A System of Time-Varying Models for Predictive Regressions: Theory and Application

A thesis submitted for the degree of Doctor of Philosophy

By

Deshui Yu

Supervisors: Prof. Jiti Gao and Dr. Hsein Kew

Department of Econometrics and Business Statistics

Monash University

Australia

Copyright notice

Abstract

A long standing problem in empirical finance has been the difficulty in predicting stock returns. In a well-known paper, Welch and Goyal (2008) find significant evidence of in-sample predictability using numerous economic and financial variables but find weak or no significant evidence of out-of-sample predictability. A typical linear predictive regression model has the form: $y_t = \alpha + \beta x_{t-1} + e_t, t = 2, ..., T$, where y_t is stock return, x_{t-1} is a financial variable and e_t is a martingale difference sequence.

The predictor, $x_t = \gamma + \rho x_{t-1} + \varepsilon_t$, is assumed to follow a stationary first-order autoregressive process with errors, ε_t, that are correlated with e_t. This correlation is generally found to be nonzero when using dividend-price ratio or dividend yield as a predictor (see Campbell and Yogo (2006)). Because of this nonzero correlation, the predictive regression model suffers from embedded endogeneity problem and, as a result, Stambaugh (1999) demonstrates that the ordinary least-squares (OLS) estimator of the predictive coefficient β is biased in finite samples. To remove this embedded endogeneity, Amihud and Hurvich (2004) introduce a linear projection method and show that this method no longer produces a biased estimate of β.

In an empirical analysis, Campbell and Yogo (2006) find that a number of popular predictors are highly persistent since the observed sample first-order autocorrelation is close to, but less than, unity. Thus, Cai et al. (2015), Elliott and Stock (1994) and others assume that the predictor follows an integrated process with $\rho = 1$ or a local-to-unity process with $\rho = 1 + c/T$ and $c < 0$ when examining the statistical properties of the OLS estimator of the slope parameter β. This approach, however, creates the unbalanced problem because the regressand y_t is stock returns, whose observed behaviour

resembles a stationary process whilst the predictor x_{t-1} is assumed to be nonstationary.

Paye and Timmermann (2006) raise the possibility of instability in the predictive relation between the stock return and the predictor. They consider a time-varying predictive regression model of the form $y_t = \alpha_t + \beta_t x_{t-1} + e_t$, where α_t and β_t vary over time with structural breaks. By contrast, Cai (2007) and Chen and Hong (2012) assume that the parameters α_t and β_t are smooth deterministic functions of t. This thesis proposes a time-varying autoregressive model of the form $x_t = \gamma_t + \rho_t x_{t-1} + \varepsilon_t$ where γ_t and ρ_t are smooth deterministic functions of t. When $|\rho_t| < 1$ uniformly in t, the predictor x_t follows a local stationary process and this helps alleviate the unbalanced problem arising from modeling the predictor as a non-stationary process. We then consider a semiparametric time-varying coefficients predictive regression model with possibly local stationary predictors of the form $y_t = \alpha_t + \beta_t x_{t-1} + \lambda\varepsilon_t + \eta_t$ where η_t is uncorrelated with both ε_t and x_{t-1} so that the embedded endogeneity can be removed from the model. This predictive regression model that uses the linear projection method then has a nonparametric component and a parametric component. We apply the profile method to estimate this model. For estimating the time-varying coefficients of the autoregressive model and of the predictive regression model, we employ the local linear method.

The main theoretical findings of the thesis are: (1) the rate of $\sqrt{Th}$ at which the nonparametric time-varying estimators converge towards normal distributions and (2) the $\sqrt{T}$ rate at which the parametric estimator converges towards a normal distribution. For empirical analysis, we find that the 14 popular predictors from Welch and Goyal (2008) fit the time-varying AR(1) models. Based on the time-varying models, we find that the 14 predictors generally contain (in-sample) predictive content of future equity premium. In the out-of-sample analysis, we find that the financial ratios have significant forecasting power for the equity premium at both short and longer horizons, and the predictability is also economically significant.

Declaration

I, Deshui Yu, declare that this thesis titled, *A System of Time-Varying Models for Predictive Regressions: Theory and Application* and the work presented in it are my own. I confirm that:

- This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution.
- To the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature: Deshui Yu

Print Name: Deshui Yu

Date: 18/11/2020

Acknowledgments

I thank my supervisors, Professor Jiti Gao and Dr Hsein Kew, deeply for their energy and enthusiasm for my Ph.D. study. They provided great help in sustaining me during the long and tough slog of completing the confirmation, mid-term candidate review, and final review. Without their help, I would not have finished this thesis.

Professor Jiti Gao gave me invaluable guidance and unparalleled support, as well as many ideas in financial econometrics for both my current and future research. He encouraged me to pursue a Ph.D. when I was undertaking my master research project at Monash under his supervision. Without his encouragement, I would not have found the beauty of econometrics and research. Dr Hsein Kew helped me greatly with the empirical research, including instruction on data collection, program coding, seminar presentation, and improving my writing skills. These fundamental skills were invaluable for my PhD and will serve me well in future research.

I would also like to extend my sincere thanks to all the Department of Econometrics and Business Statistics staff at Monash University. I especially thank Professor Farshid Vahid, Professor Param Silvapulle, and Dr Wei Wei. As my milestone panel members, I benefited greatly from their valuable suggestions and comments. I also Professor Donald Poskitt and Professor Mervyn Silvapulle, from whom I learned many advanced techniques in statistical inference and econometric theory from their first-year Ph.D. courses: *Foundation of Econometrics* and *Statistical Inference*.

Finally, I wish to express my deep appreciation to my parents, girlfriend, and friends for their support, understanding, and love day and night.

Capstone Editing provided copyediting and proofreading services, according to the

guidelines laid out in the university-endorsed national ‘Guidelines for Editing Research Theses’.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Whether stock excess returns are predictable has attracted much attention in recent finance and econometrics literature because predicting stock returns has important implications for both industrial practitioners and academic researchers. From a practitioner's standpoint, dynamic or static asset allocation requires accurate forecasts of stock returns to construct optimal mean-variance portfolios that play a crucial role in improving investment performance. Thus, investors and financial analysts use a broad set of predictors, including fundamental and technical indicators, to predict future stock returns.

From an academic standpoint, the ability to forecast returns has important implications for tests of market efficient hypothesis, that is, whether investors can gain extra returns by using any particular information available up to today. Understanding stock return predictability also helps link the time-series variation in expected returns to the cross-section of average stock returns in asset pricing models (Cochrane, 2011). For example, one can test whether the commonly accepted risk factors that can largely explain the cross-sectional average stock return also reflect significant forecasting power for stock returns movements.

Forecasting stock returns has a long tradition in finance. Much of the early literature tested whether past returns predict future returns by using first-order autoregressive models, denoted as AR(1). In an AR(1) regression, a positive autoregressive coefficient implies the 'momentum effect' in stocks where past good returns signal good future returns. In contrast, a negative coefficient indicates overreaction or mean reversion. In the empirical studies, the autoregressive coefficient has usually been found to be statistically insignificant. For example, Fama and French (1988) show that past returns forecast subsequent returns in industry portfolios, and such autocorrelation of returns becomes stronger at longer horizons. Such "mean reversion" in stock returns has an important implication that the variance of stock returns rises less than linearly with forecasting horizon, so stocks are more predictable in the long run. However, the autocorrelation of stock returns seems to be weak or even non-existent in recent years.

This traditional view has been upended by an important change. Empirical tests have expanded to other predictive variables, such as financial valuation ratios, behavioural factors, and macroeconomic variables, instead of only past returns. A standard predictive regression is specified as follows:[1]

$$y_t = \alpha + \beta x_{t-1} + e_t, \tag{1.1}$$

$$x_t = \gamma + \rho x_{t-1} + \varepsilon_t, \tag{1.2}$$

where y_t is the equity premium or excess return at time t, x_{t-1} is a lagged financial or macroeconomic variable at time $t-1$, which is commonly formulated by an AR(1) process. The innovations (ε_t, e_t) from (1.1) and (1.2) are usually assumed to be independently and identically distributed (i.i.d.) normal with the covariance matrix, $(\varepsilon_t, e_t)^\top \sim N(\mathbf{0}, \Sigma)$, with $\Sigma = \begin{pmatrix} \sigma_e^2 & \sigma_{e\varepsilon} \\ \sigma_{\varepsilon e} & \sigma_\varepsilon^2 \end{pmatrix}$. The parameter of interest is the slope or forecasting coefficient, β, and x_t is said to predict stock returns if the forecasting coefficient β is statistically different from zero.

A long list of financial and macroeconomic variables have been employed to predict the equity risk premium. Fama and French (1989) study the predictability of excess returns by using the default premium and term premium. They find that both pre-

[1]The models are routinely used in the literature to test stock return predictability. See, for example, Amihud and Hurvich (2004), Campbell and Yogo (2006), and Stambaugh (1999).

dictors are rich in information concerning the variation in future excess returns. Fama (1990) shows that stock returns are predictable by lagged industrial production for both higher-frequency monthly data and lower-frequency annual data. Some other studies use lagged financial valuation ratios as predictors of the equity premium. In particular, Campbell and Shiller (1988) propose a present-value model that confirms the stock return predictability using the lagged dividend-price ratio and other valuation ratios. The dividend-price ratio is used as a predictor in subsequent studies. Hodrick (1992) and Campbell and Yogo (2006) provide new tests for return predictability and reinforce the results of Campbell and Shiller (1988). Lewellen (2004) studies, in addition to the dividend-price ratio, the predictive power of the earnings-to-price ratio and the book-to-market ratio. In sum, these predictors document significant β using the traditional predictive regression model (1.1).

However, it is important to note that in the empirical work, the usual approach to testing return predictability is based on in-sample fitting. Welch and Goyal (2008) show that a long list of predictors, including the popularly used financial valuation ratios and other business cycle variables, cannot produce consistently better out-of-sample forecasts based on the predictive regression model (1.1) than the historical average benchmark forecasts. Thus, their study raises the concern that these predictive regression models may not be able to help investors benefit from market-timing investment strategies. Due to the lack of real-time forecasting ability, numerous studies have documented several major econometric issues about estimation and inference in traditional predictive regressions that may lead to poor out-of-sample forecasting performance: (1) the forecasting relationship between equity premium and lagged predictor variables is not stable over time (Chen and Hong, 2012; Lettau and Nieuwerburgh, 2008; Paye and Timmermann, 2006), (2) the ordinary least-square (OLS) estimator of β is usually biased in the finite samples if the innovations from equations (1.1) and (1.2) are correlated, (i.e., $\sigma_{e\varepsilon} \neq 0$) (Amihud and Hurvich, 2004; Nelson and Kim, 1993; Stambaugh, 1999), and (3) statistical inference for forecasting coefficients may be problematic when the predictors are highly persistent or nonstationary (Campbell and Yogo, 2006; Torous et al., 2004; Kostakis et al., 2015; Phillips, 2015).

First, much of extant studies in stock return prediction literature show that the hypothesis of constant coefficient parameters in predictive regressions is routinely rejected—the forecasting relationship between stock returns and lagged predictors changes over time. For instance, Paye and Timmermann (2006) report significant evidence of parameter instability in predictive regression models of stock returns for the United States (US) and other G-7 countries by conducting several structural break tests. Moreover, Lettau and Nieuwerburgh (2008) report 30-year rolling regressions of annual stock returns on lagged dividend-price ratios. The OLS estimates for the forecasting coefficients vary between 0 and 0.5, and the associated R^2 ranges from close to zero to 30%. Lewellen (2004) also finds that strong parameter instability. For 1946–1972, the slope estimate is 0.84, and for 1973–2000, the slope estimate is 0.64. Given the evidence of parameter instability, a limitation of the conventional model (1.1) is that constant β cannot capture potential time-variation in return predictability. Therefore, parameter instability in predictive regressions is still one of the most serious challenges in stock return forecasting. In particular, Spiegel (2008) asks whether academics can 'produce an empirical model that allows for economic changes over time that is also capable of determining the "right" parameter values in time to help investors?'

Time-varying return predictability plays an important role in asset pricing theory. For example, Dangl and Halling (2012), Henkel et al. (2011), and Rapach et al. (2010), find empirical evidence that the predictive ability of equity premium is stronger during bad times than during good times. Why does predictability concentrate on bad times? One possible explanation is provided by investors' time-varying risk aversion. For example, Campbell and Cochrane (1999), Cochrane (2017), and Lettau and Ludvigson (2001) show that predictable variation in the equity risk premium, which reflects how economic agents rationally response to investment opportunities over time, is likely to be driven by the time-varying shifts in investors' risk aversion. Therefore, the predictive content of equity premium links to the overall economic conditions in the manner that investors are usually more risk-averse in bad times (or recessions), requiring high returns to compensate for high risks (Cochrane, 2017). Alternatively, Cujean and Hasler (2017) provide a general equilibrium model of investors' disagreement to rationalise the fact that stock return predictability concentrates in bad times. They show that the investors' disagreement spikes in bad times, thereby leading to stronger predictability because stock returns react to past news in bad times more heavily than that in good times. Therefore, predictive regressions with constant coefficients do not allow the strength of the evidence for stock return predictability to vary over time.

Second, the OLS estimates of forecasting coefficients are biased in finite samples when we account for the correlation between stock return and predictor (i.e., $\sigma_{e\varepsilon} \neq 0$) as is the case for dividend-price ratio, earning-price ratio, and treasury bill rate (see Campbell and Yogo (2006), Table 4). For example, the shocks to dividend-price ratio are negatively correlated with the shocks to stock return. An increase in stock price raises stock return ($r_t = \frac{P_t+D_t-P_{t-1}}{P_{t-1}}$) and lowers the dividend-price ratio ($DP_t = \frac{D_t}{P_t}$). Stambaugh (1999) shows that if the correlation coefficient between the two innovations ε_t and e_t is negative, then the OLS estimate for the slope coefficient, β, is upward biased in finite samples, although the bias term goes to zero as sample size goes to infinity. Amihud and Hurvich (2004), Nelson and Kim (1993), Stambaugh (1999), and others show that the OLS estimator of β, under the assumptions that (1) the errors e_t and ε_t follow a bivariate normal distribution and (2) the predictor x_t is stationary, has the following bias expression:

$$E[\widehat{\beta} - \beta] = \frac{\sigma_{e\varepsilon}}{\sigma_\varepsilon^2} E[\widehat{\rho} - \rho] = -\frac{\sigma_{e\varepsilon}}{\sigma_\varepsilon^2} \frac{(1+3\rho)}{T} + O\left(T^{-2}\right),$$

where $\widehat{\beta}$ and $\widehat{\rho}$ are OLS estimator of β and ρ, respectively. If the covariance between e_t and ε_t, $\sigma_{e\varepsilon} \neq 0$, then $\widehat{\beta}$ is biased in finite samples because in the AR(1) model (1.2), $\widehat{\rho}$ is biased in finite samples. This is so-called the 'Stambaugh bias'.[2] Specifically, the small sample bias increases with the persistence parameter, ρ, and decreases with the sample size, T. Stambaugh (1999) shows that the return predictability from using the dividend-price ratio disappears after correcting for this bias.

Third, perhaps the most common issue in predictive regressions is the uncertainty about the time-series properties of the selected predictors, in particular, their persistence.

[2] In Cai and Wang (2014), this issue is also called the 'embedded endogeneity'.

Many studies assume that financial ratios, such as dividend-price ratio, earning-price ratio, or book-to-market ratio, are stationary. According to finance theory, many predictors should be stationary, but empirically they are only slowly mean-reverting or highly persistent. When examining the time-series properties of the financial ratios by using various standard unit root tests, we often cannot reject the hypothesis that the financial ratios follow random walks. The stationarity assumption on predictors has a non-negligible impact on return predictability tests because persistent or nonstationary predictors may cause problematic inference that leads to inconclusive results in predictive regressions. Alternatively, the predictive regression is potentially unbalanced since we regress the stock return, which resembles a stationary series, on the predictor, which resembles a non-stationary series (Kasparis et al., 2015; Phillips, 2015). The unbalanced regressions produce time-series characteristics in the right-hand-side predictive variable that are very different from those of the left-hand-side stock return.

The impact of persistent predictors on predictive regressions can be interpreted from both economic and econometric perspectives. From an economic perspective, the extremely high persistence or nonstationarity in valuation ratios violates the plausibility of finance valuation theory (i.e., the present model of Campbell and Shiller (1988)). Imposing stationarity for financial ratios, such as the dividend-price ratio, is crucial in the present-value identity, $dp_t = \mathrm{E}_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} - \mathrm{E}_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}$.[3] The present-value identity implies that the dividend-price ratio is approximated as the discounted value of future expected returns and future expected dividend growth rates. Hence, the present-value model provides a useful accounting identity to understand the link between stock prices, fundamentals (dividends) and expected returns. Especially, Cochrane (2008) argues that the dividend-price ratio should predict either expected future stock returns or dividend growth.

Up to the 1980s, the dividend-price ratio appears stationary, and the present-value relation is found to be accurate (Campbell and Shiller, 1988). However, looking at more recent data, the dramatic stock market boom since the 1990s decreases the dividend-

[3]We denote all lower-case letters variables in logs, where r_{t+1} is the continuously compounded stock return, $dp_t = d_t - p_t$ is the log dividend-price ratio, and $\Delta d_{t+1} = d_{t+1} - d_t$ is the log dividend growth rate.

price ratio and increases its persistence (Lettau et al., 2008). As a result, the standard unit root tests no longer reject the unit root hypothesis at conventional significance levels. This has led to doubt regarding the validity of the log-linear approximation. For example, Campbell (2008) points out that the present-value relation breaks down when the dividend-price ratio contains a unit root because the unconditional mean does not exist. Although the dividend-price ratio is stationary but extremely persistent, the Taylor series expansion still provides a poor approximation of present-value relation because of high persistence and volatility. Thus, failure to satisfy the stationary assumption can lead to puzzling results under the present-value framework.[4]

From an econometric perspective, statistical inference for regressions with persistent predictors might be problematic. If the predictor x_t, say dividend-price ratio, contains a unit root, then the t-statistic for β under the null hypothesis of $\beta = 0$ would converge to a nonstandard distribution, as a function of Brownian motions, rather than a conventional normal distribution (Chan and Wei, 1987). If this is the case, the standard errors, p-value and t-statistic, under the stationary assumption, used to test a null hypothesis will be misleading. Though the dividend-price ratio is stationary but highly persistent (ρ is very close to 1), the asymptotic normal distribution provides a poor approximation to the actual finite-sample distribution of t-statistics (Campbell and Yogo, 2006; Elliott and Stock, 1994; Stambaugh, 1999). The stationary assumption probably leads to problematic statistical inference when we use conventional critical values from a normal distribution. Phillips (2015) provides a comprehensive survey of the econometric issues in predictive regressions.

1.2 Recent literature

Time-varying coefficients models. To allow the coefficients to be time-varying, one group of literature uses nonparametric time-varying modelling. Robinson (1989, 1991)

[4]Present-value relations suggest that the dividend-price ratio should predict future stock return and/or dividend growth. Otherwise, the dividend-price ratio is just a constant, which is not. (see Ang and Bekaert (2007) and Cochrane (2008) for example). However, much of the existing studies show that the dividend-price ratio actually contains little power for predicting stock returns and almost none for predicting dividend growth.

proposes a nonparametric time-varying coefficient time series model with a time trend

$$y_t = \alpha_t + \beta_t x_t + e_t \tag{1.3}$$

where α_t and β_t are unknown deterministic functions of time for trend and slope respectively. A major advantage of nonparametric modelling is that we impose minimum assumptions about β_t, and we are not required to specify a particular functional form for β_t.

The nonparametric time-varying model has attracted much attention in recent studies. Cai (2007) studies the nonparametric time-varying model with a stationary assumption and develops local constant and local linear estimation approach. Chen and Hong (2012) propose a consistent test for smooth structural changes and abrupt structural breaks in the time-varying coefficients model and find strong evidence against stability stock return predictability using several financial and economic predictors. However, the predictors in these studies are assumed to be stationary. Cai et al. (2015) model the predictor x_{t-1} as a local-to-unity process, NI(1), since the commonly used predictors, such as the dividend-price ratio and the earnings-price ratio, are highly persistent. Phillips et al. (2017) consider a multivariate time-varying coefficients model with a cointegrating relationship between y_t and the regressors.

On the other hand, state-space models are also used in literature. The state space approach allows the parameters to be stochastic, so that the coefficients follow a random walk process. For example, Dangl and Halling (2012) specify the following predictive regression model

$$r_t = \beta_t x_{t-1} + u_t \quad \text{with} \quad \beta_t = \beta_{t-1} + v_t,$$

where $u_t \sim N\left(0, \sigma^2\right)$ and $v_t \sim N\left(0, V_t\right)$. Restricting V_t to limit the parameter space, the state space time-varying parameter model can be, routinely, estimated using the Kalman filter and maximum likelihood methods. Dangl and Halling (2012) employ Bayesian model averaging to predict the monthly U.S. equity premium. They find that forecasts significantly outperform the historical average. Compared to the historical mean forecasts, Henkel et al. (2011) find that the return predictions using the state–space model are usually stronger during recessions than during expansions or normal periods.

Linear projection approaches. Stambaugh (1999) provides an expression of the bias of the OLS estimator of ρ, $E[\widehat{\rho}-\rho] = -\frac{1+3\rho}{T} + O(T^{-2})$. Therefore, a simple plug-in bias-corrected estimator of β implied from the exact bias expression is given by $\widehat{\beta}^c = \widehat{\beta} + \widehat{\phi}^c(1+3\widehat{\rho})/T$, where $\widehat{\beta}$ and $\widehat{\rho}$ are OLS estimators in (1.1) and (1.2), and $\hat{\phi}^c = \sum \widehat{e}_t\widehat{\varepsilon}_t / \sum \widehat{\varepsilon}_t^2$, where $\widehat{e}_t$ and $\widehat{\varepsilon}_t$ are the residuals from the regressions (1.1) and (1.2), respectively. Regarding the Stambaugh bias, Lewellen (2004) introduces a new test to improve the ability of financial ratios to predict stock returns. β is strongly correlated with dividend-price ratio's autoregressive coefficient, $E[\hat{\beta}-\beta] = \sigma_{e\varepsilon}/\sigma_\varepsilon^2 \cdot E[\hat{\rho}-\rho]$, so information conveyed by the autocorrelation in the dividend-price ratio helps produce more powerful tests of predictability. The bias-adjusted estimator of Lewellen (2004) is

$$\widehat{\beta}_{\text{adj}} = \widehat{\beta}_{OLS} - \lambda(\hat{\rho} - 0.9999),$$

where $\lambda = \sigma_{e\varepsilon}/\sigma_\varepsilon^2$. If the predictor is stationary, the maximal information conveyed by the autoregressive process is $\lambda(\hat{\rho} - 0.9999)$, where 0.9999 means that the persistence is allowed to be extremely close to, but less than unity. Specifically, Lewellen (2004) shows two positive effects as results of conveying the dividend-price ratio's information into empirical tests: (1) the slope estimate is often larger than the standard bias-adjusted estimate, and (2) the variance of the estimate is much lower. In combination, the two effects can substantially raise the power of the empirical predictability tests.

To implement the tests, we must also estimate λ from a regression $\widehat{e}_t = \lambda\widehat{\varepsilon}_t + \eta_t$, where $\widehat{e}_t$ and $\widehat{\varepsilon}_t$ are residuals obtained from the predictive regression model and AR(1) model, and η_t is assumed to be an i.i.d. process. However, (Amihud and Hurvich, 2004) show that the OLS estimator $\widehat{\lambda}$ is inefficient. To have a more efficient bias-corrected estimator, Amihud and Hurvich (2004) use a constant predictive model and propose a linear projection approach between two innovations to remove the Stambaugh bias from the predictive regression models, which is given by $e_t = \lambda\varepsilon_t + \eta_t$, where $\eta_t \sim i.i.d.$. Plugging the linear projection function into the predictive regression (1.1), the predictive regression model has the following expression, which matches a multivariate linear regression model:

$$y_t = \alpha + \beta x_{t-1} + \lambda \varepsilon_t + \eta_t. \tag{1.4}$$

Amihud and Hurvich (2004) show that the OLS estimator $\widehat{\beta}$ in the above model is unbiased. Since ε_t is the error term and is unobservable, a two-step estimation procedure is commonly used. In step 1, we run the AR(1) model for x_t and obtain bias-corrected residuals $\widehat{\varepsilon}_t$. In step 2, we replace ε_t with the corrected residuals $\widehat{\varepsilon}_t$ in model (1.4) and then estimate this model to obtain an estimate for β. Cai and Wang (2014) also consider the linear projection approach (1.4) in which the predictor is extend to be a local-to-unity process. Further, Cai et al. (2015) relax the constant coefficient assumption by considering a time-varying predictive regression model that is a pure nonparametric time-varying coefficient model:

$$y_t = \alpha_t + \beta_t x_{t-1} + \lambda_t \varepsilon_t + \eta_t.$$

Cai et al. (2015) estimate the time-varying coefficients by using a local linear kernel estimation.

Predictive regressions with persistent predictors. Recent studies (e.g., Cai and Wang (2014), Campbell and Yogo (2006), Kostakis et al. (2015), and Torous et al. (2004)) allow the predictor x_t to follow a local-to-unity process, where the autoregressive coefficient ρ follows

$$\rho = 1 + c/T, \quad \text{for} \quad c \leq 0.$$

The local-to-unity framework assumes that the process is stationary in finite samples, but behaves like a random walk asymptotically. The local-to-unity process is often used to model variables with a high degree of persistence, such as dividend-price ratio, earning-price ratio, or book-to-market ratio. While providing flexibility in modelling, the local-to-unity process assumption allows the predictors to exhibit persistence without necessarily being nonstationary containing unit roots in finite samples. The unknown local-to-unity parameter, c, measures the departure of this autoregressive coefficient from unity. In particular, Campbell and Yogo (2006) use this assumption to re-examine the return predictability.

However, practical implementation of the above local-to-unity assumption has two main drawbacks. First, predictability tests based on the local-to-unity assumption may be misleading when the predictors are non-persistent (ρ is close to zero), for example, long-term bond return and default return spread. Phillips (2014, 2015) shows that the statistical tests using the local-to-unity framework (e.g., Campbell and Yogo (2006)) perform well if the predictor's autoregressive root is very close to one, but perform poorly when ρ is far less than unity. The confidence intervals for c that are used in the procedure are invalid and are biased asymptotically when the true value of ρ is close to zero. Specifically, the validity of the method requires each predictive variable to be at least as persistent as a local-to-unity process. However, such a restrictive assumption is almost impossible to be empirically tested (Kostakis et al., 2015). Phillips and Lee (2013), Phillips (2015), and others point out that the unknown local to unity parameter, c, may not be consistently estimable in real applications.

Second, because of the problems associated with the construction of multidimensional confidence intervals for c, the methodology is restricted to single-predictor predictive regression models (Kostakis et al., 2015). The local-to-unity assumption imposes a severe restriction. There is no natural extension to test joint forecasting power for stock returns by combining more than one predictive variable (see, e.g., Phillips (2014, 2015)). The single-predictor forecasting framework can only accommodate testing the predictability of each financial variable separately, which may omit information from other variables. For example, Cochrane (2011) finds that combining the dividend-price ratio and the consumption to wealth ratio, *cay*, proposed by Lettau and Ludvigson (2001), has stronger predictability for expected returns than using the dividend-price ratio alone. Further, Møller and Sander (2017) find that the dividend yield and earnings yield are jointly strong predictors of dividend growth, motivated by the partial-adjustment dividend model of Lintner (1956).

Recent studies in econometrics literature consider new estimation and inference methods in predictive regressions. The first line of research considers non- and semi-parametric predictive models. Kasparis et al. (2015) propose a nonparametric model for predictive regressions and develop a unifying framework for inference in predictive regressions where the predictor can be stationary or nonstationary. Lee et al. (2014) further consider a series of nonparametric and semiparametric regression models that capture possibly nonlinearity in linear models and improve the predictors' predictive ability. Lee et al. (2015) consider imposing bagging to smooth the CT's constraint regression in a nonparametric regression of local historical average (LHA model) and show that the proposed LHA model beats the sample average benchmark out-of-sample framework.

The second line of research considers balanced predictive regressions. Predictive regressions are usually unbalanced because the dependent variable, equity premium, is stationary, but the predictors are usually persistent. The conventional t-test based approaches typically lead to a spurious inference in unbalanced regressions. Ren et al. (2019) add the second lag of the predictors to the right-hand side of the regression that reduces the persistence level to achieve balance.

The third line of research focuses on predictive quantile regression models. Quantile regression is an appealing technique that potentially aims to detect predictability at conditional quantiles of the equity premium than at the mean. Lee (2016) develops IVX-based methods for inference and prediction in quantile regressions that allow for persistent predictors. Moreover, Fan and Lee (2019) provides an improved inference for predictive quantile regressions with persistent predictors, which further account for conditionally heteroskedastic errors.

1.3 Overview and thesis contribution

This thesis contributes to the literature by relaxing the constant coefficients in the autoregressive model (1.2). If it is found that the autoregressive-coefficient are time-varying, then ignoring them leads to potentially misleading inference about the predictability of stock return. We develop a nonparametric estimation method for the time-varying autoregressive coefficient. In an empirical study, we find that the time series behaviours of the 14 popular predictors considered in Welch and Goyal (2008) are well approximated by a time-varying autoregressive process. We implement a nonparametric estimator to estimate the time-varying coefficients in the predictive regression model that uses the linear projection method to correct for the Stambaugh bias, and derive their asymptotic distributions. We apply this method to reanalyse the predictability of the equity premium without imposing the restriction that the autoregressive parameter of the predictor is constant over time. Our non-parametric model suggests that there is significant variation over time in the strength of stock return predictability. In particular, the predictability is stronger during recessions than otherwise.

Another contribution of this thesis is to apply the time-varying predictive regression model to examine the out-of-sample stock returns forecasting problem at short and long horizons. Compared with constant-coefficients predictive regression models, we find that dividend-price ratio and other financial ratios can produce significant out-of-sample predictive power by using the time-varying model. We next briefly discuss the contributions of each of the main chapters.

In Chapter 2, we use the linear projection method as in Amihud and Hurvich (2004) to remove the Stambaugh bias and we allow the coefficients in both the predictive regression model and autoregressive model to vary over time. Specifically we consider a system of time-varying models for predictive regressions:

$$y_t = \alpha(\tau_t) + \beta(\tau_t) x_{t-1} + e_t, \tag{1.5}$$

$$x_t = \gamma(\tau_t) + \rho(\tau_t) x_{t-1} + \varepsilon_t, \tag{1.6}$$

$$e_t = \lambda \varepsilon_t + \eta_t, \tag{1.7}$$

where $\alpha(\tau_t)$ and $\gamma(\tau_t)$ are time-varying trends and $\beta(\tau_t)$ and $\rho(\tau_t)$ are time-varying slope coefficients. Model (1.5) allows the relationship between stock returns and the lagged predictor to be time-varying, and the trend of return is also allowed to be nonlinear with unknown functional form. Model (1.6) is a time-varying AR(1) process for the predictor, x_t, that allows for time-variation in the degree of persistence in the predictor, x_t. Following Amihud and Hurvich (2004), Cai and Wang (2014), and Lewellen (2004), model (1.7) is a linear projection that links the innovations e_t and ε_t from y_t and x_t, respectively. This linear projection conveys information from the autocorrelation in predictors that helps produce more powerful tests of predictability. In Chapter 2, we develop nonparametric local linear estimation for the time-varying parameters in our models, and we fully establish the asymptotic theory for the local linear estimators. If the coefficients in models (1.5) and (1.6) are constant over time, these models reduce to predictive regression models considered in Amihud and Hurvich (2004), Campbell and Yogo (2006), and Stambaugh (1999).

Chapter 2 has two main contributions. First, it proposes a time-varying AR(1) model for the predictor. This model covers an important class of locally stationary processes (Dahlhaus, 1996, 1997). Under the condition that $|\rho_t| < 1$ uniformly in t and ρ_t is a smooth deterministic function of time, the predictor behaves like a stationary process in a local neighbourhood of each chosen time point, although the predictor may exhibit nonstationary behaviour globally. This model nests as a special case the constant coefficient autoregressive model.

Second, we technically extend the Phillips and Solo (1992) device in which they offer an alternative method of driving asymptotics for linear processes (with constant coefficients) using the Beveridge–Nelson decomposition. According to Dahlhaus (1996), a time-varying AR(1) model can be re-expressed as an infinite-order time-varying moving average, hereafter, time-varying MA(∞), under some the regularity conditions of locally stationary processes $x_t = \mu_t + \sum_{j=0}^{\infty} \varphi_{j,t}\varepsilon_{t-j}$, where $\mu_t = \sum_{j=0}^{\infty} \varphi_{j,t}\gamma_{t-j}$ is a deterministic function of time, and $\sum_{j=0}^{\infty} \varphi_{j,t}\varepsilon_{t-j}$ is a time-varying MA(∞) process. We develop a time-varying extension of Beveridge-Nelson (tv-BN) decomposition for the time-varying MA(∞) process. The tv-BN device plays an important role in building asymptotic theory for nonparametric estimators, similar to Phillips and Solo (1992).

In Chapter 3, we implement the time-varying models to test predictability for quarterly U.S. aggregate stock returns in the sample period from 1927:I to 2018:IV using a list of commonly employed variables from Welch and Goyal (2008). This chapter focuses on in-sample predictability tests because the proposed methodologies aim to make a more reliable statistical inference for the time-varying structure in coefficients and predictors' unknown persistence degrees. Additionally, we extend the empirical analysis for time-varying predictive regressions with a single predictor into a multi-predictor framework, and we apply a time-varying VAR(1) for the multiple predictors. Multi-predictor predictive regression models have become increasingly popular on the econometrics side of return predictability literature, see, for example Amihud et al. (2008), Chen and Hong (2012), Kostakis et al. (2015), and Xu (2020). We present empirical results in two ways: (1) univariate time-varying predictive regressions with time-varying AR(1) process proposed in Chapter 2 and (2) multivariate time-varying predictive regressions with time-varying VAR(1) predictors.

The contribution of Chapter 3 comes from the empirical results using the proposed time-varying models. First, using the proposed time-varying models (1.5)–(1.7), we find evidence of instability in the estimates of ρ_t for the large majority of predictors considered in Welch and Goyal (2008). For each predictor, the estimates of ρ_t vary smoothly over time and are less than one in absolute value over the sample period. We also find that the goodness of fit of the autoregressive regression improves when we allow its coefficients to vary over time. We also find evidence of instability in the forecasting coefficient β_t for the vast majority of predictive regression models.

Second, we find that the estimated coefficients $\widehat{\beta}_t$ of the list of 14 predictors suggested are time-varying. The OLS estimates of coefficient in the linear regression model generally lie outside of the confidence intervals of the time-varying local linear estimates. More importantly, the 14 popularly used predictors indeed contain predictive content of stock returns. Like the dividend-price ratio and book-to-market ratio, the fundamental factors have strong predictability over the majority of the subsamples. The predictability is especially strong during economic recession periods, such as the oil shock of 1973-1974 and the global financial crisis of 2007-2009.

In Chapter 4, we use the time-varying predictive regression models to forecast future stock returns out-of-sample. It is essential to point out that the usual approach to testing predictability relies on the in-sample fitting in the empirical work. However, significant in-sample predictability is not a guarantee of out-of-sample forecasting performance. Critically, Welch and Goyal (2008) show that the popularly used predictors are unable to deliver consistently superior out-of-sample forecasts of future excess returns relative to simple benchmark forecasts based on the historical means. In Chapter 4, we predict future quarterly excess returns out-of-sample from $J = 1$ to $J = 12$, where J denotes the forecast horizon. Although the dividend-price ratio shows little forecasting power for future returns at short horizons, it seems to explain increasingly more of the variability in returns at longer horizons (see, e.g., Cochrane (2008, 2011) and Fama and French (1988)).

Chapter 4 reports out-of-sample forecasting results by using the time-varying predictive regression models. The time-varying predictive regressions model produces positive R_{os}^2 for dividend-price ratio and dividend yield at both short and longer horizons, and produce better forecasts using the list of predictors from Welch and Goyal (2008) than the traditional linear regression models. Therefore, we conclude that the commonly used predictors show more significant out-of-sample predictability for stock returns using the time-varying models than the traditional models.

In Chapter 5, we remarkably conclude the thesis and present a closely related topic as a direction for future research. We propose a time-varying autoregressive distributed lag model to forecast economic activities using the 14 predictors considered in this thesis. According to Cochrane (2008), the equity premium forecasts are more plausibly related to macroeconomic risk if equity premium predictors can also forecast business cycles. Stock and Watson (2003) show that the forecasting power of individual economic variables for output growth can be highly unstable over time.

The technical proofs of the main theoretical results and additional empirical results are relegated to Appendices A and B.

Chapter 2

Univariate Time-Varying Predictive Models: Theory

In this chapter, we provide theoretical treatment of the system of time-varying models for predictive regressions. To estimate the time-varying coefficients, we employ the nonparametric local linear kernel estimation method. The main benefit of kernel smoothing is that we do not need to impose any prior functional form on the nonparametric functions; it is entirely data-driven. We build the convergence of the time-varying AR process, x_t, and establish the asymptotic theory of local linear estimators for nonparametric parameters and the model's constant parameter. The main theoretical finding is the $\sqrt{Th}$ rate of convergence for the estimators of the nonparametric time-varying parameters and the $\sqrt{T}$ rate of convergence for the non-time-varying estimator. A Monte Carlo experiment is carried out to examine the finite sample properties of the estimators. All the mathematical proofs of the asymptotic results are relegated to Appendix A.

2.1 Model specification and motivation

The objective of this chapter is to deal with the surveyed econometric issues in Chapter 1. To simultaneously take into account parameter instability, inference with persistent predictors, and innovations' correlation, this chapter proposes a system of time-varying models for predictive regressions:

$$y_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + e_t, \tag{2.1}$$

$$x_t = \gamma(\tau_t) + \rho(\tau_t)x_{t-1} + \varepsilon_t, \tag{2.2}$$

$$e_t = \lambda\varepsilon_t + \eta_t, \tag{2.3}$$

where y_t is the stock excess return at time t, x_{t-1} is a lagged predictor variable at time $t-1$, $\alpha(\tau_t)$ and $\gamma(\tau_t)$ are time-varying trends, $\beta(\tau_t)$ and $\rho(\tau_t)$ are time-varying slope coefficients, and η_t is assumed to be an i.i.d. process with zero mean and finite variance, σ^2_η. Model (2.1) allows the relationship between stock returns and the lagged predictor to be time-varying, and the trend of return is also relaxed to be nonlinear with unknown functional form. Extending on previous studies (Cai, 2007; Cai et al., 2015; Chen and Hong, 2012), we propose a time-varying AR process for x_t, which is new to the literature. Model (2.2) is a time-varying AR(1) process for the predictor that allows time-variation in the trend and autocorrelation of predictor, x_t. Following Amihud and Hurvich (2004), Cai and Wang (2014), and Lewellen (2004), we use a linear projection equation (2.3) to link the innovations e_t and ε_t from y_t and x_t. This linear projection equation conveys information from the autocorrelation in predictors (2.2) that helps produce more powerful tests of predictability. Similar to Amihud and Hurvich (2004) and Cai and Wang (2014), we substitute the linear projection equation (2.3) to the predictive regression model (2.1) and obtain a semiparametric model

$$y_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + \lambda\varepsilon_t + \eta_t, \tag{2.4}$$

where $\alpha(\tau_t)$ and $\beta(\tau_t)$ are the nonparametric parameters, and λ is the parametric parameter. We use a semiparametric profile approach to estimate both parametric and nonparametric parameters. This is standard in semiparametric regressions (see, e.g., Chen et al. (2012) and Li et al. (2011b)).

In the nonparametric time-varying models, we do not make parametric assumptions about the nature of the forecasting relations between stock return y_t and the lagged predictor x_{t-1} under investigation, or the autocorrelation between x_t and x_{t-1}. The conventional linear predictive regressions rely on some critical assumptions regarding the functional form of the relation between the variables being examined. For example, the linear predictive regression analysis assumes that the relationship between the dependent and independent variables is linear. As shown in Chapter 1, forecasting coefficients usually exhibit instability over time, and thus the nonparametric models can help uncover nonlinear relations between variables that are missed by using parametric techniques.

Our models cover and extend many existing models. For example, when the coefficients $\alpha(\tau_t)$, $\beta(\tau_t)$, $\gamma(\tau_t)$, and $\rho(\tau_t)$ are constant, then our time-varying predictive system models are the conventional predictive models in Amihud and Hurvich (2004), Campbell and Yogo (2006), Lewellen (2004), and many others. When x_t is assumed to be stationary, model (2.1) becomes the trending time-varying coefficients model as considered by Cai (2007) and Chen and Hong (2012). When x_t is supposed to be a local-to-unity process, model (2.1) has been studied by Cai et al. (2015). When x_t is a random walk, then model (2.1) is covered by the multivariate time-varying model in Phillips et al. (2017). If we remove $\beta(\tau_t)x_{t-1}$, the semiparametric model (2.4) reduces to the models discussed in Gao and Hawthorne (2006).

In sum, this chapter has two main contributions that extend the literature. First, to the best of our knowledge, this study is the first to consider a time-varying AR process as the data-generating process (DGP) for predictors in predictive regression literature. The model (2.2) can be used to model many commonly used explanatory predictors with varying degrees of persistence. In particular, the time-varying AR(1) model (2.2) covers an important class of locally stationary processes (Dahlhaus, 1996, 1997) under two key conditions: (1) the nonparametric functions $\alpha(\tau_t)$ and $\rho(\tau_t)$ are twice continuously differentiable with uniformly bounded fourth order derivatives, and (2) the time-varying auto-coefficient satisfies $|\rho(\tau_t)| < 1, \forall t$. Another contribution of this chapter is the technical proposal of a time-varying extension of the Beveridge and Nelson (1981) decomposition (BN) lemma. Phillips and Solo (1992) offer an alternative method of driving asymptotics for linear processes (with constant coefficients) using the

BN lemma. According to Dahlhaus (1996), a time-varying AR model can be re-expressed as a time-varying MA(∞) under the regularity conditions of locally stationary processes. Thus, the time-varying AR(1) model has the following representation

$$x_t = \mu_t + \sum_{j=0}^{\infty} \varphi_{j,t}\varepsilon_{t-j},$$

where $\mu_t = \sum_{j=0}^{\infty} \varphi_{j,t}\gamma_{t-j}$ is a deterministic function of time, and $\sum_{j=0}^{\infty} \varphi_{j,t}\varepsilon_{t-j}$ is a time-varying MA(∞) process. We develop a tv-BN decomposition for the time-varying MA(∞) process. The tv-BN device plays an important role in building asymptotic theory for nonparametric estimators, similar to Phillips and Solo (1992).

The motivation of the time-varying AR(1) model is two-folded. First, a time-varying AR model behaves smoothly in a neighbourhood of time point $\tau \in [\frac{1}{T}, \ldots, \frac{T}{T}]$, and the observations around this point show approximately stationary behaviour. Nonetheless, the first- and second-order structure of a stochastic process are still variant concerning translations along the time axis globally. Time-varying coefficients in the AR(1) model are flexible and more realistic than either I(0) or I(1) for most of the financial variables. The locally stationary process does not require stationarity over the whole sample and also does not equate to fast and dramatic movements akin to a random walk. Chen and Hong (2012) also mention that nonstationary time series may still not be stationary, even after taking difference or trending components are removed. Locally stationary time series models fill this gap and provide a new perspective into modelling financial time series.

Second, the time-varying AR model has important economic implications. The time-varying AR(1) models capture time variation in the steady-state of the dividend-price ratio or other persistent predictors. For example, Lettau and Nieuwerburgh (2008) shows that structural changes (e.g., breaks in 1954 and 1994) are identified by real data in the steady-state mean of the dividend-price ratio that leads to a highly persistent pattern. Lettau et al. (2008) show that low-frequency changes in economic volatility, such as consumption volatility and GDP volatility, are positively associated with the low-frequency movements in the dividend-price ratio. The decline in macroeconomic risk (or the volatility of the aggregate economy) seems to explain the decrease in the financial ratios after 1990. On the other hand, Chen et al. (2012) argue that dividend smoothing, as a choice of corporate policy, makes the dividend yield more persistent, and therefore makes either stock returns or dividend growth less predictable.

Although the popularly used financial valuation ratios, such as the dividend-price ratio or the earning-price ratios should, according to finance theory, be stationary, empirically, most predictors are only slowly mean-reverting or highly persistent (Campbell and Yogo, 2006; Elliott and Stock, 1994). When examining the time-series properties of the financial ratios, unit root tests usually reject the stationarity hypothesis. However, financial economists argue against the view that the valuation ratios are random walks from the finance theory perspective. In theory, a rational present-value model of Campbell and Shiller (1988) implies no bubble in the stock prices. The no bubble condition in stock prices tends to suggest that the dividend-price ratio does not contain a unit root. As an aside, Giglio et al. (2016) test the no bubble condition of asset prices in the United Kingdom and Singapore housing markets using a data range longer than 700 years and find no evidence of bubbles. Therefore, we propose a locally stationary process that permits the presence of slow-evolving or persistent component in the predictors and does not contain unit roots.

We interpret this point by a simple analytical example. Suppose a time series x_t has the form of $x_t = \alpha_t + \rho_t x_{t-1} + \varepsilon_t$, where ε_t is assume to be i.i.d. with zero mean and finite variance σ^2. **Case 1: I(0)**. When $\alpha_t \equiv \alpha$ and $\rho_t \equiv \rho \in (-1, 1)$, the time series x_t nests a stationary process, I(0). In a linear process form, $x_t = \alpha \sum_{j=0}^{\infty} \rho^j + \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}$. The mean of x_t is $E(x_t) = \frac{\alpha}{1-\rho}$, a constant. The variance can be $var(x_t) = \frac{1}{1-\rho^2}\sigma^2$, a constant, given a finite sample size from 1 to t. However, these properties are not realistic. Lettau and Nieuwerburgh (2008) find evidence of structural breaks in the steady-state of dividend-price ratio. **Case 2: LS**: when $\alpha_t \equiv \alpha(t/T)$ and $\rho_t \equiv \rho(t/T) \in (-1, 1)$, the time series x_t covers a class of locally stationary processes (Dahlhaus, 1996). In a time-varying MA(∞) representation, we have $x_t = \sum_{j=0}^{\infty} \varphi_{j,t}\alpha_t + \sum_{j=0}^{\infty} \varphi_{j,t}\varepsilon_{t-j}$. The mean of x_t is $x_t = \sum_{j=0}^{\infty} \varphi_{j,t}\alpha_t$ and the variance of x_t given a finite sample is $var(x_t) = (\sum_{j=0}^{\infty} \varphi_{j,t})^2\sigma^2$. The variance is time-varying instead of constant, but bounded in some constant since we have $(\sum_{j=0}^{\infty} \varphi_{j,t})^2 < M$, for some constant M.[1] Therefore, a time-varying AR(1) process allows for time-varying changes in the mean and variance of x_t, but they are bounded instead of explosive. **Case 3. I(1)**: when $\alpha_t \equiv \alpha$ and $\rho_t \equiv 1$, the time series x_t nests a random walk with drift, I(1). The mean of x_t is $x_t = x_0 + \alpha t$, and the variance of x_t given a finite sample is $var(x_t) = t\sigma^2$. The mean and variance of the random walk process are absolutely unbounded. However, most financial variables including the dividend-price ratio are mean-reverting.

From the above illustration, the time-varying AR(1) process does not require stationarity over the whole sample and also does not equate to fast and dramatic movements akin to a random walk process. These properties do not contradict finance theory that financial valuation ratios are not explosive or random walks over the local sample, and also flexibly capture the various degrees of persistence in the financial ratios. The time-varying AR(1) process thus provides a better description of most financial variables in predictive regressions than the usual I(0) or I(1) processes.

