0||+||(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'G_0-(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'\hat{\Psi}\alpha^p||\\
&\le||(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'(Y-G_0)||+||(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'(G_0-\hat{\Psi}\alpha^p)||
\end{aligned}$$

Using the Cauch-Schwarz inequality, the property of an idempotent matrix, Lemma 6.6 and Assumption 3.5, we obtain

$$E[||\varepsilon'\hat{\Psi}(\hat{\Psi}'\hat{\Psi}/M)^{-\frac{1}{2}}/M||^2]=trE[\hat{\Psi}(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}\varepsilon\varepsilon']/M=O_p(\frac{1}{M}).$$

Therefore, it follows that

$$\begin{aligned}
&||(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'(Y-G_0)||=||(\hat{\Psi}\hat{\Psi}/M)^{-1}\hat{\Psi}'\varepsilon/M||=|\varepsilon'\hat{\Psi}(\hat{\Psi}'\hat{\Psi}/M)^{-1}(\hat{\Psi}'\hat{\Psi}/M)^{-1}\hat{\Psi}'\varepsilon/M^2|^{\frac{1}{2}}\\
&\le||\varepsilon'\hat{\Psi}(\hat{\Psi}'\hat{\Psi}/M)^{-\frac{1}{2}}/M||=O_p(\sqrt{\frac{1}{M}}).
\end{aligned}$$

Recalling the construction of g_0 in Equation (**??**), we have

$$\begin{aligned}
E|g_0(x_t,x_{t+1})-\Psi'_{p,t}\alpha^p|^2 &\le CE|f(x_t)-f_p(x_{t+1})|^2+CE|m(x_{t+1})[f(x_{t+1})-f_p(x_{t+1})]|^2\\
&=O_p(K^{-2(1+min(d,p))}).
\end{aligned}$$

By the Cauchy-Schwarz inequality and the fact that $\hat{\Psi}(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'$ is idempotent, we have

$$||(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'(G_0 - \Psi\alpha^p)|| = ||(G_0 - \Psi\alpha^p)'\hat{\Psi}(\hat{\Psi}'\hat{\Psi})^{-1}(\frac{\hat{\Psi}'\hat{\Psi}}{M})^{-1}\hat{\Psi}'(G_0 - \Psi\alpha^p)/M|^{\frac{1}{2}}$$
$$= O_p(K^{-(1+min(d,p))}).$$

Let $\hat{v} \equiv \Psi - \hat{\Psi}$, which is the estimated residual from the first stage OLS regression. The first order condition implies that $\Phi'\hat{v} = 0$. Given $\hat{\Psi} = \Phi(\Phi'\Phi)^{-1}\Phi'\Psi$, it is easy to show that $\hat{\Psi}'\hat{v} = \Psi'\Phi(\Phi'\Phi)^{-1}\Phi'\hat{v} = \mathbf{0}$. Then it immediately follows that

$$(\hat{\Psi}'\hat{\Psi})^{-1}\hat{\Psi}'(\Psi - \hat{\Psi})\alpha^p = (\hat{\Psi}'\hat{\Psi})^{-1}\Psi'\Phi(\Phi'\Phi)^{-1}\Phi'\hat{v} = \mathbf{0}.$$

Therefore, we conclude that

$$\int [\hat{f}(x)_p - f(x)]^2 dF(x) = O_P[(\frac{1}{M} + K^{-2(1+min(d,p))})] = O_p(K/T).$$

Lemma 6.7. *Define* $V_T(x) = var[\frac{1}{\sqrt{M}}\varphi(x)'\Phi'\varepsilon] = \frac{1}{M}\varphi(x)'E(\Phi'\varepsilon\varepsilon'\Phi)\varphi(x)$. *Then as* $T \to \infty$,

$$\sqrt{\frac{1}{V_T}}\varphi(x)'M^{-\frac{1}{2}}\Phi'\varepsilon \xrightarrow{d} N(0,1).$$

Proof of Lemma 6.7 We prove the asymptotic normality by applying the martingale difference sequence central limit theorem by Brown (1971). First, we prove that for all fixed $x \in \mathbf{X}$, V_T is well-defined. There exists $c \in R^p$ with $||c|| = 1$, we have

$$V_T(x) \equiv \varphi(x)'E[\Phi_t\Phi_t'\varepsilon_{t+1}^2]\varphi(x)/M \geq \varphi_{t+1}(x)'\lambda_{\min}E[\Phi_t\Phi_t'\varepsilon_{t+1}^2]\varphi_{t+1}(x)/M$$
$$= \varphi_{t+1}(x)'\varphi_{t+1}(x)E[c'\Phi_t'\Phi_t\varepsilon_{t+1}^2 c]/M = O(1).$$