2.2 Estimation procedure

In this section, we estimate the model by nonparametric kernel estimation methods that are well-developed and increasingly popular in econometrics literature. See, e.g., Cai (2007), Chen et al. (2012), and Phillips et al. (2017). Kernel smoothing, such as local constant or local linear estimation methods, can be treated as *local weighted least squares*, which puts higher weights on the local information in estimation around any particular time point and lessens the information far away from the time point. Some studies (see, e.g.g, Cai (2007)) show that the local linear estimation is superior to the local constant estimation in theory and application. In particular, the local linear estimator has a much smaller bias near the boundary of time-varying function support, and it produces a much more accurate approximation of the time-varying functions than that of the local linear estimator. Therefore, the focus of this thesis is only on the local linear estimation

[1] We show the time-varying MA(∞) representation later in Chapter 2.

method. [2]

Plugging the linear projection function (2.3) into the predictive regression equation (2.1), we have the following regressions in matrix form,

$$y_t = \theta(\tau_t)^\top z_{t-1} + \lambda \varepsilon_t + \eta_t, \tag{2.5}$$

$$x_t = \phi(\tau_t)^\top z_{t-1} + \varepsilon_t, \tag{2.6}$$

where $z_t = (1, x_{1,t})^\top$, $\theta(\tau_t) = (\alpha(\tau_t), \beta(\tau_t))^\top$ is a 2×1 vector of time-varying coefficients, $\phi(\tau_t) = (\gamma(\tau_t), \rho(\tau_t))^\top$ is a 2×1 vector of time-varying coefficients, and η_t and ε_t are scalar error terms. As previously mentioned, applying the linear projection between two innovations can remove the Stambaugh bias from the predictive model. Model (2.5) is a semiparametric model that contains both time-varying parameters, $\alpha(\tau_t)$ and $\beta(\tau_t)$, and a constant parameter, λ. Since ε_t is an unobservable error term, we propose a two-step estimation procedure.

Step 1 First, we estimate the time-varying AR(1) model (2.6) to find the estimate of $\phi(\tau_t)$, and obtain the estimated residual $\widehat{\varepsilon}_t$. Recall the time-varying AR(1) model

$$x_t = \phi(\tau_t)^\top z_{t-1} + \varepsilon_t. \tag{2.7}$$

By assuming that each element in $\phi(\tau_t)$ has continuous second derivatives, they can be approximated by a linear matrix function using Taylor expansion at any fixed and re-scaled time point $\tau_t = t/T$ for $\tau_t \in [0, 1]$; that is, for any deterministic time-varying function $f(\tau_t)$, we have $f(\tau_t) \approx f(\tau) + f'(\tau)(\tau_t - \tau)$. Such a specification of nonparametric functions makes it possible to construct a consistent local linear estimation of the time-varying coefficient functions, via increasingly intense sampling of data points at each particular time point $\tau \in [0, 1]$. We define the objective function as the weighted

[2] The kernel estimation method has become increasingly popular in empirical finance. For example, Ang and Kristensen (2012) apply a nonparametric methodology for estimating and testing alphas and betas across time in conditional factor models and test for the constancy of the potentially time-varying conditional alphas and betas. Boons et al. (2020) use a kernel method to estimate time-varying coefficients regression with exponential weights over an expanding window to measure the time-varying relationship between inflation risk premium and stock return. This method is equivalent to the local constant estimation, but estimate the unknown time-varying parameter at time t using all past information up to t rather than local intervals. Further, Chen et al. (2018) propose extending the heterogeneous autoregressive model of realised volatility by Corsi (2008) to a nonparametric time-varying coefficients model.

sum of squared residuals

$$\sum_{t=1}^{T}\left[x_t-\left(\phi(\tau)+\phi'(\tau)\left(\tau_t-\tau\right)\right)^{\top}z_{t-1}\right]^{\top}\left[x_t-\left(\phi(\tau)+\phi'(\tau)\left(\tau_t-\tau\right)\right)^{\top}z_{t-1}\right]K\left(\frac{\tau_t-\tau}{h}\right),$$

where $\phi'(\tau_t)$ is defined as the vector of derivatives of $\phi(\tau_t)$. Moreover, h, is the bandwidth satisfying the condition that $T\to\infty$, $h\to 0$ and $Th\to\infty$, and $K(\cdot)$ is the kernel function. Generally, a kernel function $K(u)$ is nonnegative and satisfies the following basic properties $\int K(u)du=1$, $\int uK(u)du=0$, and $\int u^2K(u)du=\kappa<\infty$. Minimising the objective function with respect to $\phi(\tau)$ gives the local linear estimator of $\phi(\tau)$,

$$\widehat{\phi}(\tau)=\begin{pmatrix}1&0&0&0\\0&1&0&0\end{pmatrix}\begin{pmatrix}S_{T,0}(\tau)&S_{T,1}^{\top}(\tau)\\S_{T,1}(\tau)&S_{T,2}(\tau)\end{pmatrix}^{-1}\begin{pmatrix}T_{T,0}(\tau)\\T_{T,1}(\tau)\end{pmatrix},$$

where we denote $S_{T,l}(\tau)=\frac{1}{Th}\sum_{t=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right)^{l}z_tz_t^{\top}$ for $l=0,1,2$, and $T_{T,l}(\tau)=\frac{1}{Th}\sum_{t=1}^{T}K\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right)^{l}z_{t-1}x_t$ for $l=0,1$. The local linear estimator can be re-expressed in a matrix form,

$$\widehat{\phi}(\tau)=[\boldsymbol{I}_2,\boldsymbol{0}_2]\left(D(\tau)^{\top}W(\tau)D(\tau)\right)^{-1}D(\tau)^{\top}W(\tau)X, \tag{2.8}$$

where $W(\tau)=diag\left(K\left(\frac{2-\tau T}{Th}\right),....,K\left(\frac{T-\tau T}{Th}\right)\right)$ is the $(T-1)\times(T-1)$ kernel matrix, I_2 is the 2×2 identity matrix, 0_2 is the 2×2 null matrix,

$$D(\tau)=\begin{pmatrix}z_1^{\top}&z_1^{\top}\frac{2-\tau T}{Th}\\ \vdots&\vdots\\ z_{T-1}^{\top}&z_{T-1}^{\top}\frac{T-\tau T}{Th}\end{pmatrix},\quad\text{and}\quad X=\begin{pmatrix}x_2\\ \vdots\\ x_T\end{pmatrix}.$$

Kernel smoothing is a local estimation around each τ using observations nearby, similar to a weighted least squares method. The residuals are computed as $\widehat{\varepsilon}_t=x_t-\widehat{\phi}(\tau_t)^{\top}z_{t-1}$.

Step 2. We estimate the predictive regression model (2.5). Replacing ε_t in the time-varying predictive model, we have (2.1) with

$$y_t=\theta(\tau_t)^{\top}z_{t-1}+\lambda\widehat{\varepsilon}_t+\eta_t.$$

We estimate this semiparametric predictive regression model by the profile likelihood estimation method, which allows for a simultaneous estimation for all parameters. The profile approach is standard in semiparametric econometrics literature (see, e.g., Chen et al. (2012), Li et al. (2011a), and others). Assuming that λ is known, we minimise the following objective function with respect to $\theta(\tau)$

$$\sum_{t=1}^{T} \left[(y_t - \lambda \widehat{\varepsilon}_t) - \left(\theta(\tau) + \theta'(\tau)\,(\tau_t - \tau) \right)^\top z_{t-1} \right]^2 K\left(\frac{\tau_t - \tau}{h} \right). \tag{2.9}$$

Similarly, the local linear estimator is given by

$$\widetilde{\theta}(\tau) = [\boldsymbol{I}_2, \boldsymbol{0}_2] \left(D(\tau)^\top W(\tau) D(\tau) \right)^{-1} D(\tau)^\top W(\tau)(Y - \widehat{\varepsilon}\lambda) = s(\tau)(Y - \widehat{\varepsilon}\lambda), \tag{2.10}$$

where $Y = (y_1, ..., y_T)^\top, \widehat{\varepsilon} = (\widehat{\varepsilon}_2, ..., \widehat{\varepsilon}_T)^\top, s(\tau) = [\boldsymbol{I}_2, \boldsymbol{0}_2] S(\tau)$, and $S(\tau) = \left(D(\tau)^\top W(\tau) D(\tau) \right)^{-1}$ $D(\tau)^\top W(\tau)$. The local linear estimator $\widetilde{\theta}(\tau)$ is infeasible because it is a linear function of the unknown parameter λ. A key step in the profile approach is that we use the local linear estimator to approximate the unknown parameter $\theta(\tau)$. The constant parameter λ, then, can be estimated by OLS. minimising the following objective function

$$L(\lambda) = \sum_{t=2}^{T} \left(y_t - \left(s(\tau_t)(Y - \widehat{\varepsilon}\lambda) \right)^\top z_{t-1} - \lambda \widehat{\varepsilon}_t \right)^2, \tag{2.11}$$

with respect to λ, we obtain the OLS estimator $\widehat{\lambda}$

$$\widehat{\lambda} = \left(\widetilde{X}^\top \widetilde{X} \right)^{-1} \widetilde{X}^\top \widetilde{Y}, \tag{2.12}$$

where $\widetilde{Y} = (I - \widetilde{s})Y$, $\widetilde{X} = (I - \widetilde{s})\widehat{\varepsilon}$ with $\widetilde{s} = \left(z_1^\top s(\tau_2), ..., z_{T-1}^\top s(\tau_T) \right)^\top$. Putting $\widehat{\lambda}$ (2.12) into (2.10), we obtain a feasible local linear estimator for $\theta(\tau)$

$$\widehat{\theta}(\tau) = [\boldsymbol{I}_2, \boldsymbol{0}_2] \left(D(\tau)^\top W(\tau) D(\tau) \right)^{-1} D(\tau)^\top W(\tau)(Y - \widehat{\varepsilon}\widehat{\lambda}) = s(\tau)(Y - \widehat{\varepsilon}\widehat{\lambda}). \tag{2.13}$$

2.3 Main asymptotic results

Although Dahlhaus (1996) defines locally stationary processes using a spectral density approach, in the context of applied research, our goal is parametric inference for nonstationary time series models that may be defined purely in the time domain. In the meantime, Dahlhaus (1996) first proves that an autoregressive process with time varying coefficients covers an important class of locally stationary processes under two conditions. Following Dahlhaus (1996), we make the assumptions for locally stationary processes.

Assumption 2.1. *Locally stationarity assumptions for the time-varying AR(1) model* $x_t = \gamma(\tau_t) + \rho(\tau_t)x_{t-1} + \varepsilon_t$*:*

(1) the smoothing functions $\gamma(\tau)$ *and* $\rho(\tau)$ *are twice continuously differentiable with respect to* τ *with uniformly bounded fourth order derivatives.*

(2) the time-varying autoregressive coefficient satisfies $|\rho(\tau)| < 1$.

Assumption 2.1 allows the parameters to change smoothly over time and hence rules out the possibility of multiple sudden structural breaks. It also rules out the presence of unit root in the predictors. Under Assumption 2.1, a tv-AR(1) model can be alternatively expressed as a first order difference equation and therefore has a solution that satisfies a time-varying MA(∞) representation. The coefficients of the time-varying MA(∞) model satisfy $\max_{t\geq 1}\sum_{j=0}^{\infty}|\varphi_{j,t}| < \infty$, uniformly in t. This assumption implies that the time series is stationary locally at a particular time point τ because of $|\rho(\tau)| < 1$, but it can be nonstationary globally as the unconditional mean and variance of x_t are still time-dependent. This framework can be applied to a time-varying AR(1) model used in my thesis so that $x_t = \gamma_t + \rho_t x_{t-1} + \varepsilon_t$ is expressed as a summation of infinite trending terms μ_t and infinite moving-average terms ε_t,

$$x_t = \mu_t + \sum_{j=0}^{\infty} \varphi_{j,t}\varepsilon_{t-j}, \tag{2.14}$$

where $\mu_t = \sum_{j=0}^{\infty}\varphi_{j,t}\gamma_{t-j}$ is a deterministic function of time; $\varphi_{0,t} = 1$ and $\varphi_{j,t} = \rho_t\rho_{t-1}\cdots\rho_{t-j+1}$ for $j \geq 1$; and ε_t is error term (see, Dahlhaus (1996) and Künsch (1995) for the discussion of this point). We make further technical conditions that are useful for establishing laws of large number (LLN) for the time-varying MA(∞) process and central limit theorems (CLT) for our nonparametric estimators.

Assumption 2.2. *Assumptions for time-varying MA(∞) processes:*

(1) μ_t and $\varphi_{j,t}$ are continuous function with bounded second derivative for all t and j.

(2) $\sup_{0\leq\tau\leq 1}\sum_{j=0}^{\infty} j|\varphi_j(\tau)| < \infty$, and $\sup_{0\leq\tau\leq 1}\sum_{j=0}^{\infty}|\varphi_j(\tau)\varphi_{j+h}(\tau)| < \infty$ for $h \geq 0$.

(3) There exists some constant M such that $\sum_{j=1}^{\infty} j|\varphi_j(u) - \varphi_j(v)| \leq M|u-v|$, for $0 \leq u, v \leq 1$.

As shown in Künsch (1995), Assumption 2.2 is needed when we re-express the time-varying AR process as the time-varying MA(∞) process. In addition, Assumption 2.2 implies the following results that are useful in deriving tv-BN decomposition

$$\max_{t\geq 1}\sum_{j=0}^{\infty}|\varphi_{j,t}| < \infty, \quad \max_{t\geq 1}\sum_{j=0}^{\infty} j^{1/2}|\varphi_{j,t}| < \infty, \quad \text{and} \quad \max_{t\geq 1}\sum_{j=0}^{\infty}\varphi_{j,t}^2 < \infty. \tag{2.15}$$

Our approach relies on a time-varying version of the BN decomposition (Beveridge and Nelson, 1981) to offer a simple approach to the LLN and CLT for the time-varying MA(∞) process. The idea is very similar to that of Phillips and Solo (1992) where they apply BN lemma to build the asymptotics for linear processes (with constant coefficients). The tv-BN decomposition has shared the benefits mentioned in Phillips and Solo (1992) as a powerful device in reducing time series asymptotics to known theorems for common cases, such as i.i.d. and martingale difference sequences. We propose the time-varying BN lemma that will play a fundamental role in our approach. We denote a time-dependent lag operator $C_t(L) = \sum_{j=0}^{\infty}\varphi_{j,t}L^j$ at a given time t, where $\varphi_{j,t}$ satisfies Assumption 2.2. We establish first- and second-order BN devices.

Lemma 2.1 (First-order BN lemma)**.**

(1) $C_t(L) = \sum_{j=0}^{\infty}\varphi_{j,t}L^j$ can be decomposed into two components, $C_t(L) = C_t(1) - (1-L)\widetilde{C}_t(L)$, where $\widetilde{C}_t(L) = \sum_{j=0}^{\infty}\left(\widetilde{\varphi}_{j,t}L^j\right)$ with $\widetilde{\varphi}_{j,t} = \sum_{h=j+1}^{\infty}\varphi_{j,t}$. The first-order tv-BN decomposition of x_t is $x_t = \mu_t + C_t(1)\varepsilon_t - (\widetilde{\varepsilon}_t - \widetilde{\varepsilon}_{t-1})$, where $\widetilde{\varepsilon}_t = \widetilde{C}_t(L)\varepsilon_t = \sum_{j=0}^{\infty}\left(\widetilde{\varphi}_{j,t}L^j\right)\varepsilon_t$.

(2) In the tv-BN decomposition, we have $\sum_{j=1}^{\infty}\left|\widetilde{\varphi}_{j,t}\right| < \infty$, and $|\widetilde{\varphi}_{t+1}(\tau) - \widetilde{\varphi}_t(\tau)| = O\left(T^{-1}\right)$.

Lemma 2.2 (Second-order BN lemma)**.**

(1) Under the assumptions, $B_{j,t}(L) = B_{j,t}(1) - (1-L)\widetilde{B}_{j,t}(L)$*, where* $\widetilde{B}_{j,t}(L) = \sum_{j=0}^{\infty} \widetilde{b}_{h,t} L^j$ *and* $\widetilde{b}_{j,t} = \sum_{i=j+1}^{\infty} b_{h,i,t} = \sum_{i=j+1}^{\infty} \varphi_{i,t}\varphi_{i+h,t}$*.*

(2) x_t^2 *can be expressed as* $x_t^2 = \mu_t^2 + 2\mu_t \sum_{j=0}^{\infty} \varphi_{j,t}\varepsilon_{t-j} + \left(\sum_{j=0}^{\infty} \varphi_{j,t}^2\right) \varepsilon_{t-j}^2 + \tilde{Q}_t + \tilde{D}_t$*, where* $\tilde{Q}_t = 2\sum_{h=1}^{\infty} B_{h,t}(1)\varepsilon_t\varepsilon_{t-h} = 2\sum_{h=1}^{\infty} \left(\sum_{j=0}^{\infty} \varphi_{j,t}\varphi_{j+h,t}\right) \varepsilon_t\varepsilon_{t-h}$*, and* $\tilde{D}_t = 2\sum_{h=1}^{\infty} \widetilde{B}_{h,t}(L)\varepsilon_t\varepsilon_{t-h} - 2\sum_{h=1}^{\infty} \widetilde{B}_{h,t-1}(L)\varepsilon_{t-1}\varepsilon_{t-1-h}$*.*

Lemma 2.1 and Lemma 2.2 demonstrate the first- and second-order time-varying BN decomposition. Such techniques produce moving-average based representations that are crucial in deriving the convergence of x_t, for example, $\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t - \tau) v_t^l x_t \xrightarrow{p} A_{T\tau}\tilde{\sigma}_l$, for $l = 0, 1, 2, 3$, in Lemma A.1, where $v_t = \frac{\tau_t - \tau}{h}$, $A_{T\tau} = \mu_{T\tau}$, and $\tilde{\sigma}_l = \int_{-\infty}^{+\infty} K\left(\frac{\tau_t - \tau}{h}\right)(v)v^l dv \cdot I(0 < \tau < 1) + \int_{\infty}^{0} K\left(\frac{\tau_t - \tau}{h}\right)(v)v^l dv \cdot I(\tau = 0) + \int_{0}^{-\infty} K\left(\frac{\tau_t - \tau}{h}\right)^2 (v)v^l dv \cdot I(\tau = 1)$. These properties would help derive the asymptotic theory throughout this thesis.

Assumption 2.3. *Regularity conditions for nonparametric local linear estimators:*

(1) Define the information set $I_{t-1} = \sigma(x_{t-2}, ..., x_1; \varepsilon_{t-1}, ..., \varepsilon_1)$*, and we impose some conditions on conditional moments of the error* ε_t *such that* $E[\varepsilon_t | I_{t-1}] = 0$*,* $E[\varepsilon_t^2 | I_{t-1}] = \sigma_\varepsilon^2$*,* $E[\varepsilon_t^4 | I_{t-1}] = \mu_4$*, and* $E[\varepsilon_t^\delta | I_{t-1}] < \infty$*, for some* $\delta \geq 2$*.*

(2) The kernel function $K(\cdot)$ *is Lipschitz continuous and symmetric with compact support [-1,1], satisfying* $sup_r |K(r)|^p = \|K\|^p < \infty$*.*

(3) The bandwidth h *satisfies as* $T \to \infty$*,* $h \to 0$ *and* $Th \to \infty$*, and simultaneously* $Th^8 \to 0$*,* $\frac{\sqrt{Th}}{\log(T)} \to \infty$ *and* $\frac{T^{1-\frac{2}{\delta}}h}{\log(T)} \to \infty$*.*

(4) The time-varying trend function $\alpha(\cdot)$ *and the coefficient function* $\beta(\cdot)$ *have continuous derivatives of up to the second order.*

The above assumptions are commonly imposed in nonparametric estimation literature (see, e.g., (Robinson, 1989), Cai (2007) and Li et al. (2011a)). For example, (2) in Assumption 2.3 is a mild condition on the kernel function and many commonly–used kernels, including the Epanechnikov kernel, $K(u) = \frac{3}{4}\left(1 - u^2\right) \mathrm{I}\{|u| \leq 1\}$, satisfy this assumption. Further, the compact support condition for the kernel function can be relaxed at the cost of more tedious proofs. Assumption 2.3 (4) is a mild common condition on the smoothness of the functions involved in the nonparametric local linear fitting.

We denote $\boldsymbol{S}_T^{-1}(\tau) = \begin{bmatrix} S_{T,0}(\tau) & S'_{T,1}(\tau) \\ S_{T,1}(\tau) & S_{T,2}(\tau) \end{bmatrix}^{-1}$, where

$$S_{T,l}(\tau) = \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\tau_t - \tau}{h}\right)\left(\frac{\tau_t - \tau}{h}\right)^l \boldsymbol{Z}_t\boldsymbol{Z}'_t \tag{2.16}$$

$$= \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\tau_t - \tau}{h}\right)\left(\frac{\tau_t - \tau}{h}\right)^l \begin{bmatrix} 1 & x_t \\ x_t & x_t^2 \end{bmatrix} \quad \text{for} \quad l = 0, 1, 2. \tag{2.17}$$

The decomposition of estimation errors of $\widehat{\phi}(\tau)$ is given by

$$\widehat{\phi}(\tau) - \phi(\tau) - \tilde{B}_T(\tau) - o(h^2) = \widetilde{T}(\tau), \quad \text{for} \quad \tau \in [0, 1], \tag{2.18}$$

in which $\widetilde{T}(\tau) = [\mathrm{I}_2, 0_2]S_T(\tau)^{-1}\widetilde{J}_T(\tau)$, and $\tilde{B}_T(\tau) = \frac{h^2}{2}[\mathrm{I}_2, 0_2]S_T^{-1}\widetilde{L}(\tau)\phi''(\tau)$. We define $\phi''(\tau)$ as the second-order derivative of $\phi(\tau)$, $\widetilde{L}(\tau) = (S_{T,2}(\tau), S_{T,3}(\tau))^\top$ and $\widetilde{J}_T(\tau) = [\widetilde{J}_{T,o}(\tau), \widetilde{J}_{T,1}(\tau)]^\top$ where $\widetilde{J}_{T,l}(\tau) = \frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t - \tau)v_t^l Z_t\varepsilon_t$, for $l = 0, 1$.

Theorem 2.1. *Under the assumptions, the asymptotic distribution of $\widehat{\phi}(\tau)$ at a point τ*

$$\sqrt{Th}\left(\widehat{\phi}(\tau) - \phi(\tau) - \bar{B}_T(\tau) - o_p(h^2)\right) \xrightarrow{d} N\Big(0, \quad \boldsymbol{V}_\varepsilon(\tau)\Big), \tag{2.19}$$

The asymptotic variance is $\boldsymbol{V}_\varepsilon(\tau) = [I_2, 0_2]\Sigma_S^{-1}(\tau)\Pi(\tau)\Sigma_S^{-1}(\tau)^\top[I_2, 0_2]^\top$ with $\Sigma_S(\tau) = \begin{bmatrix} \tilde{\sigma}_1 & A_{T\tau}\tilde{\sigma}_1 \\ A_{T\tau}\tilde{\sigma}_1 & \sigma^2 B_{0,T\tau}\tilde{\sigma}_2^2 \end{bmatrix}$ and $\Pi(\tau) = \begin{pmatrix} \pi_0(\tau) & \pi_1(\tau)^\top \\ \pi_1(\tau) & \pi_2(\tau) \end{pmatrix}$ for $l = 0, 1, 2$, where we define $\pi_l(\tau) = \begin{pmatrix} \sigma_\varepsilon^2\tilde{\sigma}_l^2 & 0 \\ 0 & \sigma_\varepsilon^2 B_{0,T\tau}(1)\tilde{\sigma}_l^2 \end{pmatrix}$, $A_{T\tau} = \mu_{T\tau}$, $B_{0,T\tau}(1) = \sum_{j=0}^{\infty}\varepsilon_{j,T\tau}^2$, $\tilde{\sigma}_l = \int_{-\infty}^{+\infty} K\left(\frac{\tau_t-\tau}{h}\right)(v)v^l dv \cdot I(0 < \tau < 1) + \int_{\infty}^{0} K\left(\frac{\tau_t-\tau}{h}\right)(v)v^l dv \cdot I(\tau = 0) + \int_{0}^{-\infty} K\left(\frac{\tau_t-\tau}{h}\right)^2 (v)v^l dv \cdot I(\tau = 1)$, and $\tilde{\sigma}_l^2 = \int_{-\infty}^{+\infty} K\left(\frac{\tau_t-\tau}{h}\right)^2 (v)v^l dv \cdot I(0 < \tau < 1) + \int_{\infty}^{0} K\left(\frac{\tau_t-\tau}{h}\right)^2 (v)v^l dv \cdot I(\tau = 0) + \int_{0}^{-\infty} K\left(\frac{\tau_t-\tau}{h}\right)^2 (v)v^l dv \cdot I(\tau = 1)$.

Local approximation of a locally stationary process induces the asymptotic theory of local linear estimators to be standard. Compared with Cai (2007), his model and our model share the same time-varying coefficient regression model, but he uses a stationary assumption for regressor x_t. Theorem 2.1 implies that the local linear estimator with stationary or locally stationary regressors shares the same convergence rate $\sqrt{Th}$ but with different asymptotic bias and variances. The focus of this chapter is only on the case that the regressor x_t is univariate for simplicity, and the same manner can be extended to the multivariate case. To estimate $V_\varepsilon(\tau)$, it suffices to estimate $\Sigma_S(\tau)$ and $\Pi(\tau)$. A consistent estimate of $\Sigma_S(\tau)$ is just the sample moment $S_T(\tau)$. $\widehat{\Pi}(\tau) = \tilde{\sigma}_1 \left(\frac{1}{\sqrt{Th}} K\left(\frac{\tau_t - \tau}{h}\right) z_{t-1}\widehat{\varepsilon}_t\right)^\top \left(\frac{1}{\sqrt{Th}} K\left(\frac{\tau_t - \tau}{h}\right) z_{t-1}\widehat{\varepsilon}_t\right)$, where $\widehat{\varepsilon}_t = x_t - \widehat{\phi}(\tau_t)^\top z_{t-1}$. Thus, $\widehat{V}_\varepsilon(\tau) = [\mathrm{I}_2, 0_2]\widehat{\Sigma}_S^{-1}(\tau)\widehat{\Pi}(\tau)\widehat{\Sigma}_S^{-1}(\tau)^\top[\mathrm{I}_2, 0_2]^\top$, see, for example, Cai (2007).

Theorem 2.2. *Under the assumptions, the central limit theorem for $\widehat{\lambda}$ is:*

$$\sqrt{T}\left(\widehat{\lambda} - \lambda\right) \xrightarrow{d} N\Big(0, \quad V_\lambda\Big),$$

where the asymptotic variance $V_\lambda = \sigma_\eta^2 \sigma_\varepsilon^{-2}$.

The above theorem shows that the profile estimator of λ can achieve the root-T convergence rate in the classical linear regression models, consistent with Cai and Wang (2014) who propose a linear predictive regression model with a projection function as well. Further, our estimator is also consistent with some prior literature on the profile estimation of semiparametric time-series model containing time-varying and constant coefficients (see, for example, Li et al. (2011b), in which they restrict the explanatory variables to be stationary). Economically, testing the significance of λ determines whether there is a correlation between the innovations in stock returns and the predictors. If λ is significant, the projection function would be able to convey information from the time-varying AR process for predictors to make more precise estimation of the time-varying predictive regression equation. In contrast, the time-varying predictive regression with the linear projection approach will be very similar to the regression without the linear projection. The covariance matrix can be consistently estimated, $\widehat{V}_\lambda = \widehat{\sigma_\eta^2}/\widehat{\sigma_\varepsilon^2}$, where $\widehat{\sigma_\varepsilon^2}$ is the squared residuals from the time-varying AR model (2.5) and $\widehat{\sigma_\eta^2}$ is the squared residuals from the time-varying predictive regression model (2.6).

Theorem 2.3 (CLT). *Under the assumptions, the asymptotic distribution of $\widehat{\theta}(\tau)$ at a point τ is:*

$$\sqrt{Th}\left(\widehat{\theta}(\tau) - \theta(\tau) - \bar{B}_T(\tau) - o_p(h^2)\right) \xrightarrow{d} N\Big(0, \quad \boldsymbol{V}_\eta(\tau)\Big), \tag{2.20}$$

where the asymptotic variance is $\boldsymbol{V}_\eta(\tau) = [I_2, 0_2]\Sigma_S^{-1}(\tau)\widetilde{\Pi}(\tau)\Sigma_S^{-1}(\tau)^\top[I_2, 0_2]^\top$, and $\widetilde{\Pi}(\tau) = \begin{pmatrix} \widetilde{\pi}_0(\tau) & \widetilde{\pi}_1(\tau)^\top \\ \widetilde{\pi}_1(\tau) & \widetilde{\pi}_2(\tau) \end{pmatrix}$ with $\widetilde{\pi}_l(\tau) = \begin{pmatrix} \sigma_\eta^2\tilde{\sigma}_l^2 & 0 \\ 0 & \sigma_\eta^2\sigma^2 B_{0,T\tau}(1)\tilde{\sigma}_l^2 \end{pmatrix}$ for $l = 0, 1, 2$.

Theorem 2.3 in the semiparametric predictive regression model shares the same limiting distribution at the same convergence rate $\sqrt{Th}$ with the local linear estimators Theorem 2.1 in the time-varying AR(1) model. The result implies that the nonparametric parts dominate the asymptotic theory of time-varying estimators, and the only difference is the information from the error terms. The standard errors of $\widehat{\beta}_t$ in the model with the linear projection are smaller than that in the model without the linear projection, because some information from ε_t have been conveyed to the predictive regressions. Similarly, to estimate $\boldsymbol{V}_\eta(\tau)$, it suffices to estimate $\Sigma_S(\tau)$ and $\widetilde{\Pi}(\tau)$. A consistent estimate of $\Sigma_S(\tau)$ is just the sample moment $S_T(\tau)$. $\widehat{\widetilde{\Pi}}(\tau) = \tilde{\sigma}_1\left(\frac{1}{\sqrt{Th}}K\left(\frac{\tau_t-\tau}{h}\right)z_{t-1}\widehat{\eta}_t\right)^\top\left(\frac{1}{\sqrt{Th}}K\left(\frac{\tau_t-\tau}{h}\right)z_{t-1}\widehat{\eta}_t\right)$, where $\widehat{\eta}_t = y_t - \widehat{\phi}(\tau_t)^\top z_{t-1} - \widehat{\lambda}\widehat{\varepsilon}_t$. Thus, $\widehat{\boldsymbol{V}}_\eta(\tau) = [\mathrm{I}_2, 0_2]\widehat{\Sigma}_S^{-1}(\tau)\widehat{\widetilde{\Pi}}(\tau)\widehat{\Sigma}_S^{-1}(\tau)^\top[\mathrm{I}_2, 0_2]^\top$.

2.4 Simulation study

In this section, we conduct a Monte Carlo simulation study to illustrate the finite sample performance of the proposed two-step estimation procedure in

$$y_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + \lambda\widehat{\varepsilon}_t + \eta_t \tag{2.21}$$

$$x_t = \gamma(\tau_t) + \rho(\tau_t)x_{t-1} + \varepsilon_t, \tag{2.22}$$

in which $\widehat{\varepsilon}_t$ is the residuals obtained by estimating (2.22). The Epanechnikove kernel $K(u) = \frac{3}{4}\left(1-u^2\right)\mathrm{I}\{|u| \leq 1\}$ is used throughout the simulation study. The parameters are estimated under three different bandwidths around the optimal choice: $T^{-1/5}$, $0.5 *$

$T^{-1/5}$, $T^{-1/5}$, and $1.5 * T^{-1/5}$. We set sample sizes, $T = 100, 300, 500$ and 800, and number of replications, $M = 1000$. The estimation performance of both the local linear estimators of time-varying parameters, $\alpha(\tau_t)$, $\beta(\tau_t)$, $\gamma(\tau_t)$ and $\rho(\tau_t)$ is evaluated by the mean absolute error (MAE):

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T} |\widehat{f}(\tau_t) - f(\tau_t)|. \tag{2.23}$$

We calculate the sample mean and sample standard deviation of the 1000 MAEs for the nonparametric local linear estimators. Moreover, we evaluate the OLS estimator of constant parameter, λ, by calculating bias and standard deviation of 1000 OLS estimates,

$$\text{Bias} = \frac{1}{M}\sum_{j=1}^{M}\left(\widehat{\lambda}_j - \lambda\right), \quad \text{SD} = \sqrt{\frac{1}{M-1}\sum_{j=1}^{M}\left[\widehat{\lambda}_j - \left(\frac{1}{M}\sum_{i=1}^{M}\widehat{\lambda}_i\right)\right]^2}. \tag{2.24}$$

We consider three cases in simulations: **Case 1**. We set true DGPs of the time-varying parameter $\gamma(\tau)$ and $\rho(\tau)$ bounded in $[-1, 1]$ in the model (2.21) $\gamma(\tau) = -0.1 \times \sin(2\pi\tau)$ and $\rho(\tau) = 0.2 \times \cos(2\pi\tau)$, where the error ε_t is randomly drawn from the standard normal distribution $\text{N}(0, 0.25)$. We set true DGPs of the time-varying parameter $\alpha(\tau)$ and $\beta(\tau)$ in the model (2.22) $\alpha(\tau) = 0.5\tau + \exp\left(-16(\tau - 0.5)^2\right) - 0.5$ and $\beta(\tau) = 0.3 \times \left(0.5\tau + \exp\left(-14(\tau - 0.5)^2\right) - 0.5\right)$, where the error η_t is randomly drawn from the standard normal distribution $\text{N}(0, 0.25)$. We set $\lambda = 0.1$ in order to mimic the empirical correlation between stock returns and inflation. **Case 2**. We set true DGPs of the time-varying parameter $\gamma(\tau)$ and $\rho(\tau)$ bounded in $[-1, 1]$ in the model (2.21) $\gamma(\tau) = -0.2 \times \sin(2\pi\tau)$ and $\rho(\tau) = 0.5 \times \cos(2\pi\tau)$, where the error ε_t is randomly drawn from the standard normal distribution $\text{N}(0, 0.25)$. We set true DGPs of the time-varying parameter $\alpha(\tau)$ and $\beta(\tau)$ in the model (2.22) $\alpha(\tau) = \tau + \exp\left(-16(\tau - 0.5)^2\right) - 0.5$ and $\beta(\tau) = 0.3 \times \left(0.5\tau + \exp\left(-14(\tau - 0.5)^2\right) - 0.5\right)$, where the error η_t is randomly drawn from the standard normal distribution $\text{N}(0, 0.25)$. We set $\lambda = -0.9$ in order to mimic the empirical correlation between stock returns and financial ratios. **Case 3**. We set true DGPs of the time-varying parameter $\gamma(\tau)$ and $\rho(\tau)$ bounded in $[-1, 1]$ in the model (2.21) $\gamma(\tau) = -0.2 \times \sin(2\pi\tau)$ and $\rho(\tau) = 0.9 \times \cos(2\pi\tau)$ in order to mimic the high persistence in the dividend-price ratio, where the error ε_t is randomly drawn from the standard normal distribution $\mathrm{N}(0, 0.25)$. We set true DGPs of the time-varying parameter $\alpha(\tau)$ and $\beta(\tau)$ in the model (2.22) $\alpha(\tau) = \tau + \exp\left(-16(\tau - 0.5)^2\right) - 0.5$ and $\beta(\tau) = 0.3 \times \left(0.5\tau + \exp\left(-14(\tau - 0.5)^2\right) - 0.5\right)$, where the error η_t is randomly drawn from the standard normal distribution $\mathrm{N}(0, 0.25)$. We set $\lambda = 0.5$.

Table 2.1: Monte Carlo Simulation for Case 1

$h = 0.5T^{-1/5}$	T = 100	T=300	T=500	T = 800
γ	0.0459	0.0272	0.0218	0.0164
	0.0151	0.0082	0.0062	0.0042
ρ	0.1660	0.1012	0.0813	0.0613
	0.0503	0.0288	0.0225	0.0163
α	0.0623	0.0401	0.0326	0.0250
	0.0156	0.0097	0.0077	0.0056
β	0.1865	0.1079	0.0855	0.0630
	0.0594	0.0330	0.0249	0.0168
λ	-0.0025	0.0025	-0.0006	-0.0003
	0.1057	0.0605	0.0450	0.0294
$h = T^{-1/5}$	T = 100	T = 300	T = 500	T = 800
γ	0.0420	0.0277	0.0227	0.0176
	0.0137	0.0081	0.0065	0.0050
ρ	0.1348	0.0823	0.0652	0.0480
	0.0548	0.0298	0.0235	0.0171
α	0.1131	0.0852	0.0745	0.0639
	0.0105	0.0062	0.0051	0.0040
β	0.1569	0.0934	0.0733	0.0532
	0.0606	0.0320	0.0250	0.0181
λ	-0.0030	-0.0028	-0.0002	0.0001
	0.1178	0.0622	0.0471	0.0320
$h = 1.5T^{-1/5}$	T = 100	T = 300	T = 500	T = 800
γ	0.0440	0.0342	0.0301	0.0261
	0.0108	0.0062	0.0049	0.0037
ρ	0.1234	0.0787	0.0632	0.0476
	0.0507	0.0305	0.0233	0.0162
α	0.1633	0.1327	0.1187	0.1046
	0.0090	0.0046	0.0035	0.0024
β	0.1655	0.0983	0.0788	0.0592
	0.0669	0.0350	0.0249	0.0148
λ	-0.0054	0.0017	-0.0009	-0.0036
	0.1243	0.0643	0.0494	0.0345

Note. This table reports mean (first row) and standard deviation (second row) of the MAEs for the nonparametric local linear estimators across bandwidth choices: $0.5T^{-1/5}$, $1.5T^{-1/5}$, and $1.5T^{-1/5}$, and reports bias and SD for the for parametric OLS estimator. True DGPs are generated by case 1.

Table 2.2: Monte Carlo Simulation for Case 2

$h = 0.5T^{-1/5}$	T = 100	T = 300	T = 500	T = 800
γ	0.0454	0.0277	0.0219	0.0178
	0.0155	0.0084	0.0062	0.0049
ρ	0.1653	0.1022	0.0815	0.0661
	0.0506	0.0301	0.0224	0.0177
α	0.0613	0.0397	0.0323	0.0267
	0.0158	0.0094	0.0075	0.0060
β	0.1845	0.1071	0.0848	0.0680
	0.0585	0.0316	0.0235	0.0187
λ	-0.0023	0.0019	0.0002	-0.0013
	0.1047	0.0593	0.0481	0.0337
$h = T^{-1/5}$	T = 100	T = 300	T = 500	T = 800
γ	0.0423	0.0345	0.0231	0.0190
	0.0138	0.0087	0.0066	0.0053
ρ	0.1351	0.0908	0.0641	0.0487
	0.0542	0.0299	0.0231	0.0184
α	0.1135	0.0805	0.0747	0.0691
	0.0106	0.0068	0.0053	0.0038
β	0.1586	0.0943	0.0738	0.0570
	0.0611	0.0310	0.0245	0.0197
λ	-0.0015	-0.0023	0.0002	-0.0013
	0.1138	0.0584	0.0467	0.0322
$h = 1.5T^{-1/5}$	T = 100	T = 300	T = 500	T = 800
γ	0.0438	0.0323	0.0301	0.0260
	0.0106	0.0055	0.0051	0.0037
ρ	0.1244	0.1034	0.0639	0.0485
	0.0552	0.0363	0.0240	0.0193
α	0.1630	0.1234	0.1187	0.1131
	0.0089	0.0052	0.0035	0.0020
β	0.1640	0.1030	0.0794	0.0627
	0.0696	0.0306	0.0256	0.0207
λ	0.0026	-0.0011	-0.0008	0.0003
	0.1248	0.0655	0.0500	0.0355

Note. This table reports mean (first row) and standard deviation (second row) of the MAEs for the nonparametric local linear estimators across bandwidth choices: $0.5T^{-1/5}$, $1.5T^{-1/5}$, and $1.5T^{-1/5}$, and report bias and SD for the for parametric OLS estimator. True DGPs are generated by case 2.

The results of the simulation study case 1 are summarized in Table 2.1. For the case 1, the true value of λ is equal to 0.1 in order to mimic the empirical evidence by inflation as the predictor. The overall simulation results indicate that the local linear estimators are not very sensitive to different bandwidth choices, especially when we increase the sample size from 100 to 800. Under all bandwidth choices, 0.5h$_{op}$, h$_{op}$,

Table 2.3: Monte Carlo Simulation for Case 3

$h = 0.5T^{-1/5}$	T = 100	T = 300	T = 500	T = 800
γ	0.1144	0.0552	0.0271	0.0175
	0.0359	0.0187	0.0077	0.0033
ρ	0.1763	0.1093	0.0835	0.0656
	0.0549	0.0326	0.0248	0.0158
α	0.1277	0.0690	0.0366	0.0149
	0.0195	0.0384	0.0097	0.0097
β	0.1807	0.1065	0.0835	0.0566
	0.0603	0.0319	0.0248	0.0158
λ	-0.0056	0.0009	0.0005	0.0003
	0.1092	0.0608	0.0464	0.0314
$h = T^{-1/5}$	T = 100	T = 300	T = 500	T = 800
γ	0.1103	0.0652	0.0418	0.0364
	0.0123	0.0334	0.0071	0.0062
ρ	0.1874	0.1340	0.0993	0.0841
	0.0508	0.0314	0.0224	0.0188
α	0.1013	0.1278	0.0694	0.0620
	0.0150	0.0357	0.0066	0.0057
β	0.1598	0.0897	0.0872	0.0735
	0.0511	0.0308	0.0254	0.0209
λ	-0.0121	-0.0045	0.0012	0.0009
	0.1048	0.0594	0.0468	0.0363
$h = 1.5T^{-1/5}$	T = 100	T = 300	T = 500	T = 800
γ	0.1336	0.0753	0.1184	0.0845
	0.0811	0.0249	0.0200	0.0200
ρ	0.2532	0.1859	0.1644	0.1360
	0.0446	0.0251	0.0205	0.0175
α	0.1494	0.1634	0.1393	0.0927
	0.0136	0.0352	0.0258	0.0208
β	0.1486	0.0895	0.0758	0.0584
	0.0525	0.0296	0.0215	0.0215
λ	-0.0619	-0.0155	-0.0119	-0.0102
	0.1045	0.0609	0.0452	0.0223

Note. This table reports mean (first row) and standard deviation (second row) of the MAEs for the nonparametric local linear estimators across bandwidth choices: $0.5T^{-1/5}$, $1.5T^{-1/5}$, and $1.5T^{-1/5}$, and reports bias and SD for the for parametric OLS estimator. True DGPs are generated by case 3.

and 1.5h$_{op}$, local linear estimators have smaller means, and standard deviations of MAE and OLS estimators have smaller bias and standard deviation when the sample size increases from 100 to 800. In time-varying AR(1) model, the local linear estimators $\widehat{\gamma}(\cdot)$ and $\widehat{\rho}(\cdot)$ have favorable finite sample properties in terms of bias and standard deviation. When we choose the optimal bandwidth, h$_{op}$, there is a decline in mean of MAE for

$\widehat{\gamma}(\cdot)$, for example, 0.1131 for $T = 100$, and -0.0639 for $T = 800$, and also a persistent decline in the standard deviation of MAE, 0.0105 for $T = 100$, and 0.004 for $T = 800$. For asymptotic properties, we can see that these two quantities rapidly decrease as the sample size increases from 250 to 800. Furthermore, an attractive property is that even though the true DGP of coefficient parameters are constant, the local linear estimates still perform well as the sample size gets large.

The results of the simulation study case 2 are summarized in Table 2.2. For case 2, the true value of λ is equal to -0.9 in order to mimic the empirical evidence using the dividend-price ratio as the predictor. The overall simulation results indicate that the the local linear estimators are not very sensitive to different bandwidth choices, particularly when we increase the sample size from 100 to 800. Under all bandwidth choices, $0.5h_{op}$, h_{op}, and $1.5h_{op}$, the local linear estimates have smaller means, and standard deviations of MAE and OLS estimators have smaller bias and standard deviation when the sample size increases from 100 to 800.

The last case is similar to case 2, but sets the persistence parameter $\rho(\tau) = 0.9 \times \cos(2\pi\tau)$ to mimic the high persistence in empirical predictors, such as dividend-price ratio or earning-price ratio. The results of the simulation study case 3 are summarized in Table 2.3. We again observe that local linear estimates have smaller means, and standard deviations of MAE and OLS estimators have smaller bias and standard deviation when the sample size increases from 100 to 800. It should be noticed that the biases for all the parameters are slightly bigger than the results in either Table 2.1 or Table 2.2, suggesting the finite sample bias increases with the persistence of the predictor.

2.5 Summary

In this chapter, we study a time-varying coefficient predictive regression model with a time-varying AR process for the predictor to re-examine the predictability of the equity premium. In particular, we show that under some conditions, any tv-AR(p) model can be written as a time-varying MA(∞) process, which is a class of locally stationary processes. We developed a nonparametric local linear method for estimating the trend function and coefficient functions where, similar to Phillips and Solo (1992), we extend the BN lemma to a time-varying version and then fully establish the asymptotic properties for those estimators.

Chapter 3

In-sample Predictive Regression Analysis

In this chapter, we implement the proposed time-varying predictive methodologies to study the predictability of US aggregate market returns. The empirical analysis focuses on in-sample predictability because the proposed models aim to make a reliable statistical inference for the aforementioned econometric issues. Moreover, we extend the time-varying models for predictive regressions with a single predictor into a multivariate framework. The multi-predictor predictive regression models are more general in econometrics side of return-forecasting literature (see, for example, Amihud et al. (2008), Kostakis et al. (2015), and Xu (2020)). We consider a time-varying VAR(1) model for the multiple predictors. One predictor in time-varying VAR(1) model is allowed to be not only autocorrelated but also determined simultaneously with many other predictors. For example, the dividend-price ratio and earning-price ratio are highly persistent and also highly correlated since stock price shares the same denominator for both ratios.

3.1 Predictive regression model

Amihud et al. (2008) propose a multi-predictor predictive regression model

$$y_t = \alpha + \beta^\top x_{t-1} + e_t, \tag{3.1}$$

$$x_t = \gamma + \rho x_{t-1} + u_t. \tag{3.2}$$

where y_t is a scalar dependent variable, $x_t = \left(x_{1,t}, ..., x_{p,t}\right)^\top$ is a $p \times 1$ vector of predictive variables, α is a scalar intercept, $\beta = \left(\beta_1, ..., \beta_p\right)^\top$ is a $p \times 1$ vector of forecasting coefficients, e_t is a scalar error term, $\gamma = \left(\gamma_1, ..., \gamma_p\right)^\top$ is a $p \times 1$ vector of intercepts, $\rho = \left(\rho_1, ..., \rho_p\right)$ is a $p \times p$ matrix, with $\rho_j = \left(\rho_{1,j}, ..., \rho_{p,j}\right)^\top$ for $j = 1, 2, ..., p$, $u_t = \left(u_{1,t}, ..., u_{p,t}\right)^\top$ is a $p \times 1$ vector of error terms so that the vector $\left(e_t, u_t^\top\right)^\top$ are independent multivariate normal with zero mean. Amihud et al. (2008) also consider a linear projection function such that there exists a $p \times 1$ vector $e_t = \lambda^\top u_t + \eta_t$, where η_t is normal random variable with mean zero, independent of both u_t and x_t.

The system of predictive models (3.1) and (3.2) is a multi-predictor extension of the single-predictor predictive regression considered in Amihud and Hurvich (2004). The models of Amihud et al. (2008) have several important implications for empirical research. Previous empirical research focuses on the single-predictor forecasting models in which the stock returns are predicted by a lagged predictor, including dividend-price ratio, earning-price ratio, book-to-market ratio, interest rate, term spread, and others. The single-predictor forecasting framework can only accommodate testing the predictability of each financial variable separately, which may omit information from other variables. For example, Cochrane (2011) finds that combining the dividend-price ratio and the consumption to wealth ratio (Lettau and Ludvigson, 2001) has stronger predictability for expected returns than using the two predictors alone. Further, Møller and Sander (2017) find that the dividend yield and earnings yield are strong predictors of dividend growth according to the Lintner's dividend model. Moreover, the predictor variables are autoregressive, and some are correlated with other predictors. For example, dividend-price ratio and earning-price ratio serving as stock valuation ratios are highly correlated as the two ratios share the same denominator–stock price. The VAR(1) model (3.2) is a natural choice for the case that one predictor, in addition to being autoregressive, is also determined simultaneously with other predictors.

This chapter is largely motivated by the constant-coefficients multi-predictor models of Amihud et al. (2008). We propose a nonparametric time-varying extension of the multi-predictor predictive regressions (Amihud et al., 2008). We specify a time-varying multi-predictor predictive regression model with a multi-dimensional time-varying VAR(1) process for the predictors as follows:

$$y_t = \alpha(\tau_t) + \beta(\tau_t)^\top x_{t-1} + e_t, \tag{3.3}$$

$$x_t = \gamma(\tau_t) + \rho(\tau_t) x_{t-1} + \varepsilon_t, \tag{3.4}$$

where $\alpha(\tau_t)$ is a scalar time-varying trend function, $\beta(\tau_t) = \left(\beta_1(\tau_t), ..., \beta_p(\tau_t)\right)^\top$, is a $p \times 1$ vector of time-varying coefficients, $\gamma(\tau_t) = \left(\gamma_1(\tau_t), ..., \gamma_p(\tau_t)\right)^\top$, a $p \times 1$ vector of time-varying trend functions, and $\rho(\tau_t) = \left(\rho_1(\tau_t), ..., \rho_p(\tau_t)\right)$ is a $p \times p$ matrix, where each $\rho_j(\tau_t)$ is a $p \times 1$ vector, $\rho_j(\tau_t) = \left(\rho_{1,j}(\tau_t), ..., \rho_{p,j}(\tau_t)\right)^\top$ for $j = 1, 2, ..., p$. We specify that the time-varying function f_t depends on the sample size T, $f_t = f(\tau_t)$. This formulation is necessary to provide the asymptotic justification for any nonparametric smoothing estimators in models (3.3) and (3.4). In the models, e_t is a scalar error term and $\varepsilon_t = \left(\varepsilon_{1,t}, ..., \varepsilon_{p,t}\right)^\top$ is a $p \times 1$ vector of error terms. Following Amihud et al. (2008), we allow the error terms e_t and ε_t to be contemporarily correlated:

$$e_t = \lambda^\top \varepsilon_t + \eta_t, \tag{3.5}$$

where η_t is independently and identically distributed.

Differing to Amihud et al. (2008), this chapter's main contribution is to consider a time-varying AR(1)(VAR(1)) model for the predictor(s), and this has not been considered in the existing literature on stock return prediction. The nonparametric time-varying coefficients functions allow for the time-varying forecasting relationship between future stock returns and predictors. The nonparametric specification of time-varying coefficients does not impose parametric functional form on the regression coefficients and it is entirely data-driven.

Our model has two important practical implications. First, the model (3.4) allows the coefficients to change over time. Therefore, the persistence in each predictor is allowed to be time-varying. The time-varying persistence of financial ratios is not uncommon due to the changes in the steady-state growth rate of economic fundamentals or changes in the expected return of equity. It is also important to note that variables in the VAR system are allowed to be contemporaneously correlated (e.g., dividend yield and earning-price ratio). The time-varying VAR(1) model also allows for the correlation between any two predictors to be time-varying. Second, the time-varying VAR(1) model allows for the time-variation in the steady-state mean of predictors. Shifts in the steady-state mean of predictors have important economic impacts on return predictability. For instance, Lettau and Nieuwerburgh (2008) point out that change in the steady-state mean of financial ratios over the sample period can explain a lot of predictability puzzles. Lettau and Nieuwerburgh (2008) show permanent changes in the mean of several price ratios by structural break tests. In real-time, however, the turning point of the changes in the steady-state of financial ratios or the timing of new means of financial ratios are difficult to detect. The time-varying VAR(1) model uses only local information and therefore can reflect the gradual movements in the steady-state of predictor variables at each time.

We study the predictability of U.S. aggregate stock returns with a sample period from 1927:I to 2018:IV using a list of commonly employed variables (Welch and Goyal, 2008). Our thesis focuses on in-sample predictability because the proposed methodology aims to robustify in-sample inference for time-varying structure in coefficients and predictors' unknown persistence properties. We estimate the time-varying coefficients by nonparametric local linear estimation methods. Kernel smoothing identifies the time variation of conditional parameters and provides period-by-period estimates of conditional parameters on fixed windows equally weighting all observations in that window Ang and Kristensen (2012). We present our empirical results in two ways: (1) univariate time-varying predictive regressions with time-varying AR(1) predictor in line of Welch and Goyal (2008), and (2) multivariate time-varying predictive regressions with time-varying VAR(1) predictors in line of Amihud et al. (2008).

The main results are summarised here. First, we find significant time-variation in the coefficients of the 14 popular predictors (Welch and Goyal, 2008) in the time-varying AR(1) models. The estimated coefficients of individual predictors change smoothly over time, and the absolute values of $\widehat{\rho}(\tau_t)$ are generally less than one, showing the validity of local stationarity. We also find that the time-varying AR(1) models fit real data better than the linear AR(1) models with constant coefficients, measured by in-sample R^2. Second, we find that the estimated coefficients $\widehat{\beta}(\tau_t)$ of all predictors are time-varying. The OLS estimates of coefficient in the linear regression model generally lie outside of the confidence intervals of the time-varying local linear estimates. More importantly, the 14 popularly used predictors indeed contain predictive content of stock returns. Like the dividend-price ratio and book-to-market ratio, the fundamental factors have strong predictability over the majority of the subsamples. The predictability is incredibly strong during economic recession periods, such as the oil shock of 1973–1974 and the global financial crisis of 2007–2009. INF seems to be the weakest predictor among the 14 predictors, but predictability remains significant over about half of the full sample.

3.2 Estimation procedure

In this section, we estimate the model by nonparametric kernel smoothing which is well-developed and increasingly popular in econometrics literature. See, e.g., Cai (2007), Li et al. (2011a), Chen et al. (2012), Cai et al. (2015), and Phillips et al. (2017). Kernel smoothing, such as local constant or local linear estimation methods, can be treated as *locally weighted least squares*, which puts higher weights on the local information in estimation around any particular time point. The benefits of employing the local linear method in the nonparametric regression model in the current literature are its attractive statistical properties compared to the local constant method, such as efficiency, bias reduction, and adaptation of boundary effects. In this case, we use local linear fitting throughout the empirical chapter.

Plugging the multivariate linear projection equation (3.5) into the time-varying predictive regression (3.3), we have the following regressions in matrix form,

$$y_t = \theta(\tau_t)^\top z_{t-1} + \lambda^\top \varepsilon_t + \eta_t, \tag{3.6}$$

$$x_t = \phi(\tau_t)^\top z_{t-1} + \varepsilon_t, \tag{3.7}$$

where $z_t = (1, x_{1,t}, ..., x_{p,t})^\top$ is the $(p+1) \times 1$ vector, $\theta(\tau_t) = (\alpha(\tau_t), \beta_1(\tau_t), ..., \beta_p(\tau_t))^\top$ is the $(p+1) \times 1$ vector, $\phi(\tau_t) = \left(\gamma(\tau_t), \rho_1(\tau_t), ..., \rho_p(\tau_t)\right)^\top$ is the $(p+1) \times p$ matrix, and η_t is assumed to be an i.i.d. process with zero mean and finite variance, σ_η^2. When $p = 1$, the models (3.6) and (3.7) reduce to the system of univariate predictive regressions in Chapter 2. We use the local linear estimation throughout this thesis. For a more comprehensive and rigorous treatment, see Chen et al. (2012) and Li et al. (2011b). Since ε_t is unobservable, a two-step estimation approach is employed. **Step 1**: Estimate the time-varying VAR(1) model (3.7) to obtain the local linear estimates of $\phi(\tau_t)$ and the residual $\widehat{\varepsilon}_t$. **Step 2**: Estimate the time-varying predictive regression model (3.6) and obtain $\widehat{\theta}(\tau_t)$ and $\widehat{\lambda}$.

First, we construct local linear estimators for the time-varying coefficients. Recall the time-varying VAR(1) model:

$$x_t = \phi(\tau_t)^\top z_{t-1} + \varepsilon_t. \tag{3.8}$$

By assuming that each element in $\phi(\tau_t)$ has continuous second derivatives, then $\phi(\tau_t)$ can be approximated by a linear matrix function using Taylor expansion at any fixed and re-scaled time point τ in [0,1]. That is:

$$\gamma_j(\tau_t) \approx \gamma_j(\tau) + \gamma_j'(\tau)(\tau_t - \tau), \tag{3.9}$$

$$\rho_{ij}(\tau_t) \approx \rho_{ij}(\tau) + \rho_{ij}'(\tau)(\tau_t - \tau), \quad \text{for all} \quad i \quad \text{and} \quad j. \tag{3.10}$$

We define the objective function as the weighted sum of squared residuals:

$$\sum_{t=2}^{T} \left[x_t - \left(\phi(\tau) + \phi'(\tau)\left(\tau_t - \tau\right)\right) z_{t-1}\right]^\top \left[x_t - \left(\phi(\tau) + \phi'(\tau)\left(\tau_t - \tau\right)\right) z_{t-1}\right] K\left(\frac{\tau_t - \tau}{h}\right), \tag{3.11}$$

where $\phi'(\tau_t)$ is the first-order derivative of $\phi(\tau_t)$. Moreover, h is the bandwidth satisfying the condition that $T \to \infty$, $h \to 0$ and $Th \to \infty$, and $K(\cdot)$ is the kernel function. The Gaussian kernel function is used throughout this chapter.

Minimising the objective function with respect to $\phi(\tau)$ gives the local linear estimator of $\phi(\tau)$,

$$\widehat{\phi}(\tau) = [\boldsymbol{I}_{p+1}, \mathbf{0}_{p+1}] \left(D(\tau)^\top W(\tau) D(\tau)\right)^{-1} D(\tau)^\top W(\tau) X, \tag{3.12}$$

where $W(\tau) = \operatorname{diag}\left(K\left(\frac{2-\tau T}{Th}\right), ..., K\left(\frac{T-\tau T}{Th}\right)\right)$, $\boldsymbol{I}_{p+1}$ is the $(p+1) \times (p+1)$ identity matrix, $\mathbf{0}_{p+1}$ is the $(p+1) \times (p+1)$ null matrix,

$$D(\tau) = \begin{pmatrix} z_1^\top & z_1^\top \frac{2-\tau T}{Th} \\ \vdots & \vdots \\ z_{T-1}^\top & z_{T-1}^\top \frac{T-\tau T}{Th} \end{pmatrix}, \quad \text{and} \quad X = \begin{pmatrix} x_{1,2} & \cdots & x_{p,2} \\ \vdots & \ddots & \vdots \\ x_{1,T} & \cdots & x_{p,T} \end{pmatrix}. \tag{3.13}$$

Kernel smoothing is a local estimation using observations around each τ, similar to a rolling kernel-weighted least squares. Sometimes we call the method as localised weighted least-squares in the sense of least-square principal. The residuals are calculated as $\widehat{\varepsilon}_t = x_t - \widehat{\phi}^\top(\tau_t) z_{t-1}$.

In step 2, we replace ε_t in the time-varying predictive model (3.6) with the residual, $\widehat{\varepsilon}_t$, and obtain a semiparametric predictive regression model,

$$y_t = \theta(\tau_t)^\top z_{t-1} + \lambda^\top \widehat{\varepsilon}_t + \eta_t. \tag{3.14}$$

By assuming that each element in $\theta(\tau_t)$ has continuous second derivatives, they can be approximated by a linear matrix function using Taylor expansion at any fixed τ, satisfying $0 \leq \tau \leq 1$. That is, $\theta_j(\tau_t) \approx \theta_j(\tau) + \theta_j'(\tau)(\tau_t - \tau)$ for any j-th element in θ. Assuming $\lambda^\top$ is known, we minimise the following objective function with respect to

$$\sum_{t=2}^{T} \left[\left(y_t - \lambda^\top \varepsilon_t \right) - \left(\theta(\tau) + \theta'(\tau) \left(\tau_t - \tau \right) \right)^\top z_{t-1} \right]^2 K \left(\frac{\tau_t - \tau}{h} \right). \tag{3.15}$$

The local linear estimator is given by

$$\widetilde{\theta}(\tau) = [\boldsymbol{I}_{p+1}, \mathbf{0}_{p+1}] \left(D(\tau)^\top W(\tau) D(\tau) \right)^{-1} D(\tau)^\top W(\tau)(Y - \widehat{\varepsilon}\lambda) = s(\tau)(Y - \widehat{\varepsilon}\lambda), \tag{3.16}$$

where $Y = (y_2, ..., y_T)^\top$ is a $(T-1) \times 1$ vector, $\widehat{\varepsilon} = (\widehat{\varepsilon}_2, ..., \widehat{\varepsilon}_T)^\top$ is a $(T-1) \times p$ matrix, $\widehat{\varepsilon}_t = \left(\widehat{\varepsilon}_{1,t}, ..., \widehat{\varepsilon}_{p,t} \right)^\top$, and $s(\tau) = [\boldsymbol{I}_{p+1}, \mathbf{0}_{p+1}] S(\tau)$ with $S(\tau) = \left(D(\tau)^\top W(\tau) D(\tau) \right)^{-1} D(\tau)^\top$ $W(\tau)$. The local linear estimator $\widetilde{\theta}(\tau)$ is a linear function of λ. A profile approach is that we use the local linear estimator to approximate the unknown parameter $\theta(\tau)$. The constant parameter λ, then, can be estimated by OLS. By minimising the following objective function

$$L(\lambda) = \sum_{t=2}^{T} \left(y_t - \widetilde{\theta}(\tau)^\top z_{t-1} - \lambda^\top \widehat{\varepsilon}_t \right)^2, \tag{3.17}$$

with respect to λ, we obtain the OLS estimator $\widehat{\lambda}$

$$\widehat{\lambda} = \left(\widetilde{X}^\top \widetilde{X} \right)^{-1} \widetilde{X}^\top \widetilde{Y}, \tag{3.18}$$

where $\widetilde{Y} = (I - \widetilde{s})Y, \quad \widetilde{X} = (I - \widetilde{s})\widehat{\varepsilon}$, and $\widetilde{s} = \left((z_1^\top s(\tau_2))^\top, ..., (z_{T-1}^\top s(\tau_T)^\top) \right)^\top$. Putting $\widehat{\lambda}$ into the infeasible local linear estimator $\widetilde{\theta}(\tau)$, we obtain a feasible local linear estimator

$$\widehat{\theta}(\tau) = [\boldsymbol{I}_{p+1}, \mathbf{0}_{p+1}] \left(D(\tau)^\top W(\tau) D(\tau) \right)^{-1} D(\tau)^\top W(\tau)(Y - \widehat{\varepsilon}\widehat{\lambda}) = s(\tau)(Y - \widehat{\varepsilon}\widehat{\lambda}). \tag{3.19}$$

In this thesis, we construct point-wise confidence intervals for the time-varying coefficients. The point-wise confidence intervals for the time-varying coefficients allow us to see how return predictability changes over time, which is the objective of this thesis. Hence, we aim to answer when stock returns are predictable. Instead, Chen and Hong

(2012) test the overall stability and significance of local linear coefficients in time-varying models but assume that predictors are stationary. Cai et al. (2015) further propose a L_2 test statistic to test the stability of the coefficients in a predictive model with nonstationary regressors. Testing overall predictability with locally stationary predictors is important to answer whether stock returns are predictable.

3.3 Empirical results

3.3.1 Data and summary statistics

To illustrate whether return predictability exists, we use the U.S. quarterly equity premium analysed in Welch and Goyal (2008), which is from 1927:I to 2018:IV available at Amit Goyal's website (`http://www.hec.unil.ch/agoyal/`). Stock returns are measured as continuously compounded returns on the S&P 500 index, including dividends, and the Treasury bill rate is used to compute the equity premium or excess return. For the predictor variables used to predict the equity premium, we use the fourteen popular predictors studied in Welch and Goyal (2008). We note that other predictors proposed in recent studies do not make our list because we can make a direct comparison between our time-varying results and the results in Welch and Goyal (2008). The brief descriptions of these predictors are listed here:

1. **Log dividend-price ratio [DP]**: log of a 12-month moving average sum of dividends paid on the S&P 500 index minus the log of stock prices (S&P 500 index).

2. **Log dividend yield [DY]**: log of a 12-month moving average sum of dividends paid on the S&P 500 index minus the log of lagged stock prices.

3. **Log earning-price ratio [EP]**:difference between the log of earnings on the S&P 500 index and the log of prices, where earnings are measured using a 12-month moving sum.

4. **Log dividend-payout ratio [DE]**: log of a 12-month moving sum of dividends minus the log of stock prices.

5. **Book-to-market ratio [BM]**: book-to-market value ratio for the Dow Jones Industrial Average (DJIA).

6. **Stock variance [SVAR]**: monthly sum of squared daily returns on the S&P 500 index.

7. **Net equity expansion [NTIS]** : ratio of a 12-month moving sum of net equity issues by NYSE-listed stocks to the total end-of-year market capitalisation of NYSE stocks.

8. **T-bill rate [TBL]**: interest rate on a three-month Treasury bill (secondary market).

9. **Long-term yield [LTY]**: long-term government bond yield.

10. **Long-term return [LTR]**: return on long-term government bonds.

11. **Term spread [TMS]**: difference between the long-term yield and the T-bill rate.

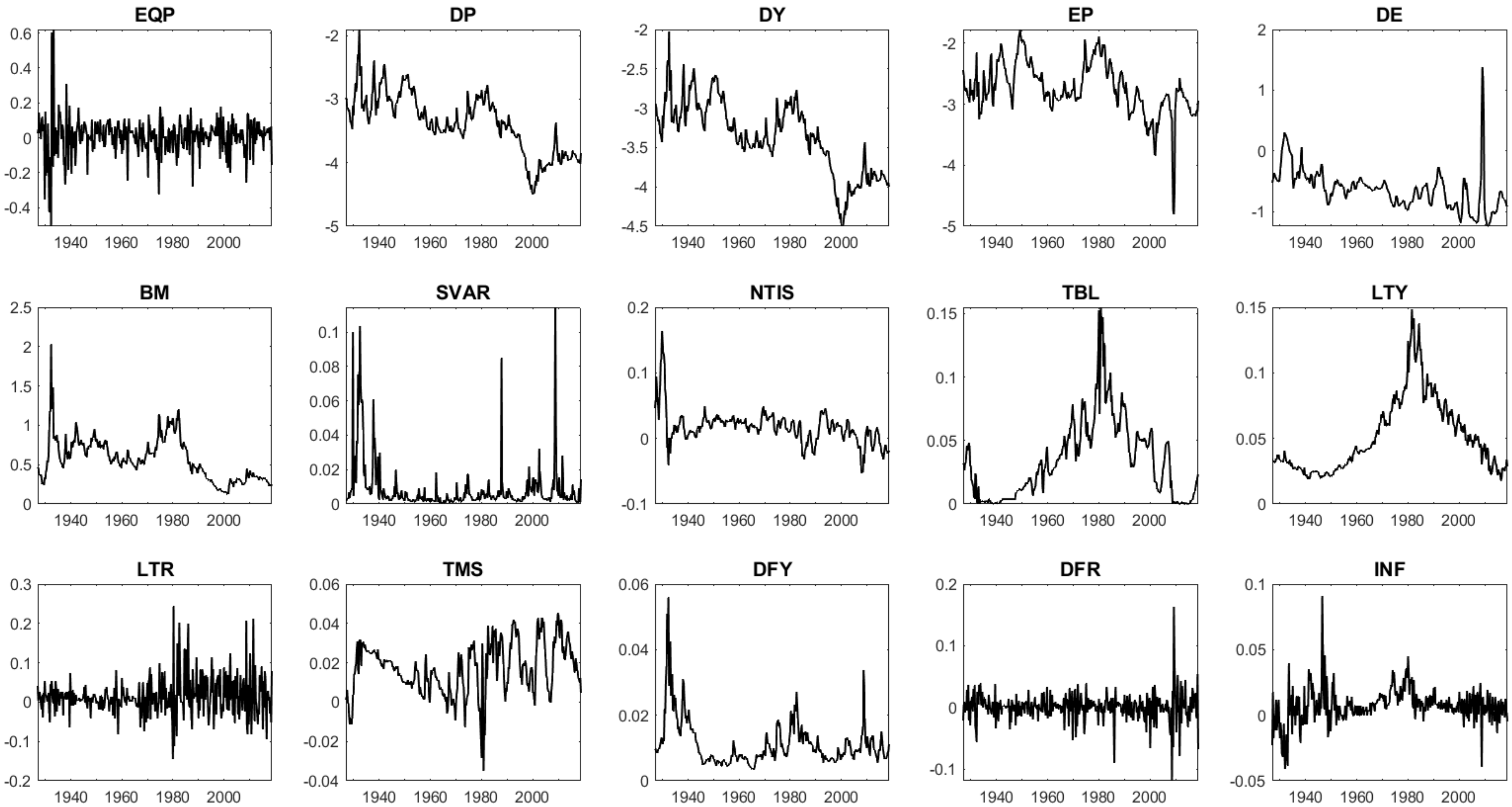
12. **Default yield spread [DFY]**: difference between Moodys BAA- and AAA- rated corporate bond yields.

13. **Default return spread [DFR]**: long-term corporate bond return minus the long-term government bond return.

14. **Inflation [INF]**: the Consumer Price Index (All Urban Consumers) from 1927 to 2017 from the Bureau of Labor Statistics.

Figure 3.1 plots the time series of stock return and the 14 predictors. Stock returns seem to be very stationary, but some of the financial ratios and macroeconomic variables display high persistence. The financial valuation ratios, including DP, DY, EP, and BM, are highly persistent, and hence the stationary assumption could be misleading for these predictors. Conversely, NTIS, LTR, DFR, and INF seem to be stationary and, therefore, the local-to-unity assumption may not be a good model for these stationary variables. The fact that persistence varies dramatically across the predictors motivates the time-varying AR(1) process.

Figure 3.1: Time Series Graphs of the Equity Premium and 14 Economic Predictors from 1927:I to 2018:IV



Note. This figure shows the time series plots of the quarterly log excess return and the 14 economic variables defined in the text. The sample period is from 1927:I to 2018:IV.

Table 3.1: Summary Statistics for the Equity Premium and 14 Predictors, 1927:I - 2018:IV

Panel A: 1927:I to 2018:IV							
	Mean	SD.	Skew.	Kurt.	Max	Min	Per.
EQP	0.006	0.104	0.034	10.945	0.620	-0.506	-0.030
DP	-3.379	0.466	-0.167	2.654	-1.904	-4.493	0.973
DY	-3.364	0.460	-0.251	2.585	-2.033	-4.497	0.975
EP	-2.741	0.422	-0.644	5.826	-1.775	-4.807	0.936
DE	-0.637	0.334	1.623	9.544	1.380	-1.244	0.930
BM	0.570	0.268	0.846	4.840	2.028	0.125	0.944
SVAR	0.009	0.015	4.294	24.107	0.114	0.000	0.621
NTIS	0.017	0.026	1.518	10.330	0.164	-0.053	0.936
TBL	0.034	0.031	1.107	4.386	0.155	0.000	0.963
LTY	0.051	0.028	1.111	3.696	0.148	0.018	0.987
LTR	0.014	0.047	1.063	7.096	0.244	-0.145	-0.042
TMS	0.017	0.013	-0.308	3.342	0.045	-0.035	0.853
DFY	0.011	0.007	2.503	11.898	0.056	0.003	0.898
DFR	0.001	0.022	0.172	14.149	0.163	-0.118	-0.074
INF	0.007	0.013	0.401	9.123	0.091	-0.041	0.501
Panel B: 1990:I to 2018:IV							
	Mean	SD.	Skew.	Kurt.	Max	Min	Per.
EQP	0.010	0.077	-0.870	4.114	0.178	-0.258	0.056
DP	-3.913	0.274	0.151	2.781	-3.253	-4.493	0.943
DY	-3.897	0.274	0.033	2.826	-3.232	-4.497	0.951
EP	-3.114	0.366	-2.261	10.203	-2.566	-4.807	0.864
DE	-0.800	0.422	2.827	13.639	1.380	-1.244	0.881
BM	0.294	0.085	0.084	2.781	0.520	0.125	0.929
SVAR	0.008	0.012	6.359	53.865	0.114	0.001	0.504
NTIS	0.005	0.021	-0.430	3.019	0.046	-0.053	0.934
TBL	0.027	0.022	0.253	1.778	0.079	0.000	0.959
LTY	0.051	0.019	0.141	2.087	0.091	0.018	0.960
LTR	0.020	0.054	0.588	4.046	0.212	-0.095	-0.019
TMS	0.024	0.012	-0.090	1.926	0.045	-0.002	0.919
DFY	0.010	0.004	3.339	19.169	0.034	0.006	0.800
DFR	0.000	0.030	0.632	11.071	0.163	-0.118	-0.061
INF	0.006	0.008	-1.347	9.629	0.025	-0.039	-0.099

Note. This table reports the mean, standard deviation (SD.), skewness (Skew.), kurtosis (Kurt.), maximum (Max.), minimum (Min.), and Persistence (Per.) for the quarterly log excess return and 14 popular predictors returns suggested by Welch and Goyal (2008) for the U.S. aggregate stock market. The persistence parameter is the estimated slope coefficient from an AR(1) model.

Summary statistics, including mean, standard deviation, skewness, kurtosis, maximum, and minimum, are reported in Table 3.1 for the quarterly equity premium and the 14 predictors over the full sample (1927:I - 2018:IV) in panel A and a recent sub-period

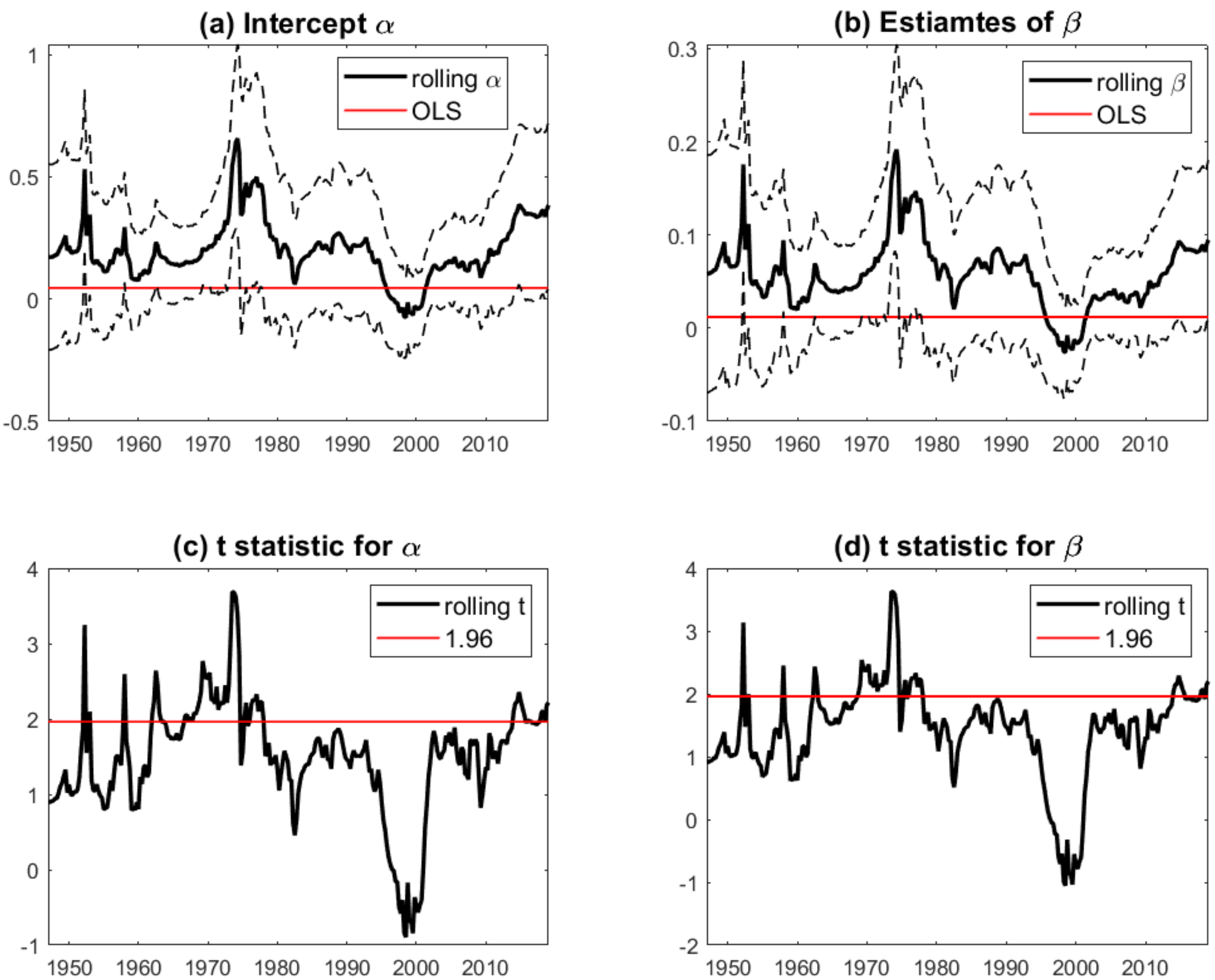
(1990:I - 2018:IV) in panel B. The equity premium has a quarterly mean of 0.6%, while the standard deviation is 10.46%, suggesting the high volatility in the U.S. stock market returns. The reported skewness of 0.1464 and kurtosis of 11.34 mean that stock returns are slightly right-skewed and have an extremely fat tail. In 1990:I - 2018:IV, the equity premium has a higher mean, 1%, but a smaller standard deviation, 7.7%, than those in the full sample. For the predictor variables, the sample means and standard deviations are different in the two periods. The sample mean of DP is from -3.379 to -3.913, but its standard deviation is from 46.6% to 27.4%. These findings are broadly consistent with the empirical evidence on the U.S. stock market in previous literature.

Financial ratios (DP, DY, EP, DE, and DM) are persistent, as their AR(1) coefficients are generally above 0.93 for the period 1927:I to 2018:IV reported in panel A, and the persistence tends to be smaller in a more recent period of 1990:I to 2018:IV, reported in panel B. For example, the persistence in DP drops from 0.973 to 0.943, and the persistence in DE drops from 0.93 to 0.881. Macroeconomic variables like LTR and DFR are, in contrast, stationary as their persistence is close to zero. Interestingly, INF is positively autocorrelated over the full sample with persistence equal to 0.501, but the persistence becomes -0.099 in the period of 1990:I to 2018:IV. The magnitude and sign of the AR(1) coefficient of INF are very different before and after 1990. These summary statistics reveal an essential fact that both the mean and persistence of the popular predictors change over time.

3.3.2 Rolling regression analysis

The empirical evidence of parameter instability is widespread in finance. It is quite common to use only the most recent observations to estimate the parameters (the so-called 'rolling regression' method) and see whether parameter instability exists. In the rolling regressions, one produces a sequence of OLS estimates of coefficients in a linear regression model using a fixed number of the most recent data. This section adopts a 20-years rolling window approach for the traditional predictive model and the AR(1) model with constant coefficients employing a list of popularly used predictive variables (Welch and Goyal, 2008). We plot rolling estimates with 95% point-wise confidence intervals for both intercept and slope coefficients. The main reason for using rolling regression is to detect instability of parameters over time. The rolling approach happens to be one useful visual inspection that can help determine whether a coefficient in the model is stable or changing over time, either suddenly or gradually.

Figure 3.2: 20-years Rolling Estimation on the Predictive Regression Model $r_t = \alpha + \beta x_{t-1} + e_t$ using Dividend-Price Ratio, 1947:I to 2018:IV



Note. This figure displays the 20-years rolling regression results for the traditional predictive model $r_t = \alpha + \beta x_{t-1} + e_t$, where r_t is the equity premium at time t, and x_{t-1} is the dividend-price ratio at time $t-1$. Sample data is quarterly from 1927:I to 2018:IV. Panels (a) and (b) plot the rolling OLS estimates for α and β, respectively, with 95% confidence bands assuming that x_t is stationary. Panels (c) and (d) plot the associated t-statistics of the rolling estimates for β and α, respectively, with the critical value 1.96 at 5% significance level under the null of no predictability.

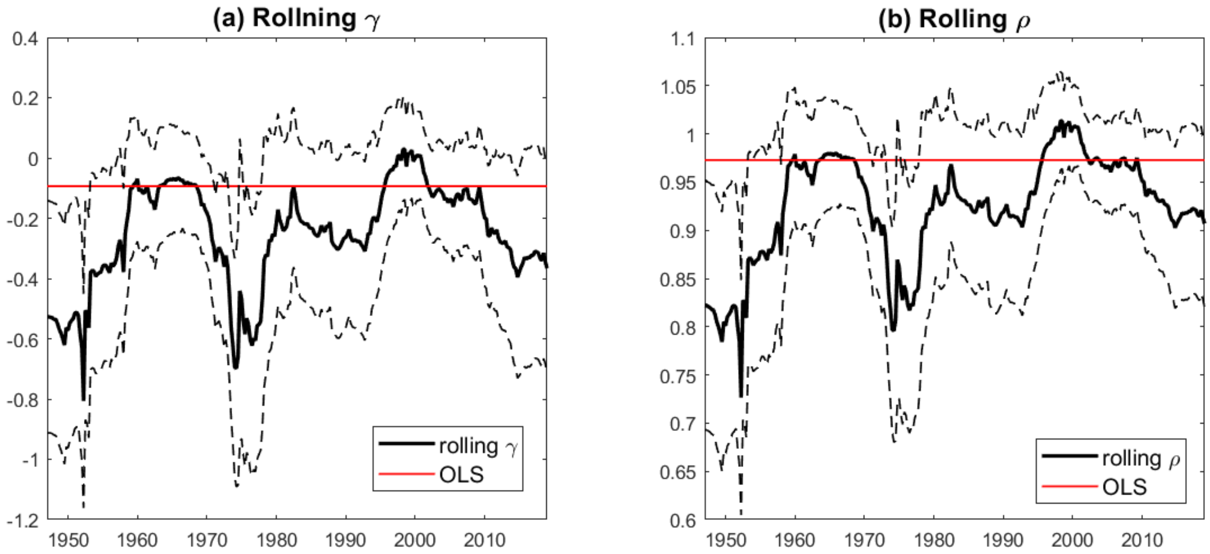
Figure 3.2 displays the 20-years rolling estimation results for the traditional predictive model $r_t = \alpha + \beta x_{t-1} + e_t$ where r_t is the equity premium at t and we use the dividend-price ratio as the predictor. Our sample data is from 1927:I to 2018:IV, so the

20-years rolling estimates of the coefficients start from 1947:I. Panel (b) in Figure 3.2 reports the rolling OLS estimates for β with 95% confidence band assuming that x_t is stationary and the full sample OLS estimates. Panel (d) plots the t-statistics of the rolling estimates of β under the null of no predictability, $\beta = 0$.[1] The rolling estimates of β are around 0.18 in the late 1950s, bigger than the OLS estimate of β using the full sample. After that, $\widehat{\beta}$ dramatically increases during the oil-shock period of 1973–1974, in which the confidence interval of the rolling estimates is statistically different from the full-sample OLS estimate. For this period, $\widehat{\beta}$ is also statistically significant at 5% as the associated t-statistic is clearly above 1.96, as shown in panel (d). The estimated coefficient of $\widehat{\beta}$ drops from 1990 to 2000 and becomes lower than the OLS estimate near 2000. In the subsequent period, $\widehat{\beta}$ gradually rises until the end of the data sample. Overall, the figure documents strong time-variation in the coefficients, consistent with the results in Lettau and Nieuwerburgh (2008) who also use the rolling regression approach to detect the coefficient instability. The t-statistics reported in panel (d) indicate that stock returns predictability using the dividend-price ratio is significant during the oil shock. This evidence has led some researchers to conclude that although the dividend-price ratio does not forecast stock returns over most sub-samples, the predictability is not absent everywhere. Predictability exists in some periods, like the oil-shock period of 1973–1974, consistent with Goyal and Welch (2003).

Panel (a) in Figure 3.2 reports the rolling OLS estimates for α (with 95% confidence band) assuming that x_t is stationary and the OLS estimates using the full sample, while Panel (c) plots the t-statistics of the rolling estimates of α under the null of $\alpha = 0$. Similar to the case of β, α also exhibits an obvious time-varying pattern, as shown in panels (a) and (c). The intercept parameter shows structural changes in the periods of 1953-1955, 1970-1974, and 1998-2000. The associated t-statistics are substantially above 1.96 during the periods. It is necessary to note that the rolling regression results are informal visual inspection used to detect the instability of the coefficient of β, rather than making formal statistical inference. The rolling window results for all 14 predictors are provided in

[1] The standard errors are computed by assuming that the predictors are stationary. The stationary assumption apparently makes a fuzzy inference, but the goal of the rolling regressions is to make a visual diagnostic of parameter instability in regressions, rather than a formal statistic inference.

Figure 3.3: 20-years Rolling Estimation on the AR Model $x_t = \gamma + \rho x_{t-1} + \varepsilon_t$ using Dividend-Price Ratio, 1947:I to 2018:IV.



Note. This figure plots the 20-years rolling regression results for the autoregressive model, $x_t = \gamma + \rho x_{t-1} + \varepsilon_t$, where x_t is the dividend-price ratio at t. Data are quarterly from 1927:I to 2018:IV. Panels (a) and (b) plot the rolling OLS estimates, for γ and ρ, respectively, with 95% confidence band assuming that x_t is stationary.

Appendix B.

Some prior studies have documented the instability of forecasting, but little attention has been paid to the instability in AR(1) processes for predictors. We plot the 20-years rolling estimation results using quarterly data from 1927:I to 2018:IV for the traditional AR(1) model $x_t = \gamma + \rho x_{t-1} + \varepsilon_t$ in Figure 3.3, where x_t is the dividend-price ratio. Panel (a) reports the rolling OLS estimates of γ (with 95% confidence band) assuming that x_t is stationary compared with the OLS estimates using the full sample. γ is significant in all sub-samples, as the t-statistics are larger than critical values 1.96 at 5% level of significance. This result is not surprising because the dividend-price ratio is highly persistent (ρ is close to one), and much of the variation in the dividend-price ratio comes from its lags. Persistence parameter ρ jumps from 1953 to 1960, stays high (close to one), but quickly drops after 1973. ρ rebounds back to a high level (above 0.9) and slightly move around the full-sample $\widehat{\rho}$.

In panel (a), we observe a similar pattern for estimates of γ as in the case of ρ. γ is negatively significant only before 1958 and between 1970 and 1978. The rolling estimates are not statistically different for the full-sample OLS estimate, as the full-sample OLS estimate lies inside the upper and lower confidence bounds. The results here suggest that the persistence of the dividend-price ratio is also time-varying. We leave additional rolling window results for all other predictors in Appendix B.

3.3.3 Univariate time-varying regressions

In this section, we study the predictability of U.S. market return using the nonparametric time-varying predictive regression models. To relate our model to the literature on forecasting stock returns, we compare the following three models:
Model (i): a conventional linear predictive regression model, $y_t = \alpha + \beta x_{t-1} + e_t$.
Model (ii): a pure time-varying predictive regression model, $y_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + e_t$.
Model (iii): a time-varying model with the linear projection method, $y_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + \lambda\varepsilon_t + \eta_t$.
By comparing models (i) and (ii), we are able to understand how much predictability is gained by allowing the constant parameters to be time-varying. Comparing model (ii) and model (iii) allows us to see the importance and necessity of introducing a linear projection function to a model.

In the nonparametric estimation, we choose the Gaussian density as the kernel function, and we choose bandwidth $h = T^{-1/5}$, where T denotes the sample size. Bandwidth selection for nonparametric hypothesis testing is a nontrivial problem and has been widely studied. As commented by Gao and Gijbels (2008) and Zhang et al. (2012), among others, there exists no uniform guidance for an optimal choice, especially if the data are dependent. On the positive side, we show in Appendix B that the empirical local linear estimation results are not quite sensitive to the benchmark choice of bandwidth, $h = T^{-1/5}$.

First, we compare models (i) and (ii) using in-sample fitting, which is one way to judge the rationality of the time-varying AR(1) models from a statistical perspective. Specifically, we report the R^2 statistics for the univariate time-varying AR(1) model $x_t = \gamma(\tau_t) + \rho(\tau_t)x_{t-1} + \varepsilon_t$ (tv-AR) and the constant AR(1) model $x_t = \gamma + \rho x_{t-1} + \varepsilon_t$ using the 14 predictors. In Table 3.2, the calculated R^2 of the time-varying AR models are

Table 3.2: R^2 (%) for the Univariate Time-Varying AR(1) Model and the Constant AR(1) Model

	DP	DY	EP	DE	BM	SVAR	NTIS
Constant AR(1) model	94.58	94.82	87.68	86.40	88.79	38.40	87.48
Time-varying AR(1) model	94.99	95.27	88.70	86.99	89.78	46.04	88.10
	TBL	LTY	LTR	TMS	DFY	DFR	INF
Constant AR(1) model	92.78	97.33	0.00	72.91	80.58	0.54	26.25
Time-varying AR(1) model	93.27	97.58	4.88	74.67	81.75	3.41	37.43

Note. This table reports R^2 for the univariate time-varying AR(1) model $x_t = \gamma(\tau_t) + \rho(\tau_t)x_{t-1} + \varepsilon_t$ and the constant AR(1) model $x_t = \gamma + \rho x_{t-1} + \varepsilon_t$ (cons AR) using the 14 individual predictors. Data is quarterly from 1927:I to 2018:IV.

consistently higher than the values of R^2 for the constant AR models. The time-varying AR(1) models, therefore, are able to capture the dynamics of the predictors better than the constant AR(1) models.

Table 3.3: Regression Analysis on the Linear Projection Function

			$\widehat{\sigma}^2$ $(\times 10^3)$		$adj\bar{R}^2$ (%)	
	$\widehat{\lambda}$	t-stat	Model (ii)	Model (iii)	Model (ii)	Model (iii)
DP	-0.95	-68.47	10.27	0.65	0.72	96.70
DY	0.08	1.55	10.24	10.14	0.98	1.46
EP	-0.40	-12.83	9.82	6.77	1.43	58.85
DE	-0.11	-2.49	10.34	10.16	0.77	1.21
BM	-0.96	-28.52	9.47	3.05	1.22	68.84
SVAR	-3.42	-7.36	10.44	8.95	0.76	9.00
NTIS	1.63	2.79	9.72	9.58	1.03	1.19
TBL	-0.60	-0.95	10.10	10.05	0.72	1.12
LTY	-1.86	-1.52	10.31	10.23	0.66	1.65
LTR	0.03	0.22	10.52	10.52	0.84	1.41
TMS	0.61	0.75	10.13	10.11	0.59	0.59
DFY	-16.85	-11.06	9.92	7.61	0.33	7.79
DFR	1.42	5.96	10.17	9.26	0.42	2.92
INF	0.29	0.55	10.50	10.50	0.83	0.92

Note. This table reports the estimation results of the linear projection function term λ in the 14 univariate time-varying predictive regressions. The first two columns report the estimates of λ with t-statistics. The rest of the columns report the estimated variances, $\widehat{\sigma}^2 = \frac{1}{T}\sum_{t=2}^{T}(y_t - \hat{y}_t)$, and R^2 on the regressions without the linear projection functions, $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + e_t$ and with the linear projection function, $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + \lambda\varepsilon_t + \eta_t$.

We examine whether λ is significant. Table 3.3 reports the estimation results of the linear projection function term λ in the 14 univariate time-varying predictive regressions in the settings for models (ii) and (iii). The first two columns report the estimates of λ and the associated t-statistics in the model (iii). If λ is significant, we conclude that there is a significant correlation between the innovations. As shown in Table 3.3, the correlation between returns' shock and predictors' shock is statistically significant for DP, EP, BM, SVAR, DFY, and DFR at 5% level, as their t-statistics are far above the critical value of 1.96. Thus, the linear projection method for these variables will deliver useful information on the time-varying AR(1) models to the time-varying predictive regression. Lewellen (2004) discusses this point in the conventional linear predictive regression system. First, the use of the linear projection method improves the efficiency of the time-varying estimators. Results shown in columns 3 and 4 of Table 3.3 confirm this argument. For the cases where λ is statistically different from zero, the variance of residuals, $\widehat{\sigma}^2$, on the regressions with linear projection functions is much less than the regressions without linear projection functions. Second, the values of adjusted R^2 are much higher when using linear projection approaches. In contrast, when we predict stock returns by using DY, TBL, LTY, LTR, TMS, and INF, λ is statistically insignificant, indicating that the shocks to these variables are uncorrelated with the shocks to stock returns. As a result, introducing a linear projection function to the time-varying model (ii) does not change the variance of residuals and R^2. For example, when we look at INF, the estimated λ is 0.29 with the t-statistic of 0.55, and the values of adjusted R^2 and $\widehat{\sigma}_\eta^2$ of regressions with λ are 0.83 and 10.5, which are close to the corresponding regressions without λ, 0.92 and 10.5.

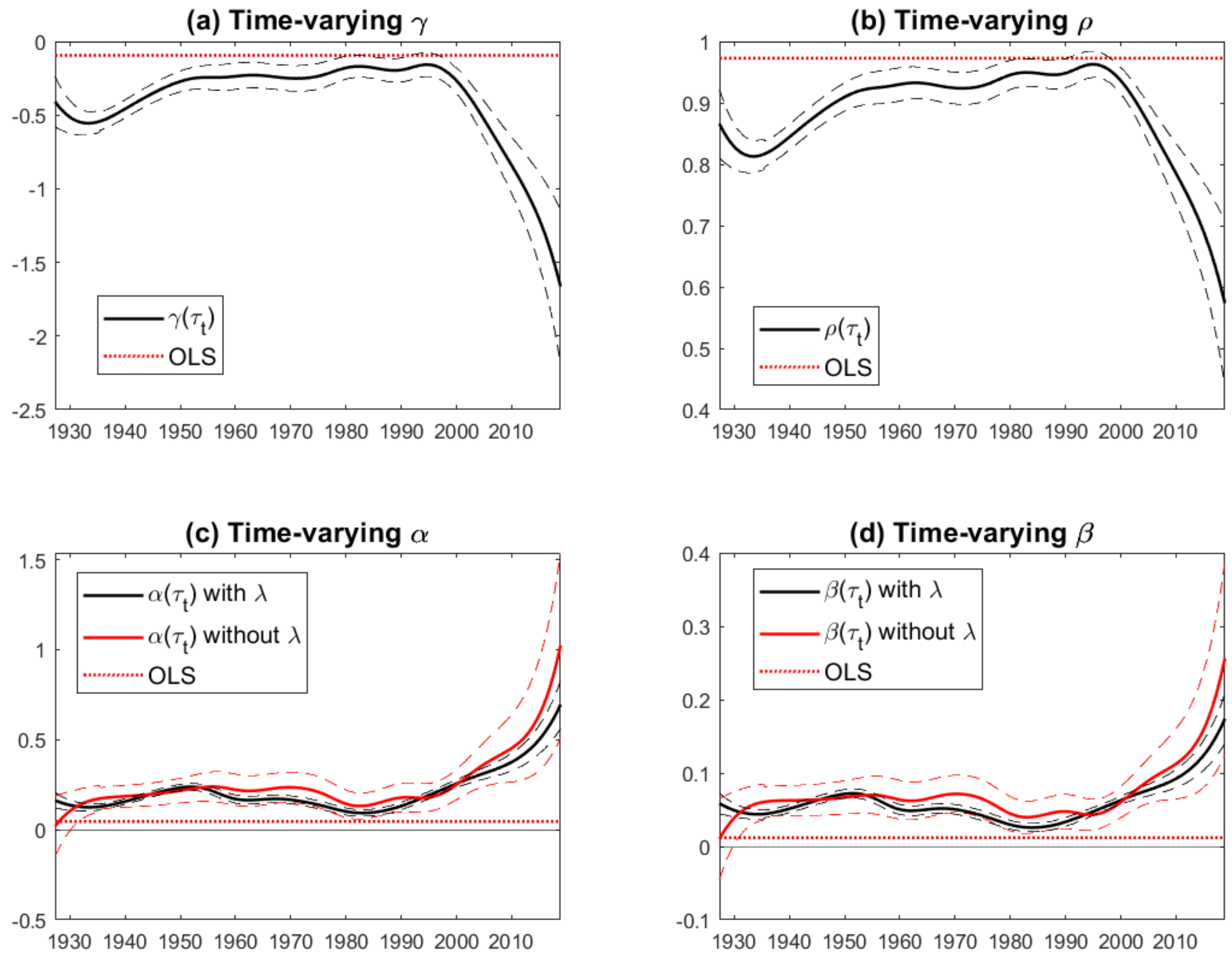
We next investigate the predictability of stock returns. First, we analyse the predictability by dividend-price ratio (DP) in detail, as it is also arguably the most popular predictor among the set of variables in other studies motivated by the finance theory discussed in Campbell and Shiller (1988), and whether DP predicts stock returns remains controversial in the literature. Panels (a) and (b) of Figure 3.4 show the local linear estimation results for the time-varying AR(1) model $x_t = \gamma(\tau_t) + \rho(\tau_t)x_{t-1} + \varepsilon_t$, and panels (c) and (d) plot the local linear estimation results for the pure time-varying predictive model without the linear projection function $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + e_t$ and with the linear projection function $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + \lambda\varepsilon_t + \eta_t$. Sample data is quarterly from 1927:I to 2018:IV.