Because $E(\frac{1}{\sqrt{V_T}}\varphi(x)'M^{-\frac{1}{2}}\Phi_t\varepsilon_{t+1}|I_t) = 0$, $\{\frac{1}{\sqrt{V_T}}\varphi(x)'M^{-\frac{1}{2}}\Phi_t\varepsilon_{t+1}\}$ is a martingale difference sequence for all x and t. Second, we want to establish the Lindeberg condition given each

x. By the Minkowski and triangular inequalities, we have

$$V_T(x)^{-1} \sum_{t=1}^{T} E\big\{[\varphi(x)'\Phi_t\varepsilon_{t+1}]^2 I\{[\varphi(x)'\Phi_t\varepsilon_{t+1}]^2 \geq \epsilon V_T\}\big\}$$
$$\leq V_T(x)^{-1} \sum_{t=1}^{T} (\epsilon V_{pT})^{-\frac{\delta}{2}} E|\varphi(x)'\Phi_t\varepsilon_{t+1}]|^{2+\delta} = o_p(1).$$

The second condition that we need to verify that $\frac{1}{M}\sum_{t=0}^{T-1} \varphi(x)'\Phi_t\varepsilon_{t+1}^2\Phi_t'\varphi(x) - V_T = o_p(1)$. Given the fact that

$$E||\frac{\Phi'\varepsilon\varepsilon'\Phi}{M} - E\frac{\Phi'\varepsilon\varepsilon'\Phi}{M}||^2 = \sum_{i=1}^{K+p}\sum_{j=1}^{K+p} E[\frac{1}{M}\sum_{t=0}^{T-1}\phi_{i,t}\phi_{j,t}\varepsilon_{t+1}^2 - E\frac{1}{M}\sum_{t=0}^{T-1}\phi_{i,t}\phi_{j,t}\varepsilon_{t+1}^2]^2$$
$$= \sum_{i=1}^{K+p}\sum_{j=1}^{K+p}[\frac{1}{M}E(\phi_{i,t}^2\phi_{j,t}^2\varepsilon_{t+1}^4) + \frac{2}{M^2}\sum_{1<k<m<T-1} cov(\phi_{i,t}\phi_{j,t}\varepsilon_{t+1}^2, \phi_{i,m}\phi_{j,m}\varepsilon_{m+1}^2)]$$
$$\leq \sum_{i=1}^{K+p}\sum_{j=1}^{K+p}[\frac{1}{M}\sqrt{E(\phi_{i,t}^4\phi_{j,t}^4)}\sqrt{E[\varepsilon_{t+1}^8|]}$$
$$+ \frac{2^{(4+2\delta)/(4+\delta)}(4+\delta)/\delta}{M^2}\sum_{i=1}^{K+p}\sum_{j=1}^{K+p}\sum_{\tau=1}^{T-1}\alpha(\tau)^{\frac{\delta}{4+\delta}}[E\phi_{i,t}^{4+\delta}\phi_{j,t}^{4+\delta}]^{\frac{2}{4+\delta}}[E|\varepsilon_{t+1}|^{8+2\delta}]^{\frac{2}{4+\delta}}$$
$$= O(\frac{1}{M}).$$

It immediately follows that

$$|\frac{1}{M}\sum_{t=0}^{T-1}\varphi(x)'\Phi_t\varepsilon_{t+1}^2\Phi_t'\varphi(x) - V_T| = |tr\{\varphi(x)'(\frac{\Phi'\varepsilon\varepsilon'\Phi}{M} - E\frac{\Phi'\varepsilon\varepsilon'\Phi}{M})\varphi\}|$$
$$= |tr(\frac{\Phi'\varepsilon\varepsilon'\Phi}{M} - E\frac{\Phi'\varepsilon\varepsilon'\Phi}{M})\varphi(x)\varphi(x)'| = O_p(\lambda_{\max}|\frac{\Phi'\varepsilon\varepsilon'\Phi}{M} - E\frac{\Phi'\varepsilon\varepsilon'\Phi}{M}|) = o_p(1).$$

It follows that $\frac{1}{\sqrt{V_T}}\varphi(x)'\Phi'\varepsilon \xrightarrow{d} N(0,1)$ by Brown (1971). $\square$

Proof of theorem 3.2.

$$MD_T = A_T^{-2}(x) = \varphi(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}E(\Phi'\varepsilon\varepsilon'\Phi/M)P_T^{-1}Q_T(Q_T'P_T^{-1}Q_T)^{-1}\varphi(x).$$

As $T \to \infty$, based on Lemma 6.3, we can drive a useful relationship between A_T and V_T that they are of the same order. Considering results from the 2SLS series regression, we have

$$\sqrt{M}\varphi(x)'(\hat{\alpha}-\alpha)=\sqrt{M}\varphi(x)'(\Psi'\Phi(\Phi'\Phi)^{-1}\Phi'\Psi)^{-1}\Psi'\Phi(\Phi'\Phi)^{-1}\Phi'\varepsilon.$$

Hence, by Lemmas 6.5, 6.7, we have

$$\begin{aligned}
&|\sqrt{M}\varphi(x)'(\hat{\alpha}-\alpha)-\varphi(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}M^{-\frac{1}{2}}\Phi'\varepsilon|\\
&=|\varphi(x)'\{[\frac{\Psi'\Phi}{M}(\frac{\Phi'\Phi}{M})^{-1}\frac{\Phi'\Psi}{M}]^{-1}\frac{\Psi'\Phi}{M}(\frac{\Phi'\Phi}{M})^{-1}-(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}\}M^{-\frac{1}{2}}\Phi'\varepsilon|\\
&\le\lambda_{\max}\{[\frac{\Psi'\Phi}{M}(\frac{\Phi'\Phi}{M})^{-1}\frac{\Phi'\Psi}{M}]^{-1}\frac{\Psi'\Phi}{M}(\frac{\Phi'\Phi}{M})^{-1}-(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}\}|M^{-\frac{1}{2}}\varphi(x)'\Phi'\varepsilon|V_T^{-\frac{1}{2}}V_{pT}^{\frac{1}{2}}\\
&=o_p(1).
\end{aligned}$$

It implies that $\sqrt{M}\varphi(x)'(\hat{\alpha}-\alpha)$ and $\varphi(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}M^{-\frac{1}{2}}\Phi'\varepsilon$ have the same limiting distribution. It is sufficient to derive the limiting distribution of the latter.

We apply Brown's (1971) CLT theorem for martingale difference sequences. It is easy to show that $E(A_T\varphi(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}\Phi_t\varepsilon_{t+1}|I_t)=0$ for all $t=0,\cdots,T-1$. Define $\lambda_{it}=\varphi(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}e_i$ where $e_i\in R^p$ has the $i-$th element equal to 1 and 0 otherwise. By the Minkowski and Markov inequalities, the properties of trace, we have

$$\begin{aligned}
&D_T(x)^{-1}M^{-1}\sum_{t=1}^{T}E\{[\sum_{i=1}^{K+p}\lambda_{iT}\phi_{i,t}\varepsilon_{t+1}]^2I\{[\sum_{i=1}^{K+p}\lambda_{iT}\phi_{i,t}\varepsilon_{t+1}]^2\ge\epsilon MD_T(x)\}\}\\
&\le D_T^{-1}(x)M^{-1}\sum_{t=1}^{T}(\epsilon MD_T)^{-\frac{\delta}{2}}E|\sum_{i=1}^{p}\lambda_{it}\phi_{i,t}\varepsilon_{t+1}|^{2+\delta}\le D_T^{-1-\frac{\delta}{2}}(x)M^{-\frac{\delta}{2}}\{\sum_{i=1}^{p}\lambda_{iT}[E|\phi_{i,t}\varepsilon_{t+1}|^{2+\delta}]^{\frac{1}{2+\delta}}\}^{2+\delta}\\
&=O(1)[D_T]^{-1-\frac{\delta}{2}}M^{-\frac{\delta}{2}}|tr[\varphi(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}P_T^{-1}Q_T(Q_T'P_T^{-1}Q_T)^{-1}\varphi(x)]|^{1+\frac{\delta}{2}}\\
&=O(1)[D_T]^{-1-\frac{\delta}{2}}M^{-\frac{\delta}{2}}[D_T]^{1+\delta/2}=o_p(1).
\end{aligned}$$

It is straightforward to show that $var(A_T\varphi(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}M^{-\frac{1}{2}}\Phi'\varepsilon)=1$. Thus using the same reasonings as in Lemma 6.7, we can show that

$$\frac{1}{M}\sum_{t=0}^{T-1}\varphi^p(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}\Phi_{pt}\varepsilon_{t+1}^2\Phi_{pt}'P_T^{-1}Q_T(Q_T'P_T^{-1}Q_T)^{-1}\varphi^p(x)-D_{pT}=o_p(1).$$