Panels (a) and (b) of Figure 3.4 display the local linear estimates for $\gamma(\tau_t)$ and $\rho(\tau_t)$ with 95% point-wise confidence bands.[2] As discussed in Chapter 2, a time-varying AR process that is said to be locally stationary must satisfy two conditions: (1) the time-varying functions move smoothly over time and (2) $|\rho(\tau_t)| < 1$ for all t. First, the local linear estimates of $\gamma(\tau_t)$ and $\rho(\tau_t)$ change smoothly over the sample period 1927:I to 2017:IV, thus the results fit the smoothness condition. Second, the point-wise confidence interval for the time-varying coefficient of $\rho(\tau_t)$ lies safely below one, thus satisfying the second condition of locally stationary processes. Therefore, we conclude that DP does not violate the conditions of local stationarity. Interestingly, $\widehat{\rho}(\tau_t)$ maintains around 0.9 from the 1940s to the late 1990s and sharply drops after the boom stock market of the 1990s. Therefore, there are two different pictures of DP dynamics before and after the 1990s. The strong time-variation in DP's persistence is captured by neither the conventional stationary AR(1) models or random walks, but instead by the time-varying AR(1) models.

Panels (c) and (d) of Figure 3.4 plot the local linear estimates for $\beta(\tau_t)$ and $\alpha(\tau_t)$, respectively, with 95% confidence bands in the time-varying predictive regression model and the OLS estimates of constant coefficients in linear regression model. The red line represents the estimated time-varying coefficients from the pure time-varying model without linear projection function, $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + e_t$, and the black line represents the estimated time-varying coefficients with the linear projection function $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + \lambda\varepsilon_t + \eta_t$. It can be seen that the confidence interval for $\widehat{\beta}(\tau_t)$ in both models (ii) and (iii) lies above zero, suggesting the time-varying predictability is significant over time. $\widehat{\beta}(\tau_t)$ is stable from 1927 to 1990, and the value fluctuates around 0.2. However, the predictability becomes particularly strong during and after 1990, while the estimated $\beta(\tau_t)$ the post-2000 is almost triple than that in 1990. Thus, DP shows different ability to predict stock returns before and after 1990, reflecting the predictabil-

[2]We compute the standard errors of the local linear estimators in the univariate time-varying predictive regression models by using the consistent estimators, discussed in Chapter 2.

Figure 3.4: Univariate Time-Varying Regression Results Using Dividend-Price Ratio (DP), 1947:I to 2018:I



Note. This figure presents the local linear estimation results for the time-varying AR(1) model $x_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + e_t$ in panels (a) and (b), and the pure time-varying predictive model without the linear projection function $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + e_t$ and without the linear projection function $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + \lambda\varepsilon_t + \eta_t$ in panels (c) and (d). x_{t-1} is the dividend-price ratio at $t-1$. Data is quarterly from 1927:I to 2018:IV.

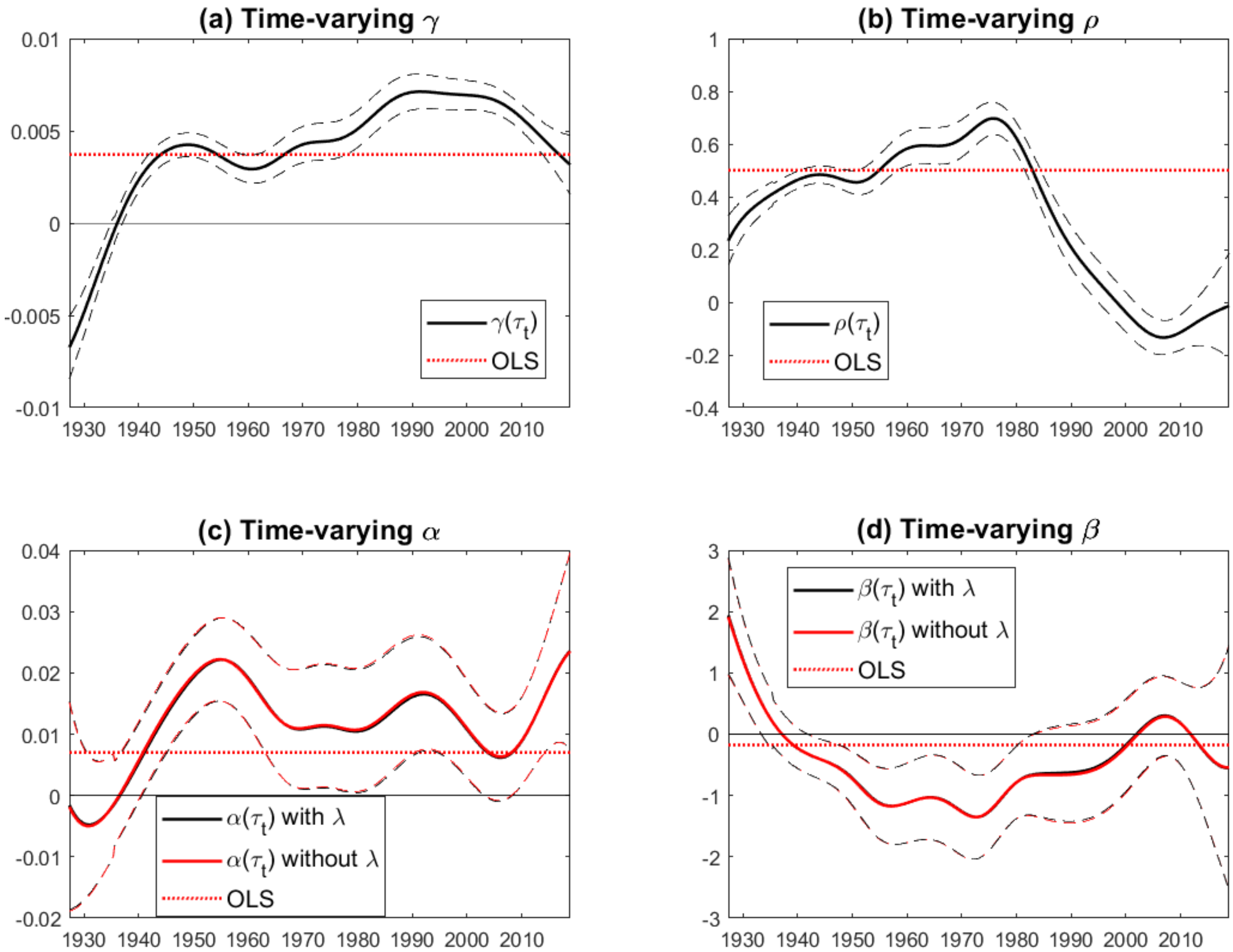
ity actually changes over time instead of remaining a constant. Similarly, as reported in Figure 3.4, the estimates of $\alpha(\tau_t)$ also vary over time and are significant at 5% level.

More importantly, we can see that the 95% confidence bands become narrower for a pure time-varying model than a time-varying model using the linear projection method. The main reason for the narrower confidence interval is that the introduction of the linear projection function to the predictive regression efficiently reduces the standard error of the local linear estimators; that is, although the level of time-varying coefficients is similar, their t-statistics are quite different. Our time-varying models provide more substantial evidence of return predictability, as DP in much of the our current results suggests that the return predictability is marginally significant. Our results reinforce the central message of Lewellen (2004) that when the innovations of stock returns and predictors are correlated, the linear approach can hugely improve stock return predictability.

Moreover, the time-varying predictability measured by $\beta(\tau_t)$ seems to be negatively related to the persistence in DP measured by $\rho(\tau_t)$. From 1927– 1930, the persistence of DP decreases and, conversely, the return predictability increases. From 2000–2018, predictability dramatically increases, but there is a sharp drop in the persistence of DP. This finding closely relates to some previous studies. Why do DP's persistence and its ability to predict stock returns seem opposed? Favero et al. (2011) argue that a slow-moving movement, as a source of persistence, in DP deteriorates its forecasting power of DP for future stock returns. Chen et al. (2012) further find that dividend smoothing is a source that makes the dividend-price ratio more persistent and makes either stock returns or dividend growth less predictable.

We next study the predictability by considering inflation (INF) as the predictor. Figure 3.5 plot the local linear estimates for $\gamma(\tau_t)$, $\rho(\tau_t)$, $\alpha(\tau_t)$, and $\beta(\tau_t)$, with 95% confidence interval. As in the case of DP, the confidence intervals for the linear estimates of $\gamma(\tau_t)$ and $\rho(\tau_t)$ using INF do not cover the full-sample OLS estimate from the traditional linear predictive regression model (i). So, the coefficients $\gamma(\tau_t)$ and $\rho(\tau_t)$ are time-varying instead of constant. Moreover, INF's persistence varies with time, high in the post-1980 period and low in the post-1980 period, but the confidence interval for $\widehat{\rho}_t$ lies below one over the full sample. Therefore, modelling INF as a time-varying AR(1) process also nicely fits the conditions of a locally stationary process. Turning to the local linear estimates $\widehat{\alpha}(\tau_t)$ and $\widehat{\beta}(\tau_t)$ in panels (c) and (d), INF shows weaker predictability for stock returns than DP. INF negatively predicts future stock return from around 1945 to 1982, as $\widehat{\beta}(\tau_t)$ is statistically less than zero. Conversely, INF tends to positively predict stock returns between 1927 to 1933, covering the Great Depression. After about 1985, INF shows no predictability anymore. From around 1940, the 95% lower confidence bands of $\widehat{\alpha}(\tau_t)$ is almost significant over time, except for a short period drop during the global financial crisis, 2007–2009.

Figure 3.5: Univariate Time-Varying Regression Results Using Inflation (INF), 1947:I to 2018:I

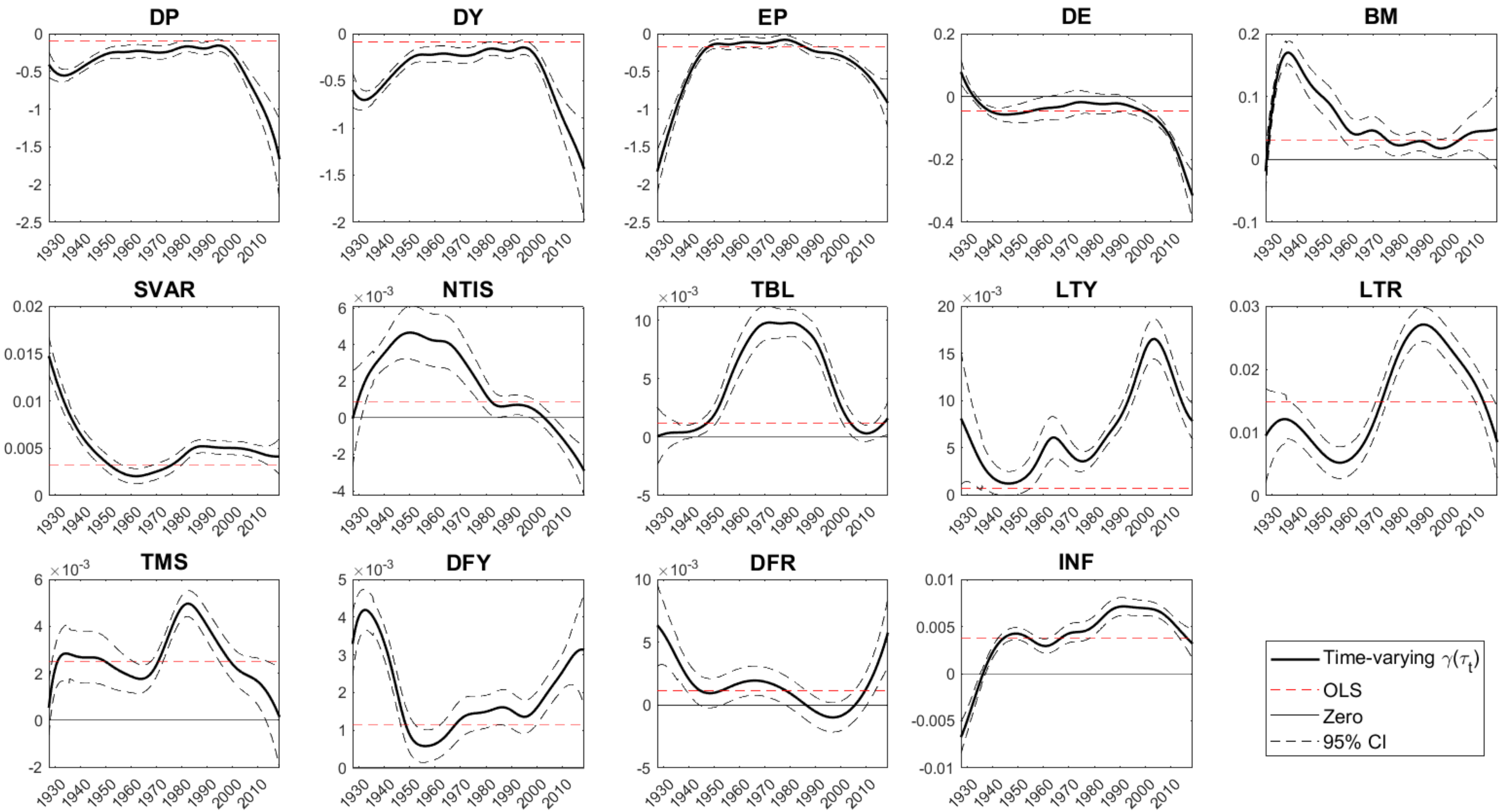


Note. This figure presents the local linear estimation results for the time-varying AR(1) model $x_t = \gamma(\tau_t) + \rho(\tau_t)x_{t-1} + \varepsilon_t$ in panels (a) and (b), and the pure time-varying predictive model without the linear projection function $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + e_t$ and without the linear projection function $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + \lambda\widehat{\varepsilon}_t + \eta_t$ in panels (c) and (d). x_{t-1} is the inflation rate at $t-1$. Data is quarterly from 1927:I to 2018:IV.

Interestingly, the local linear estimates $\widehat{\alpha}(\tau_t)$ and $\widehat{\beta}(\tau_t)$ (and the associated 95% confidence bands) with and without using the linear projection approach are largely overlapped. This suggests that it makes little difference to the results whether we use the linear projection method or not. This is because λ in Table 3.3 is insignificant, and the linear projection function conveys almost zero information from the time-varying AR(1) model.

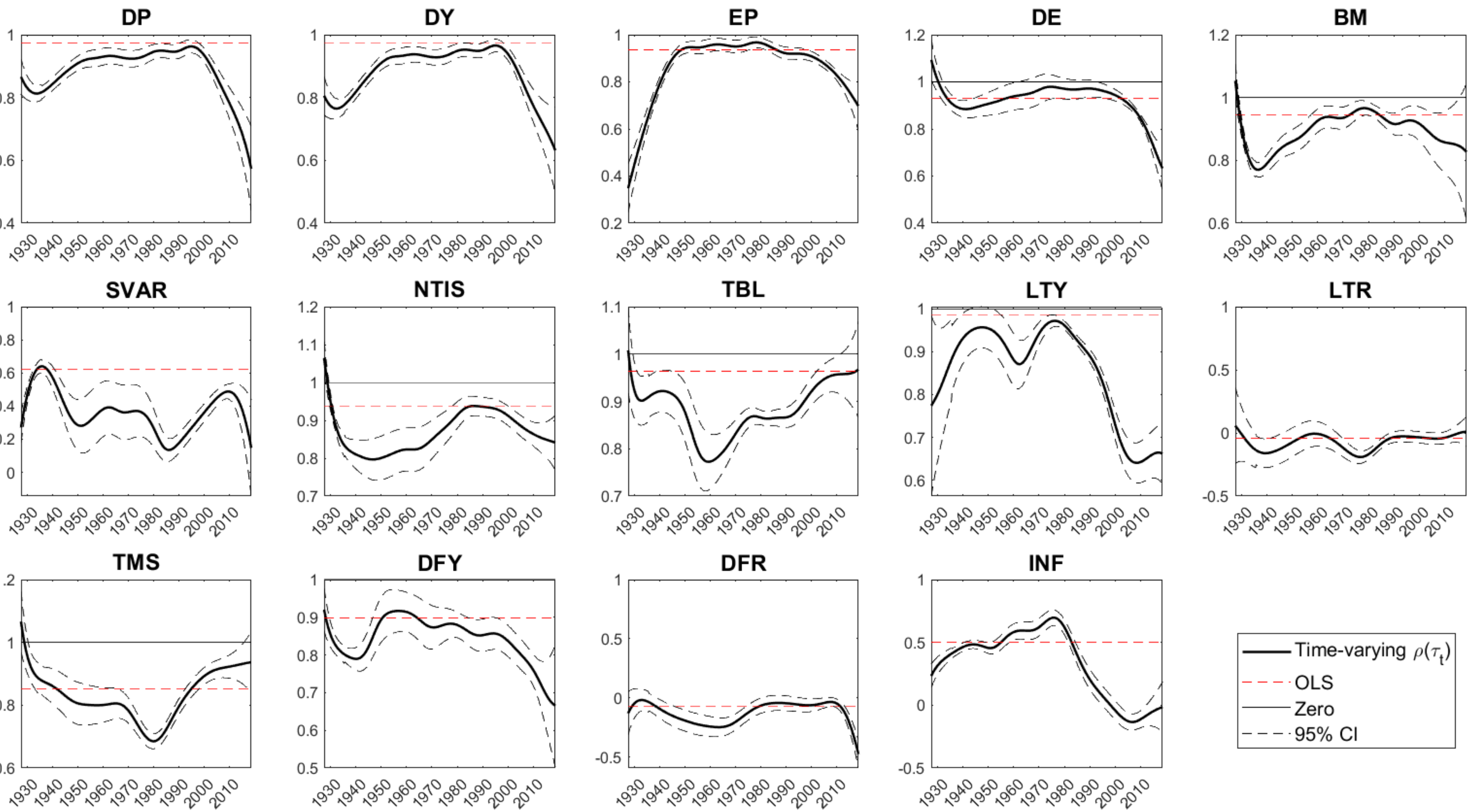
Figure 3.6 and Figure 3.7 display the comprehensive results on the instability of the univariate time-varying AR models for all 14 predictors. Figure 3.6 shows the lo-

Figure 3.6: $\widehat{\gamma}(\tau_t)$ in the Univariate Time-Varying AR Model, 1927:I to 2018:IV



Note. This figure shows the local linear estimates $\widehat{\gamma}(\tau_t)$ in the AR model $x_t = \gamma(\tau_t) + \rho(\tau_t)x_{t-1} + \varepsilon_t$ with pointwise 5% confidence bands for all 14 predictors. Data is quarterly from 1927:I to 2018:IV.

Figure 3.7: $\widehat{\rho}(\tau_t)$ in the Univariate Time-Varying AR Model, 1927:I to 2018:IV



Note. This figure shows the local linear estimates $\widehat{\rho}(\tau_t)$ in the AR model $x_t = \gamma(\tau_t) + \rho(\tau_t)x_{t-1} + \varepsilon_t$ with 5% pointwise confidence bands for all 14 predictors. Data is quarterly from 1927:I to 2018:IV.

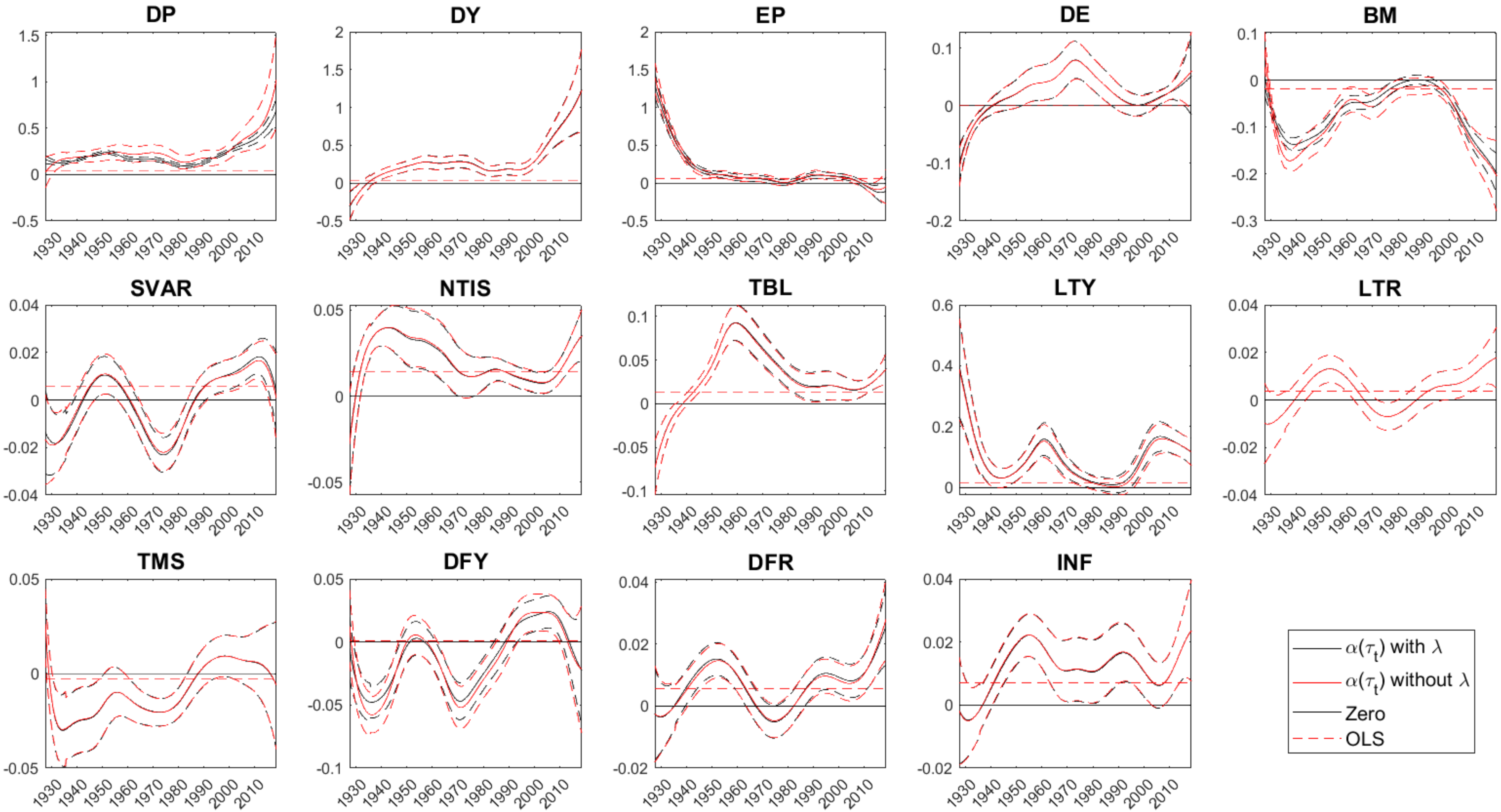
cal linear estimates $\widehat{\gamma}(\tau_t)$ in the AR model $x_t = \gamma(\tau_t) + \rho(\tau_t)x_{t-1} + \varepsilon_t$ with the point-wise 95% confidence bands for all 14 predictors. Comparing with the OLS estimates $\widehat{\gamma}$ using the full sample, the local linear estimates $\widehat{\gamma}(\tau_t)$ show dramatic variation over time, since $\widehat{\gamma}$ is outside of the 95% confidence bands of the time-varying estimators at the majority of sub-samples. These results suggest that the commonly used 14 predictor variables contain significant slow-moving components or nonlinear time trend. The time-varying shifts in the steady-state of variables are consistent with Lettau and Nieuwerburgh (2008).

Instability of persistence parameter $\rho(\tau_t)$ is also supported by the evidence in Figure 3.7. The local linear estimates of $\rho(\tau_t)$ move smoothly over time for all predictors, and absolute values of $\rho(\tau_t)$ are always less than one. The results imply that the predictors, such as DP, DY, EP, and BM, can be nonstationary in a global sample, but they are stationary locally around a particular time point. The results ensure that the predictors modeled by the time-varying AR model can satisfy the two conditions of local stationarity. Therefore, the time-varying AR(1) models are better than the conventional linear regression models in terms of statistical performance and economic rationality.

In addition, the confidence bands of the regressions with the linear projection equation are much narrower than those of the regressions without the linear projection equation when the innovations from stock returns and predictors are correlated (see, e.g., DP, EP, and BM). For economic variables, like LTY, LTR, or INF, the confidence bands(red and black dashed lines) are generally overlapped. Therefore, the linear projection approach is robust, regardless of whether the innovations are correlated.

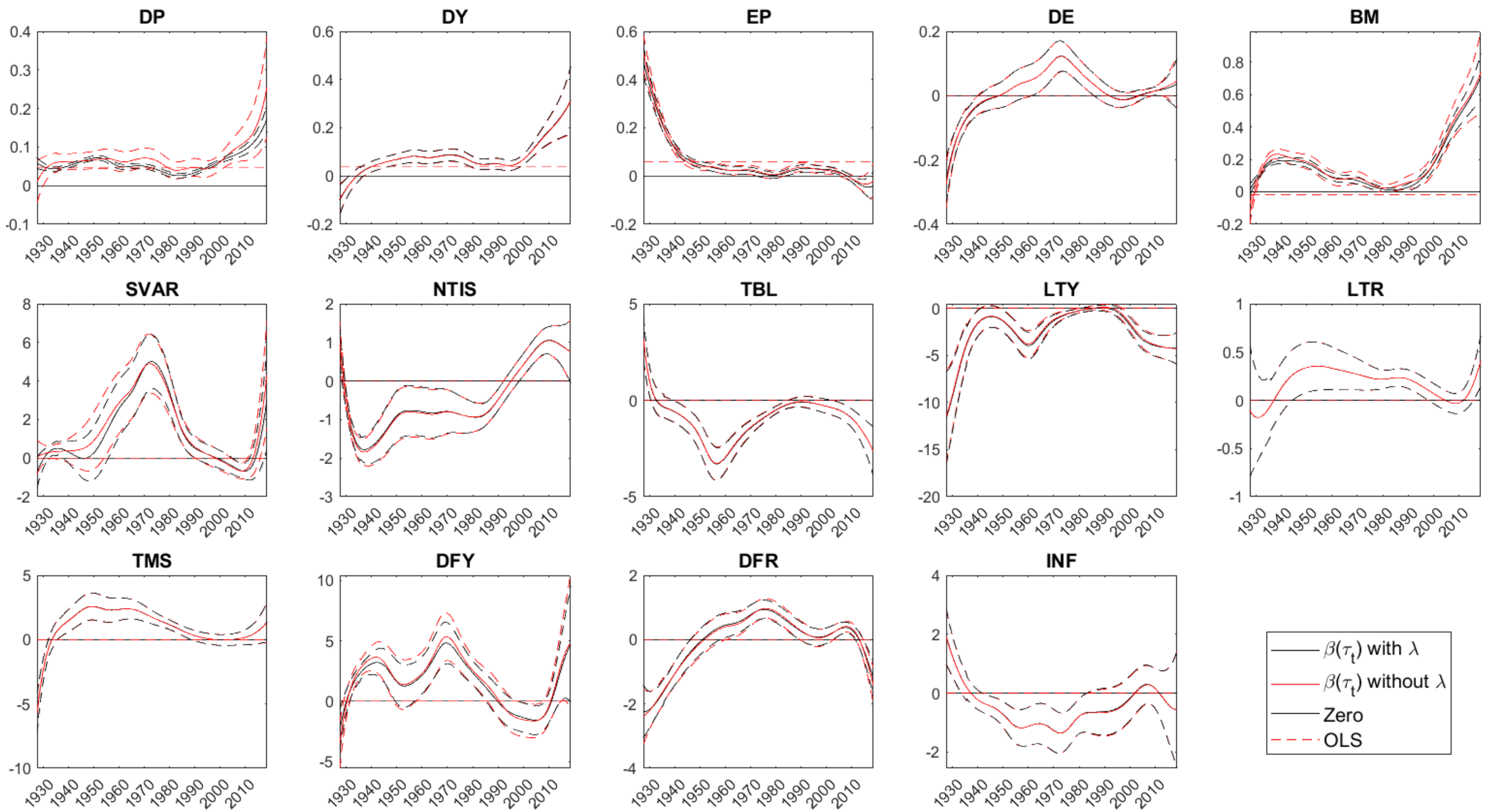
Figure 3.8 shows the local linear estimates $\widehat{\alpha}(\tau_t)$ in the time-varying predictive regression model with the 95% point-wise confidence bands for the 14 individual predictors with and without the linear projection function. The red line represents the estimated time-varying coefficients in model $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + e_t$, and the black line represents the estimated time-varying coefficients with linear projection function $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + \lambda\varepsilon_t + \eta_t$. The local linear estimates of $\alpha(\tau_t)$ change substantially in magnitude over time. Except for NTIS and INF, all variables have significant

Figure 3.8: Time-Varying $\widehat{\alpha}(\tau_t)$ in the Univariate Predictive Regression Model



Note. This figure shows the local linear estimates $\widehat{\alpha}(\tau_t)$ with point-wise 5% confidence bands for all 14 predictors in the time-varying model without linear projection function, $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + e_t$, and with linear projection function $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + \lambda\widehat{\varepsilon}_t + \eta_t$ Data is quarterly from 1927:I to 2018:IV.

Figure 3.9: Time-varying $\widehat{\beta}(\tau_t)$ in the univariate predictive regression models



Note. The figure shows the local linear estimates $\widehat{\beta}(\tau_t)$ in the PR model $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + e_t$ with pointwise 5% confidence bands for all 14 predictors. Data is quarterly from 1927:I to 2018:IV.

trending functions $\alpha(\tau_t)$. Moreover, the 95% confidence bands of the coefficients in pure time-varying models are narrower than those in the time-varying model with the linear projection function, in particular, when λ is statistically significant. If the correlation between the errors from the predictive model and the AR model is non-zero, then the linear projection method would be able to deal with the Stambaugh bias or embedded endogeneity and, therefore, improve the efficiency of time-varying local linear estimators in predictive regressions.

We also comprehensively examine the predictive performance employing the set of 14 predictors. Figure 3.9 shows the local linear estimates $\widehat{\beta}(\tau_t)$ in the time-varying predictive regression model with the 95% point-wise confidence bands for the 14 individual predictors without the linear projection function, $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + e_t$, and with the linear projection function, $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + \lambda\widehat{\varepsilon}_t + \eta_t$. $\beta(\tau_t)$ is the key parameter of testing return predictability. First, we find that $\widehat{\beta}(\tau_t)$ of all predictors are time-varying. The full sample OLS estimates of coefficient in the linear regression model generally lie outside of the confidence intervals of the time-varying local linear estimates. Second, the 14 popularly used predictors, to some degree, contain predictive content of equity premium. The fundamental factors, like DP, DY, EP, and BM, have strong predictability over the sample, since the estimated coefficients $\widehat{\beta}(\tau_t)$ are statistically different from zeros, especially during economic recession periods, such as the oil shock and the global financial crisis. INF seems to be the weakest predictor of equity premium among the list of 14 predictors, with predictability disappearing in more than half of the full sample.

3.3.4 Multivariate predictive regressions

We utilize the proposed multiple time-varying predictive models to re-examine the stock return predictability of a particular combination of predictors motivated by a theoretical or empirical perspective. We use the following combinations: (1) DE and TBL, (2) DP and NTIS, and (3) DFY and INF. The pair (1) is when the two variables seem stable over time but experience sudden changes during the Great Depression 1929–1933 and the Global Financial Crisis 2007–2009. The pair (2) combines a traditional valuation ratio and a financial market indicator of volume, and we are allowed to see how these financial variables contribute to return predictability. The last pair uses two popular macroeconomic variables to see how economic activities relate to predictable stock returns variation. We conduct multiple regressions with two predictors only in this thesis for simplicity. It can be extended to regressions that include more predictor variables.

We use a bootstrap simulation procedure to approximate the confidence interval of the local linear estimates. The full statistical theory of the local linear estimators in the multi-predictor predictive system is left for future work. The 95% confidence bands of $\alpha(\tau)$ and $\beta(\tau)$ at point $0 \le \tau \le 1$ are defined by

$$\left[\widehat{\alpha}(\tau) - 1.96 \times sd(\widehat{\alpha}(\tau)), \quad \widehat{\alpha}(\tau) + 1.96 \times sd(\widehat{\alpha}(\tau))\right], \tag{3.20}$$

$$\left[\widehat{\beta}_j(\tau) - 1.96 \times sd(\widehat{\beta}_j(\tau)), \quad \widehat{\beta}_j(\tau) + 1.96 \times sd(\widehat{\beta}_j(\tau))\right], \tag{3.21}$$

where $\widehat{\alpha}(\tau)$ and $\widehat{\beta}(\tau)$ are the local linear estimates of $\alpha(\tau)$ and $\beta(\tau)$, and we estimate $sd(\widehat{\beta}(\tau))$ using the two-point wild bootstrapping method. The bootstrap procedure for tv-VAR model is described as follows.

$$\left[\widehat{\gamma}_j(\tau) - 1.96 \times sd(\widehat{\gamma}_j(\tau)), \quad \widehat{\gamma}_j(\tau) + 1.96 \times sd(\widehat{\gamma}_j(\tau))\right] \tag{3.22}$$

$$\left[\widehat{\rho}_{ij}(\tau) - 1.96 \times sd(\widehat{\rho}_{ij}(\tau)), \quad \widehat{\rho}_{ij}(\tau) + 1.96 \times sd(\widehat{\rho}_{ij}(\tau))\right] \tag{3.23}$$

$$\left[\widehat{\lambda} - 1.96 \times sd(\widehat{\lambda}), \quad \widehat{\lambda} + 1.96 \times sd(\widehat{\lambda})\right], \tag{3.24}$$

The 95% confidence bands of $\gamma(\tau)$ and $\rho(\tau)$ at point $0 \le \tau \le 1$ can be defined by

1. Estimate models. We estimate γ_t and ρ_t by the local linear smoothing method introduced in the main text and denote the resulting estimates by $\widehat{\gamma}_t$ and $\widehat{\rho}_t$ for $t = 2, 3, ..., T$. After obtaining the residuals from the time-varying AR model, we estimate α_t, β_t and λ by using the semiparametric profile method and denote the resulting estimates by $\widehat{\alpha}_t$, $\widehat{\beta}_t$, and$\widehat{\lambda}$ for $t = 2, 3, ..., T$.

2. Re-sample residuals. For each $t = 2, 3, ..., T$ we generate the bootstrapping residu-

als $\{\varepsilon_t^*\}$ in the time-varying VAR(1) model by $\varepsilon_t^* = \widehat{\varepsilon}_t v_t^*$ where v_t^* is drawn from the two-point distribution, $Pr\left(v^* = -\frac{\sqrt{5}-1}{2}\right) = \frac{\sqrt{5}+1}{2\sqrt{5}}$ and $Pr\left(v^* = -\frac{\sqrt{5}+1}{2}\right) = \frac{\sqrt{5}-1}{2\sqrt{5}}$. For each $t = 2, 3, ..., T-1$, we generate the bootstrapping residuals $\{\eta_t^*\}$ in the semiparametric predictive regression by $\eta_t^* = \widehat{\eta}_t u_t^*$ where u_t^* is drawn from the two-point distribution, $Pr\left(u_t^* = -\frac{\sqrt{5}-1}{2}\right) = \frac{\sqrt{5}+1}{2\sqrt{5}}$ and $Pr\left(u_t^* = -\frac{\sqrt{5}+1}{2}\right) = \frac{\sqrt{5}-1}{2\sqrt{5}}$.

3. Re-generate $\{x_t^*\}$. Construct $x_2^* = \widehat{\gamma}_1 + \widehat{\rho}_1 x_1 + \varepsilon_1^*$ where $\widehat{\gamma}_1$, $\widehat{\rho}_1$ are calculated from step 1 and $\{\varepsilon_t^*\}$ is calculated from step 2. Then, we generate $x_3^* = \widehat{\gamma}_2 + \widehat{\rho}_2 x_2 + \varepsilon_2^*$. We keep generating recursively until $T-1$, and we obtain the full series $\{x_t^*\}$.

4. Re-generate $\{y_t^*\}$. We generate $\{y_t^*\}$ by $y_t^* = \widehat{\alpha}_t + \widehat{\beta}_t x_{t-1} + \widehat{\lambda}\widehat{\varepsilon}_t + \eta_t^*$, where $\widehat{\alpha}_t$, $\widehat{\beta}_t$ and $\widehat{\lambda}$ are calculated from step 1 and $\{\eta_t^*\}$ is calculated from step 2.

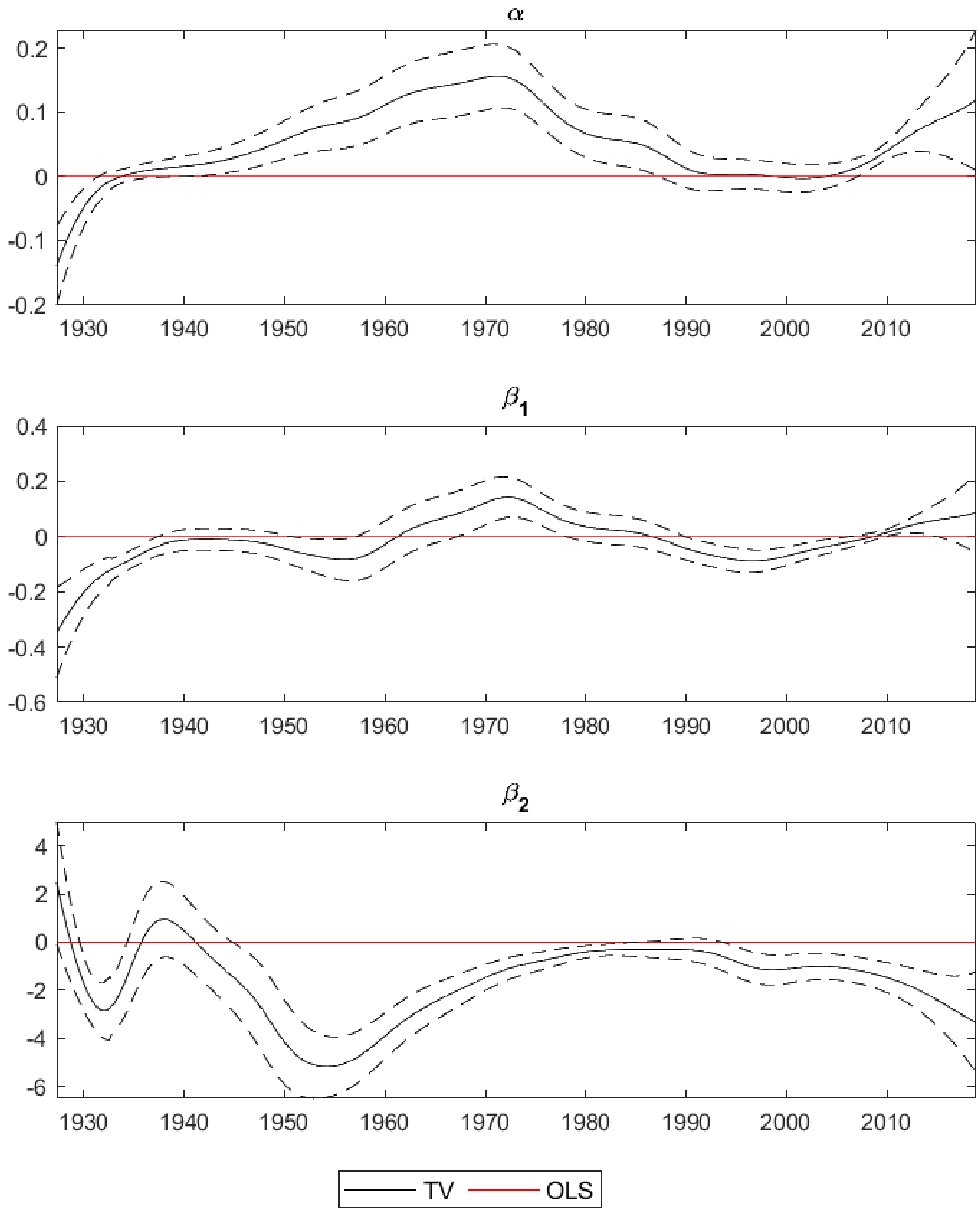
5. Use $\{y_t^*\}$ and $\{x_t^*\}$ to run the time-varying VAR regression and semiparametric predictive regression to estimate all the parameters over time, and repeat step 2 to step 4 for 2000 times.

We obtain $\widehat{\gamma}_t^{(1)}, \widehat{\gamma}_t^{(2)}, ..., \widehat{\gamma}_t^{(2000)}$ and $\widehat{\rho}_t^{(1)}, \widehat{\rho}_t^{(2)}, ..., \widehat{\rho}_t^{(2000)}$ for $t = 2, 3, ..., T$. Therefore, the point-wise standard deviation of $\widehat{\gamma}(\tau_t)$ and $\widehat{\rho}(\tau_t)$ for each t can be directly calculated from the above estimates. By using bootstrap, we also obtain $\widehat{\alpha}_t^{(1)}, \widehat{\alpha}_t^{(2)}, ..., \widehat{\alpha}_t^{(2000)}$, $\widehat{\beta}_t^{(1)}, \widehat{\beta}_t^{(2)}, ..., \widehat{\beta}_t^{(2000)}$, and $\widehat{\lambda}^{(1)}, \widehat{\lambda}^{(2)}, ..., \widehat{\lambda}^{(2000)}$ for $t = 2, 3, ..., T$. Therefore, the point-wise standard deviation of $\widehat{\alpha}(\tau)$, $\widehat{\beta}(\tau)$, and $\widehat{\lambda}$ for each t can be directly calculated from the above estimates.

The simultaneous confidence bands for the coefficient functions proposed in Zhang and Peng (2010) for the case where the observations are independent do not appear to work well in our setting, where the data are dependent and locally stationary, and they are generally much wider than the point-wise confidence bands. Therefore, in the empirical study, we report only the point-wise confidence bands. This bootstrapping is widely used in the empirical estimation in non- and semiparametric models (see, e.g., Chen et al. (2018) and Dong et al. (2017)).

Figure 3.10 shows the time-varying local linear estimates of $\alpha(\tau_t)$, $\beta_1(\tau_t)$, and $\beta_2(\tau_t)$ in the time-varying multi-predictor model. Data is quarterly from 1927:I to 2018:IV. Both

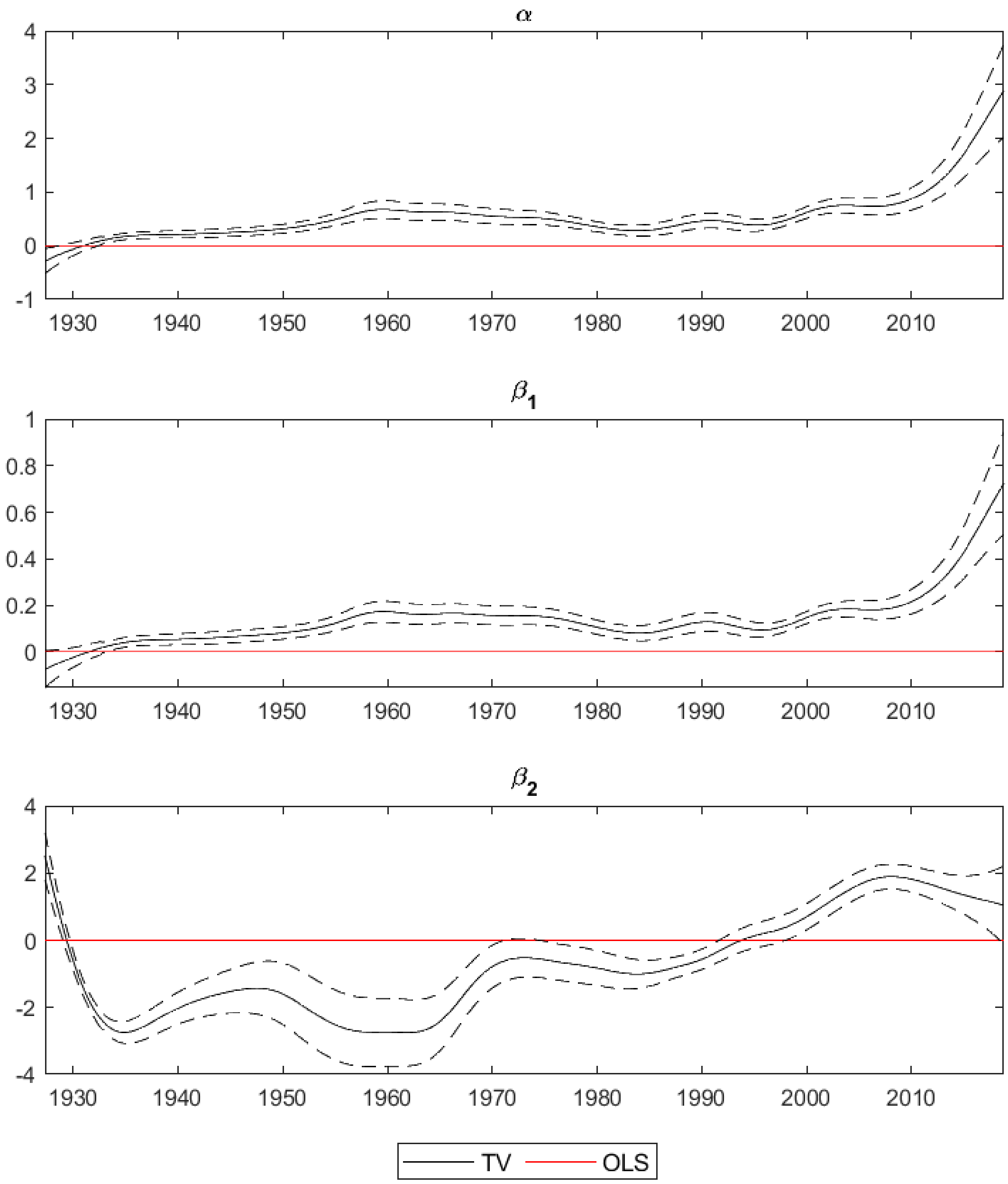
Figure 3.10: Local Linear Estimates in the Multi-Predictor Time-Varying Predictive Regression Model Using Combination (1): DE and TBL



Note. This figure shows the time-varying local linear estimates in the time-varying multi-predictor model $r_t = \alpha(\tau_t) + \beta_1(\tau_t)DE_{t-1} + \beta_2(\tau_t)TBL_{t-1} + \lambda^\top \widehat{\varepsilon}_t + \eta_t$. Data is quarterly from 1927:I to 2018:IV.

DE and TBL can predict stock returns. In particular, DE positively predicts the equity premium around two economic recessions: the oil shock of 1973-1973 and the global financial crisis of 2007-2009. In the rest of the sub-samples, DE negatively predicts returns. The fact suggests a strong time-variation in $\beta_1(\tau_t)$. TBL has a negative correlation

Figure 3.11: Local Linear Estimates in the Multi-Predictor Time-Varying Predictive Regression Model Using Combination (2): DP and NTIS

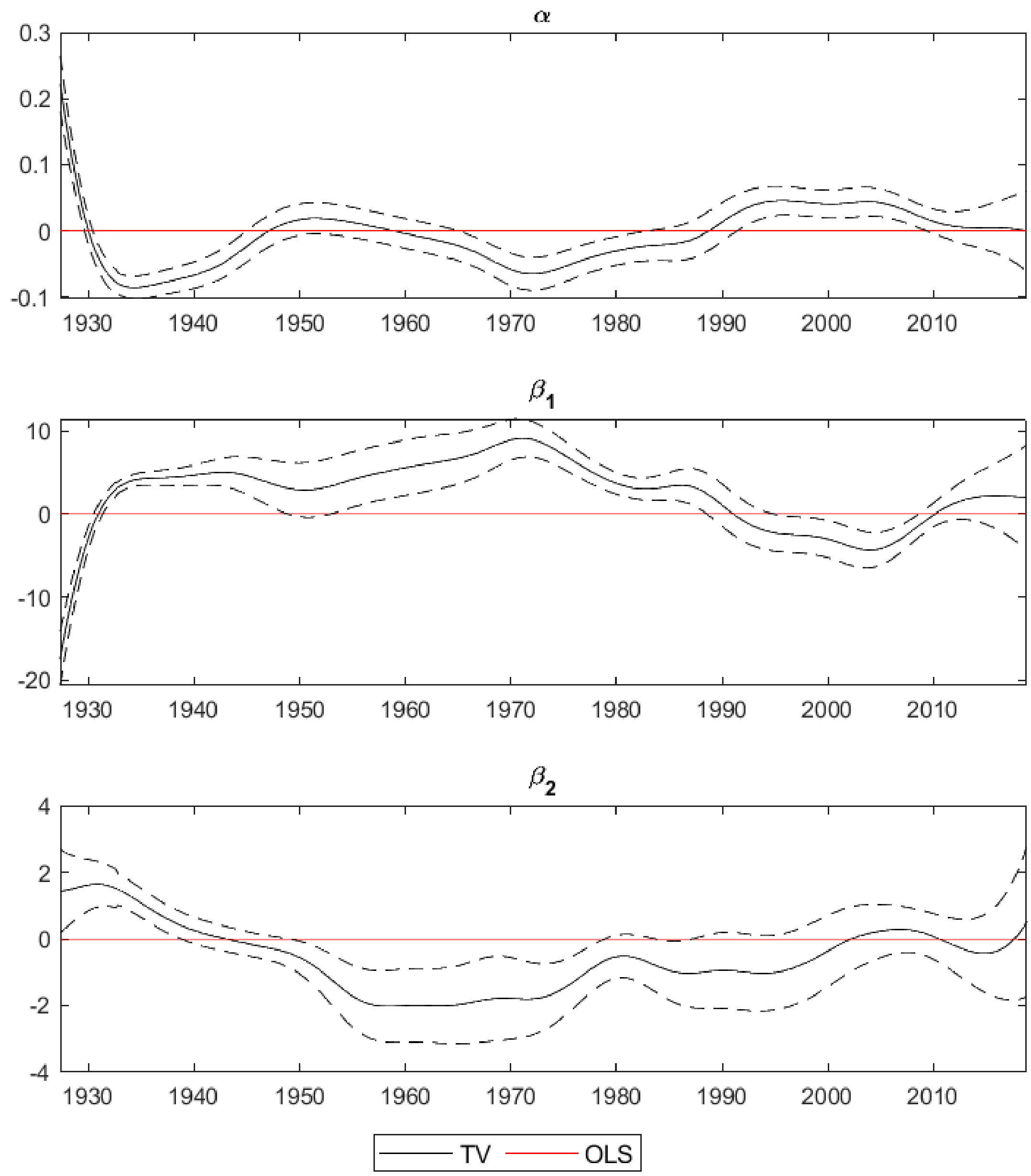


Note. This figure shows the time-varying local linear estimates in the time-varying multi-predictor model $r_t = \alpha(\tau_t) + \beta_1(\tau_t)DP_{t-1} + \beta_2(\tau_t)NTIS_{t-1} + \lambda^\top \widehat{\varepsilon}_t + \eta_t$. Data is quarterly from 1927:I to 2018:IV.

with the stock return before 1995 but tends to correlate with the stock return after 1995 positively. So, $\beta_2(\tau_t)$ and $\beta_2(\tau_t)$ are really time-varying parameter, consistent with the univariate predictive regression results using DE and TBL alone.

Figure 3.11 shows the time-varying local linear estimates of $\alpha(\tau_t)$, $\beta_1(\tau_t)$, and $\beta_2(\tau_t)$ in the time-varying multi-predictor model $r_t = \alpha(\tau_t) + \beta_1(\tau_t)DP_{t-1} + \beta_2(\tau_t)NTIS_{t-1} +$

Figure 3.12: Local Linear Estimates in the Multi-Predictor Time-Varying Predictive Regression Model Using Combination (3): DFY and INF



Note. The figure shows the time-varying local linear estimates in the time-varying multi-predictor model $r_t = \alpha(\tau_t) + \beta_1(\tau_t)DFY_{t-1} + \beta_2(\tau_t)INF_{t-1} + \lambda^\top \widehat{\varepsilon}_t + \eta_t$. Data is quarterly from 1927:I to 2018:IV.

$\lambda^\top \widehat{\varepsilon}_t + \eta_t$. Data is quarterly from 1927:I to 2018:IV. Figure 3.12 shows the time-varying local linear estimates of $\alpha(\tau_t)$, $\beta_1(\tau_t)$, and $\beta_2(\tau_t)$in the time-varying multi-predictor model $r_t = \alpha(\tau_t) + \beta_1(\tau_t)DFY_{t-1} + \beta_2(\tau_t)INF_{t-1} + \lambda^\top \widehat{\varepsilon}_t + \eta_t$. Data is quarterly from 1927:I to 2018:IV. Similarly, we find that (i) the variables have ability to predict equity premium, while their performance are consistent to the results in the univariate time-varying predictive regressions using these variables along; and (ii) the local linear estimates of $\alpha(\tau_t)$, $\beta_1(\tau_t)$, and $\beta_2(\tau_t)$ in the two pair of predictors change dramatically over time. The confidence intervals do not cover the correspond OLS estimates in the linear predictive regression models.

3.4 Summary

In summary, in Chapter 3, we implement the time-varying models to test predictability for quarterly U.S. aggregate stock returns in the sample period 1927:I to 2018:IV using a list of commonly employed variables from Welch and Goyal (2008). First, we find significant time-variation in the coefficients of the 14 popular predictors (Welch and Goyal, 2008) in the time-varying AR(1) models. The estimated coefficients of individual predictors change smoothly over time, and the absolute values of $\widehat{\rho}_t$ are generally less than one, indicating the validity of local stationarity. We also find that the time-varying AR(1) models fit real data better than the linear AR(1) models with constant coefficients (i.e., higher R^2). Second, we find that the estimated coefficients $\widehat{\beta}(\tau_t)$ of all predictors are time-varying. The OLS estimates of coefficient in the linear regression model generally lie outside of the confidence intervals of the time-varying local linear estimates. The contribution of Chapter 3 comes from the empirical results using the proposed time-varying models. More importantly, the 14 popularly used predictors certainly contain predictive content of stock returns. Like the DP and BM, the fundamental factors have strong predictability in the majority of the subsamples, and the predictability is especially strong during economic recession periods, such as the oil shock of 1973-1974 and the global financial crisis of 2007-2009. INF seems to be the weakest predictor among the 14 predictors, but predictability remains significant in about half of the full sample. Finally, we extend the time-varying models for predictive regressions with a single predictor into a multi-predictor framework and we apply a time-varying VAR(1) for multiple predictors. For example, DP and INF also show significant predictability, similar to the DP and INF cases in the univariate regressions.

Chapter 4

Out-of-sample Equity Premium Forecasting

There is an impressive body of empirical evidence documenting that stock returns are predictable by various lagged financial valuation ratios, including dividend-price ratio, earning-price ratio, and book-to-market ratio (see Cochrane, 2008; Fama and French, 1988; Kothari and Shanken, 1997; Lamont, 1998, e.g.). However, the predictability is statistically tenuous. Phillips (2015) outlines some of pitfalls in predictive regression models. This includes model uncertainty/parameter instability (Avramov, 2002; Dangl and Halling, 2012; Paye and Timmermann, 2006), misleading inference (Campbell and Yogo, 2006; Kostakis et al., 2015; Lewellen, 2004), and small sample bias or embedded endogeneity (Amihud and Hurvich, 2004; Nelson and Kim, 1993; Stambaugh, 1999). We have addressed these issues in Chapter 3, and we find the in-sample predictability is truly time-varying and significant in the system of time-varying models for predictive regressions including single predictor and multiple predictors.

It is essential to point out that in early empirical work, the usual approach to testing predictability relies on the in-sample fitting. However, out-of-sample tests are more relevant for investors to assess real-time return forecasts under the assumption that predictive models are well-specified, and the parameters are constant over time. In-sample statistical tests provide more efficient parameter estimation and thus more precise return forecasts. However, these assumptions are unlikely to hold in practice. Empirically, Welch and Goyal (2008) show that the popularly used predictors are unable to deliver consistently superior out-of-sample forecasts of future excess returns relative to simple benchmark forecasts based on the historical means. As a result, the predictive regressions may not help investors obtain accurate forecasts of return consistently over time and also may not signal an economically significant degree of equity premium forecastability in terms of increased utility gains for an investor with mean-variance preference. Thus, the commonly used financial ratios, like dividend-price ratio, earning-price ratio, and book-to-market ratio have been proposed based on good in-sample fitting, but their out-of-sample predictive qualities are still in question.

Several recent papers propose new methods to improve the accuracy of out-of-sample forecasting. For example, Campbell and Thompson (2008) conduct an out-of-sample forecasting exercise using various popular predictors, but modifications that reveal the effectiveness of theoretically motivated restrictions (e.g., positivity and monotonicity) are imposed on the predictive regression models. They show that many predictive regressions can provides better out-of-sample forecast than the historical mean model. Out-of-sample explanatory power is small but is economically meaningful for mean-variance investors as the utility gains are broadly consistent with the out-of-sample R^2_{os} statistic. Dangl and Halling (2012) apply a Bayesian econometric method allowing for gradual changes of coefficients (random walk process). For monthly returns of the S&P 500 index, they demonstrate statistical and economic evidence of out-of-sample predictability; relative to an investor using the historical mean, an investor using their Bayesian methodology could earn consistently positive utility gains (between 1.8% and 5.8% per year over different periods). Besides, as shown in their empirical analysis, models with gradually varying coefficients are strongly supported by the data. Rapach et al. (2010) recommend combining individual forecasts (e.g., mean, median, trimmed mean, and DMSPEs) that deliver statistically and economically significant out-of-sample gains relative to the historical average consistently over time.

A critical question is still open: to what extent can we rely on an individual predictor to predict future stock returns out-of-sample? As emphasised by Spiegel (2008) in a special issue of *The Review of Financial Studies* on the theme of equity premium predictability, '*Can our empirical models accurately forecast the equity premium any better than the historical mean*?' This chapter answers this question. We predict future equity premium out-of-sample at one and longer horizons. The motivation is straight. Although the dividend-price ratio shows little forecasting power for future returns at short horizons, it seems to explain increasingly more of the variability in returns at longer horizons (see, e.g., Cochrane (2008, 2011) and Fama and French (1988)). The 2013 Nobel for Economic Sciences introduction says: "There is no way to predict whether the price of stocks and bonds will go up or down over the next few days or weeks. But it is quite possible to foresee the broad course of the prices of these assets over longer time periods, such as the next three to five years..." We use a time-varying predictive regression model proposed in Chapter 2 and Chapter 3 to investigate out-of-sample equity premium predictability at short and long horizons.

4.1 Forecasting models

In this section, we introduce the long horizon time-varying predictive regression model, and briefly describe the methodology for estimating the nonparametric parameters. The time-varying models capture the time-varying forecasting relationship between future excess returns and lagged predictors. The time-varying forecasting model is specified as

$$r_{t+1,t+J} = \alpha_{J,t} + \beta_{J,t} x_t + e_{t+J}, \tag{4.1}$$

where $r_{t+1,t+J} = \sum_{j=1}^{J} r_{t+j}$ representing long horizon continuously compounded stock returns in excess of the risk-free rate from $t+1$ to $t+J$, x_t is the lagged predictor at t, $\alpha_{J,t}$ and $\beta_{J,t}$ are the nonparametric time-varying coefficients representing the slow-evolving components of long-horizon returns and predictability of long-horizon returns by the predictor x_t, and e_{t+J} is the error term.

We estimate the time-varying coefficients θ_t, where we denote $\theta_{J,t} = (\alpha_{J,t} \quad \beta_{J,t})'$, by the nonparametric local linear kernel estimation method, which is a purely data-driven mechanism, as we do in Chapter 2 and Chapter 3. We specify the time-varying coefficient as an unknown deterministic function of time in a standardised form as $\theta_t = \theta(\tau_t)$, for $t = 1, 2, ..., T$, where $\tau_t = t/T$ and $\theta(\cdot)$ is a continuous and differentiable function defined on [0,1] (see, e.g., (Robinson, 1989) for a discussion on this point). At point $\tau \in [0, 1]$, we apply the Taylor series expansion in its neighbourhoods so that the standardized time-varying function can be approximated by a linear function that $\theta_i(\tau_t) \approx \theta_i(\tau) + \theta_i'(\tau)(\tau_t - \tau)$ for $t = 1, 2, ..., T$ and $i = 1, 2$. The local linear estimator of $\theta(\tau_t)$ at a re-scaled time point τ is expressed as

$$\widehat{\theta}(\tau) = (\mathbf{I}_2, \mathbf{0}_2) \left(\sum_{t=1}^{T} K_h(\tau_t - \tau) z_t z_t' \right)^{-1} \sum_{t=1}^{T} K_h(\tau_t - \tau) z_t r_{t+1,t+J}, \tag{4.2}$$

where $z_s = (1, x_t)'$, $K_h(u) = K(u/h)/h$ in which $K(\cdot)$ is a standard kernel function, and h is the bandwidth parameter. In this chapter, We use the Epanechnikov kernel function $K(u) = \frac{3}{4}(1 - u^2) I\{|u| \leq 1\}$ and choose $h = N^{-1/5}$, where N denotes the sample size of each iteration of out-of-sample forecasting process, and $\theta(\frac{T+J}{T}) = \theta(\frac{T}{T}) + \theta'(\frac{T}{T}) \left(\frac{T+J}{T} - \frac{T}{T}\right)$.

The local constant kernel estimation method is also widely applied. However, a vast number of studies (e.g., Cai (2007)), show that the local linear estimation is superior to the local constant estimation in theory and application. In particular, the local linear estimator has a much smaller bias near the boundary of the trend function support, and it produces a much more accurate approximation of the trend function than that of the local linear method. Therefore, the focus of this chapter is only on the nonparametric local linear estimation method.

We follow Chen and Hong (2012) to reflect the data in the boundary region and deal with boundary issues. Suppose we observe data sample for a dependent variable Y and an independent variable X from time period 1 to T. For the future period, we artificially generate $(Y_i, X_i) = (Y_{2T-i}, X_{2T-i})$, for $T + 1 \leq i \leq T + \lfloor Th \rfloor$, where $\lfloor Th \rfloor$ denotes the integer part of Th. We then estimate the parameters $\widehat{\theta}_t$ at each t for $1 \leq t \leq T + \lfloor Th \rfloor$, using the synthesised data from combining the original sample data from 1 to T and the artificially generate data from $T + 1$ to $T + \lfloor Th \rfloor$. We call the estimators using the synthesised data as 'reflection estimators'. To the best of our knowledge, although this approach has been used to estimate time-varying coefficients in Chen and Hong (2012) for testing instability of return predictability (in-sample analysis), this approach has not previously been used for out-of-sample forecasting.

There is one major reason for us to use the reflection approach. Although the convergence rate of nonparametric local linear estimator at a boundary point is the same as in the interior region, it is not true for the asymptotic variance. The asymptotic variance of the local linear estimator at a boundary point tends to be larger because fewer observations are available to the local linear estimators in the boundary regions (Cai, 2007). This boundary issue may lead to the predicted values of the dependent variable Y at the end of the sample being very different to those in the middle. This is very similar to the Hamilton (2018)'s critical augment on the Hodrick–Prescott filter. It is important to note that in the out-of-sample experiment, the real-time forecasts rely on the time-varying coefficients of only the last time point of each recursively expanding sample period instead of middle or other time points. The reflection approach makes the boundary behaviour of the local linear estimator similar to that at interior points.

We choose some competing models. Following Welch and Goyal (2008), we use the historical mean of past returns as the benchmark forecast. Based on the same benchmark model, we compare the forecasting performance of the time-varying predictive regression models with their constant-coefficient counterparties, or the traditional predictive regression models

$$r_{t+1,t+J} = \alpha + \beta x_t + e_{t+J}. \tag{4.3}$$

To sum up, we will evaluate out-of-sample predictability by using the time-varying model (4.1) with the reflection adjustment, and the traditional predictive regression model (4.3).

4.2 Statistical evaluation

In this section, we provide the out-of-sample R-square statistic, denoted as R^2_{os}, (see Campbell and Thompson, 2008; Welch and Goyal, 2008, for example) to quantify the forecast accuracy using the predictive values $\widehat{r}_{t+1,t+J}$ and $\bar{r}_{t+1,t+J}$, where $\widehat{r}_{t+1,t+J}$ is a forecast based on our predictive regression models using past information up to t only and $\bar{r}_{t+1,t+J}$ is the sample average served as a benchmark forecast. The spirit of an out-of-sample experiment is that we use only information available at the time to produce future forecasts so that these forecasts do not have a look-ahead bias. The R^2_{os} is akin to the familiar R^2_{IS} and formulated as follows:

$$R^2_{os} = 1 - \frac{\sum_{s=1}^{n_2} \left(r_{n_1+s,n_1+J+s} - \hat{r}_{n_1+s,n_1+J+s}\right)^2}{\sum_{s=1}^{n_2} \left(r_{n_1+s,n_1+J+s} - \bar{r}_{n_1+s,n_1+J+s}\right)^2}, \tag{4.4}$$

where we divide the total T observations into in-sample period with first n_1 observation and out-of-sample period with $n_2 = T - n_1$ observations. Following Welch and Goyal (2008), we use the historical mean of past returns as a benchmark, i.e., $\bar{r}_{t,t+J} = \frac{1}{t}\sum_{i=1}^{t} r_i$. The historical mean model is a stringent benchmark in out-of-sample tests, where the traditional predictive regression forecasts based on individual predictors typically fail to outperform the historical mean (Welch and Goyal, 2008). The decision rule is simple: when $R^2_{os} > 0$, the predictive regression forecast is more accurate than the historical mean model, and therefore out-of-sample predictability exists and vice versa. Such benchmark forecasts are also available by regressing returns on a constant. Intuitively, a positive R^2_{os} indicates that the model forecasts have lower mean squared forecast errors (MSFE) than the benchmark forecasts, suggesting the existence of return predictability.

We compute the univariate predictive regression forecasts of the equity premium for the 14 predictors studies in Chapter 3. We use the data from 1927:I to 1946:IV as the initial in-sample estimation period, and recursively compute out-of-sample forecasts from 1947:I to 2018:IV. The selection of forecast evaluation period is somewhat arbitrary.

The forecasts employ a recursive estimation window, meaning that the initial estimation sample is from 1927:I to 1946:IV, with additional observations used as they become available. Forecasting model parameters are also frequently estimated with a rolling window, which drops earlier observations as additional observations become available. Rolling estimation windows are typically justified by appealing to structural breaks, although a rolling window generally will not be an optimal estimation window in the presence of breaks. Pesaran and Timmermann (2007) and Clark and McCracken (2009) show that, from an MSFE perspective, it can be optimal to employ pre-break data when estimating forecasting models, a manifestation of the classic bias efficiency trade-off. More generally, they demonstrate that the optimal window size is a complicated function of both timing and size of the breaks. Since these parameters are difficult to estimate precisely, recursive estimation windows frequently perform better in terms of MSFE than rolling windows or windows selected based on structural break tests when forecasting stock returns.

We further conduct statistical tests to investigate whether the model-based forecast has a significantly lower MSFE than the historical average benchmark forecast, which is equivalent to testing that

$$H_0 : R^2_{os} \leq 0 \quad \text{vs} \quad H_A : R^2_{os} > 0. \tag{4.5}$$

The well-known method of Diebold and Mariano (1995) for testing the null hypothesis of equal MSFE has an asymptotic standard normal distribution when comparing forecasts from non-nested models. However, Clark and West (2007) and McCracken (2007), show that this statistic has a nonstandard distribution when comparing forecasts from nested models. The time-varying predictive regression models nest the historical mean model when $\alpha_t = \alpha$ and $\beta_t = 0$. Clark and West (2007) produce a modified statistic, denoted as CW MSFE-adjusted statistic, for comparing both nested models that has an asymptotic distribution well approximated by the standard normal.

In this thesis, we use the CW MSFE-adjusted statistic that provides a convenient method for assessing statistical significance when comparing nested models obviating the need to look up a new set of critical values for each application. We start by obtaining the following series in the out-of-sample evaluation period,

$$f_{t+1} = (r_{t+1} - \bar{r}_{t+1})^2 - \left[(r_{t+1} - \hat{r}_{t+1})^2 - (\bar{r}_{t+1} - \hat{r}_{t+1})^2 \right]. \tag{4.6}$$

We regress $\{f_t\}_{s=m+q_0}^{T-1}$ on a constant

$$f_t = c + \varepsilon_t, \tag{4.7}$$

and the t-statistic corresponding to the constant, c, as the CW test statistic. To account for the heteroscedasticity and serial correlation in residuals due to the overlapping data, we use the Newey-West adjusted standard errors (Newey and West, 1987) to compute the MSFE-adjusted t-statistics with k-lags, consistent with Neely et al. (2014). A p-value for a one-sided (upper-tail) test is then obtained with standard normal distribution. A small p-value thus indicates that a proposed model significantly outperforms the benchmark model. A vast number of recent studies of stock return predictability report the CW MSFE-t statistic, including, for example, Rapach et al. (2010), Dangl and Halling (2012), and Neely et al. (2014).

In order to improve out-of-sample performance, Campbell and Thompson (2008) suggest imposing sign restrictions on the estimated slope parameter $\widehat{\beta}$ in (4.1). In particular, when forming a forecast, if $\widehat{\beta}$ has an unexpected sign, they set $\widehat{\beta} = 0$ in and if $\widehat{r}_{t+1,t+J} < 0$ they set the forecast value equals to zero. Following Campbell and Thompson (2008), we also put their economic restrictions on our time-varying predictive regression model (4.2).

Table 4.1 and Table 4.2 give results for constant coefficients predictive regression forecasts and time-varying coefficients predictive regression forecasts, respectively. In Panel A of Table 4.1, we report the R^2_{os} statistics for the predictive regression model (4.3) without imposing the Campbell (2008) restrictions, while Panel B of Table 4.1 presents results for the predictive regression model that imposes the Campbell and Thompson (2008) restrictions. The R^2_{os} statistics in Panel A suggest that forecasts obtained from the predictive regression models generally fail to outperform the historical average benchmark for all predictors across all horizons. For 6 of the 14 predictors, the R^2_{os} statistics in Panel B are positive for some forecast horizons. This suggests that imposing the Campbell and Thompson (2008) sign restrictions will, in general, improve the predictive regression forecasts when we compare the results in Panel A with those in Panel B.

Both Panels A and B of Table 4.2 show that the R^2_{os} statistics are positive for DP, DY and EP across different horizons. The R^2_{os} statistics in Panel B are slightly greater than their Panel A counterparts. For TBL, LTR and TMS, the R^2_{os} statistics turn from negative to positive as the horizon increases. This suggests that our proposed time-varying coefficients predictive regression forecasts outperform the historical average benchmark. Also the Campbell (2008) sign restrictions slightly improve the forecasting performance of the time-varying coefficients predictive models. Equation 4.1 reports the p-values from the CW MSFE-adjusted statistics. Both Panels A and B of (4.1) show that the positive R^2_{os} statistics (reported in Table 4.2) are significantly greater than zero at conventional levels. These results suggest that the following predictors DP, DY, EP, TBL, LTR and TMS display statistically significant out-of-sample predictive ability across different horizons.

Table 4.1: R^2_{os} by the Traditional Predictive Regression Model (4.3), Evaluation Period: 1947:I to 2018:IV.

Panel A: Unrestricted predictive regression forecasts														
	DP	**DY**	**EP**	**DE**	**BM**	**SVAR**	**NTIS**	**TBL**	**LTR**	**LTY**	**TMS**	**DFY**	**DFR**	**INF**
J = 1	-1.34%	-0.13%	-6.66%	-3.45%	-7.30%	-0.34%	-3.02%	-0.27%	-2.47%	-0.51%	**0.01%**	-1.02%	-6.23%	-0.35%
J = 2	-0.35%	**0.00%**	-6.48%	-5.22%	-6.95%	-0.91%	-5.83%	-0.53%	-3.92%	-0.10%	**0.04%**	-0.91%	-5.00%	-0.37%
J = 3	-0.32%	-2.72%	-5.88%	-5.25%	-8.68%	-1.32%	-9.30%	-1.16%	-5.41%	-2.32%	**0.37%**	-1.70%	-7.67%	-1.26%
J = 4	-3.50%	-4.37%	-9.81%	-4.31%	-17.63%	-2.05%	-14.91%	-2.78%	-6.29%	-1.61%	**0.41%**	-3.40%	-5.39%	-3.98%
J = 5	-5.41%	-7.36%	-9.96%	-3.92%	-19.84%	-2.80%	-16.98%	-5.06%	-8.70%	-4.14%	-0.50%	-4.78%	-8.39%	-4.68%
J = 6	-8.74%	-13.47%	-11.62%	-4.05%	-23.59%	-3.59%	-21.95%	-6.75%	-9.53%	-2.88%	-1.13%	-7.15%	-7.67%	-5.74%
J = 7	-13.99%	-16.64%	-14.69%	-4.78%	-28.82%	-4.32%	-26.35%	-11.34%	-11.75%	-4.26%	-4.05%	-10.33%	-9.32%	-7.23%
J = 8	-17.63%	-20.95%	-14.32%	-6.59%	-32.62%	-5.43%	-29.60%	-18.04%	-15.00%	-5.34%	-8.63%	-13.22%	-7.98%	-8.55%
J = 9	-21.90%	-24.68%	-13.69%	-9.19%	-37.21%	-6.65%	-33.94%	-25.55%	-18.67%	-6.11%	-13.87%	-16.31%	-7.80%	-9.68%
J = 10	-25.55%	-29.47%	-13.18%	-12.24%	-42.70%	-7.66%	-35.68%	-33.97%	-23.03%	-6.90%	-19.21%	-20.69%	-8.81%	-9.72%
J = 11	-29.92%	-33.51%	-14.27%	-15.10%	-46.50%	-10.63%	-34.58%	-43.74%	-28.07%	-8.83%	-25.11%	-23.75%	-10.02%	-10.70%
J = 12	-32.98%	-36.85%	-16.21%	-16.51%	-51.21%	-12.65%	-35.94%	-50.02%	-30.19%	-9.90%	-30.04%	-28.22%	-10.86%	-11.63%
Panel B: Predictive regression forecasts with Campbell and Thompson (2008) constraints														
	DP	**DY**	**EP**	**DE**	**BM**	**SVAR**	**NTIS**	**TBL**	**LTR**	**LTY**	**TMS**	**DFY**	**DFR**	**INF**
J = 1	**0.07%**	**0.86%**	-2.90%	-2.83%	-4.15%	-0.34%	-3.03%	**0.63%**	**0.64%**	-0.14%	**0.01%**	-0.89%	-5.50%	-0.35%
J = 2	**0.61%**	**0.93%**	-1.53%	-4.66%	-4.26%	-0.89%	-5.78%	**0.91%**	**0.55%**	**1.77%**	**0.17%**	-0.91%	-4.81%	-0.37%
J = 3	**0.70%**	**0.55%**	-0.64%	-5.09%	-5.72%	-1.32%	-9.28%	**0.70%**	-0.14%	-0.19%	**0.46%**	-1.70%	-7.03%	-1.26%
J = 4	-0.20%	-0.22%	-2.07%	-4.31%	-10.75%	-2.05%	-14.77%	-0.12%	-0.67%	-0.90%	**0.48%**	-3.35%	-5.27%	-3.98%
J = 5	-1.18%	-0.98%	-1.37%	-3.92%	-12.31%	-2.80%	-16.62%	-1.39%	-1.78%	-3.17%	-0.75%	-4.64%	-7.69%	-4.68%
J = 6	-2.11%	-2.30%	-0.96%	-4.05%	-14.67%	-3.59%	-21.32%	-2.55%	-2.27%	-2.77%	-1.52%	-6.48%	-6.75%	-5.74%
J = 7	-3.61%	-3.96%	-1.54%	-4.78%	-18.23%	-4.32%	-25.48%	-4.92%	-3.29%	-4.24%	-4.03%	-8.68%	-8.42%	-7.23%
J = 8	-5.71%	-6.46%	-0.89%	-6.59%	-21.73%	-5.43%	-28.88%	-8.35%	-4.53%	-5.34%	-7.64%	-10.76%	-7.56%	-8.55%
J = 9	-8.36%	-8.83%	-0.80%	-8.47%	-26.13%	-6.65%	-33.16%	-11.79%	-5.82%	-6.11%	-10.98%	-12.86%	-7.75%	-9.68%
J = 10	-10.95%	-11.42%	-0.50%	-10.52%	-30.65%	-7.66%	-34.94%	-15.15%	-7.14%	-6.90%	-13.40%	-15.42%	-8.81%	-9.72%
J = 11	-13.41%	-13.61%	-0.86%	-12.51%	-33.91%	-10.63%	-34.14%	-18.16%	-8.69%	-8.83%	-15.45%	-17.27%	-10.02%	-10.70%
J = 12	-15.34%	-15.67%	-1.51%	-13.96%	-37.96%	-12.56%	-35.46%	-19.99%	-8.86%	-9.90%	-18.12%	-19.55%	-10.86%	-11.63%

Note. This table reports the R^2_{os} of the 14 predictors from Welch and Goyal (2008) using the traditional linear predictive regression model (4.3), relative to the historical mean benchmark. Panel A reports the R^2_{os} by the unrestricted predictive regressions forecasts, and Panel B reports the R^2_{os} by the predictive regressions forecasts with Campbell and Thompson (2008) constraints. The forecast horizon, J, is from $J = 1$ to $J = 12$. The out-of-sample evaluation period is from 1947:I to 2018:IV.

Table 4.2: R^2_{os} by the Time-Varying Predictive Regression Model (4.1), Evaluation Period: 1947:I to 2018:IV

Panel A: Raw forecasts														
	DP	**DY**	**EP**	**DE**	**BM**	**SVAR**	**NTIS**	**TBL**	**LTR**	**LTY**	**TMS**	**DFY**	**DFR**	**INF**
J = 1	-0.45%	**0.11%**	-2.27%	-4.07%	-3.53%	-12.46%	-3.67%	-2.62%	-3.93%	-2.24%	-1.06%	-3.41%	-2.11%	-1.32%
J = 2	**0.29%**	**0.69%**	-2.91%	-5.56%	-4.90%	-16.77%	-6.43%	-4.51%	-6.64%	-1.22%	-1.98%	-4.39%	-3.88%	-2.45%
J = 3	**1.95%**	**2.24%**	-1.64%	-5.99%	-4.05%	-13.00%	-6.96%	-6.27%	-7.96%	-1.81%	-1.83%	-5.21%	-3.59%	-1.91%
J = 4	**3.58%**	**3.59%**	**0.16%**	-5.81%	-2.93%	-11.89%	-7.27%	-8.18%	-8.71%	-2.56%	-1.89%	-6.25%	-4.40%	-1.53%
J = 5	**5.03%**	**4.88%**	**1.51%**	-5.42%	-2.03%	-9.96%	-7.54%	-9.99%	-8.86%	-4.46%	-1.81%	-6.83%	-4.93%	-3.82%
J = 6	**6.46%**	**5.72%**	**2.28%**	-4.99%	-1.13%	-9.64%	-7.84%	-9.11%	-8.70%	-4.24%	-0.98%	-6.84%	-5.89%	-4.41%
J = 7	**7.46%**	**5.78%**	**2.72%**	-4.30%	-0.57%	-9.77%	-8.27%	-8.35%	-8.78%	-5.30%	-1.11%	-6.91%	-5.75%	-5.51%
J = 8	**7.94%**	**6.09%**	**3.13%**	-4.31%	-0.70%	-10.75%	-8.78%	-7.49%	-8.67%	-5.85%	-0.47%	-7.09%	-6.08%	-6.35%
J = 9	**8.49%**	**6.72%**	**3.80%**	-4.31%	-0.50%	-12.34%	-9.24%	-6.25%	-8.44%	-5.50%	**0.17%**	-7.05%	-6.14%	-6.58%
J = 10	**9.18%**	**7.60%**	**4.38%**	-4.11%	-0.06%	-12.61%	-9.54%	-5.02%	-8.22%	-5.08%	**1.38%**	-7.03%	-6.14%	-6.92%
J = 11	**10.13%**	**8.55%**	**4.67%**	-3.51%	**0.23%**	-13.32%	-9.93%	-3.52%	-8.19%	-5.61%	**3.00%**	-7.16%	-6.32%	-7.96%
J = 12	**11.20%**	**9.31%**	**4.92%**	-2.55%	**0.30%**	-11.72%	-10.29%	-2.26%	-7.96%	-5.62%	**3.88%**	-6.88%	-6.00%	-8.85%
Panel B: Predictive regression forecasts with Campbell and Thompson (2008) constraints														
	DP	**DY**	**EP**	**DE**	**BM**	**SVAR**	**NTIS**	**TBL**	**LTR**	**LTY**	**TMS**	**DFY**	**DFR**	**INF**
J = 1	**0.02%**	**0.98%**	-0.80%	-1.35%	-3.23%	-12.36%	-3.25%	-0.09%	-0.54%	-0.82%	-1.17%	-3.39%	-0.32%	**0.14%**
J = 2	**1.31%**	**1.67%**	-0.53%	-2.79%	-4.16%	-16.73%	-5.40%	**0.32%**	-1.24%	**0.81%**	-2.40%	-4.39%	-2.84%	**0.64%**
J = 3	**2.94%**	**3.02%**	**0.75%**	-3.51%	-3.56%	-13.00%	-5.97%	-0.17%	-1.22%	-0.45%	-2.08%	-5.21%	-3.32%	**1.53%**
J = 4	**4.38%**	**4.18%**	**2.20%**	-3.47%	-2.77%	-11.89%	-6.39%	-1.51%	-1.19%	-1.78%	-2.15%	-6.24%	-4.26%	**0.83%**
J = 5	**5.57%**	**5.15%**	**3.14%**	-3.32%	-1.90%	-9.96%	-6.84%	-3.23%	-0.68%	-4.10%	-2.00%	-6.75%	-4.93%	-1.66%
J = 6	**6.60%**	**5.50%**	**3.83%**	-3.36%	-0.92%	-9.64%	-7.34%	-2.45%	**0.04%**	-4.15%	-1.11%	-6.76%	-5.89%	-2.69%
J = 7	**7.09%**	**5.72%**	**4.30%**	-3.10%	-0.42%	-9.77%	-8.02%	-1.80%	**0.49%**	-5.24%	-1.09%	-6.74%	-5.75%	-4.42%
J = 8	**7.36%**	**6.03%**	**4.77%**	-3.53%	-0.68%	-10.75%	-8.75%	-1.28%	**1.43%**	-5.85%	-0.44%	-6.64%	-6.08%	-5.64%
J = 9	**7.97%**	**6.52%**	**5.17%**	-3.90%	-0.58%	-12.34%	-9.20%	-0.27%	**2.47%**	-5.50%	**0.20%**	-6.23%	-6.14%	-6.19%
J = 10	**8.62%**	**7.30%**	**5.54%**	-4.03%	-0.01%	-12.61%	-9.45%	**1.63%**	**3.64%**	-5.08%	**1.42%**	-5.65%	-6.14%	-6.62%
J = 11	**9.46%**	**8.15%**	**6.01%**	-3.53%	**0.38%**	-13.28%	-9.76%	**3.90%**	**4.37%**	-5.61%	**3.06%**	-5.15%	-6.32%	-7.52%
J = 12	**10.49%**	**9.09%**	**6.23%**	-2.60%	**0.40%**	-11.72%	-9.81%	**5.88%**	**5.47%**	-5.62%	**3.95%**	-4.64%	-6.00%	-8.50%

Note. This table reports the R^2_{os} of the 14 predictors from Welch and Goyal (2008) using the time-varying predictive regression model (4.1), relative to the historical mean benchmark. Panel A reports the R^2_{os} by the unrestricted predictive regressions forecasts, and Panel B reports the R^2_{os} by the predictive regressions forecasts with Campbell and Thompson (2008) constraints. The forecast horizon, J, is from $J = 1$ to $J = 12$. The out-of-sample evaluation period is from 1947 to 2018.

Table 4.3: P-values of the MSFE-Adjust t-Statistics by the Conventional Predictive Regression Model (4.3), Evaluation Period: 1947:I to 2018:IV.

Panel A: Panel A: Unrestricted predictive regression forecasts														
	DP	**DY**	**EP**	**DE**	**BM**	**SVAR**	**NTIS**	**TBL**	**LTR**	**LTY**	**TMS**	**DFY**	**DFR**	**INF**
J = 1	0.077	0.077	0.093	0.645	0.210	0.920	0.344	0.139	0.153	0.305	0.214	0.834	0.985	0.863
J = 2	0.058	0.057	0.068	0.643	0.146	0.896	0.414	0.138	0.103	0.024	0.219	0.989	0.954	0.672
J = 3	0.051	0.028	0.034	0.632	0.118	0.999	0.395	0.148	0.106	0.024	0.174	0.999	0.868	0.958
J = 4	0.035	0.028	0.017	0.656	0.103	1.000	0.391	0.131	0.106	0.064	0.118	0.943	0.888	1.000
J = 5	0.036	0.028	0.015	0.721	0.109	0.999	0.385	0.134	0.100	0.184	0.137	0.938	0.846	1.000
J = 6	0.035	0.028	0.017	0.832	0.112	0.999	0.390	0.133	0.091	0.200	0.139	0.913	0.832	1.000
J = 7	0.036	0.036	0.021	0.893	0.124	1.000	0.464	0.152	0.083	0.889	0.218	0.913	0.914	0.998
J = 8	0.044	0.042	0.025	0.928	0.164	1.000	0.505	0.168	0.065	0.983	0.296	0.934	0.990	0.999
J = 9	0.051	0.047	0.027	0.926	0.206	0.998	0.552	0.171	0.050	0.984	0.339	0.948	0.994	0.999
J = 10	0.056	0.049	0.029	0.920	0.233	1.000	0.585	0.156	0.038	0.970	0.333	0.953	0.988	0.997
J = 11	0.058	0.048	0.032	0.913	0.247	0.983	0.620	0.147	0.032	0.972	0.308	0.958	0.986	0.996
J = 12	0.057	0.048	0.035	0.909	0.260	0.976	0.721	0.144	0.027	0.978	0.321	0.958	0.990	0.996
Panel B: Predictive regression forecasts with Campbell and Thompson (2008) constraints														
CT	**DP**	**DY**	**EP**	**DE**	**BM**	**SVAR**	**NTIS**	**TBL**	**LTR**	**LTY**	**TMS**	**DFY**	**DFR**	**INF**
J = 1	0.053	0.037	0.044	0.558	0.237	0.920	0.351	0.111	0.058	0.245	0.230	0.827	0.985	0.863
J = 2	0.045	0.042	0.024	0.591	0.153	0.895	0.417	0.101	0.059	0.008	0.213	0.989	0.951	0.672
J = 3	0.045	0.034	0.016	0.619	0.126	0.999	0.404	0.124	0.083	0.010	0.168	0.999	0.874	0.958
J = 4	0.045	0.041	0.013	0.656	0.116	1.000	0.402	0.130	0.096	0.048	0.118	0.962	0.887	1.000
J = 5	0.050	0.041	0.014	0.721	0.108	0.999	0.388	0.152	0.104	0.163	0.156	0.961	0.824	1.000
J = 6	0.050	0.042	0.015	0.832	0.100	0.999	0.391	0.172	0.103	0.212	0.168	0.943	0.792	1.000
J = 7	0.052	0.050	0.017	0.893	0.107	1.000	0.457	0.204	0.102	0.891	0.241	0.939	0.892	0.998
J = 8	0.061	0.061	0.020	0.928	0.138	1.000	0.504	0.244	0.093	0.983	0.298	0.950	0.989	0.999
J = 9	0.074	0.070	0.023	0.923	0.181	0.998	0.551	0.271	0.084	0.984	0.312	0.956	0.994	0.999
J = 10	0.085	0.077	0.025	0.915	0.210	1.000	0.583	0.292	0.076	0.970	0.282	0.954	0.988	0.997
J = 11	0.093	0.082	0.026	0.907	0.225	0.983	0.622	0.304	0.072	0.972	0.245	0.955	0.986	0.996
J = 12	0.097	0.086	0.027	0.904	0.243	0.976	0.718	0.302	0.063	0.978	0.255	0.950	0.990	0.996

Note. This table reports the p-values for the Clark and West (2007) MSFE-adjusted statistic tests the null hypothesis that the historical average MSFE is less than or equal to the predictive regression MSFE against the alternative that the historical average MSFE is greater than the predictive regression MSFE (corresponding to $H_0 : R^2_{os} \leq 0$ against $HA : R^2_{os} > 0$) using the time-varying predictive regression model (4.3). Panel A reports the p-values by the unrestricted predictive regressions forecasts, and Panel B reports the p-values by the predictive regressions forecasts with Campbell and Thompson (2008) constraints. The forecast horizon, J, is from $J = 1$ to $J = 12$. The out-of-sample evaluation period is from 1947 to 2018.

Table 4.4: P-values of the MSFE-adjust statistics by the time-varying model (4.1), evaluation period: 1947:I to 2018:IV.

Panel A: Unrestricted predictive regression forecasts														
Raw	**DP**	**DY**	**EP**	**DE**	**BM**	**SVAR**	**NTIS**	**TBL**	**LTR**	**LTY**	**TMS**	**DFY**	**DFR**	**INF**
J = 1	0.091	0.034	0.520	0.590	0.586	0.204	0.625	0.058	0.277	0.279	0.028	0.704	0.224	0.290
J = 2	0.032	0.020	0.307	0.489	0.347	0.131	0.609	0.066	0.234	0.041	0.048	0.652	0.660	0.246
J = 3	0.014	0.011	0.126	0.343	0.151	0.351	0.373	0.061	0.189	0.058	0.041	0.634	0.531	0.136
J = 4	0.009	0.010	0.053	0.262	0.080	0.309	0.258	0.053	0.135	0.074	0.044	0.590	0.583	0.116
J = 5	0.008	0.010	0.033	0.193	0.059	0.272	0.242	0.042	0.099	0.201	0.048	0.538	0.543	0.180
J = 6	0.007	0.011	0.029	0.149	0.045	0.224	0.304	0.029	0.072	0.113	0.043	0.463	0.536	0.151
J = 7	0.008	0.015	0.032	0.110	0.038	0.150	0.387	0.024	0.052	0.257	0.058	0.393	0.454	0.188
J = 8	0.011	0.018	0.037	0.089	0.048	0.121	0.406	0.018	0.032	0.308	0.056	0.296	0.433	0.155
J = 9	0.013	0.019	0.041	0.076	0.062	0.131	0.405	0.013	0.019	0.295	0.047	0.206	0.384	0.155
J = 10	0.014	0.020	0.045	0.069	0.072	0.114	0.388	0.010	0.012	0.231	0.035	0.156	0.331	0.126
J = 11	0.015	0.019	0.050	0.065	0.077	0.178	0.374	0.008	0.009	0.246	0.027	0.123	0.304	0.140
J = 12	0.016	0.020	0.055	0.060	0.087	0.156	0.366	0.008	0.008	0.238	0.026	0.103	0.281	0.116
Panel B: Predictive regression forecasts with Campbell and Thompson (2008) restrictions														
CT	**DP**	**DY**	**EP**	**DE**	**BM**	**SVAR**	**NTIS**	**TBL**	**LTR**	**LTY**	**TMS**	**DFY**	**DFR**	**INF**
J = 1	0.076	0.022	0.299	0.372	0.555	0.192	0.609	0.031	0.146	0.163	0.055	0.703	0.141	0.131
J = 2	0.024	0.015	0.134	0.379	0.305	0.130	0.580	0.019	0.139	0.014	0.069	0.652	0.541	0.078
J = 3	0.012	0.010	0.054	0.302	0.147	0.351	0.354	0.025	0.122	0.032	0.047	0.634	0.522	0.058
J = 4	0.009	0.011	0.030	0.238	0.087	0.309	0.254	0.034	0.103	0.067	0.049	0.589	0.587	0.079
J = 5	0.009	0.011	0.025	0.171	0.068	0.272	0.240	0.040	0.086	0.190	0.056	0.534	0.543	0.136
J = 6	0.009	0.014	0.025	0.134	0.052	0.224	0.301	0.035	0.071	0.119	0.048	0.460	0.536	0.123
J = 7	0.010	0.017	0.028	0.101	0.044	0.150	0.383	0.031	0.056	0.259	0.060	0.390	0.454	0.169
J = 8	0.014	0.021	0.034	0.085	0.053	0.121	0.407	0.026	0.039	0.308	0.056	0.291	0.433	0.146
J = 9	0.015	0.023	0.040	0.074	0.068	0.131	0.405	0.021	0.027	0.295	0.047	0.206	0.384	0.152
J = 10	0.017	0.023	0.045	0.069	0.076	0.114	0.388	0.016	0.020	0.231	0.035	0.155	0.331	0.127
J = 11	0.018	0.024	0.049	0.066	0.079	0.178	0.377	0.014	0.017	0.246	0.027	0.126	0.304	0.139
J = 12	0.019	0.024	0.054	0.061	0.088	0.156	0.361	0.014	0.014	0.238	0.026	0.109	0.281	0.117

Note. This table reports the p-values for the Clark and West (2007) MSFE-adjusted statistic tests the null hypothesis that the historical average MSFE is less than or equal to the predictive regression MSFE against the alternative that the historical average MSFE is greater than the predictive regression MSFE (corresponding to $H_0 : R^2_{os} \leq 0$ against $HA : R^2_{os} > 0$) using the time-varying predictive regression model (4.1). Panel A reports the p-values by the unrestricted predictive regressions forecasts, and Panel B reports the p-values by the predictive regressions forecasts with Campbell and Thompson (2008) constraints. The forecast horizon, J, is from $J = 1$ to $J = 12$. The out-of-sample evaluation period is from 1947 to 2018.

4.3 Economic evaluation

In previous section, the MSFE-based tests were shown to support return predictability by using the time-varying predictive regressions models. Overwhelmingly, the R^2_{os} and the associated MSFE-adjusted tests are the most popular statistical measures of out-of-sample forecasting performance, as these statistics directly tells us whether the proposed predictive regression model can provide statistically smaller forecast error than the historical mean benchmark model. An obvious limitation of the R^2_{OS} based statistical measures is the difficulty to explicitly account for the risks by an investor to time the market over the out-of-sample period. As a result, we are unable to know how much economic values we gain from using the proposed forecasting models relative to the benchmark model. To address this, we alternatively seek the metrics in terms of economic meanings. In this section, we assess the economic value of equity risk premium predictability for an investor with mean-variance preference. Following Campbell and Thompson (2008) and Rapach and Zhou (2013), we adopt the utility-based metric of out-of-sample performance.

First, we define a mean-variance market timing strategy. Consider a mean variance investor who decides to allocate her or his wealth between the market equity portfolio and a risk-free bill at each time period t. The investor puts on the risky market portfolio with the weight:

$$\boxed{w_t = \frac{1}{\gamma}\frac{\widehat{r}_{t+1,t+J}}{\widehat{\sigma}^2_{t+1,t+J}},}$$

where γ is the risk-aversion parameter (following Neely et al. (2014), we select $\gamma = 5$), $\widehat{r}_{t+1,t+J}$ is the predicted value of quarterly equity premium on the U.S. aggregate stock market based on predictive regressions by one of the financial ratios or macroeconomic variables up to time t, and $\widehat{\sigma}^2_{t+1,t+J}$ is the predicted value of the variance of quarterly equity premium. Following Campbell and Thompson (2008), we calculate the sample variance of past five-years or equivalently 20-quarters returns at t moving window as the forecast of the variance of the equity risk premium from $t+1$ to $t+J$. We also restrict the weight w_t between 0 and 1.5 to prevent short-sales and allow an upper limit of 50 percent leverage, following Campbell and Thompson (2008). Therefore, the share $1 - w_t$ is allocated to risk-free bills.

The portfolio return at time $t + 1$ is given by

$$R_{p,t+1} = w_t r_{t+1} + rf_{t+1}, \tag{4.8}$$

where rf_{t+1} is the return on risk-free bills at time $t + 1$. Therefore, the investor realizes an average utility level for the portfolio is :

$$\mathrm{U} = \widehat{\mu}_p - 0.5\gamma\widehat{\sigma}_p^2, \tag{4.9}$$

where $\widehat{\mu}_p$ and $\widehat{\sigma}_p^2$ are the sample mean and variance, respectively, for the investment portfolio over the out-of-sample evaluation period. The utility can be interpreted as the risk-free rate of return that an investor is willing to accept instead of adopting the given risky portfolio. We compute certainty equivalent return (CER) gain (also called utility gain), denoted as ΔCER, as the differences in the realized average utilities between investors who use the economic variables against those who use the historical mean returns as the forecasts,

$$\Delta\mathrm{CER} = \mathrm{U}_{pr} - \mathrm{U}_{hm}. \tag{4.10}$$

The CER gain is then the difference between the utility for the investor when this investor uses the predictive regression model-based forecast to perform asset allocation, denoted as U_{pr}, and for the investor when this investor uses the historical mean benchmark forecast to form the portfolio, denoted as U_{hm}. In specific, we modify the CER by multiplying this difference by 400, so that it can be interpreted as the annual percentage portfolio management fee that an investor would be willing to pay to have access to the predictive regression forecast instead of relying on the historical average forecast (Rapach and Zhou, 2013). Moreover, we adjust CER for net of transactions costs. The transaction costs are calculated using the quarterly turnover measures and we assume that the transactions cost equal to 50 basis points of the aggregate portfolio value per transaction, consistent with Neely et al. (2014). In this way, the CER provides a rational measure of the economic value of return predictability, and the CER approach has been extensively used in empirical work (see, for example, Jiang et al. (2019), Neely et al. (2014) Rapach and Zhou (2013), Rapach et al. (2016), and others).