Thus we have proved that $A_T\varphi(x)'(Q_T'P_T^{-1}Q_T)^{-1}Q_T'P_T^{-1}M^{-\frac{1}{2}}\Phi'\varepsilon \xrightarrow{d} N(0,1)$ as $T\to\infty$. Because we have proved that $A_T\sqrt{M}\varphi(x)'(\hat{\alpha}-\alpha)$ has the same limiting distribution, it immediately follows that,

$$A_T\sqrt{M}\varphi(x)'(\hat{\alpha}-\alpha)\xrightarrow{d} N(0,1),$$

By the Slutsky theorem, it is sufficient to show that

$$A_T\sqrt{M}[E\hat{f}(x)-f(x)]\to 0,\ \text{as}\, p,T\to\infty.$$

Recall that the approximation error of the truncated series is $O_p(K^{-1+min(d,s)})$. Define

$$Q_T\equiv A_T\sqrt{M}[E\hat{f}(x)-f(x)]$$

Under Lemma 6.7, we have

$$||Q_T||\leq\lambda_{\max}|A_T\sqrt{M}|||E\hat{f}_p(x)-f(x)||=[\lambda_{\min}(D_T)]^{-\frac{1}{2}}\sqrt{M}O_p(K^{-1+min(d,s)})\to 0,$$

where the last equation holds due to the assumption that the approximation errors will be asymptotically negligible compared to the asymptotic variance. Therefore, we complete the proof that $A_T\sqrt{M}[E\hat{p}(x)-f(x)]\to 0$,as $T\to\infty$. □

Proof of Proposition 1: Recall that $\hat{\Psi}=\Phi(\Phi'\Phi)^{-1}\Phi'\Psi$. Hence $\hat{\Psi}'\hat{\Psi}=\hat{\Psi}'\Psi=\Psi'\hat{\Psi}=\Sigma$. Let I_{K+p} be the $(K+p)\times(K+p)$ identity matrix. Let $U\Gamma U'$ be the eigendecomposition of

$\Sigma^{-1/2}P\Sigma^{-1/2}$. Note that r_i is the ith diagonal element of Γ. Hence we have

$$trace(H) = trace(I + \lambda\Sigma^{-1/2}P\Sigma^{-1/2})^{-1} = \sum_{j=1}^{K+p} \frac{1}{1 + \lambda r_j},$$

$$(\hat{\Psi}'\hat{\Psi} + \lambda P)^{-1} = \Sigma^{-1/2}(I_{K+p} + \lambda\Sigma^{-1/2}P\Sigma^{-1/2})^{-1}\Sigma^{-1/2} = \Sigma^{-1/2}U(I_{K+p} + \lambda\tilde{D})^{-1}U'\Sigma^{-1/2}.$$

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

$$Y'\hat{Y} = Y'HY = Y'\Phi\Sigma^{-1/2}(I_{K+p} + \lambda\Sigma^{-1/2}P\Sigma^{-1/2})^{-1}\Sigma^{-1/2}\hat{\Psi}'Y = \sum_{i=1}^{K+p} \frac{1}{1 + \lambda r_i} z_{1,i}z_{2,i}. \quad (22)$$

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

$$\begin{aligned}
\hat{Y}'\hat{Y} &= Y'H'HY = Y'\hat{\Psi}(\hat{\Psi}'\hat{\Psi} + \lambda P)^{-1}\Phi'\Phi(\hat{\Psi}'\hat{\Psi} + \lambda P)^{-1}\hat{\Psi}'Y \\
&= Y'\hat{\Psi}(\Phi'\Phi)^{-1/2}[(\Phi'\Phi)^{-1/2}\hat{\Psi}'\hat{\Psi}(\Phi'\Phi)^{-1/2} + \lambda(\Phi'\Phi)^{-1/2}P(\Phi'\Phi)^{-1/2}]^{-2}(\Phi'\Phi)^{-1/2}\hat{\Psi}'Y
\end{aligned}$$

Using the same techniques as above. Define $\tilde{\Sigma} = (\Phi'\Phi)^{-1/2}\hat{\Psi}'\hat{\Psi}(\Phi'\Phi)^{-1/2}$. Let $\tilde{U}\tilde{\Gamma}\tilde{U}'$ be the eigendecomposition of the matrix $\tilde{\Sigma}^{-1/2}(\Phi'\Phi)^{-1/2}P(\Phi'\Phi)^{-1/2}\tilde{\Sigma}^{-1/2}$. Then we have

$$\hat{Y}'\hat{Y} = \sum_{i=1}^{K+p} \frac{1}{(1 + \lambda\tilde{r}_i)^2} z_{3,i}^2, \quad (23)$$

where $\tilde{r}_i$ is the ith diagonal element of $\tilde{\Gamma}$, and $z_{3,i}$ is the ith element of $Z_3 = \tilde{U}'\tilde{\Sigma}^{-1/2}(\Phi'\Phi)^{-1/2}\hat{\Psi}'Y$.

Together with equation (22) and (23), we prove that Proposition 1 holds. $\square$