Furthermore, we estimate the Sharpe ratio, denoted as SR, of realized returns generated by the portfolio, which is computed as

$$\text{SP} = \frac{\widehat{\mu}_p}{\widehat{\sigma}_p}. \tag{4.11}$$

Finally, we report average quarterly turnover. The quarterly turnover is the percentage of wealth traded each month. For the portfolios, it gives the relative average turnover; that is, average quarterly turnover for the portfolio based on the predictive regression model forecasts divided by the average quarterly turnover for the portfolio based on the historical mean forecasts.

Table 4.5 reports CER gains, ΔCER, for an investor with a relative risk coefficient of three who relies on the univariate predictive regression forecast given by the traditional predictive regression model (4.3) in Panel A, and the time-varying predictive regression model (4.1) with the reflection approach, in Panel B. In accord with the R^2_{os} statistics in Table 4.1 and Table 4.2, Table 4.5 shows that the CER gains are positive across all horizons and for all the predictors. This suggests that investors overall realizes the higher average utility gains when they use the forecasts based on the predictive regression model (4.1) than the historical average. Panel A and Panel B of Table 4.5 report CER gains using, respectively, the constant coefficients predictive regression forecasts and the time-varying coefficients predictive regression forecasts. Comparing Panel A to Panel B, we see that the CER gains in Panel B are greater than their Panel A counterparts. In particular, for DP and DY, the CER gains are above 2% for longer horizons $J \geq 3$ when using the time-varying coefficients predictive regression forecasts but below 2% when using the constant parameters predictive regression forecasts.

Table 4.6 reports Sharpe ratios for the bivariate predictive regression forecast given by the traditional predictive regression model (4.3) in Panel A, and the time-varying predictive regression model (4.1) with the reflection approach in Panel B. The results suggest that the Sharpe ratios, in general, are higher in Panel B than in Panel A.

Table 4.7 reports average turnover ratios for the bivariate predictive regression forecast given by the traditional predictive regression model (4.3) in Panel A, and the traditional predictive regression model (4.1) with the reflection approach in Panel B. As we can see, the model (4.1) produces much higher average turnover ratios than not only the benchmark model but also the traditional predictive regression model (4.3). The higher turnover ratio implies trading more often, thus indicating the higher transaction costs for portfolios. Even though there exist relatively high turnovers for 14 economic variables that generate higher transactions costs, the time-varying predictive regression model (4.1) still achieve the highest CER gains among the forecasting models. The presence of high turnover ratios and CER gains simultaneously further highlights the practical usefulness of the time-varying predictive regression models.

To conclude, this chapter has investigated the empirical out-of-sample performance of time-varying coefficients regressions to predict the equity premium over multiple horizons. We find that significant evidence that dividend-price ratio, dividend yield, earning-price ratio can predict out-of-sample across different horizons and that T-bill rate, long-term returns and term spread can predict out-of-sample at long horizons.

Table 4.5: CER, Evaluation Period: 1947:I to 2018:IV.

Panel A: Traditional linear predictive regression model														
OLS	**DP**	**DY**	**EP**	**DE**	**BM**	**SVAR**	**NTIS**	**TBL**	**LTR**	**LTY**	**TMS**	**DFY**	**DFR**	**INF**
J = 1	1.56%	1.75%	1.67%	1.34%	1.17%	1.36%	1.41%	1.81%	1.85%	1.37%	1.69%	1.27%	0.77%	1.40%
J = 2	1.87%	1.95%	1.99%	2.06%	1.61%	2.03%	2.12%	2.18%	2.13%	2.07%	2.29%	1.98%	1.70%	2.05%
J = 3	1.96%	1.91%	2.17%	2.31%	1.79%	2.21%	2.43%	2.28%	2.21%	2.11%	2.51%	2.20%	1.92%	2.24%
J = 4	1.93%	1.92%	2.27%	2.46%	1.84%	2.36%	2.64%	2.27%	2.27%	2.22%	2.57%	2.23%	2.11%	2.32%
J = 5	1.93%	1.93%	2.33%	2.56%	1.93%	2.44%	2.73%	2.25%	2.28%	2.26%	2.55%	2.31%	2.17%	2.42%
J = 6	1.93%	1.92%	2.38%	2.58%	2.01%	2.51%	2.76%	2.25%	2.32%	2.41%	2.56%	2.32%	2.30%	2.49%
J = 7	1.92%	1.93%	2.43%	2.60%	2.04%	2.58%	2.78%	2.20%	2.35%	2.53%	2.49%	2.32%	2.38%	2.54%
J = 8	1.93%	1.93%	2.47%	2.61%	2.06%	2.62%	2.79%	2.14%	2.36%	2.59%	2.40%	2.35%	2.48%	2.58%
J = 9	1.91%	1.92%	2.49%	2.60%	2.06%	2.63%	2.80%	2.10%	2.36%	2.61%	2.34%	2.37%	2.53%	2.61%
J = 10	1.90%	1.90%	2.50%	2.58%	2.05%	2.66%	2.80%	2.07%	2.36%	2.63%	2.33%	2.37%	2.56%	2.62%
J = 11	1.88%	1.89%	2.49%	2.56%	2.04%	2.59%	2.80%	2.05%	2.36%	2.60%	2.33%	2.38%	2.56%	2.63%
J = 12	1.86%	1.88%	2.48%	2.56%	2.03%	2.57%	2.80%	2.05%	2.38%	2.61%	2.31%	2.36%	2.58%	2.64%
Panel B: Time-varying predictive regression model														
TV	**DP**	**DY**	**EP**	**DE**	**BM**	**SVAR**	**NTIS**	**TBL**	**LTR**	**LTY**	**TMS**	**DFY**	**DFR**	**INF**
J = 1	1.32%	1.49%	1.45%	1.56%	1.09%	1.25%	1.16%	1.81%	1.69%	1.41%	1.95%	1.33%	1.58%	1.70%
J = 2	1.92%	1.97%	2.04%	2.12%	1.80%	1.90%	1.86%	2.26%	2.12%	2.17%	2.44%	2.01%	1.97%	2.31%
J = 3	2.15%	2.22%	2.27%	2.36%	2.08%	2.22%	2.26%	2.40%	2.29%	2.24%	2.68%	2.23%	2.22%	2.48%
J = 4	2.32%	2.34%	2.43%	2.49%	2.24%	2.38%	2.49%	2.49%	2.38%	2.30%	2.73%	2.35%	2.35%	2.58%
J = 5	2.40%	2.41%	2.52%	2.60%	2.38%	2.49%	2.59%	2.57%	2.46%	2.42%	2.78%	2.48%	2.46%	2.60%
J = 6	2.46%	2.46%	2.59%	2.67%	2.49%	2.59%	2.65%	2.66%	2.54%	2.52%	2.89%	2.57%	2.52%	2.64%
J = 7	2.49%	2.56%	2.64%	2.73%	2.57%	2.68%	2.68%	2.70%	2.60%	2.62%	2.93%	2.65%	2.62%	2.66%
J = 8	2.54%	2.61%	2.68%	2.78%	2.60%	2.76%	2.71%	2.74%	2.64%	2.69%	2.98%	2.71%	2.69%	2.73%
J = 9	2.60%	2.65%	2.70%	2.82%	2.63%	2.79%	2.75%	2.76%	2.67%	2.75%	3.00%	2.76%	2.74%	2.76%
J = 10	2.64%	2.67%	2.72%	2.84%	2.67%	2.83%	2.77%	2.78%	2.68%	2.77%	3.01%	2.79%	2.77%	2.78%
J = 11	2.65%	2.68%	2.73%	2.86%	2.70%	2.84%	2.79%	2.79%	2.69%	2.79%	3.03%	2.81%	2.78%	2.79%
J = 12	2.65%	2.69%	2.73%	2.87%	2.72%	2.86%	2.80%	2.81%	2.70%	2.80%	3.03%	2.83%	2.81%	2.83%

Note. This table reports annualized average utility gain for an investor with mean-variance preferences and relative risk-aversion coefficient of five who allocates between equities and risk-free bills using either an historical average or equity risk premium forecast based on predictive regression of the 14 predictors from Welch and Goyal (2008). Panel A reports the results based on the traditional linear predictive regression model (4.3), and Panel B reports the results based on the time-varying predictive regression model (4.1). The out-of-sample evaluation period is from 1947:I to 2018:IV.

Table 4.6: Sharpe Ratio, Evaluation Period: 1947:I to 2018:IV.

Panel A: Linear predictive regression model														
OLS	**DP**	**DY**	**EP**	**DE**	**BM**	**SVAR**	**NTIS**	**TBL**	**LTR**	**LTY**	**TMS**	**DFY**	**DFR**	**INF**
J = 1	18.55%	22.40%	19.73%	17.39%	12.11%	16.32%	18.62%	21.85%	22.93%	16.62%	20.74%	14.96%	9.40%	16.67%
J = 2	27.80%	29.73%	26.62%	25.42%	19.17%	25.91%	26.19%	31.67%	31.59%	27.16%	30.05%	25.15%	20.43%	26.61%
J = 3	33.18%	32.21%	33.00%	31.35%	24.67%	31.95%	33.13%	37.68%	35.87%	30.04%	37.46%	31.49%	26.03%	32.73%
J = 4	35.48%	34.49%	39.07%	36.42%	28.43%	37.46%	40.48%	40.59%	39.33%	34.28%	43.00%	35.19%	31.87%	36.37%
J = 5	37.37%	36.55%	42.97%	41.95%	33.04%	42.58%	45.92%	43.28%	41.92%	37.85%	47.31%	39.99%	34.98%	41.67%
J = 6	38.76%	37.67%	46.81%	46.73%	37.16%	47.48%	48.95%	46.53%	45.37%	44.43%	51.99%	43.96%	40.78%	46.56%
J = 7	39.88%	39.60%	50.74%	51.05%	40.78%	52.40%	51.95%	48.29%	48.40%	50.70%	54.48%	47.78%	46.54%	51.01%
J = 8	41.51%	41.25%	54.66%	54.93%	43.79%	57.01%	55.68%	49.35%	51.43%	56.26%	55.34%	52.34%	52.99%	55.98%
J = 9	42.40%	42.61%	57.36%	58.80%	45.40%	61.13%	59.78%	50.55%	53.68%	60.35%	56.45%	56.37%	57.29%	60.55%
J = 10	43.16%	43.18%	59.87%	61.94%	46.89%	64.53%	63.41%	51.80%	55.73%	63.68%	58.29%	59.42%	61.48%	64.50%
J = 11	43.37%	43.64%	61.23%	64.27%	48.18%	66.13%	66.46%	52.56%	57.57%	65.72%	60.03%	62.33%	64.61%	67.60%
J = 12	44.07%	44.29%	62.62%	66.52%	49.35%	68.51%	69.88%	54.25%	59.44%	68.80%	61.27%	65.03%	67.41%	70.27%
Panel B: Time-varying predictive regression model														
TV	**DP**	**DY**	**EP**	**DE**	**BM**	**SVAR**	**NTIS**	**TBL**	**LTR**	**LTY**	**TMS**	**DFY**	**DFR**	**INF**
J = 1	15.27%	17.03%	17.21%	18.79%	13.10%	16.15%	14.92%	21.75%	20.05%	17.03%	23.82%	16.53%	18.88%	20.46%
J = 2	24.86%	26.56%	26.20%	27.39%	21.86%	23.13%	22.84%	31.82%	30.53%	28.27%	32.52%	25.32%	24.76%	31.06%
J = 3	32.58%	35.55%	32.85%	34.29%	28.28%	30.01%	31.06%	37.98%	37.93%	31.99%	40.81%	31.27%	31.58%	37.09%
J = 4	39.96%	40.87%	38.98%	39.51%	33.45%	35.52%	38.67%	42.45%	42.65%	34.80%	45.55%	35.80%	36.34%	41.82%
J = 5	44.84%	45.19%	43.75%	44.84%	38.65%	40.42%	43.74%	46.67%	47.41%	40.07%	50.47%	41.49%	40.82%	44.95%
J = 6	48.98%	49.36%	48.26%	49.57%	43.70%	45.89%	48.08%	51.67%	51.82%	45.92%	56.28%	46.84%	44.54%	48.92%
J = 7	52.94%	54.50%	52.74%	53.88%	48.89%	51.91%	52.02%	55.40%	56.08%	51.12%	60.81%	51.88%	51.12%	52.36%
J = 8	57.34%	58.00%	56.81%	58.49%	52.06%	57.91%	56.37%	59.26%	59.99%	56.45%	65.98%	56.75%	56.62%	57.62%
J = 9	61.73%	61.93%	60.57%	63.13%	55.37%	62.65%	60.80%	62.30%	63.43%	61.51%	70.00%	61.15%	61.10%	61.24%
J = 10	65.14%	64.77%	64.09%	67.03%	59.23%	66.97%	64.69%	65.35%	66.19%	65.25%	72.78%	65.28%	64.85%	64.86%
J = 11	67.37%	66.92%	66.80%	70.24%	62.56%	69.65%	68.22%	67.85%	68.54%	68.81%	76.15%	68.73%	68.15%	67.89%
J = 12	69.44%	69.56%	69.43%	72.99%	65.61%	73.05%	71.94%	70.37%	70.75%	70.80%	79.04%	71.75%	71.47%	71.39%

Note. This table reports Sharpe ratio for an investor with mean-variance preferences and relative risk-aversion coefficient of five who allocates between equities and risk-free bills using either an historical average or equity risk premium forecast based on predictive regression of the 14 predictors from Welch and Goyal (2008). Panel A reports the results based on the traditional linear predictive regression model (4.3), and Panel B reports the results based on the time-varying predictive regression model (4.1). The out-of-sample evaluation period is from 1947:I to 2018:IV.

Table 4.7: Relative Turnover, Evaluation Period: 1947:I to 2018:IV.

Panel A: Linear predictive regression model														
OLS	**DP**	**DY**	**EP**	**DE**	**BM**	**SVAR**	**NTIS**	**TBL**	**LTR**	**LTY**	**TMS**	**DFY**	**DFR**	**INF**
J = 1	7.36%	9.65%	9.92%	7.89%	8.28%	6.91%	13.76%	8.33%	6.30%	33.09%	12.65%	8.26%	37.00%	10.17%
J = 2	8.83%	9.20%	10.99%	7.50%	10.04%	7.42%	13.72%	8.77%	5.94%	47.94%	12.52%	6.64%	30.83%	10.14%
J = 3	8.62%	7.05%	11.42%	7.36%	10.27%	6.86%	13.41%	8.46%	5.96%	51.22%	12.16%	6.40%	40.14%	10.30%
J = 4	7.15%	7.31%	11.14%	6.89%	9.39%	6.42%	13.87%	9.01%	6.15%	39.15%	13.39%	6.94%	31.39%	9.17%
J = 5	7.47%	7.07%	11.34%	6.41%	9.60%	6.62%	13.63%	8.94%	6.34%	38.74%	13.79%	6.98%	36.61%	9.50%
J = 6	7.26%	6.77%	11.50%	6.17%	9.30%	6.43%	14.43%	9.04%	6.57%	32.74%	14.24%	7.14%	32.18%	9.27%
J = 7	7.00%	6.90%	11.20%	6.37%	9.02%	6.05%	14.38%	8.77%	6.38%	20.90%	15.06%	7.36%	33.53%	10.62%
J = 8	7.14%	7.22%	11.15%	6.77%	9.59%	6.34%	14.70%	8.49%	5.96%	14.25%	15.88%	6.53%	24.20%	10.65%
J = 9	7.43%	6.87%	11.28%	7.07%	9.66%	6.13%	14.58%	7.96%	5.70%	10.99%	16.61%	6.70%	19.50%	10.51%
J = 10	7.17%	7.38%	11.10%	7.20%	9.35%	6.63%	14.80%	7.51%	5.53%	11.44%	16.19%	6.60%	18.14%	9.72%
J = 11	7.67%	7.28%	11.43%	7.05%	9.66%	6.27%	14.37%	6.88%	5.29%	12.92%	15.96%	6.29%	17.07%	9.36%
J = 12	7.51%	7.34%	11.40%	7.06%	9.58%	6.31%	14.26%	6.87%	5.12%	10.53%	15.96%	6.38%	14.93%	10.03%
Panel B: Time-varying predictive regression model														
TV	**DP**	**DY**	**EP**	**DE**	**BM**	**SVAR**	**NTIS**	**TBL**	**LTR**	**LTY**	**TMS**	**DFY**	**DFR**	**INF**
J = 1	12.08%	14.06%	9.96%	10.54%	13.18%	15.11%	13.58%	9.32%	7.31%	42.14%	18.41%	10.78%	36.52%	19.43%
J = 2	13.34%	14.23%	10.02%	9.67%	13.82%	15.40%	12.78%	9.86%	7.01%	41.17%	18.27%	10.77%	22.37%	20.53%
J = 3	13.74%	13.65%	10.49%	8.57%	13.87%	13.23%	11.92%	9.37%	7.31%	39.47%	17.55%	9.78%	14.12%	19.70%
J = 4	13.63%	13.59%	10.54%	8.14%	14.05%	11.97%	11.44%	8.88%	7.31%	34.75%	16.02%	9.28%	13.47%	16.75%
J = 5	13.27%	13.30%	10.95%	7.49%	13.71%	12.47%	10.67%	8.19%	7.19%	27.02%	14.92%	8.67%	12.77%	14.12%
J = 6	13.07%	13.44%	10.60%	7.36%	12.88%	12.09%	10.07%	7.63%	6.73%	28.83%	12.61%	8.37%	13.55%	15.00%
J = 7	13.13%	13.34%	10.54%	7.24%	12.71%	12.58%	9.47%	7.33%	6.52%	20.60%	9.37%	7.46%	13.39%	11.92%
J = 8	13.60%	13.26%	10.52%	7.30%	12.96%	12.40%	9.51%	7.01%	6.47%	15.63%	7.92%	6.77%	9.56%	11.93%
J = 9	13.73%	12.43%	10.56%	7.27%	12.97%	11.51%	9.27%	6.91%	6.09%	11.57%	8.02%	6.62%	7.94%	11.57%
J = 10	13.00%	12.66%	10.21%	7.46%	12.46%	10.94%	9.07%	6.88%	6.00%	13.47%	8.15%	6.68%	9.57%	11.23%
J = 11	13.30%	12.38%	10.53%	7.43%	12.13%	10.27%	8.54%	7.18%	5.79%	11.67%	8.29%	6.82%	9.33%	11.22%
J = 12	13.12%	12.02%	10.30%	7.35%	11.98%	9.32%	8.67%	7.01%	5.87%	11.73%	7.91%	6.91%	8.36%	11.21%

Note. This table reports annualized relative turnover ratio for an investor with mean-variance preferences and relative risk-aversion coefficient of five who allocates between equities and risk-free bills using either an historical average or equity risk premium forecast based on predictive regression of the 14 predictors from Welch and Goyal (2008). Panel A reports the results based on the traditional linear predictive regression model (4.3), and Panel B reports the results based on the time-varying predictive regression model (4.1). The out-of-sample evaluation period is from 1947:I to 2018:IV.

Chapter 5

Conclusion and Future Directions

5.1 Conclusion

The overall purpose of this thesis is to propose a system of time-varying models for predictive regressions to deal with the econometric issues discussed in Chapter 1. In Chapter 2, to simultaneously take into account parameter instability, inference with persistent predictors, and innovations' correlation, we propose a system of time-varying models for predictive regressions. First, we propose a time-varying AR(1) process for predictors. We first consider a time-varying AR process as the DGP for predictors in predictive regression literature. In particular, the time-varying AR(1) model covers a vital class of locally stationary processes. Second, we develop a time-varying extension of BN decomposition for the time-varying MA(∞) process.

In Chapter 3, we implement the time-varying models to test predictability for quarterly US aggregate stock returns. First, we find significant time-variation in the coefficients of the 14 popular predictors (Welch and Goyal, 2008) in the time-varying AR(1) models. The time-varying AR(1) model provides a better description of predictors than the linear AR(1) model with constant coefficients (i.e., higher R^2). Also, the estimated coefficients of individual predictors change smoothly over time, and the absolute values of $\widehat{\rho}(\tau_t)$ are generally less than, showing the validity of local stationarity. Second, we find that the estimated coefficients $\widehat{\beta}(\tau_t)$ of the 14 predictors suggested are time-varying. More importantly, the 14 popularly used predictors indeed contain predictive content of stock returns. Like the dividend-price ratio and book-to-market ratio, the valuation ratios have strong predictability for the majority of the sub-samples. The predictability is incredibly strong during economic recession periods, such as the oil shock of 1973-1974 and the global financial crisis of 2007-2009.

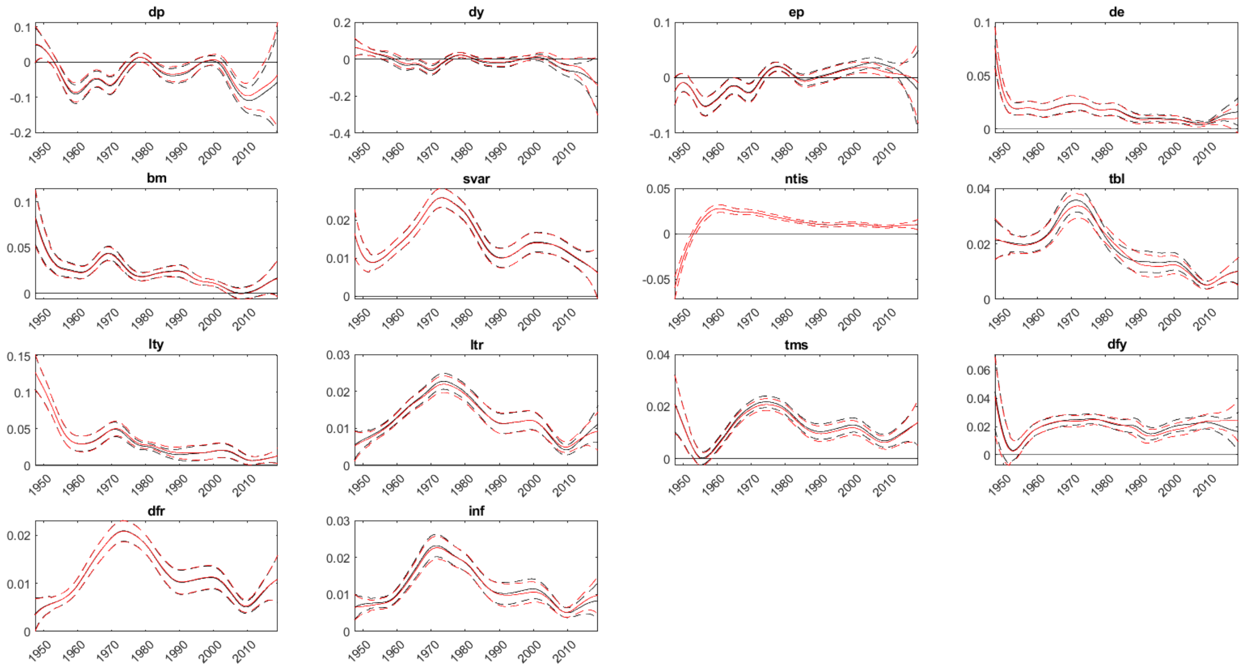
In Chapter 4, we use the time-varying predictive regression models to forecast future stock returns out-of-sample. We predict future quarterly excess returns out-of-sample. This chapter's primary contribution is to exhibit reliable out-of-sample return predictability by the Welch and Goyal (2008) predictors. To the best of our knowledge, our time-varying predictive regressions model produces the highest R_{os}^2 for the same list of predictors from Welch and Goyal (2008). We show that although seemingly useless at short horizons like $J = 1$ or $J = 2$, the time-varying models have a strong ability to predict returns at long horizons. Although non-financial variables like default yield spread and inflation mostly fail to forecast short-horizon returns, they can provide significant predictive power for stock return at long horizons, e.g., $J > 10$. Therefore, we conclude that the commonly used predictors show significant out-of-sample predictability for stock returns using the time-varying models largely missed by the traditional models.

5.2 Future direction

We present a closely related topic as a future research direction based on this thesis. We propose a time-varying autoregressive distributed lag (ARDL) model to forecast economic activities using the 14 variables considered in this thesis. The motivation is straightforward. According to Cochrane (2008), equity premium forecasts are more plausibly related to macroeconomic risk if the equity premium predictors can also forecast business cycles. Stock and Watson (2003) show that the forecasting power of individual economic variables to output growth can be highly unstable over time.

Cochrane (2008) points out that stock return forecasts are more plausibly related to macroeconomic risk if the predictors can also forecast business cycles. There is evidence that economic variables from the literature that predict future stock returns also have predictive ability for real output growth. For example, As shown by Stock and Watson

Figure 5.1: $\widehat{\alpha}(\tau_t)$ in the Time-Varying ARDL Model



Note. This figure shows the Time-varying local linear estimates $\widehat{\alpha}(\tau_t)$ in the time-varying ARDL model $y_t = \alpha(\tau_t) + \phi(\tau_t)y_{t-1} + \beta(\tau_t)x_{t-1} + \lambda\widehat{\varepsilon}_t + \eta_t$ using the 14 individual predictors.

(2003) and others, the forecasting power of individual economic variables to output growth can be highly unstable over time, very similar to the situation for time-varying stock return predictability. We, similar to Stock and Watson (2003), form macroeconomic growth forecasts using the 14 variables considered in this thesis via a time-varying autoregressive distributed lag (tv-ARDL) model:

$$y_t = \alpha(\tau_t) + \phi(\tau_t)y_{t-1} + \beta(\tau_t)x_{t-1} + \lambda\widehat{\varepsilon}_t + \eta_t \tag{5.1}$$

$$x_t = \gamma(\tau_t) + \rho(\tau_t)x_{t-1} + \varepsilon_t, \tag{5.2}$$

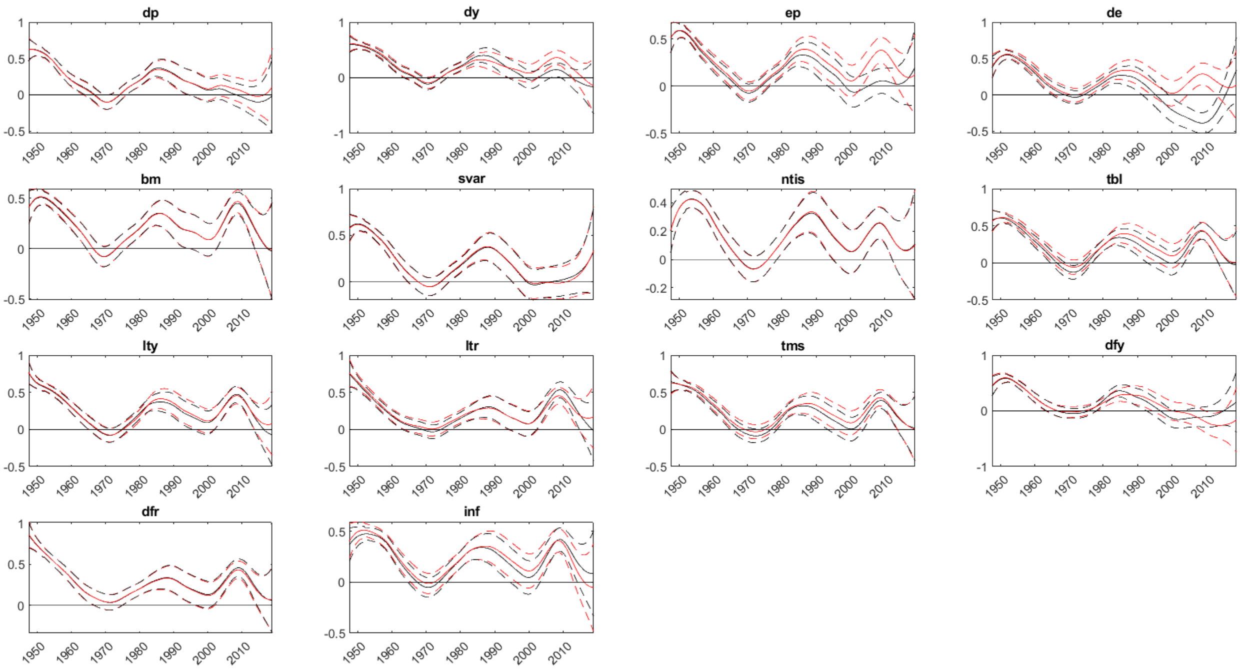
where y_t is defined as the growth rate of log GDP (seasonally adjusted), $\Delta\log(\text{GDP})$, at time t. Data are quarterly from 1947:1 to 2018:IV. The dataset can be downloaded for free from the Federal Reserve Bank of St. Louis (`http://research.stlouisfed.org/econ/mccracken/sel/`).

Figure (5.1) shows the time-varying local linear estimates $\widehat{\alpha}(\tau_t)$ in the time-varying ARDL model $y_t = \alpha(\tau_t) + \phi(\tau_t)y_{t-1} + \beta(\tau_t)x_{t-1} + \lambda\widehat{\varepsilon}_t + \eta_t$ using the 14 individual predictors. The red line represents the estimated time-varying coefficients $y_t = \alpha(\tau_t) + \phi(\tau_t)y_{t-1} + \beta(\tau_t)x_{t-1} + \lambda\widehat{\varepsilon}_t + \eta_t$, and the black line represents the estimated time-varying coefficients with the linear projection function $y_t = \alpha(\tau_t) + \phi(\tau_t)y_{t-1} + \beta(\tau_t)x_{t-1} + \eta_t$. The results are similar to the time-varying estimates of intercepts in return forecasting. $\widehat{\alpha}_t$ is significant for the entire sample. The time-varying estimates change over time and have the biggest values during the oil shock of 1973–1974. It is a convincing fact that GDP growth is a leading macroeconomic variable that reflects the economic fluctuations. The 95% confidence bands have nearly the same width for the pure time-varying model and the bias-reduced time-varying model with the linear projection function, as λ is not significant when we forecast GDP growth. Once again, if the correlation between the errors from the predictive model and the AR model is non-zero, then the linear projection method would be able to remove the "Stambaugh bias" and improve the efficiency of the time-varying local linear estimators.

Figure (5.2) shows the local linear estimates $\widehat{\phi}(\tau_t)$ in the time-varying predictive regression model (with the 95% point-wise confidence bands) for the 14 individual predictors. The red line represents the estimated time-varying coefficients $y_t = \alpha(\tau_t) + \phi(\tau_t)y_{t-1} + \beta(\tau_t)x_{t-1} + \lambda\widehat{\varepsilon}_t + \eta_t$, and the black line represents the estimated time-varying coefficients with the linear projection function $y_t = \alpha(\tau_t) + \phi(\tau_t)y_{t-1} + \beta(\tau_t)x_{t-1} + \eta_t$. Economically, the estimated local linear estimates $\widehat{\phi}(\tau_t)$ represent the time-varying autocorrelations of GDP growth rate. We confirm that there is a strong autocorrelation in GDP growth rate and the autocorrelation is time-varying.

Figure (5.3) shows the local linear estimates $\widehat{\beta}(\tau_t)$ in the time-varying predictive regression model with the 95% point-wise confidence bands for the 14 individual predictors. The red line represents the estimated time-varying coefficients $y_t = \alpha(\tau_t) + \phi(\tau_t)y_{t-1} + \beta(\tau_t)x_{t-1} + \lambda\widehat{\varepsilon}_t + \eta_t$, and the black line represents the estimated time-varying coefficients with the linear projection function $y_t = \alpha(\tau_t) + \phi(\tau_t)y_{t-1} + \beta(\tau_t)x_{t-1} + \eta_t$. β_t is the key parameter to testing the predictability of GDP growth. Except for inflation rate, 13 of the 14 popularly used predictors contain significant predictive content for future GDP growth. Our results empirically confirm the macro-finance theory that the predictability of stock return should be highly related to business cycle. Thus, the infor-

Figure 5.2: $\widehat{\phi}(\tau_t)$ in the Time-Varying ARDL Model

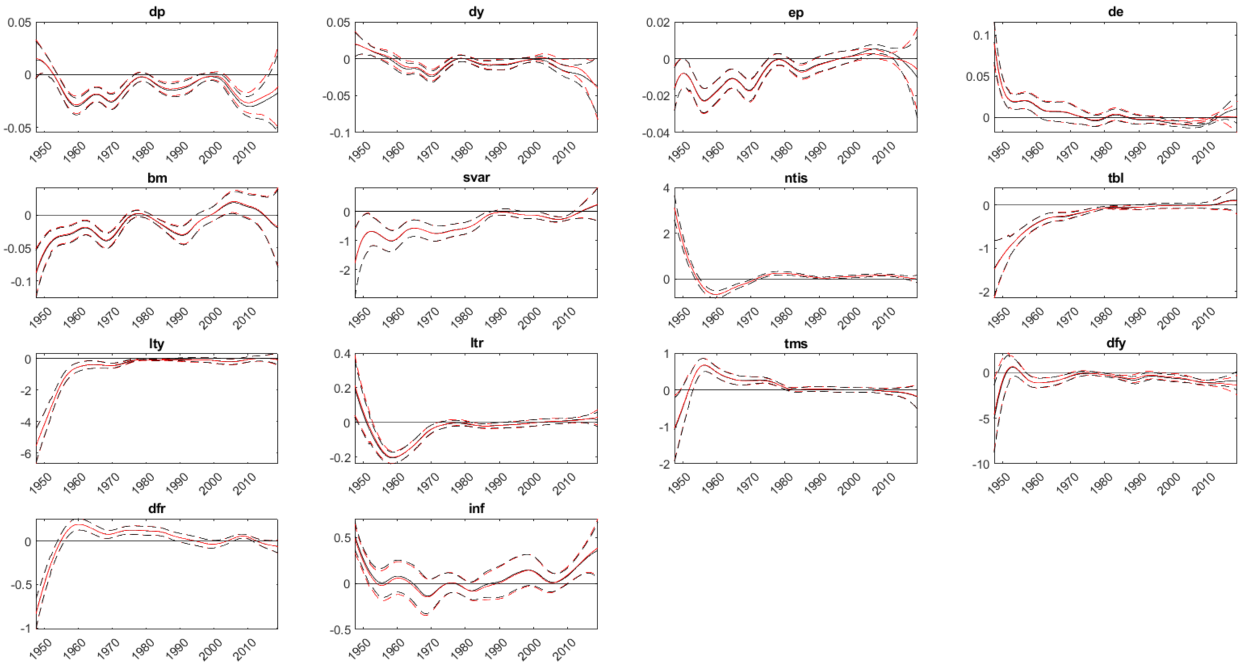


Note. The figure shows the time-varying local linear estimates $\widehat{\phi}(\tau_t)$ in the time-varying ARDL model $y_t = \alpha(\tau_t) + \phi(\tau_t)y_{t-1} + \beta(\tau_t)x_{t-1} + \lambda\widehat{\varepsilon}_t + \eta_t$ using the 14 individual predictors.

mation that is useful for predicting stock returns also should predict future economic activities that are an indicator of business cycle, such as GDP growth.

So far, we have undertaken preliminary empirical analysis for predicting economic activities using the stock return predictors. However, when we consider a lagged dependent variable as another explanatory variable or predictor in the models, the statistical theory of the local linear estimators in the resulting time-varying ARDL models has not been explored in prior studies. Thus, an important direction for future research is to fully establish the theory for the estimators in the time-varying ARDL models.

Figure 5.3: $\widehat{\beta}(\tau_t)$ in the Time-Varying ARDL Model



Note. This figure shows the time-varying local linear estimates $\widehat{\beta}(\tau_t)$ in the time-varying ARDL model $y_t = \alpha(\tau_t) + \phi(\tau_t)y_{t-1} + \beta(\tau_t)x_{t-1} + \lambda\widehat{\varepsilon}_t + \eta_t$ using the 14 individual predictors.

Appendix A

Technical Proof

A.1 Proof of Theorems

A.1.1 List of Lemmas

Before moving to the proof of main theorems, we list necessary lemmas.

Lemma A.1. *Under the assumptions, we have the following lemmas:*

(1) $\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t - \tau)v_t^l \longrightarrow \tilde{\sigma}_l$*, for* $l = 0, 1, 2, 3$*.*

(2) $\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t - \tau)^2 v_t^l \longrightarrow \tilde{\sigma}_l^2$*, for* $l = 0, 1, 2, 3$*.*

(3) $\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t - \tau)v_t^l x_t \xrightarrow{p} A_{T\tau}\tilde{\sigma}_l$*, for* $l = 0, 1, 2, 3$*.*

(4) $\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t - \tau)^2 v_t^l x_t \xrightarrow{p} A_{T\tau}\tilde{\sigma}_l^2$ *for* $l = 0, 1, 2, 3$*.*

(5) $\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t - \tau)v_t^l x_t^2 \xrightarrow{p} A_{T\tau}^2\tilde{\sigma}_l + \sigma^2 B_{0,T\tau}(1)\tilde{\sigma}_l$*, for* $l = 0, 1, 2, 3$*.*

(6) $\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t - \tau)^2 v_t^l x_t^2 \xrightarrow{p} A_{T\tau}^2\tilde{\sigma}_l^2 + \sigma^2 B_{0,T\tau}(1)\tilde{\sigma}_l^2$*, for* $l = 0, 1, 2, 3$*.*

We define $\tilde{\sigma}_l = \int_{-\infty}^{+\infty} K(v)v^l dv \cdot I(0 < \tau < 1) + \int_{\infty}^{0} K^2(v)v^l dv \cdot I(\tau = 0) + \int_{0}^{-\infty} K^2(v)v^l dv \cdot I(\tau = 1)$, $\tilde{\sigma}_l^2 = \int_{-\infty}^{+\infty} K^2(v)v^l dv \cdot I(0 < \tau < 1) + \int_{\infty}^{0} K^2(v)v^l dv \cdot I(\tau = 0) + \int_{0}^{-\infty} K^2(v)v^l dv \cdot I(\tau = 1)$, $A_{T\tau} = \mu_{T\tau}$, $B_{0,T\tau}(1) = \sum_{j=0}^{\infty}\varepsilon_{j,T\tau}^2$*, and* $I(\cdot)$ *is the indicator function.*

It will be shown that Lemma A.1 is essential to work out the main asymptotic results in next subsection.

Lemma A.2. *Under the assumptions, we have the following CLT*

$\sqrt{Th}\widetilde{J}_T(\tau) \xrightarrow{d} N\Big(0, \quad \Pi(\tau)\Big)$*, with* $\Pi(\tau) = \begin{pmatrix} \pi_1(\tau) & \pi_2(\tau) \\ \pi_2(\tau)^\top & \pi_2(\tau) \end{pmatrix}$*, where* $\widetilde{J}_T(\tau) =$

$$[\boldsymbol{I}_2, \boldsymbol{0}_2]S_T(\tau)^{-1}\widetilde{J}_T(\tau) \text{ and } \pi_l(\tau) = \begin{bmatrix} \sigma_\varepsilon^2 \tilde{\sigma}_l^2 & A_{T\tau}\tilde{\sigma}_l^2 \\ A_{T\tau}\tilde{\sigma}_l^2 & \sigma_\varepsilon^2 \sigma^2 B_{0,T\tau}(1)\tilde{\sigma}_l^2 \end{bmatrix}.$$

Lemma A.3. *Under the assumptions, we have the following lemmas:*

(1) $\sup_{0<\tau<1} \left| \frac{1}{T}\sum_{t=1}^{T} z_t K_h(\tau_t - \tau) - \mu(\tau) \right| = O_P\left(h^2 + \left(\frac{\log T}{Th}\right)^{\frac{1}{2}} \right)$.

(2) $\sup_{0<\tau<1} \left| \frac{1}{T}\sum_{t=1}^{T} z_t z_t^\top K_h(\tau_t - \tau) - \Sigma_0(\tau) \right| = O_P\left(h^2 + \left(\frac{\log T}{Th}\right)^{\frac{1}{2}} \right)$.

(3) $\sup_{0<\tau<1} |\widehat{\phi}(\tau) - \phi(\tau)| = O_P\left(h^2 + \left(\frac{\log T}{Th}\right)^{\frac{1}{2}} \right)$.

Lemma A.4. *Under the assumptions, we have the following lemmas*

(1) $\frac{1}{T}\left(\tilde{X}^\top \tilde{X}\right)^{-1} \xrightarrow{p} 1/\sigma_\varepsilon^2$.

(2) $\frac{1}{\sqrt{T}}\tilde{X}^\top (\boldsymbol{I} - \tilde{s})\,(Z_1^\top\theta(\tau_2)), ..., Z_1^\top\theta(\tau_T))^\top = o_p(1)$.

(3) $\frac{1}{\sqrt{T}}\tilde{X}^\top (\boldsymbol{I} - \tilde{s})\left(\lambda - \widehat{\lambda}\right)\varepsilon = o_p(1)$.

(4) $\frac{1}{\sqrt{T}}\tilde{X}^\top (\boldsymbol{I} - \tilde{s})\,\widehat{\lambda}\varepsilon = o_p(1)$.

(5) $\sqrt{Th}s(\tau)\widehat{\varepsilon}(\lambda - \widehat{\lambda}) = o_p(1)$.

(6) $\sqrt{Th}s(\tau)(\varepsilon - \widehat{\varepsilon})\lambda = o_p(1)$.

We define $s(\tau) = [I_2, 0_2]\left[D(\tau)'K(\tau)D(\tau)\right]^{-1} D(\tau)'K(\tau)$ *and* $\varepsilon = (\varepsilon_1, ..., \varepsilon_T)^\top$.

Lemma A.5. *Under the assumptions, we have the following CLT*

$$\sqrt{T}\left(\frac{1}{T}\tilde{X}^\top (\boldsymbol{I} - \tilde{s})\,\eta\right)^{-1} \xrightarrow{d} N\Big(0, \quad \sigma_\eta^2\sigma_\varepsilon^{-2}\Big).$$

A.1.2 Proof of Theorem 2.1

The local linear estimator of $\phi(\tau)$ is $\widehat{\phi}(\tau) = [\mathrm{I}_2, 0_2]\left[D(\tau)'K(\tau)D(\tau)\right]^{-1} D(\tau)'K(\tau)Y$, and note that $Y = (z_1'\phi(\tau_1), ..., z_T'\phi(\tau_T))^\top + (\varepsilon_1, ..., \varepsilon_T)^\top$, by considering Taylor expansion about $\phi(\tau_t)$ element by element around $\phi(\tau)$, we have

$$\phi(\tau_t) \approx \phi(\tau) + h\phi(\tau)'\left(\frac{\tau_t - \tau}{h}\right) + \frac{h^2}{2}\phi''(\tau)\left(\frac{\tau_t - \tau}{h}\right)^2 + \left(\frac{\tau_t - \tau}{h}\right)^2 o(h^2). \tag{A.1}$$

We denote $\phi''(\tau) = (\alpha''(\tau), \beta''(\tau))^\top$.

Next, we expand elements $\Phi(\tau_t)$ for $t = 1, 2, .., T$ in Y

$$
\begin{aligned}
Y =& \begin{bmatrix} z_1\{\phi(\tau) + h\phi(\tau)'\left(\frac{\tau_2-\tau}{h}\right) + \frac{h^2}{2}\phi(\tau)''\left(\frac{\tau_2-\tau}{h}\right)^2 + \left(\frac{\tau_t-\tau}{h}\right)^2 o(h^2)\} \\ \vdots \\ z_T\{\phi(\tau) + h\phi(\tau)'\left(\frac{\tau_T-\tau}{h}\right) + \frac{h^2}{2}\phi(\tau)''\left(\frac{\tau_T-\tau}{h}\right)^2 + \left(\frac{\tau_t-\tau}{h}\right)^2 o(h^2)\} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{bmatrix} \\
=& D(\tau)\left(\phi(\tau), h\phi(\tau)'\right)^\top + M(\tau)\left(\frac{h^2}{2}\phi(\tau)'' + o(h^2)\right) + \varepsilon,
\end{aligned} \tag{A.2}
$$

where $M(\tau) = \left(z_1'\left(\frac{\tau_2-\tau}{h}\right)^2, ..., z_t'\left(\frac{\tau_T-\tau}{h}\right)^2\right)^\top$. Therefore, $\widehat{\phi}(\tau)$ can be expressed as

$$
\begin{aligned}
\phi(\tau) =& [\boldsymbol{I}_2, \mathbf{0}_2]\left[D(\tau)'K(\tau)D(\tau)\right]^{-1} D(\tau)'K(\tau) \\
&\times \left(D(\tau)\left(\phi(\tau), h\phi(\tau)'\right)^\top + M(\tau)\left(\frac{h^2}{2}\phi(\tau)'' + o(h^2)\right) + \varepsilon\right).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\widehat{\phi}(\tau) - \phi(\tau) =& [\boldsymbol{I}_2, \mathbf{0}_2]\left[D(\tau)'K(\tau)D(\tau)\right]^{-1} D(\tau)'K(\tau)M(\tau)\left(h^2/2\phi(\tau)'' + o(h^2)\right) \\
&+ [\mathrm{I}_2, 0_2]\left[D(\tau)'K(\tau)D(\tau)\right]^{-1} D(\tau)'K(\tau)\varepsilon \\
=& \widetilde{B}_T(\tau) + \widetilde{T}(\tau) + o(h^2).
\end{aligned}
$$

The decomposition of estimation errors of ϕ is given by

$$
\widehat{\phi}(\tau) - \phi(\tau) - \tilde{B}_T(\tau) - o(h^2) = \widetilde{T}(\tau), \quad \text{for} \quad \tau \in [0, 1], \tag{A.3}
$$

in which $\widetilde{T}(\tau) = [\boldsymbol{I}_2, \mathbf{0}_2]S_T(\tau)^{-1}\widetilde{J}_T(\tau)$, and $\tilde{B}_T(\tau) = \frac{h^2}{2}[\mathrm{I}_2, 0_2]S_T^{-1}\widetilde{L}(\tau)\phi''(\tau)$. Furthermore, we define $\widetilde{L}(\tau) = (S_{T,2}(\tau), S_{T,3}(\tau))^\top$ and $\widetilde{J}_T(\tau) = [\widetilde{J}_{T,0}(\tau), \widetilde{J}_{T,1}(\tau)]^\top$ where $\widetilde{J}_{T,l}(\tau) = \frac{1}{Th}\sum_{t=1}^T K_h(\tau_t - \tau)v_t^l Z_t\varepsilon_t$, for $l = 0, 1$.

For the bias term, we have

$$\begin{aligned}
\widetilde{L}(\tau) &= [S_{T,0}, S_{T,1}]' \left(Z_2' \left(\frac{\tau_2-\tau}{h}\right)^2, ..., Z_t' \left(\frac{\tau_T-\tau}{h}\right)^2 \right)^\top \\
&= \frac{1}{Th} \sum_{t=1}^{T} K_h(\tau_t-\tau) v_t^l \begin{bmatrix} Z_t \\ Z_t \frac{\tau_t-\tau}{h} \end{bmatrix} Z_t' \left(\frac{\tau_t-\tau}{h}\right)^2 \\
&= \frac{1}{Th} \sum_{t=1}^{T} K_h(\tau_t-\tau) v_t^l \begin{bmatrix} Z_t Z_t' v_t^2 \\ Z_t Z_t' v_t^3 \end{bmatrix} \\
&= (S_{T,2}(\tau), S_{T,3}(\tau))^\top .
\end{aligned}$$

According to Lemma A.1, we have $S_{T,l}(\tau) \longrightarrow \Sigma_l(\tau) = \begin{bmatrix} \tilde{\sigma}_l & A_{T\tau}\tilde{\sigma}_l \\ A_{T\tau}\tilde{\sigma}_l & \sigma_\varepsilon^2 B_{0,T\tau}\tilde{\sigma}_l \end{bmatrix}$, for $l = 0, 1, 2, 3$. It follows that

$$S_T^{-1}(\tau) \xrightarrow{p} \Sigma_S^{-1}(\tau), \quad \text{and} \quad \widetilde{L}(\tau) = \begin{bmatrix} S_{T,2}(\tau) & S_{T,3}(\tau) \end{bmatrix} \longrightarrow \Sigma_L(\tau),$$

where $\Sigma_S^{-1}(\tau) = \begin{bmatrix} \Sigma_0(\tau) & \Sigma_1(\tau) \\ \Sigma_1(\tau) & \Sigma_2(\tau) \end{bmatrix}^{-1}$ and $\Sigma_L(\tau) = \begin{bmatrix} \Sigma_2(\tau) & \Sigma_3(\tau) \end{bmatrix}$.

In $\widetilde{T}(\tau) = [\mathrm{I}_2, 0_2] S_T(\tau)^{-1} \widetilde{J}_T(\tau)$, by Lemma A.2, we have the following CLT for $\widetilde{J}_T(\tau)$

$$\sqrt{Th}\widetilde{J}_T(\tau) \xrightarrow{d} N\Big(0, \quad \Pi(\tau)\Big), \quad \text{with} \quad \Pi(\tau) = \begin{pmatrix} \pi_1(\tau) & \pi_2(\tau) \\ \pi_2(\tau)^\top & \pi_2(\tau) \end{pmatrix}, \tag{A.4}$$

where $\pi_l(\tau) = \begin{bmatrix} \sigma_\varepsilon^2 \tilde{\sigma}_l^2 & A_{T\tau}\tilde{\sigma}_l^2 \\ A_{T\tau}\tilde{\sigma}_l^2 & \sigma_\varepsilon^2 \sigma^2 B_{0,T\tau}(1)\tilde{\sigma}_l^2 \end{bmatrix}$.

Therefore, it follows that

$$\begin{aligned}
\sqrt{Th}\left(\widehat{\phi}(\tau) - \phi(\tau) - \bar{B}_T - o_p(h^2)\right) &= \sqrt{Th}\widetilde{\boldsymbol{t}}_T = [\mathrm{I}_2, 0_2] S_T^{-1} \sqrt{Th}\widetilde{J}_T \\
&\xrightarrow{d} N\Big(0, \quad [\mathrm{I}_2, 0_2]\Sigma_S^{-1}(\tau)\Pi(\tau)\Sigma_S^{-1}(\tau)^\top [\mathrm{I}_2, 0_2]^\top\Big).
\end{aligned}$$

It completes the proof.

A.1.3 Proof of Theorem 2.2

Deriving the estimation error of λ and multiplying $\sqrt{T}$ on the both sides, we can get

$$\begin{aligned}\sqrt{T}\left(\widehat{\lambda}-\lambda\right) =& \sqrt{T}\left(\frac{1}{T}\tilde{X}^{\top}\tilde{X}\right)^{-1}\frac{1}{T}\tilde{X}^{\top}\left(\mathbf{I}-\tilde{s}\right)\left(Z_1^{\top}\theta(\tau_2)),...,Z_1^{\top}\theta(\tau_T)\right)^{\top}\\ &+\sqrt{T}\left(\frac{1}{T}\tilde{X}^{\top}\tilde{X}\right)^{-1}\frac{1}{T}\tilde{X}^{\top}\left(\mathbf{I}-\tilde{s}\right)\left(\lambda-\widehat{\lambda}\right)\varepsilon\\ &+\sqrt{T}\left(\frac{1}{T}\tilde{X}^{\top}\tilde{X}\right)^{-1}\frac{1}{T}\tilde{X}^{\top}\left(\mathbf{I}-\tilde{s}\right)\widehat{\lambda}\varepsilon\\ &+\sqrt{T}\left(\frac{1}{T}\tilde{X}^{\top}\tilde{X}\right)^{-1}\frac{1}{T}\tilde{X}^{\top}\left(\mathbf{I}-\tilde{s}\right)\eta\\ =&\mathbb{N}_{\lambda}(1)+\mathbb{N}_{\lambda}(2)+\mathbb{N}_{\lambda}(3)+\mathbb{N}_{\lambda}(4).\end{aligned}$$

As we can see, $\mathbb{N}_{\lambda}(2)$ is the leading term that can produce the main central limit theorem for $\hat{\lambda}$. To prove Theorem 2.2, it suffices to show the following results,

$$\mathbb{N}_{\lambda}(1)=o_p(1),\quad \mathbb{N}_{\lambda}(2)=o_p(1),\quad \mathbb{N}_{\lambda}(2)=o_p(1),\quad \mathbb{N}_{\lambda}(4)\xrightarrow{d}N\left(0,\ \frac{\sigma_{\eta}^2}{\sigma_{\varepsilon}^2}\right).$$

Recall Lemma A.4, we have the following results:

(1) $\frac{1}{T}\left(\tilde{X}^{\top}\tilde{X}\right)^{-1}\xrightarrow{p}1/\sigma_{\varepsilon}^2$.

(2) $\frac{1}{\sqrt{T}}\tilde{X}^{\top}\left(\mathbf{I}-\tilde{s}\right)\left(Z_1^{\top}\theta(\tau_2)),...,Z_1^{\top}\theta(\tau_T)\right)^{\top}=o_p(1)$.

(3) $\frac{1}{\sqrt{T}}\tilde{X}^{\top}\left(\mathbf{I}-\tilde{s}\right)\left(\lambda-\widehat{\lambda}\right)\varepsilon=o_p(1)$.

(4) $\frac{1}{\sqrt{T}}\tilde{X}^{\top}\left(\mathbf{I}-\tilde{s}\right)\widehat{\lambda}\varepsilon=o_p(1)$.

For $\mathbb{N}_{\lambda}(1)$, we have

$$\begin{aligned}\mathbb{N}_{\lambda}(1) &= \sqrt{T}\left(\frac{1}{T}\tilde{X}^{\top}\tilde{X}\right)^{-1}\frac{1}{T}\tilde{X}^{\top}\left(\mathbf{I}-\tilde{s}\right)\left(Z_1^{\top}\theta(\tau_2)),...,Z_1^{\top}\theta(\tau_T)\right)^{\top}\\ &= \left(\frac{1}{T}\tilde{X}^{\top}\tilde{X}\right)^{-1}\frac{1}{\sqrt{T}}\tilde{X}^{\top}\left(\mathbf{I}-\tilde{s}\right)\left(Z_1^{\top}\theta(\tau_2)),...,Z_1^{\top}\theta(\tau_T)\right)^{\top}\\ &= O(1)\cdot o_p(1)\xrightarrow{p}0.\end{aligned}$$

For $\mathbb{N}_\lambda(2)$, we have

$$\begin{aligned}\mathbb{N}_\lambda(2) &= \sqrt{T}\left(\frac{1}{T}\tilde{X}^\top\tilde{X}\right)^{-1}\frac{1}{T}\tilde{X}^\top(\mathbf{I}-\tilde{s})\left(\lambda-\widehat{\lambda}\right)\varepsilon \\ &= \left(\frac{1}{T}\tilde{X}^\top\tilde{X}\right)^{-1}\frac{1}{\sqrt{T}}\tilde{X}^\top(\mathbf{I}-\tilde{s})\left(\lambda-\widehat{\lambda}\right)\varepsilon \\ &= O(1)\cdot o_p(1)\xrightarrow{p}0.\end{aligned}$$

For $\mathbb{N}_\lambda(3)$, we have

$$\begin{aligned}\mathbb{N}_\lambda(3) &= \sqrt{T}\left(\frac{1}{T}\tilde{X}^\top\tilde{X}\right)^{-1}\frac{1}{T}\tilde{X}^\top(\mathbf{I}-\tilde{s})\widehat{\lambda}\varepsilon \\ &= \left(\frac{1}{T}\tilde{X}^\top\tilde{X}\right)^{-1}\frac{1}{\sqrt{T}}\tilde{X}^\top(\mathbf{I}-\tilde{s})\widehat{\lambda}\varepsilon \\ &= O(1)\cdot o_p(1)\xrightarrow{p}0.\end{aligned}$$

By Lemma A.5, we have

$$\sqrt{T}\left(\frac{1}{T}\tilde{X}^\top(\mathbf{I}-\tilde{s})\eta\right)^{-1}\xrightarrow{d}N\Big(0,\quad\sigma_\eta^2\sigma_\varepsilon^2\Big).$$

Combining with the result that $\frac{1}{T}\tilde{X}^\top\tilde{X}\xrightarrow{p}\sigma_\varepsilon^2$, therefore, for $\mathbb{N}_\lambda(4)$, we have

$$\begin{aligned}\mathbb{N}_\lambda(4) &= \sqrt{T}\left(\frac{1}{T}\tilde{X}^\top\tilde{X}\right)^{-1}\frac{1}{T}\tilde{X}^\top(\mathbf{I}-\tilde{s})\eta \\ &= \frac{1}{\sqrt{T}}\left(\frac{1}{T}\tilde{X}^\top\tilde{X}\right)^{-1}\tilde{X}^\top(\mathbf{I}-\tilde{s})\eta \\ &\xrightarrow{d}N\left(0,\quad\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}\right).\end{aligned}$$

We complete the proof of the the following results

$$\mathbb{N}_\lambda(1)=o_p(1),\quad\mathbb{N}_\lambda(2)=o_p(1),\quad\mathbb{N}_\lambda(2)=o_p(1),\quad\mathbb{N}_\lambda(4)\xrightarrow{d}N\left(0,\quad\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}\right).$$

Taken together, we can show $\sqrt{T}\left(\widehat{\lambda}-\lambda\right)\xrightarrow{d}N\left(0,\quad\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}\right)$. This completes the proof.

A.1.4 Proof of Theorem 2.3

In the semiparametric predictive regression, the feasible local linear estimator $\widehat{\theta}(\tau)$ is

$$\begin{aligned}
\widehat{\theta}(\tau) &= s(\tau)\left(Y - \widehat{\lambda}\widehat{\varepsilon}\right) \\
&= s(\tau)(Y - \lambda\varepsilon + \lambda\varepsilon - \widehat{\varepsilon}\lambda + \widehat{\varepsilon}\lambda - \widehat{\lambda}\widehat{\varepsilon}) \\
&= s(\tau)(Y - \lambda\varepsilon) + s(\tau)\widehat{\varepsilon}(\lambda - \widehat{\lambda}) + s(\tau)(\varepsilon - \widehat{\varepsilon})\lambda \\
&= \mathbb{N}_\theta(1) + \mathbb{N}_\theta(2) + \mathbb{N}_\theta(3).
\end{aligned}$$

We consider $\mathbb{N}_\theta(1)$,

$$\mathbb{N}_\theta(1) = s(\tau)(Y - \lambda\varepsilon) = s(\tau)(\alpha(\tau_t) + \beta(\tau_t) + \eta_t),$$

where $s(\tau)\varepsilon = [\mathrm{I}_2, 0_2]S_T(\tau)^{-1}\widetilde{\zeta}_T(\tau)$ with $\widetilde{\zeta}_T(\tau) = [\widetilde{\zeta}_{T,o}(\tau), \quad \widetilde{\zeta}_{T,1}(\tau)]^\top$ and

$$\widetilde{\zeta}_{T,l}(\tau) = \frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t - \tau)v_t^l Z_t\eta_t \quad \text{for} \quad l = 0, 1.$$

Recall the term that produces the CLT, $\widetilde{\boldsymbol{t}}_T = [\mathrm{I}_2, 0_2]S_T^{-1}\widetilde{\boldsymbol{\zeta}}_T$ where $[\mathrm{I}_2, 0_2]$ is non-stochastic and $\boldsymbol{S}_{\boldsymbol{T}}^{-1}(\tau) \xrightarrow{p} \Sigma_S^{-1}(\tau)$.

$$\sqrt{Th}\widetilde{\zeta}_T(\tau) = \begin{bmatrix} \sqrt{Th}\widetilde{\zeta}_1 \\ \sqrt{Th}\widetilde{\zeta}_2 \end{bmatrix} = \begin{bmatrix} \sqrt{Th}\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t - \tau)Z_t\eta_t \\ \sqrt{Th}\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t - \tau)v_t^l Z_t\eta_t \end{bmatrix},$$

and the variance-covariance matrix is

$$var(\sqrt{Th}\widetilde{\zeta}_T(\tau)) = \begin{bmatrix} var(\sqrt{Th}\widetilde{\zeta}_1(\tau)) & cov(\sqrt{Th}\widetilde{\zeta}_1(\tau), \widetilde{\zeta}_2(\tau)) \\ cov(\sqrt{Th}\widetilde{\zeta}_2(\tau), \widetilde{\zeta}_1(\tau)) & var(\sqrt{Th}\widetilde{\zeta}_2(\tau)) \end{bmatrix}.$$

We can show the elements in the covariance matrix one-by-one

$$
\begin{aligned}
var(\sqrt{Th}\widetilde{\zeta}_1) &= var\left(\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)Z_t\eta_t\right) \\
&= var\begin{bmatrix} \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)\eta_t \\ \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)x_t\eta_t \end{bmatrix} = \begin{bmatrix} var(\widetilde{\zeta}_{1,1}) & cov(\widetilde{\zeta}_{1,1},\widetilde{\zeta}_{1,2}) \\ cov(\widetilde{\zeta}_{1,2},\widetilde{\zeta}_{1,1}) & var(\widetilde{\zeta}_{1,2}) \end{bmatrix}
\end{aligned}
$$

where we define $\widetilde{\zeta}_{1,1} = \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)\eta_t$ and $\widetilde{\zeta}_{1,2} = \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)x_t\eta_t$. Now, we deal with the elements in the matrix $var(\widetilde{\zeta}_1)$, $var(\widetilde{\zeta}_{1,1}) \xrightarrow{p} \sigma_\eta^2\sigma_{l=0}^2 = \widetilde{\pi}_{1,1}$

$$
\begin{aligned}
var(\widetilde{\zeta}_{1,1}) &= E\left[\widetilde{\zeta}_{1,1} - E\left(\widetilde{\zeta}_{1,1}\right)\right]^2 \\
&= E\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)\eta_t - E\left(\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)\eta_t\right)\right]^2 \\
&= \frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t-\tau)^2 E[\eta_t^2] + 0 \\
&= \sigma_\eta^2 \int_{\frac{-\tau}{h}}^{\frac{1-\tau}{h}} K^2(v)\,\sigma_\eta^2 dv + o(1) \\
&\longrightarrow \sigma_\eta^2\sigma_{l=0}^2 = \widetilde{\pi}_{1,1}
\end{aligned}
$$

where $\tilde{\sigma}_l^2 = \int_{-\infty}^{+\infty} K^2(v)v^l dv \cdot I(0<\tau<1) + \int_{\infty}^{0} K^2(v)v^l dv \cdot I(\tau=0) + \int_{0}^{-\infty} K^2(v)v^l dv \cdot I(\tau=1)$.

$cov(\widetilde{\zeta}_{1,2},\widetilde{\zeta}_{1,1}) \xrightarrow{p} 0 = \widetilde{\pi}_{1,2}$

$$
\begin{aligned}
cov(\widetilde{\zeta}_{1,2},\widetilde{\zeta}_{1,1}) &= cov(\widetilde{\zeta}_{1,1},\widetilde{\zeta}_{1,2}) \\
&= E\left[\widetilde{\zeta}_{1,1} - E(\widetilde{\zeta}_{1,1})\right]\left[\widetilde{\zeta}_{1,2} - E(\widetilde{\zeta}_{1,2})\right] \\
&= E\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)\eta_t - E\left(\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)\eta_t\right)\right] \\
&\quad \left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)x_t\eta_t - E\left(\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)x_t\eta_t\right)\right] \\
&= E\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)\eta_t \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)x_t\eta_t\right] \\
&= \sigma_\eta^2 \frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t-\tau)^2 x_t \longrightarrow 0.
\end{aligned}
$$

$$
\begin{aligned}
var(\widetilde{\zeta}_{1,2}) &= E\left[\widetilde{\zeta}_{1,2} - E\left(\widetilde{\zeta}_{1,2}\right)\right]^2 = E\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)x_t\eta_t\right]^2 \\
&= \sigma_\eta^2 \frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t-\tau)^2 x_t^2 \\
&\longrightarrow \sigma_\eta^2\sigma^2 B_{0,T\tau}(1)\sigma_{K,l=0}^2.
\end{aligned}
$$

Hence, we have

$$
var(\widetilde{\zeta}_{1,1}) \longrightarrow \Pi_1 = \begin{bmatrix} \sigma_\eta^2\sigma_{l=0}^2 & 0 \\ 0 & \sigma_\eta^2\sigma^2 B_{0,T\tau}(1)\sigma_{K,l=0}^2 \end{bmatrix}. \tag{A.5}
$$

Second, we deal with the elements in the matrix $cov(\widetilde{\zeta}_1, \widetilde{\zeta}_2)$.

$$
\begin{aligned}
cov(\widetilde{\zeta}_1, \widetilde{\zeta}_2) &= E\left[\widetilde{\zeta}_1 - E(\widetilde{\zeta}_1)\right]\left[\widetilde{\zeta}_2 - E(\widetilde{\zeta}_2)\right]^\top \\
&= E\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)Z_t\eta_t\right]\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t Z_t\tilde{\eta}_t\right]^\top \\
&= E\begin{bmatrix} \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)\tilde{\eta}_t \\ \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)x_t\tilde{\eta}_t \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t\tilde{\eta}_t \\ \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t x_t\tilde{\eta}_t \end{bmatrix}^\top,
\end{aligned}
$$

where

$$
\begin{aligned}
&\sigma_\eta^2\frac{1}{Th}\sum_{t=1}^{T} K^2\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right) \longrightarrow \widetilde{\pi}_{3,1} = \sigma_\eta^2\sigma_{K,l=1}^2, \\
&\sigma_\eta^2\frac{1}{Th}\sum_{t=1}^{T} K^2\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right)x_t \xrightarrow{p} \widetilde{\pi}_{2,2} = 0, \\
&\sigma_\eta^2\frac{1}{Th}\sum_{t=1}^{T} K^2\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right)x_t \xrightarrow{p} \widetilde{\pi}_{2,2} = 0, \\
&\sigma_\eta^2\frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\tau_t-\tau}{h}\right)x_t^2 \xrightarrow{p} \sigma_\eta^2\sigma^2 B_{0,T\tau}(1)\sigma_{l=1}.
\end{aligned}
$$

Hence, we have

$$cov(\widetilde{\zeta}_1, \widetilde{\zeta}_2) \xrightarrow{p} \widetilde{\Pi}_2 = \begin{bmatrix} \sigma_\eta^2 \sigma_{K,l=1}^2 & 0 \\ 0 & \sigma_\eta^2 \sigma^2 B_{0,T\tau}(1) \sigma_{l=1} \end{bmatrix}. \tag{A.6}$$

Third, we deal with the elements in the matrix $var(\widetilde{\zeta}_2)$,

$$var(\widetilde{\zeta}_2) = \begin{bmatrix} var(\widetilde{\zeta}_{2,1}) & cov(\widetilde{\zeta}_{2,1}, \widetilde{\zeta}_{2,2}) \\ cov(\widetilde{\zeta}_{2,2}, \widetilde{\zeta}_{2,1}) & var(\widetilde{\zeta}_{2,2}) \end{bmatrix} = var \begin{bmatrix} \sqrt{Th} \frac{1}{Th} \sum_{t=1}^T K\left(\frac{\tau_t - \tau}{h}\right) \left(\frac{\tau_t - \tau}{h}\right) \eta_t \\ \sqrt{Th} \frac{1}{Th} \sum_{t=1}^T K\left(\frac{\tau_t - \tau}{h}\right) \left(\frac{\tau_t - \tau}{h}\right) x_t \eta_t \end{bmatrix}$$

where

$$\begin{aligned} var(\widetilde{\zeta}_{2,1}) &= E\left[\frac{1}{\sqrt{Th}} \sum_{t=1}^T K_h(\tau_t - \tau) v_t \eta_t\right]^2 \\ &= \sigma_\eta^2 \int_{\frac{-\tau}{h}}^{\frac{1-\tau}{h}} vK^2(v)\, \sigma_\eta^2 dv + o(1) \\ &\longrightarrow \sigma_\eta^2 \sigma_{l=1}^2, \end{aligned}$$

and

$$\begin{aligned} cov(\widetilde{\zeta}_{2,2}, \widetilde{\zeta}_{2,1}) &= E\left[\widetilde{\zeta}_{2,1} - E(\widetilde{\zeta}_{2,1})\right]\left[\widetilde{\zeta}_{2,2} - E(\widetilde{\zeta}_{2,2})\right] \\ &= E\left[\frac{1}{\sqrt{Th}} \sum_{t=1}^T K_h(\tau_t - \tau) v_t x_t \eta_t - E\left(\frac{1}{\sqrt{Th}} \sum_{t=1}^T K_h(\tau_t - \tau) v_t x_t \eta_t\right)\right] \\ &\times \left[\frac{1}{\sqrt{Th}} \sum_{t=1}^T K_h(\tau_t - \tau) v_t \eta\eta - E\left(\frac{1}{\sqrt{Th}} \sum_{t=1}^T K_h(\tau_t - \tau) v_t \eta\eta\right)\right] \\ &= E\left[\frac{1}{\sqrt{Th}} \sum_{t=1}^T K_h(\tau_t - \tau) v_t x_t \eta \tilde{a}_t \frac{1}{\sqrt{Th}} \sum_{t=1}^T K_h(\tau_t - \tau) v_t \eta \tilde{a}_t\right] \\ &= \sigma_\eta^2 \frac{1}{Th} \sum_{t=1}^T K_h(\tau_t - \tau)^2 E[v_t^2 x_t] + 0 \\ &\xrightarrow{p} 0, \end{aligned}$$

and

$$\begin{aligned} var(\widetilde{\zeta}_{2,2}) &= E\left[\widetilde{\zeta}_{2,2} - E\left(\widetilde{\zeta}_{2,2}\right)\right]^2 \\ &= E\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right)x_t\eta\right]^2 \\ &= \sigma_\eta^2\frac{1}{Th}\sum_{t=1}^{T} K^2\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right)^2 E[x_t^2] \\ &\longrightarrow \sigma_\eta^2\sigma^2 B_{0,T\tau}(1)\sigma_{l=2}^2. \end{aligned}$$

Hence, we have

$$var(\widetilde{\zeta}_2) \longrightarrow \widetilde{\Pi}_3 = \begin{bmatrix} \sigma_\eta^2\sigma_{l=1}^2 & 0 \\ 0 & \sigma_\eta^2\sigma^2 B_{0,T\tau}(1)\sigma_{l=2}^2 \end{bmatrix}. \tag{A.7}$$

In summary, we have that $var(\sqrt{Th}\widetilde{\zeta}_T) \xrightarrow{p} \Pi(\tau)$ and η_t is i.i.d., and it is straightforward to show the following CLT

$$\sqrt{Th}\widetilde{\zeta}_T \xrightarrow{d} N\Big(0, \quad \widetilde{\Pi}(\tau)\Big). \tag{A.8}$$

Applying Slustky's theorem, we have

$$\sqrt{Th}s(\tau)\eta = [\mathrm{I}_2, 0_2]S_T^{-1}\sqrt{Th}\widetilde{\zeta}_T \xrightarrow{d} N\Big(0, \quad [\mathrm{I}_2, 0_2]\Sigma_S^{-1}(\tau)\widetilde{\Pi}(\tau)\Sigma_S^{-1}(\tau)^\top[\mathrm{I}_2, 0_2]^\top\Big).$$

By Lemma A.4, we prove $s(\tau)\widehat{\varepsilon}(\lambda - \widehat{\lambda})$.

$$\begin{aligned} S(\tau)\widehat{\varepsilon}(\lambda-\widehat{\lambda}) &= \left(D(\tau)^\top W(\tau)D(\tau)\right)^{-1} D(\tau)^\top W(\tau)\widehat{\varepsilon}(\lambda-\widehat{\lambda}) \\ &= \left(D(\tau)^\top W(\tau)D(\tau)\right)^{-1} D(\tau)^\top W(\tau)\varepsilon(\lambda-\widehat{\lambda}) \\ &\quad + \left(D(\tau)^\top W(\tau)D(\tau)\right)^{-1} X(\tau)^\top W(\tau)(\widehat{\varepsilon}-\varepsilon)(\theta-\widehat{\lambda}) \\ &= O_P\left(\left(\frac{1}{Th_2}\right)^{\frac{1}{2}}\right) O_P\left(\frac{1}{\sqrt{T}}\right) + O_P\left(c_T\right) O_P\left(\frac{1}{\sqrt{T}}\right) \\ &= o_P\left(\frac{1}{\sqrt{Th_2}}\right). \end{aligned}$$

Third, we consider $s(\tau)(\varepsilon - \widehat{\varepsilon})\lambda$.

$$s(\tau)(\varepsilon - \widehat{\varepsilon})\lambda = \left(D(\tau)^\top W(\tau) D(\tau)\right)^{-1} D(\tau)^\top W(\tau) \left(\widehat{\phi}(\tau_t) - \phi(\tau_t)\right) Z_{t-1}^\top.\lambda$$

According to Lemma A.4, we have results: (1) $\sqrt{Th}\mathbb{N}_\theta(2) = \sqrt{Th}s(\tau)\widehat{\varepsilon}(\lambda - \widehat{\lambda}) = o_p(1)$, and (2) $\sqrt{Th}\mathbb{N}_\theta(3) = \sqrt{Th}s(\tau)(\varepsilon - \widehat{\varepsilon})\lambda = o_p(1)$.

Therefore, the asymptotic distribution of θ at a point τ is

$$\sqrt{Th}\left(\widehat{\theta}(\tau) - \theta(\tau) - \tilde{B}_T(\tau) - o_p(h^2)\right) \xrightarrow{d} N\Big(0, \quad \boldsymbol{V}_\eta(\tau_t)\Big). \tag{A.9}$$

This completes the proof.

A.2 Proof of Lemmas

Proof of Lemma 2.1. The proof of tv-BN decomposition is similar with Beveridge and Nelson (1981).

$$\begin{aligned} C_t(L) = \sum_{j=0}^{\infty} \varphi_{j,t} L^j &= \varphi_{j,t} L^1 + \varepsilon_{1,t} L^1 + \varepsilon_{1,t} L^2 + \ldots\ldots \\ &= \left(\sum_{j=0}^{\infty} \varphi_{j,t} - \sum_{j=1}^{\infty} \varphi_{j,t}\right) + \left(\sum_{j=1}^{\infty} \varphi_{j,t} - \sum_{j=2}^{\infty} \varphi_{j,t}\right) L + \ldots\ldots \\ &= \left(\sum_{j=0}^{\infty} \varphi_{j,t} - \sum_{j=1}^{\infty}\right)\left(\varphi_{j,t} + \sum_{j=1}^{\infty} \varphi_{j,t} L - \sum_{j=2}^{\infty} \varphi_{j,t} L\right) + \ldots\ldots \end{aligned}$$

Rearrange the above one, we have

$$\begin{aligned}
C_t(L) &= \sum_{j=0}^{\infty}\varphi_{j,t} - \left(\sum_{j=1}^{\infty}\varphi_{j,t} - \sum_{j=1}^{\infty}\varphi_{j,t}L\right) - \left(\sum_{j=2}^{\infty}\varphi_{j,t}L - \sum_{j=2}^{\infty}\varphi_{j,t}L^2\right) - \ldots\ldots \\
&= \sum_{j=0}^{\infty}\varphi_{j,t} - (1-L)\left(\sum_{j=1}^{\infty}\varphi_{j,t}\right) - (1-L)\left(\sum_{j=2}^{\infty}\varphi_{j,t}L\right)L + \ldots\ldots \\
&= \sum_{j=0}^{\infty}\varphi_{j,t} - (1-L)\sum_{j=0}^{\infty}\left(\sum_{h=j+1}^{\infty}\varepsilon_h(\tau_t)\right)L^j \\
&= C_t(1) - (1-L)\sum_{j=0}^{\infty}\left(\widetilde{\varphi}_{j,t}\right)L^j \\
&= C_t(1) - (1-L)\widetilde{C}_t(L).
\end{aligned}$$

Next, we have

$$\begin{aligned}
\sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t}^2 = \sum_{j=0}^{\infty}\left[\left(\sum_{h=j+1}^{\infty}\varepsilon_{h,t}\right)^2\right] &\leq \sum_{j=0}^{\infty}\left[\left(\sum_{h=j+1}^{\infty}|\varepsilon_{h,t}|\right)^2\right] \\
&= \sum_{j=0}^{\infty}\left[\left(\sum_{h=j+1}^{\infty}|\varepsilon_{h,t}|^{1/2}h^{1/4}|\varepsilon_{h,t}|^{1/2}h^{-1/4}\right)^2\right] \\
&= \sum_{j=0}^{\infty}\left[\left(\sum_{h=j+1}^{\infty}|\varepsilon_{h,t}|h^{1/2}\right)\left(\sum_{h=j+1}^{\infty}|\varepsilon_{h,t}|^{1/2}h^{-1/2}\right)\right].
\end{aligned}$$

For $\sum_{j=0}^{\infty}\sum_{h=j+1}^{\infty}|\varepsilon_{h,t}|^{1/2}h^{-1/2}$, notice that the term $|\varepsilon_{1,t}|$ appears in the sum only once, when $j=0$. The term $|\varepsilon_{2,t}|$ appears in the sum twice, when $j=0,1$. Hence, $|\varepsilon_{h,t}|$ appears when $j=0,1,...,h-1$, total h times. Therefore,

$$\sum_{j=0}^{\infty}\sum_{h=j+1}^{\infty}|\varepsilon_{h,t}|^{1/2}h^{-1/2} = \sum_{h=1}^{\infty}|\varepsilon_{h,t}|j^{-1/2}j = \sum_{h=1}^{\infty}|\varepsilon_{h,t}|j^{1/2},$$

and by 2.15 we have $\sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t}^2 \leq \sum_{h=j+1}^{\infty}|\varepsilon_{h,t}|j^{1/2} < \infty$. This completes the proof.

$$\sum_{j=0}^{\infty}|\widetilde{\varphi}_{j,t}| = \sum_{j=0}^{\infty}\left|\sum_{h=j+1}^{\infty}\varepsilon_{h,t}\right| \leq \sum_{j=0}^{\infty}\sum_{h=j+1}^{\infty}|\varepsilon_{h,t}| = \sum_{j=0}^{\infty}|\varepsilon_{h,t}|j. \tag{A.10}$$

Thus, we have

$$\sum_{j=0}^{\infty} \widetilde{\varphi}_{j,t}^2 \leq \left(\sum_{h=j+1}^{\infty} |\varepsilon_{h,t}| h^{1/2} \right) \sum_{j=0}^{\infty} \sum_{h=j+1}^{\infty} |\varepsilon_{h,t}|^{1/2} h^{-1/2} < \infty.$$

According to Assumption 2.2, $\sum_{j=0}^{\infty} |\widetilde{\varphi}_{j,t}| < \infty$ holds.

$$\begin{aligned}
\sum_{j=0}^{\infty} \varphi_{j,t} \varepsilon_{t-j} &= C_t(L)\varepsilon_t \\
&= \left[C_t(1) - (1-L)\widetilde{C}_t(L) \right] \varepsilon_t \\
&= C_t(1)\varepsilon_t - (\widetilde{\varepsilon}_t - \widetilde{\varepsilon}_{t-1}),
\end{aligned}$$

where $\widetilde{\varepsilon}_t$ is defined as $\widetilde{\varepsilon}_t = \widetilde{C}_t(L)\varepsilon_t = \sum_{j=0}^{\infty} \left(\widetilde{\varphi}_{j,t} L^j \right) \varepsilon_t$. This completes the proof. ■

Proof of Lemma 2.2. Since ε_t is i.i.d. and $\sum_{j=0}^{\infty} |\varphi_{j,t}| < \infty$ then

$$\begin{aligned}
E\left[\frac{1}{T} \sum_{t=1}^{T} C_t(1)\varepsilon_t \right] &= E\left[\frac{1}{T} \sum_{t=1}^{T} \sum_{j=0}^{\infty} \varphi_{j,t} \varepsilon_t \right] \\
&= \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{j=0}^{\infty} \varphi_{j,t} \right) E[\varepsilon_t] \\
&= 0,
\end{aligned}$$

and due to $\sum_{t=1}^{\top} C_t(1)^2 < \infty$, then

$$\begin{aligned}
E\left[\frac{1}{T} \sum_{t=1}^{T} C_t(1)\varepsilon_t \right]^2 &= E\left[\frac{1}{T^2} \sum_{t=1}^{\top} \sum_{s=1}^{\top} C_t(1) C_s(1) \varepsilon_t \varepsilon_s \right] \\
&= \frac{1}{T^2} \sum_{t=1}^{\top} C_t^2 E\left[\varepsilon_t^2 \right] \longrightarrow 0, \quad \text{as} \quad T \to \infty.
\end{aligned}$$

Therefore, we have $\frac{1}{T} \sum_{t=1}^{T} C_t(1)\varepsilon_t \xrightarrow{p} 0.$ ■

Proof of Lemma 2.2.

$$\left(\sum_{j=1}^{\infty} \varphi_{j,t} \varepsilon_{t-j} \right)^2 = \sum_{j=1}^{\infty} \varphi_{j,t}^2 \varepsilon_{t-j}^2 + 2 \sum_{i=1}^{\infty} \sum_{j=1, i \neq j}^{\infty} \varphi_{i,t} \varphi_{j,t} \varepsilon_{t-i} \varepsilon_{t-j},$$

(due to the change of variable $i = h + h$, so that $h = 1, 2...$)

$$\left(\sum_{j=1}^{\infty} \varphi_{j,t}\varepsilon_{t-j}\right)^2 = \sum_{j=1}^{\infty} \varphi_{j,t}^2\varepsilon_{t-j}^2 + 2\sum_{h=1}^{\infty}\sum_{j=0}^{\infty} \varphi_{j,t}\varphi_{j+h,t}\varepsilon_{t-j}\varepsilon_{t-j-h}$$
$$= B_{0,t}(L)\varepsilon_t^2 + 2\sum_{h=1}^{\infty} B_{h,t}(L)\varepsilon_t\varepsilon_{t-h},$$

where $B_{h,t}(L) = \sum_{j=0}^{\infty} b_{h,j,t}L^j = \sum_{j=0}^{\infty} \varphi_{j,t}\varphi_{j+h,t}L^j$, for $h = 1, 2...$ Similarly, the second-order tv-BN decomposition of $B_{h,t}(L)$ is $B_{h,t}(L) = B_{h,t}(1) - (1-L)\widetilde{B}_{h,t}(L)$, where $\widetilde{B}_{h,t}(L) = \sum_{j=0}^{\infty} \widetilde{b}_{h,t}L^j$, and $\widetilde{b}_{h,t} = \sum_{i=j+1}^{\infty} b_{h,i,t} = \sum_{i+j+1}^{\infty} \varphi_{i,t}\varphi_{i+h,t}$. Therefore, we have

$$(\sum_{j=1}^{\infty} \varphi_{j,t}\varepsilon_{t-j})^2 = B_{0,t}(L)\varepsilon_t^2 + 2\sum_{h=1}^{\infty} B_{h,t}(L)\varepsilon_t\varepsilon_{t-h}$$
$$= B_{0,t}(L)\varepsilon_t^2 + 2\sum_{h=1}^{\infty}\left(B_{h,t}(1) - (1-L)\widetilde{B}_{h,t}(L)\right)\varepsilon_t\varepsilon_{t-h}$$
$$= B_{0,t}(L)\varepsilon_t^2 + 2\sum_{h=1}^{\infty} B_{h,t}(1)\varepsilon_t\varepsilon_{t-h} - 2\sum_{h=1}^{\infty} \widetilde{B}_{h,t-1}(L)\varepsilon_{t-1}\varepsilon_{t-1-h}.$$

This completes the proof of Lemma 2.2 . ■

Proof of Lemma A.1. The first moment of $\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t - \tau)v_t^l$ is

$$E\left[\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t^l\right] = \frac{1}{Th}\int_0^1 K\left(\frac{u-\tau}{h}\right)\left(\frac{u-\tau}{h}\right)^l du + o(1)$$
$$= \int_{-\frac{\tau}{h}}^{\frac{1-\tau}{h}} K(v)v^l dv$$
$$= \tilde{\sigma}_l + o(1).$$

This completes the proof of (1) of Lemma A.1. We can similarly prove (2) of Lemma A.1. Consider (3) of Lemma A.1, we have

$$\frac{1}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l x_t = \frac{1}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l\mu_t$$
$$+ \frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t^l C_t(1)\varepsilon_t$$
$$+ \frac{1}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l(\widetilde{\varepsilon}_t - \widetilde{\varepsilon}_{t-1}),$$

we show the following results

$$\frac{1}{Th}\sum_{t=2}^{\top} K_h(\tau_t-\tau)v_t^l\mu_t \xrightarrow{p} A_{Tu}\tilde{\sigma}_l \tag{A.11}$$

$$\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t^l C_t(1)\varepsilon_t \xrightarrow{p} 0, \tag{A.12}$$

$$\frac{1}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l(\widetilde{\varepsilon}_t-\widetilde{\varepsilon}_{t-1}) \xrightarrow{p} 0, \quad \text{for} \quad l=0,1,2,3. \tag{A.13}$$

To show (A.11), denote $\tilde{a}_t=\mu_t$, the first moment

$$\begin{aligned}
E\left[\frac{1}{Th}\sum_{t=2}^{\top} K_h(\tau_t-\tau)v_t^l\tilde{a}_t\right] =&\frac{1}{Th}\int_0^1 K\left(\frac{u-\tau}{h}\right)\left(\frac{u-\tau}{h}\right)^l A_{u\cdot T}du+o(1)\\
=&\int_{-\frac{\tau}{h}}^{\frac{1-\tau}{h}} K(v)v^l A_{T\tau+Thv}dv+o(1)\\
=&A_{T\tau}\sigma_l+o(1)
\end{aligned}$$

and the second moment is

$$\begin{aligned}
&E\left[\frac{1}{Th}\sum_{t=2}^{\top} K_h(\tau_t-\tau)v_t^l\tilde{a}_t-A_{T\tau+Thv}\sigma_l\right]^2\\
&=E\left[\frac{1}{T^2h^2}\sum_{t=2}^{T} K_h(\tau_t-\tau)(v)^2\left(\frac{\tau_t-\tau}{h}\right)^{2l}A_t^2\right]\\
&+E\left[\frac{2}{T^2h^2}\sum_{s=2}^{T}\sum_{t>s}^{T} K_sK_h(\tau_t-\tau)\left(\frac{\tau_t-\tau}{h}\right)^l\left(\frac{\tau_s-\tau}{h}\right)^l A_s\tilde{a}_t\right]\\
&\longrightarrow 0,
\end{aligned}$$

as A_t and A_s are deterministic. To show (A.12), the first moment is

$$E\left[\frac{1}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l C_t(1)\varepsilon_t\right]=\frac{1}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l C_t(1)E\left[\varepsilon_t\right]=0, \tag{A.14}$$

and the second moment is

$$
\begin{aligned}
& E\left[\frac{1}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l C_t(1)\varepsilon_t\right]^2 \\
&= E\left[\frac{1}{T^2h^2}\sum_{t=2}^{T} K_h(\tau_t-\tau)^2\left(\frac{\tau_t-\tau}{h}\right)^{2l} C_t(1)^2\varepsilon_t^2\right] \\
&+ E\left[\frac{2}{T^2h^2}\sum_{s=2}^{T}\sum_{t>s}^{T} K_sK_h(\tau_t-\tau)\left(\frac{\tau_t-\tau}{h}\right)^{l}\left(\frac{s-\tau T}{Th}\right)^{l} C_s(1)C_t(1)\varepsilon_s\varepsilon_t\right] \\
&= \frac{1}{T^2h^2}\sum_{t=2}^{T} K_h(\tau_t-\tau)^2 C_t(1)^2 E\left[\varepsilon_t^2\right] + 0 \\
&= \frac{1}{Th^2}\int_0^1 K^2\left(\frac{u-\tau}{h}\right)\left(\frac{u-\tau}{h}\right)^{2l} C_{u\cdot T}(1)^2\sigma^2 du + o(1) \\
&= \frac{\sigma^2}{Th}\int_{-\frac{\tau}{h}}^{\frac{1-\tau}{h}} K(v)^2 v^{2l} C_{T\tau+Thv}(1)^2 dv + o(1) \\
&= C_{\tau\cdot T}(1)^2\frac{\sigma^2}{Th}\tilde{\sigma}_l^2 \\
&\longrightarrow 0.
\end{aligned} \tag{A.15}
$$

Before giving the proof of A.13, we first introduce some lemma about a_t. In Phillips and Solo (1992), $\frac{1}{T}\sum_{t=1}^{T}(\widetilde{\varepsilon}_t-\widetilde{\varepsilon}_{t-1})$ reduces to $\frac{1}{T}\left(\widetilde{\varepsilon}_T-\widetilde{\varepsilon}_0\right)$ because the terms $\widetilde{\varepsilon}_1$ to $\widetilde{\varepsilon}_{T-1}$ can be cancelled out. However, the same arguments or techniques are not applicable for kernel smoothing since the weights (kernel functions $K_h(\tau_t-\tau)$) to each individual $\widetilde{\varepsilon}_t$ are obviously different. Thus, no terms in this particular case can be removed. To deal with the convergence of $\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t^l(\widetilde{\varepsilon}_t-\widetilde{\varepsilon}_{t-1})$, we let $a_t = (\widetilde{\varepsilon}_t-\widetilde{\varepsilon}_{t-1})$, and then

$$
\begin{aligned}
\tilde{a}_t &= \sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t}\varepsilon_{t-j} - \sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t-1}\varepsilon_{t-1-j} \\
&= \sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t}\varepsilon_{t-j} - \sum_{j=1}^{\infty}\widetilde{\varphi}_{j-1,t-1}\varepsilon_{t-j} \\
&= \widetilde{\varphi}_{0,t}\varepsilon_t + \sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t}\varepsilon_{t-j} - \sum_{j=1}^{\infty}\widetilde{\varphi}_{j-1,t-1}\varepsilon_{t-j} \\
&= \widetilde{\varphi}_{0,t}\varepsilon_t + \sum_{j=1}^{\infty}\widetilde{\alpha}_{j,t}\varepsilon_{t-j},
\end{aligned}
$$

where $\widetilde{\alpha}_{j,t} = \widetilde{\varphi}_{j,t} - \widetilde{\varphi}_{j-1,t-1}$. Furthermore, we can assume that the initial values $\widetilde{\alpha}_{0,t} = \widetilde{\varphi}_{0,t}$

so that a_t can be represented by another MA(∞) process with a single summation,

$$\tilde{a}_t = \widetilde{\varphi}_{0,t}\varepsilon_t + \sum_{j=1}^{\infty} \widetilde{\alpha}_{j,t}\varepsilon_{t-j} = \sum_{j=0}^{\infty} \widetilde{\alpha}_{j,t}\varepsilon_{t-j}, \tag{A.16}$$

in which $\widetilde{\alpha}_{j,t}$ is a series of newly defined coefficients. This transformation simplifies a lot of calculation in the proof. Under Assumption 2.1 and Assumption 2.2, we have some properties of $\widetilde{\alpha}_{j,t}$: $\sum_{j=0}^{\infty} |\widetilde{\alpha}_{j,t}| < \infty$, and $\sum_{j=0}^{\infty} \widetilde{\alpha}_{j,t}^2 < \infty$ uniformly in $t \geq 1$. Thus, the expected value of a_t is zero, $E[\tilde{a}_t] = 0$,

$$E[\tilde{a}_t] = \sum_{j=0}^{\infty} \widetilde{\alpha}_{j,t}\varepsilon_{t-j} = \sum_{j=0}^{\infty} \widetilde{\alpha}_{j,t} E\left[\varepsilon_{t-j}\right] = 0.$$

The variance of a_t is finite, $E[a_t - E(a_t)]^2 < \infty$.

$$\begin{aligned}
E[\tilde{a}_t]^2 &= E\left[\sum_{j=0}^{\infty} \widetilde{\alpha}_{j,t}^2 \varepsilon_{t-j}^2 + 2\sum_{i=0}^{\infty} \sum_{j=0, j\neq i}^{\infty} \widetilde{\alpha}_{i,t}\widetilde{\alpha}_{j,t}\varepsilon_{t-i}\varepsilon_{t-j}\right] \\
&= \sum_{j=0}^{\infty} \widetilde{\alpha}_{j,t}^2 E\left[\varepsilon_{t-j}^2\right] + 2\sum_{i=0}^{\infty} \sum_{j=0, j\neq i}^{\infty} \widetilde{\alpha}_{i,t}\widetilde{\alpha}_{j,t} E\left[\varepsilon_{t-i}\varepsilon_{t-j}\right] \\
&= \sum_{j=0}^{\infty} \widetilde{\alpha}_{j,t}^2 \sigma^2 < \infty. \qquad \text{(A.17)}
\end{aligned}$$

The covariance between $\tilde{a}_t$ and $\tilde{a}_t$, for $t \neq s$, converges to zero in probability, $E[a_t a_s] \longrightarrow 0$, as $T \to \infty$ and $|t - s| \to \infty$.

$$\begin{aligned}
E[\tilde{a}_t \tilde{a}_s] &= E\left[\sum_{j=0}^{\infty} \widetilde{\alpha}_{j,t}\varepsilon_{t-j} \sum_{i=0}^{\infty} \widetilde{\alpha}_{i,s}\varepsilon_{s-i}\right] \\
&= \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} \widetilde{\alpha}_{j,t}\widetilde{\alpha}_{i,t} E\left[\varepsilon_{t-j}\varepsilon_{s-i}\right] \\
&= \sigma^2 \sum_{j=0}^{\infty} \widetilde{\alpha}_{j,t}\widetilde{\alpha}_{s-t+j,s} \\
&= \sigma^2 \gamma_t(h) \longrightarrow 0 \quad as \quad h \to \infty, \qquad \text{(A.18)}
\end{aligned}$$

where $h = t - s \neq 0$. For more discussion about long-run covariances, please see Gao and Anh (2000).

Therefore, the first moment

$$
\begin{aligned}
&E\left[\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l(\widetilde{\varepsilon}_t-\widetilde{\varepsilon}_{t-1})\right]\\
&=E\left[\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\left(\sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t}\varepsilon_{t-j}-\sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t-1}\varepsilon_{t-1-j}\right)\right]\\
&=E\left[\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t}\varepsilon_{t-j}\right]-E\left[\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t-1}\varepsilon_{t-1-j}\right]\\
&=\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t}E\left[\varepsilon_{t-j}\right]-\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t-1}E\left[\varepsilon_{t-1-j}\right]\\
&=0,
\end{aligned}
\tag{A.19}
$$

and the second moment is

$$
\begin{aligned}
&E\left[\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\tilde{a}_t\right]^2\\
&=E\left[\frac{1}{T^2h^2}\sum_{t=2}^{T}K_h(\tau_t-\tau)^2v_t^{2l}\tilde{a}_t^2\right]+E\left[\frac{1}{T^2h^2}\sum_{s=2}^{T}\sum_{t>s}^{T}K_sK_h(\tau_t-\tau)v_s^lv_t^l\tilde{a}_s\tilde{a}_t.\right]\\
&=\frac{1}{T^2h^2}\sum_{t=2}^{T}K_h(\tau_t-\tau)^2v_t^{2l}E\left[\tilde{a}_t^2\right]+\frac{1}{T^2h^2}\sum_{s=2}^{T}\sum_{t>s}^{T}K_sK_h(\tau_t-\tau)v_s^lv_t^lE\left[\tilde{a}_t\tilde{a}_s\right]\\
&\longrightarrow 0.
\end{aligned}
\tag{A.20}
$$

Provided that (A.19) and (A.20) hold, we have

$$
\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)^2v^lx_t\xrightarrow{p}A_{Tu}\tilde{\sigma}_l,
$$

by Slustky's theorem. Thence, we complete the proof of (3) and (4) of Lemma A.1.

According to the second-order tv-BN decomposition Lemma 2.2, the kernel-weighted sum of x_t^2, $\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^lx_t^2$ now can be decomposed into three components

$$
\begin{aligned}
\frac{1}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l x_{t-1}^2 =& \frac{1}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l \left(\sum_{j=0}^{\infty}\varphi_{j,t}^2\right)\varepsilon_t^2 \\
&+\frac{1}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l \tilde{Q}_t \\
&+\frac{1}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l \tilde{D}_t \\
&+\mu_t+2\mu_t\sum_{j=0}^{\infty}\varphi_{j,t}\varepsilon_{t-j}.
\end{aligned}
$$

To prove Lemma A.1 (4), (5), and (6), it suffices to show the following results:

$$\frac{1}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l(\mu_t)^2 \xrightarrow{p} A_{T\tau}^2\sigma_l, \tag{A.21}$$

$$\frac{2}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l\left(\mu_t\sum_{j=0}^{\infty}\varphi_{j,t}\varepsilon_{t-j}\right) \xrightarrow{p} 0, \tag{A.22}$$

$$\frac{1}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l\left(\sum_{j=0}^{\infty}\varphi_{j,t}^2\right)\varepsilon_t^2 \xrightarrow{p} \sigma^2 B_{0,T\tau}(1)\tilde{\sigma}_l, \tag{A.23}$$

$$\frac{1}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l\tilde{Q}_t \longrightarrow 0, \tag{A.24}$$

$$\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t^l\tilde{D}_t \longrightarrow 0, \tag{A.25}$$

for $l=0,1,2,3$, where we define $B_{0,T\tau}(1)=\sum_{j=0}^{\infty}\varepsilon_{j,T\tau}^2$ and $\tilde{\sigma}_l=\int_{-\infty}^{+\infty}K(v)v^l dv\cdot I(0<\tau<1)+\int_{\infty}^{0}K(v)v^l dv\cdot I(\tau=0)+\int_{0}^{-\infty}K(v)v^l dv\cdot I(\tau=1)$. The first moment is

$$
\begin{aligned}
E\left[\frac{1}{Th}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t^l(\mu_t)^2\right] =& \frac{1}{Th}\int_0^1 K\left(\frac{u-\tau}{h}\right)\left(\frac{u-\tau}{h}\right)^l A_{u\cdot T}^2 du+o(1) \\
=& \int_{-\frac{\tau}{h}}^{\frac{1-\tau}{h}} K(v)v^l A_{T\tau+Thv}^2 dv+o(1) \\
=& A_{T\tau}^2\sigma_l+o(1),
\end{aligned}
$$

and similarly we have the second moment

$$
\begin{aligned}
&E\left[\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\left(\mu_t\right)^2-A_{T\tau}^2\sigma_l\right]^2\\
=&E\left[\left(\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^lA_t^2\right)^2\right]-E\left[2A_{T\tau}^2\sigma_l\left(\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^lA_t^2\right)^2\right]+(A_{T\tau}^2\sigma_l)^2\\
=&E\left[\frac{1}{T^2h^2}\sum_{t=2}^{T}K_h(\tau_t-\tau)(v)^2\left(\frac{\tau_t-\tau}{h}\right)^{2l}A_t^4\right]\\
&+E\left[\frac{2}{T^2h^2}\sum_{s=2}^{T}\sum_{t>s}^{T}K_sK_h(\tau_t-\tau)\left(\frac{\tau_t-\tau}{h}\right)^l\left(\frac{\tau_s-\tau}{h}\right)^lA_s^2A_t^2\right]\\
&-2A_{T\tau}^2\sigma_l\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^lA_t^2+\left(A_{T\tau}^2\sigma_l\right)^2\\
&\longrightarrow 0.
\end{aligned}
$$

Therefore, result (A.21) holds.

We have

$$
\begin{aligned}
&\frac{2}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\left(\mu_t\sum_{j=0}^{\infty}\varphi_{j,t}\varepsilon_{t-j}\right)\\
=&\frac{2}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\mu_t\left(\sum_{j=0}^{\infty}\varphi_{j,t}\varepsilon_{t-j}\right)\\
=&\frac{2}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\mu_t\left(C_t(1)\varepsilon_t-(\widetilde{\varepsilon}_t-\widetilde{\varepsilon}_{t-1})\right)\\
=&\frac{2}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\tilde{a}_t\left(C_t(1)\varepsilon_t\right)-\frac{2}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\tilde{a}_t\left((\widetilde{\varepsilon}_t-\widetilde{\varepsilon}_{t-1})\right)\\
=&\frac{2}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\mu_tC_t(1)\varepsilon_t-\frac{2}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\mu_t\left(\widetilde{\varepsilon}_t-\widetilde{\varepsilon}_{t-1}\right).
\end{aligned}
$$

Since we have shown that $\frac{1}{Th}\sum_{t=1}^{T}K_h(\tau_t-\tau)v_t^lC_t(1)\varepsilon_t \xrightarrow{p} 0$, $\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l(\widetilde{\varepsilon}_t-\widetilde{\varepsilon}_{t-1}) \xrightarrow{p} 0$, for $l=0,1,2,3$, and A_t is deterministic function of time, thus we immedieately have

$$
\frac{2}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\tilde{a}_tC_t(1)\varepsilon_t \xrightarrow{p} 0 \quad\text{and}\quad \frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\tilde{a}_t(\widetilde{\varepsilon}_t-\widetilde{\varepsilon}_{t-1}) \xrightarrow{p} 0, \tag{A.26}
$$

Since

$$E\left[\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\tilde{a}_tC_t(1)\varepsilon_t\right]=\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\tilde{a}_tC_t(1)E\left[\varepsilon_t\right]=0,$$

and

$$\begin{aligned}
&E\left[\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\tilde{a}_tC_t(1)\varepsilon_t\right]^2\\
&=E\left[\frac{1}{T^2h^2}\sum_{t=2}^{T}K_h(\tau_t-\tau)^2\left(\frac{\tau_t-\tau}{h}\right)^{2l}A_t^2C_t(1)^2\varepsilon_t^2\right]\\
&+E\left[\frac{2}{T^2h^2}\sum_{s=2}^{T}\sum_{t>s}^{T}K_sK_h(\tau_t-\tau)\left(\frac{\tau_t-\tau}{h}\right)^{l}\left(\frac{s-\tau T}{Th}\right)^{l}\tilde{a}_tA_sC_s(1)C_t(1)\varepsilon_s\varepsilon_t\right]\\
&=\frac{1}{T^2h^2}\sum_{t=2}^{T}K_h(\tau_t-\tau)^2A_t^2C_t(1)^2E\left[\varepsilon_t^2\right]+0\\
&=\frac{1}{Th^2}\int_0^1K^2\left(\frac{u-\tau}{h}\right)\left(\frac{u-\tau}{h}\right)^{2l}A_{uT}^2C_{u\cdot T}(1)^2\sigma^2du+o(1)\\
&=\frac{\sigma^2}{Th}\int_{-\frac{\tau}{h}}^{\frac{1-\tau}{h}}K(v)^2v^{2l}A_{T\tau+Thv}^2C_{T\tau+Thv}(1)^2dv+o(1)\\
&=A_{\tau\cdot T}^2C_{\tau\cdot T}(1)^2\frac{\sigma^2}{Th}\tilde{\sigma}_l^2\\
&\longrightarrow 0,
\end{aligned}$$

thus (A.21) holds. In addition, we have

$$\begin{aligned}
&E\left[\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\tilde{a}_t(\widetilde{\varepsilon}_t-\widetilde{\varepsilon}_{t-1})\right]\\
&=E\left[\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\tilde{a}_t\left(\sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t}\varepsilon_{t-j}-\sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t-1}\varepsilon_{t-1-j}\right)\right]\\
&=E\left[\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^lA_t\sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t}\varepsilon_{t-j}\right]-E\left[\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^lA_t\sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t-1}\varepsilon_{t-1-j}\right]\\
&=\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^lA_t\sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t}E\left[\varepsilon_{t-j}\right]-\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\tilde{a}_t\sum_{j=0}^{\infty}\widetilde{\varphi}_{j,t-1}E\left[\varepsilon_{t-1-j}\right]\\
&=0,
\end{aligned}$$

and also

$$
\begin{aligned}
E\left[\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l a_t\right]^2 &= E\left[\frac{1}{T^2h^2}\sum_{t=2}^{T}K_h(\tau_t-\tau)^2 v_t^{2l}A_t^2\tilde{a}_t^2\right] \\
&+ E\left[\frac{1}{T^2h^2}\sum_{s=2}^{T}\sum_{t>s}^{T}K_sK_h(\tau_t-\tau)v_s^l v_t^l A_s\tilde{a}_t a_s a_t.\right] \\
&= \frac{1}{T^2h^2}\sum_{t=2}^{T}K_h(\tau_t-\tau)^2 v_t^{2l}A_t^2 E\left[\tilde{a}_t^2\right] \\
&+ \frac{1}{T^2h^2}\sum_{s=2}^{T}\sum_{t>s}^{T}K_sK_h(\tau_t-\tau)v_s^l v_t^l A_s\tilde{a}_t E\left[a_s a_t\right] \\
&\longrightarrow 0.
\end{aligned}
$$

Thus, (A.22) holds. By (A.21) and (A.22), we therefore, can show

$$
\frac{2}{Th}\sum_{t=2}^{\top}K_h(\tau_t-\tau)v_t^l\left(\mu_t\sum_{j=0}^{\infty}\varphi_{j,t}\varepsilon_{t-j}\right)\xrightarrow{p}0.
$$

thus (A.22) holds. Since

$$
\begin{aligned}
E\left[\frac{1}{Th}\sum_{t=1}^{\top}K_h(\tau_t-\tau)v_t^l\left(\sum_{j=0}^{\infty}\varphi_{j,t}^2\right)\varepsilon_t^2\right] &= \frac{1}{Th}\sum_{t=1}^{\top}K_h(\tau_t-\tau)v_t^l B_{0,t}(1)E\left[\varepsilon_t^2\right] \\
&= \frac{1}{Th}\int_0^1 K\left(\frac{u-\tau}{h}\right)\left(\frac{u-\tau}{h}\right)^l B_{0,Tu}(1)\sigma^2 du + o(1) \\
&= \sigma^2\int_{\frac{-\tau}{h}}^{\frac{1-\tau}{h}}K(v)\,B_{0,T(\tau+vh)}(1)d(v) + o(1) \\
&= \sigma^2 B_{0,T\tau}(1)\int_{\frac{-\tau}{h}}^{\frac{1-\tau}{h}}K(v)\,v^l d(v) + o(1) \\
&= \sigma^2 B_{0,T\tau}(1)\tilde{\sigma}_l + o(1),
\end{aligned}
$$

and its second moment is

$$
\begin{aligned}
&E\left[\frac{1}{Th}\sum_{t=2}^{\top}K_h(\tau_t-\tau)v_t^l\left(\sum_{j=0}^{\infty}\varphi_{j,t}^2\right)\varepsilon_t^2-\sigma^2B_{0,T\tau}(1)\tilde{\sigma}_l\right]^2\\
=&E\left[\frac{1}{Th}\sum_{t=2}^{\top}K_h(\tau_t-\tau)v_t^l\left(\sum_{j=0}^{\infty}\varphi_{j,t}^2\right)\varepsilon_t^2\right]^2\\
&-2\sigma^2B_{0,T\tau}(1)\tilde{\sigma}_lE\left[\frac{1}{Th}\sum_{t=2}^{\top}K_h(\tau_t-\tau)v_t^l\left(\sum_{j=0}^{\infty}\varphi_{j,t}^2\right)\varepsilon_t^2\right]\\
&+\left(\sigma^2B_{0,T\tau}(1)\tilde{\sigma}_l\right)^2\\
=&\frac{1}{T^2h^2}\sum_{t=2}^{\top}K_h(\tau_t-\tau)^2v_t^{2l}\left(\sum_{j=0}^{\infty}\varphi_{j,t}^2\right)^2E\left[\varepsilon_t^4\right]\\
&+\frac{1}{T^2h^2}\sum_{t=2}^{\top}\sum_{s=2,s\neq t}^{\top}K_h(\tau_t-\tau)K_sv_t^lv_s^l\sum_{j=0}^{\infty}\varphi_{j,t}^2\sum_{j=0}^{\infty}\varepsilon_{j,s}^2E\left[\varepsilon_t^2\varepsilon_s^2\right]\\
&-2\sigma^2B_{0,T\tau}(1)\tilde{\sigma}_l\left(\sigma^2B_{0,T\tau}(1)\tilde{\sigma}_l+o(1)\right)+\left(\sigma^2B_{0,T\tau}(1)\tilde{\sigma}_l\right)^2,
\end{aligned}
$$

where

$$
\begin{aligned}
&\frac{1}{T^2h^2}\sum_{t=2}^{\top}K_h(\tau_t-\tau)^2v_t^{2l}\left(\sum_{j=0}^{\infty}\varphi_{j,t}^2\right)^2E\left[\varepsilon_t^4\right]\\
&\quad=\frac{1}{T^2h^2}\sum_{t=2}^{\top}K_h(\tau_t-\tau)^2v_t^{2l}B_{0,T\tau}(1)^2\kappa^4\\
&\quad=\frac{\kappa^4}{Th^2}\int_0^1K^2\left(\frac{u-\tau}{h}\right)\left(\frac{u-\tau}{h}\right)^{2l}B_{0,Tu}(1)^2du+o(1)\\
&\quad=\frac{\kappa^4}{Th}\int_{\frac{-\tau}{h}}^{\frac{1-\tau}{h}}K^2(v)\,v^{2l}B_{0,Tu+Tvh}(1)^2d(v)+o(1)\\
&\quad=\frac{\kappa^4}{Th}B_{0,T\tau}(1)^2\int_{\frac{-\tau}{h}}^{\frac{1-\tau}{h}}K^2(v)\,v^{2l}d(v)+o(1)\\
&\quad=\frac{\kappa^4}{Th}B_{0,T\tau}(1)^2\tilde{\sigma}_l^2\\
&\quad\longrightarrow 0,
\end{aligned}
$$

and $\frac{1}{T^2h^2}\sum_{t=2}^{\top}\sum_{s=2,s\neq t}^{\top}K_h(\tau_t-\tau)K_sv_t^lv_s^l\sum_{j=0}^{\infty}\varphi_{j,t}^2\sum_{j=0}^{\infty}\varepsilon_{j,s}^2E\left[\varepsilon_t^2\varepsilon_s^2\right]\longrightarrow 0$. Thus, (A.23) holds.

Since

$$\begin{aligned}
E\left[\frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)v_t^l\tilde{Q}_t\right] &= \frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)v_t^l E\left[2\sum_{h=1}^{\infty} B_{h,t}(1)\varepsilon_t\varepsilon_{t-h}\right] \\
&= \frac{2}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)v_t^l \sum_{h=1}^{\infty} B_{h,t}(1)E\left[\varepsilon_t\varepsilon_{t-h}\right] \\
&= 0,
\end{aligned}$$

and the second moment is

$$\begin{aligned}
E\left[\frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)v_t^l\tilde{Q}_t\right]^2 &= E\left[\frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t^{2l}\tilde{Q}_t^2\right] \\
&+ E\left[\frac{2}{T^2h^2}\sum_{s=1}^{T}\sum_{t=1,t\neq s}^{T} K_h(\tau_t-\tau)v_t^l K_s v_s^l \tilde{Q}_t Q_s\right].
\end{aligned}$$

We show next that in the above formula, the variance of $\tilde{Q}_t$ is finite.

$$\begin{aligned}
E[\tilde{Q}_t^2] &= E\left[2\sum_{h=1}^{\infty} B_{h,t}(1)\varepsilon_t\varepsilon_{t-h}\right]^2 \\
&= 4\sigma^2\left[\sum_{h=1}^{\infty}\left(\sum_{j=0}^{\infty}\varphi_{j,t}\varphi_{j+h,t}\right)^2\right] \\
&\leq 4\sigma^2\left[\sum_{h=1}^{\infty}\left(\sum_{j=0}^{\infty}\varphi_{j,t}^2\right)\sum_{j=0}^{\infty}\varphi_{j+h,t}^2\right] \\
&= 4\sigma^2\left[\left(\sum_{j=0}^{\infty}\varphi_{j,t}^2\right)\sum_{h=1}^{\infty}\sum_{j=0}^{\infty}\varphi_{j+h,t}^2\right] \\
&= 4\sigma^2\left[\left(\sum_{j=0}^{\infty}\varphi_{j,t}^2\right)\sum_{j=1}^{\infty} j\varphi_{j,t}^2\right] \\
&< \infty.
\end{aligned}$$

Hence,

$$E[\tilde{Q}_t]^2 = 4\sigma^2 E\left[\sum_{h=1}^{\infty} B_{h,t}(1)\varepsilon_{t-h}\right]^2 \leq 4\sigma^2\left[\left(\sum_{j=0}^{\infty}\varphi_{j,t}^2\right)\sum_{j=1}^{\infty} j\varphi_{j,t}^2\right] < \infty.$$

The covariance between $\tilde{Q}_t$ and Q_s for $t \neq s$ is

$$\begin{aligned} E[\tilde{Q}_t Q_s] &= E\left[2\sum_{h=1}^{\infty} B_{h,t}(1)\varepsilon_t\varepsilon_{t-h} 2\sum_{h=1}^{\infty} B_{h,s}(1)\varepsilon_s\varepsilon_{s-h}\right] \\ &= 4E\left[\varepsilon_t\varepsilon_s\right] E\left[\sum_{h=1}^{\infty} B_{h,t}(1)\varepsilon_{t-h} \sum_{h=1}^{\infty} B_{h,s}(1)\varepsilon_{s-h}\right] \\ &= 0. \end{aligned}$$

Thus, we have

$$\begin{aligned} E\left[\frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v^{2l}\tilde{Q}_t^2\right] &= \frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v^{2l} 4\sigma^4\left[\sum_{h=1}^{\infty} B_{h,t}(1)^2\right] \\ &= \frac{4\sigma^4}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v^{2l} F_{h,t} \\ &= \frac{4\sigma^4}{Th^2}\int_0^1 K_h(\tau_t-\tau)^2 v^{2l} F_{j,Tu} du + o(1) \\ &= \frac{4\sigma^4}{Th}\int_{\frac{-\tau}{h}}^{\frac{1-\tau}{h}} K^2(v)\, F_{j,T(\tau+vh)} v^{2l} d(v) + o(1) \\ &= F_{j,T\tau}\frac{4\sigma^4}{Th}\int_{\frac{-\tau}{h}}^{\frac{1-\tau}{h}} K^2(v)\, v^{2l} d(v) + o(1) \\ &= F_{j,T\tau}\frac{4\sigma^4}{Th}\tilde{\sigma}_l^2 + o(1) \\ &\longrightarrow 0. \end{aligned}$$

For the covariance term

$$\begin{aligned} E\left[\frac{2}{T^2h^2}\sum_{s=1}^{T}\sum_{t=1,t\neq s}^{T} K_h(\tau_t-\tau) v_t^l K_s v_s^l \tilde{Q}_t Q_s\right] &= \frac{2}{T^2h^2}\sum_{s=1}^{T}\sum_{t=1,t\neq s}^{T} E\left[\tilde{Q}_t Q_s\right] \\ &= 0. \end{aligned}$$

Since

$$E\left[\frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau) v_t^l \tilde{Q}_t\right] = 0,$$

and

$$E\left[\frac{1}{Th}\sum_{t=1}^{\top} K\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right)^l \tilde{Q}_t\right]^2 \longrightarrow 0,$$

then Equation A.24 holds.

We apply the same technique once again to transform $\widetilde{D}_t$ into a single sum. That is,

$$
\begin{aligned}
\widetilde{D}_t &= \sum_{h=1}^{\infty} \widetilde{B}_{h,t}(L)\varepsilon_t\varepsilon_{t-h} - \sum_{h=1}^{\infty} \widetilde{B}_{h,t-1}(L)\varepsilon_{t-1}\varepsilon_{t-1-h} \\
&= \sum_{h=1}^{\infty}\left(\sum_{j=0}^{\infty} \varphi_{j,t}\varphi_{j+h,t}\varepsilon_{t-j}\varepsilon_{t-j-h} - \sum_{j=0}^{\infty} \varphi_{j,t}\varphi_{j+h,t}\varepsilon_{t-1-j}\varepsilon_{t-1-j-h}\right) \\
&= \sum_{h=1}^{\infty}\left(\varepsilon_{0,t}\varepsilon_{h,t}\varepsilon_t\varepsilon_{t-h} - \sum_{j=1}^{\infty}\left(\varphi_{j,t}\varphi_{j+h,t} - \varepsilon_{j-1,t}\varepsilon_{j-1+h,t}\right)\varepsilon_{t-j}\varepsilon_{t-j-h}\right) \\
&= \sum_{h=1}^{\infty}\left(\varepsilon_{0,t}\varepsilon_{h,t}\varepsilon_t\varepsilon_{t-h} - \sum_{j=1}^{\infty} \phi_{h,j,t}\varepsilon_{t-j}\varepsilon_{t-j-h}\right) \\
&= \sum_{h=1}^{\infty}\left(\sum_{j=0}^{\infty} \widetilde{\phi}_{h,j,t}\varepsilon_{t-j}\varepsilon_{t-j-h}\right),
\end{aligned}
\tag{A.27}
$$

where we define $\widetilde{\phi}_{h,j,t} = \varphi_{j,t}\varphi_{j+h,t} - \varepsilon_{j-1,t}\varepsilon_{j-1+h,t}$, and it can be assumed that the initial values $\varepsilon_{0,t}\varepsilon_{h,t}\varepsilon_t\varepsilon_{t-h} = \widetilde{\phi}_{0,t}$. We have

$$
\sum_{j=0}^{\infty} |\widetilde{\phi}_{j,t}| < \infty, \quad \sum_{j=0}^{\infty} \widetilde{\phi}_{j,t}^2 < \infty,
\tag{A.28}
$$

uniformly in $t \geq 1$. Even if $\tilde{D}_t$ has been merged as (A.27), there are still double summations which can be further simplified as

$$
\begin{aligned}
\tilde{D}_t &= \sum_{h=1}^{\infty}\left(\sum_{j=0}^{\infty} \widetilde{\phi}_{h,j,t}\varepsilon_{t-j}\varepsilon_{t-j-h}\right) \\
&= \sum_{h=1}^{\infty} \sum_{k=-\infty=1}^{\top} \widetilde{\phi}_{h,t-k,t}\varepsilon_k\varepsilon_{k-h}, \qquad (k = t-j) \\
&= \sum_{k=-\infty}^{\top}\left(\sum_{h=1}^{\infty} \widetilde{\phi}_{h,t-k,t}\varepsilon_{k-h}\right)\varepsilon_k \\
&= \sum_{k=-\infty}^{\top} \eta_{t,k}\varepsilon_k,
\end{aligned}
$$

where $\eta_{t,k} = \sum_{h=1}^{\infty} \widetilde{\phi}_{h,t-k,t}\varepsilon_{k-h}$ that is always independent with ε_k due to the fact that

$h \geq 1$ and therefore $cov(\varepsilon_k, \varepsilon_{k-h}) = 0$. Obviously, we can have the following proposition.

$$
\begin{aligned}
E\left[\frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t - \tau) v_t^l \widetilde{D}_t\right] &= E\left[\frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t - \tau) v_t^l \sum_{k=-\infty}^{\top} \eta_{t,k}\varepsilon_k\right] \\
&= \frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t - \tau) v_t^l \sum_{k=-\infty}^{\top} E\left[\eta_{t,k}\varepsilon_k\right] \\
&= 0,
\end{aligned}
$$

and

$$
\begin{aligned}
E\left[\frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t - \tau) v_t^l \tilde{D}_t\right]^2 &= E\left[\frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t - \tau)^2 v_t^{2l} \tilde{D}_t^2\right] \\
&+ E\left[\frac{2}{T^2h^2}\sum_{s=1}^{T}\sum_{t=1,t\neq s}^{T} K_h(\tau_t - \tau) v_t^l K_s v_s^l D_s \tilde{D}_t\right] \\
&= E\left[\frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t - \tau)^2 v_t^{2l} \left(\sum_{k=-\infty}^{\top} \eta_{t,k}\varepsilon_k\right)^2\right] \\
&+ E\left[\frac{2}{T^2h^2}\sum_{s=1}^{T}\sum_{t=1,t\neq s}^{T} K_h(\tau_t - \tau) v_t^l K_s v_s^l \sum_{k=-\infty}^{s} \eta_{s,k}\varepsilon_k \sum_{k=-\infty}^{\top} \eta_{t,k}\varepsilon_k\right],
\end{aligned}
$$

where

$$
\begin{aligned}
&E\left[\frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t - \tau)^2 v_t^{2l} \left(\sum_{k=-\infty}^{\top} \eta_{t,k}\varepsilon_k\right)^2\right] \\
&= \frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t - \tau)^2 v_t^{2l} \sum_{k=-\infty}^{\top} \eta_{t,k}^2\sigma^2 \\
&= \frac{\sigma^2}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t - \tau)^2 v_t^{2l} \sum_{k=-\infty}^{\top} \left(\sum_{h=1}^{\infty} \widetilde{\phi}_{h,t-k,t}\varepsilon_{k-h}\right)^2 \\
&= \frac{\sigma^4}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t - \tau)^2 v_t^{2l} \sum_{j=0}^{\infty} \left(\sum_{h=1}^{\infty} \widetilde{\phi}_{h,j,t}^2\right).
\end{aligned}
$$

We denote $H_{j,t} = \sum_{j=0}^{\infty}\sum_{h=1}^{\infty}\widetilde{\phi}_{h,j,t}^{2}$ and then

$$
\begin{aligned}
E\left[\frac{1}{T^2h^2}\sum_{t=1}^{\top}K_h(\tau_t-\tau)^2 v_t^{2l}\tilde{D}_t^2\right] &= \frac{\sigma^4}{T^2h^2}\sum_{t=1}^{\top}K_h(\tau_t-\tau)^2 v_t^{2l}H_t \\
&= H_{j,T\tau}\frac{\sigma^4}{Th}\int_{\frac{-\tau}{h}}^{\frac{1-\tau}{h}}K^2(v)\,v^{2l}d(v)+o(1) \\
&= H_{j,T\tau}\frac{\sigma^4}{Th}\tilde{\sigma}_l^2 \\
&\longrightarrow 0. \qquad\qquad (A.29)
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
&E\left[\frac{2}{T^2h^2}\sum_{s=1}^{T}\sum_{t=1,t\neq s}^{T}K_h(\tau_t-\tau)v_t^lK_sv_s^l\sum_{k=-\infty}^{s}\eta_{s,k}\varepsilon_k\sum_{k=-\infty}^{\top}\eta_{t,k}\varepsilon_k\right] \\
&\frac{2}{T^2h^2}\sum_{s=1}^{T}\sum_{t=1,t\neq s}^{T}K_h(\tau_t-\tau)v_t^lK_sv_s^lE\left[\sum_{k=-\infty}^{s}\eta_{s,k}\varepsilon_k\sum_{k=-\infty}^{\top}\eta_{t,k}\varepsilon_k\right] \\
&=0,
\end{aligned}
$$

because when $t\neq k$ we always have the result that $E\left[\sum_{k=-\infty}^{s}\eta_{s,k}\varepsilon_k\sum_{k=-\infty}^{\top}\eta_{t,k}\varepsilon_k\right]=0$. Therefore, $E\left[\frac{1}{Th}\sum_{t=1}^{\top}K_h(\tau_t-\tau)v_t^l\tilde{D}_t\right]^2 \xrightarrow{p} 0$ and therefore (A.25) holds. The results (A.21), (A.22), (A.23), (A.24), and (A.25) are sufficient to complete the proof of (5) and (6) of Lemma A.1. ■

Proof of Lemma A.2. It is clear that the asymptotic distribution of the local linear estimator $\boldsymbol{\theta}(\tau)$ depends on the error term ε_t with the assumption. Recall the term that produces the CLT, $\widetilde{T}_T=[\mathrm{I}_2,0_2]S_T^{-1}\widetilde{J}_T$ where $[\mathrm{I}_2,0_2]$ is non-stochastic and $\boldsymbol{S}_{\boldsymbol{T}}^{-1}(\tau)\xrightarrow{p}\Sigma_S^{-1}(\tau)$.

$$
\sqrt{Th}\widetilde{J}_T=\begin{bmatrix}\sqrt{Th}\widetilde{J}_1\\ \sqrt{Th}\widetilde{J}_2\end{bmatrix}=\begin{bmatrix}\sqrt{Th}\frac{1}{Th}\sum_{t=1}^{T}K_h(\tau_t-\tau)Z_t\varepsilon_t\\ \sqrt{Th}\frac{1}{Th}\sum_{t=1}^{T}K_h(\tau_t-\tau)v_t^lZ_t\varepsilon_t\end{bmatrix},
$$

and the variance-covariance matrix is

$$
var(\sqrt{Th}\widetilde{J}_T)=\begin{bmatrix}var(\sqrt{Th}\widetilde{J}_1) & cov(\sqrt{Th}\widetilde{J}_1,\widetilde{J}_2)\\ cov(\sqrt{Th}\widetilde{J}_2,\widetilde{J}_1) & var(\sqrt{Th}\widetilde{J}_2)\end{bmatrix}.
$$

We can show the elements in the var-cov matrix one-by-one

$$
\begin{aligned}
var(\sqrt{Th}\widetilde{J}_1) &= var\left(\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)Z_t\varepsilon_t\right)\\
&= var\begin{bmatrix} \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)\varepsilon_t \\ \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)x_t\varepsilon_t \end{bmatrix} = \begin{bmatrix} var(\widetilde{J}_{1,1}) & cov(\widetilde{J}_{1,1},\widetilde{J}_{1,2}) \\ cov(\widetilde{J}_{1,2},\widetilde{J}_{1,1}) & var(\widetilde{J}_{1,2}) \end{bmatrix},
\end{aligned}
$$

where we define $\widetilde{J}_{1,1} = \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)\varepsilon_t$ and $\widetilde{J}_{1,2} = \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)x_t\varepsilon_t$. Now, we deal with the elements in the matrix $var(\widetilde{J}_1)$, $var(\widetilde{J}_{1,1}) \xrightarrow{p} \sigma_\varepsilon^2\sigma_{l=0}^2 = \pi_{1,1}$

$$
\begin{aligned}
var(\widetilde{J}_{1,1}) &= E\left[\widetilde{J}_{1,1} - E\left(\widetilde{J}_{1,1}\right)\right]^2\\
&= E\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)\varepsilon_t - E\left(\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)\varepsilon_t\right)\right]^2\\
&= \frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t-\tau)^2 E[\varepsilon_t^2] + 0\\
&= \sigma_\varepsilon^2 \int_{\frac{-\tau}{h}}^{\frac{1-\tau}{h}} K^2(v)\,\sigma_\varepsilon^2 dv + o(1)\\
&\longrightarrow \sigma_\varepsilon^2\sigma_{l=0}^2 = \pi_{1,1},
\end{aligned}
$$

where $\tilde{\sigma}_l^2 = \int_{-\infty}^{+\infty} K^2(v)v^l dv \cdot I(0<\tau<1) + \int_{\infty}^{0} K^2(v)v^l dv \cdot I(\tau=0) + \int_{0}^{-\infty} K^2(v)v^l dv \cdot I(\tau=1)$.

$cov(\widetilde{J}_{1,2}, \widetilde{J}_{1,1}) \longrightarrow 0 = \pi_{1,2}$

$$
\begin{aligned}
cov(\widetilde{J}_{1,2}, \widetilde{J}_{1,1}) &= cov(\widetilde{J}_{1,1}, \widetilde{J}_{1,2}) \\
&= E\left[\widetilde{J}_{1,1} - E(\widetilde{J}_{1,1})\right]\left[\widetilde{J}_{1,2} - E(\widetilde{J}_{1,2})\right] \\
&= E\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t - \tau)\varepsilon_t - E\left(\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t - \tau)\varepsilon_t\right)\right] \\
&\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t - \tau)x_t\varepsilon_t - E\left(\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t - \tau)x_t\varepsilon_t\right)\right] \\
&= E\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t - \tau)\varepsilon_t \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t - \tau)x_t\varepsilon_t\right] \\
&= E\left[\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t - \tau)^2 x_t\varepsilon_t^2\right] \\
&= \sigma_\varepsilon^2 \frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t - \tau)^2 x_t \\
&\longrightarrow 0 = \pi_{1,2}.
\end{aligned}
$$

$$
\begin{aligned}
var(\widetilde{J}_{1,2}) = E\left[\widetilde{J}_{1,2} - E\left(\widetilde{J}_{1,2}\right)\right]^2 &= E\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t - \tau)x_t\varepsilon_t\right]^2 \\
&= \sigma_\varepsilon^2 \frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t - \tau)^2 x_t^2 \\
&\longrightarrow \sigma_\varepsilon^2\sigma^2 B_{0,T\tau}(1)\sigma_{K,l=0}^2 = \pi_{1,3}.
\end{aligned}
$$

Hence, we have

$$
var(\widetilde{J}_{1,1}) \longrightarrow \Pi_1 = \begin{bmatrix} \sigma_\varepsilon^2\sigma_{l=0}^2 & 0 \\ 0 & \sigma_\varepsilon^2\sigma^2 B_{0,T\tau}(1)\sigma_{K,l=0}^2 \end{bmatrix}. \tag{A.30}
$$

Second, we deal with the elements in the matrix $cov(\widetilde{J}_1, \widetilde{J}_2)$.

$$
\begin{aligned}
cov(\widetilde{J}_1, \widetilde{J}_2) &= E\left[\widetilde{J}_1 - E(\widetilde{J}_1)\right]\left[\widetilde{J}_2 - E(\widetilde{J}_2)\right]^\top \\
&= E\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau) Z_t\varepsilon_t\right]\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau) v_t Z_t\varepsilon_t\right]^\top \\
&= E\begin{bmatrix} \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)\varepsilon_t \\ \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)x_t\varepsilon_t \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t\varepsilon_t \\ \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t x_t\varepsilon_t \end{bmatrix}^\top
\end{aligned}
$$

where

$$
\begin{aligned}
&\sigma_\varepsilon^2 \frac{1}{Th}\sum_{t=1}^{T} K^2\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right) \longrightarrow \pi_{3,1} = \sigma_\varepsilon^2\sigma_{K,l=1}^2 \\
&\sigma_\varepsilon^2 \frac{1}{Th}\sum_{t=1}^{T} K^2\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right) x_t \xrightarrow{p} \pi_{2,2} = 0 \\
&\sigma_\varepsilon^2 \frac{1}{Th}\sum_{t=1}^{T} K^2\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right) x_t \xrightarrow{p} \pi_{2,2} = 0 \\
&\sigma_\varepsilon^2 \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\tau_t-\tau}{h}\right) x_t^2 \xrightarrow{p} \sigma_\varepsilon^2\sigma^2 B_{0,T\tau}(1)\sigma_{l=1}.
\end{aligned}
$$

Hence, we have

$$
cov(\widetilde{J}_1, \widetilde{J}_2) \xrightarrow{p} \varepsilon = \begin{bmatrix} \sigma_\varepsilon^2\sigma_{K,l=1}^2 & 0 \\ 0 & \sigma_\varepsilon^2\sigma^2 B_{0,T\tau}(1)\sigma_{l=1} \end{bmatrix}. \tag{A.31}
$$

Third, we deal with the elements in the matrix $var(\widetilde{J}_2)$,

$$
\begin{aligned}
var(\widetilde{J}_2) &= var\left(\frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right) Z_t\varepsilon_t\right) \\
&= var\begin{bmatrix} \sqrt{Th}\frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right)\varepsilon_t \\ \sqrt{Th}\frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right)x_t\varepsilon_t \end{bmatrix} \\
&= \begin{bmatrix} var(\widetilde{J}_{2,1}) & cov(\widetilde{J}_{2,1}, \widetilde{J}_{2,2}) \\ cov(\widetilde{J}_{2,2}, \widetilde{J}_{2,1}) & var(\widetilde{J}_{2,2}) \end{bmatrix},
\end{aligned}
$$

where

$$
\begin{aligned}
var(\widetilde{J}_{2,1}) &= E\left[\widetilde{J}_{2,1} - E\left(\widetilde{J}_{2,1}\right)\right]^2 \\
&= E\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t\varepsilon_t\right]^2 \\
&= \sigma_\varepsilon^2 \int_{\frac{-\tau}{h}}^{\frac{1-\tau}{h}} vK^2\left(v\right)\sigma_\varepsilon^2 dv + o(1) \\
&\longrightarrow \sigma_\varepsilon^2\sigma_{l=1}^2,
\end{aligned}
$$

$$
\begin{aligned}
cov(\widetilde{J}_{2,2},\widetilde{J}_{2,1}) &= E\left[\widetilde{J}_{2,1} - E(\widetilde{J}_{2,1})\right]\left[\widetilde{J}_{2,2} - E(\widetilde{J}_{2,2})\right] \\
&= E\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t x_t\varepsilon_t - E\left(\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t x_t\varepsilon_t\right)\right] \\
&\times\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t\varepsilon_t - E\left(\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t\varepsilon_t\right)\right] \\
&= E\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t x_t\varepsilon_t \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K_h(\tau_t-\tau)v_t\varepsilon_t\right] \\
&= \sigma_\varepsilon^2\frac{1}{Th}\sum_{t=1}^{T} K_h(\tau_t-\tau)^2 v_t^2 x_t + 0 \\
&\longrightarrow 0,
\end{aligned}
$$

and

$$
\begin{aligned}
var(\widetilde{J}_{2,2}) &= E\left[\widetilde{J}_{2,2} - E\left(\widetilde{J}_{2,2}\right)\right]^2 \\
&= E\left[\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} K\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right)x_t\varepsilon_t\right]^2 \\
&= \sigma_\varepsilon^2\frac{1}{Th}\sum_{t=1}^{T} K^2\left(\frac{\tau_t-\tau}{h}\right)\left(\frac{\tau_t-\tau}{h}\right)^2 x_t^2 \\
&\longrightarrow \sigma_\varepsilon^2\sigma^2 B_{0,T\tau}(1)\sigma_{l=2}^2.
\end{aligned}
$$

Hence, we have

$$var(\widetilde{J}_2) \longrightarrow \Pi_3 = \begin{bmatrix} \sigma_\varepsilon^2 \sigma_{l=1}^2 & 0 \\ 0 & \sigma_\varepsilon^2 \sigma^2 B_{0,T\tau}(1) \sigma_{l=2}^2 \end{bmatrix}. \tag{A.32}$$

Since we have results (A.30), (A.31) and (A.32), we have the following property

$$var(\sqrt{Th}\widetilde{J}_T) \xrightarrow{p} \Pi(\tau).$$

We denote that $\boldsymbol{\lambda}$ is a 4×1 vector of constants, $\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 \end{bmatrix}^\top$, where λ_i can be any number. To show the asymptotic distribution, we follow Gao and Anh (1999) to prove two conditions.

Condition 1: $\sum_{t=1}^{\top} E\left[\left(\boldsymbol{\lambda}^\top \boldsymbol{a}_t(\tau)\right)^2 |\Omega_{t-1}\right] \xrightarrow{p} \boldsymbol{\lambda}^\top \Pi(\tau) \boldsymbol{\lambda}$.

$\boldsymbol{\lambda}^\top \widetilde{J}_T$ now becomes a scalar so that we can use the CLT for martingale difference sequence, we first define $\boldsymbol{a}_t(\tau) = \begin{bmatrix} \frac{1}{\sqrt{Th}} K\left(\frac{\tau_t - \tau}{h}\right) Z_t \varepsilon_t \\ \frac{1}{\sqrt{Th}} K\left(\frac{\tau_t - \tau}{h}\right)\left(\frac{\tau_t - \tau}{h}\right) Z_t \varepsilon_t \end{bmatrix}$, and hence we want to show that

$$\sqrt{Th}\boldsymbol{\lambda}^\top \widetilde{J}_T = \sum_{t=1}^{\top} \boldsymbol{\lambda}^\top \boldsymbol{a}_t(\tau) \xrightarrow{d} N\Big(0, \quad \boldsymbol{\lambda}^\top \Pi(\tau) \boldsymbol{\lambda}\Big).$$

Thus, it is obvious that

$$\begin{aligned}
&\sum_{t=1}^{\top} E\left[\left(\boldsymbol{\lambda}^\top \boldsymbol{a}_t(\tau)\right)^2 |\Omega_{t-1}\right] = \boldsymbol{\lambda}^\top \sum_{t=1}^{\top} E\left[\boldsymbol{a}_t(\tau)\boldsymbol{a}_t^\top(\tau)|\Omega_{t-1}\right] \boldsymbol{\lambda} \\
&= \boldsymbol{\lambda}^\top \begin{bmatrix} E\left[\frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 Z_t Z_t^\top \varepsilon_t^2|\Omega_{t-1}\right] & E\left[\frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t Z_t Z_t^\top \varepsilon_t^2|\Omega_{t-1}\right] \\ E\left[\frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t Z_t Z_t^\top \varepsilon_t^2|\Omega_{t-1}\right] & E\left[\frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t^2 Z_t Z_t^\top \varepsilon_t^2|\Omega_{t-1}\right] \end{bmatrix} \boldsymbol{\lambda} \\
&= \boldsymbol{\lambda}^\top \begin{bmatrix} \frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 Z_t Z_t^\top E\left[\varepsilon_t^2|\Omega_{t-1}\right] & \frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t Z_t Z_t^\top E\left[\varepsilon_t^2|\Omega_{t-1}\right] \\ \frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t Z_t Z_t^\top E\left[\varepsilon_t^2|\Omega_{t-1}\right] & \frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t^2 Z_t Z_t^\top E\left[\varepsilon_t^2|\Omega_{t-1}\right] \end{bmatrix} \boldsymbol{\lambda}.
\end{aligned} \tag{A.33}$$

Now we show the general term in the above that

$$
\begin{aligned}
\frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t^l Z_t Z_t^{\top} &= \frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t^l \begin{bmatrix} 1 & x_t \\ x_t & x_t^2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t^l & \frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t^l x_t \\ \frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t^l x_t & \frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t^l x_t^2 \end{bmatrix} \\
&\xrightarrow{p} \begin{bmatrix} \tilde{\sigma}_l^2 & A_{T\tau}\tilde{\sigma}_l^2 \\ A_{T\tau}\tilde{\sigma}_l^2 & \sigma^2 B_{0,T\tau}(1)\tilde{\sigma}_l^2 \end{bmatrix}, \text{for} \quad l=0,1,2.
\end{aligned}
$$

Since we assume $E\left[\varepsilon_t^2|\Omega_{t-1}\right]=\sigma_\varepsilon^2$, we have

$$
\frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 Z_t Z_t^{\top} E\left[\varepsilon_t^2|\Omega_{t-1}\right] \xrightarrow{p} \begin{bmatrix} \sigma_\varepsilon^2\sigma_{l=0}^2 & \sigma_{l=0}^2 A_{T\tau} \\ \sigma_{l=0}^2 A_{T\tau} & \sigma_\varepsilon^2\sigma^2 B_{0,T\tau}(1)\sigma_{l=0}^2 \end{bmatrix} = \Pi_1(\tau)
$$

$$
\frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t^l Z_t Z_t^{\top} E\left[\varepsilon_t^2|\Omega_{t-1}\right] \xrightarrow{p} \begin{bmatrix} \sigma_\varepsilon^2\sigma_{l=1}^2 & A_{T\tau}\sigma_{l=1}^2 \\ A_{T\tau}\sigma_{l=1}^2 0 & \sigma_\varepsilon^2\sigma^2 B_{0,T\tau}(1)\sigma_{l=1}^2 \end{bmatrix} = \Pi_2(\tau)
$$

$$
\frac{1}{Th}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t^2 Z_t Z_t^{\top} E\left[\varepsilon_t^2|\Omega_{t-1}\right] \xrightarrow{p} \begin{bmatrix} \sigma_\varepsilon^2\sigma_{l=2}^2 & A_{T\tau}\sigma_{l=2}^2 \\ A_{T\tau}\sigma_{l=2}^2 & \sigma_\varepsilon^2\sigma^2 B_{0,T\tau}(1)\sigma_{l=2}^2 \end{bmatrix} = \Pi_3(\tau).
$$

Let's return to (A.33),

$$
\begin{aligned}
&\sum_{t=1}^{\top} E\left[\left(\boldsymbol{\lambda}^{\top}\boldsymbol{a}_t(\tau)\right)^2|\Omega_{t-1}\right] \\
&= \boldsymbol{\lambda}^{\top}\begin{bmatrix} \sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 Z_t Z_t^{\top} E\left[\varepsilon_t^2|\Omega_{t-1}\right] & \sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t Z_t Z_t^{\top} E\left[\varepsilon_t^2|\Omega_{t-1}\right] \\ \sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t Z_t Z_t^{\top} E\left[\varepsilon_t^2|\Omega_{t-1}\right] & \sum_{t=1}^{\top} K_h(\tau_t-\tau)^2 v_t^2 Z_t Z_t^{\top} E\left[\varepsilon_t^2|\Omega_{t-1}\right] \end{bmatrix}\boldsymbol{\lambda} \\
&= \boldsymbol{\lambda}^{\top}\begin{bmatrix} \Pi_1(\tau) & \Pi_2(\tau) \\ \Pi_2(\tau) & \Pi_3(\tau) \end{bmatrix}\boldsymbol{\lambda} \\
&= \boldsymbol{\lambda}^{\top}\Pi(\tau)\boldsymbol{\lambda}.
\end{aligned}
$$

Hence, the first condition holds.

Condition 2: $\sum_{t=1}^{\top} E\left[\left(\boldsymbol{\lambda}^{\top}\boldsymbol{a}_t(\tau)\right)^4|\Omega_{t-1}\right] \xrightarrow{p} 0.$

The conditional fourth moment

$$
\begin{aligned}
\sum_{t=1}^{\top} E\left[\left(\boldsymbol{\lambda}^{\top}\boldsymbol{a}_t(\tau)\right)^4 |\Omega_{t-1}\right] &= \sum_{t=1}^{\top} E\left[\left(\boldsymbol{\lambda}^{\top}\boldsymbol{a}_t(\tau)\boldsymbol{a}_t^{\top}(\tau)\boldsymbol{\lambda}\right)^2 |\Omega_{t-1}\right] \\
&= \sum_{t=1}^{\top} E\left[\boldsymbol{\lambda}^{\top}\boldsymbol{a}_t(\tau)\boldsymbol{a}_t^{\top}(\tau)\boldsymbol{\lambda}\boldsymbol{\lambda}^{\top}\boldsymbol{a}_t(\tau)\boldsymbol{a}_t^{\top}(\tau)\boldsymbol{\lambda}|\Omega_{t-1}\right],
\end{aligned}
$$

where we know that $\boldsymbol{\lambda}^{\top}\boldsymbol{a}_t(\tau)\boldsymbol{a}_t^{\top}(\tau)\boldsymbol{\lambda}\boldsymbol{\lambda}^{\top}\boldsymbol{a}_t(\tau)\boldsymbol{a}_t^{\top}(\tau)\boldsymbol{\lambda}$ is just a scalar. According to the multinomial formula, $\left(\boldsymbol{\lambda}^{\top}\boldsymbol{a}_t(\tau)\right)^4 = \left(\lambda_1\widetilde{J}_1 + \lambda_2\widetilde{J}_2 + \lambda_3\widetilde{J}_3 + \lambda_4\widetilde{J}_4\right)^4$. Indeed, there are lots of terms in the above. However, $\widetilde{J}_1, \widetilde{J}_2, \widetilde{J}_3$, and $\widetilde{J}_4$ are different only in x_t and v_t^4. The general terms of the long expression come to:

$$
\begin{aligned}
&E\left[\frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^4\varepsilon_t^4 v_t^l x_t|\Omega_{t-1}\right] \\
&E\left[\frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^4\varepsilon_t^4 v_t^l x_t^2|\Omega_{t-1}\right] \\
&E\left[\frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^4\varepsilon_t^4 v_t^l x_t^l\Omega_{t-1}\right].
\end{aligned}
$$

In section 3.1 we assume that ε_t has finite conditional fourth moment and $\xi_t = x_t\varepsilon_t$ is a m.d.s.. The first-order and second-order tv-BN decomposition, therefore, can quickly show that

$$
E\left[\frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^4\varepsilon_t^4 v_t^l x_t|\Omega_{t-1}\right] = \frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^4 v_t^l x_t E\left[\varepsilon_t^4|\Omega_{t-1}\right] \xrightarrow{p} 0 \tag{A.34}
$$

and

$$
E\left[\frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^4\varepsilon_t^4 v_t^l x_t^2|\Omega_{t-1}\right] = \frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^4\varepsilon_t^4 v_t^l x_t^2 E\left[\varepsilon_t^4|\Omega_{t-1}\right] \xrightarrow{p} 0. \tag{A.35}
$$

As a result, the first 3 elements in $left(\boldsymbol{\lambda}^{\top}\boldsymbol{a}_t(\tau)^4$ will converge to zero in probability by Slustky's theorem.

Now, we expand x_t^4 and take expectation

$$x_t^4 = \left(\sum_{j=0}^{\infty} \varphi_{j,t}\varepsilon_{t-j}\right)^4 = \sum_{j_1=0}^{\infty}\sum_{j_2=0}^{\infty}\sum_{j_3=0}^{\infty}\sum_{j_4=0}^{\infty} \varepsilon_{j_1,t}\varepsilon_{j_2,t}\varepsilon_{j_3,t}\varepsilon_{j_4,t}\varepsilon_{t-j_1}\varepsilon_{t-j_2}\varepsilon_{t-j_3}\varepsilon_{t-j_4},$$
$$E[x_t^4] = \sum_{j_1=0}^{\infty}\sum_{j_2=0}^{\infty}\sum_{j_3=0}^{\infty}\sum_{j_4=0}^{\infty} \varepsilon_{j_1,t}\varepsilon_{j_2,t}\varepsilon_{j_3,t}\varepsilon_{j_4,t}E\left[\varepsilon_{t-j_1}\varepsilon_{t-j_2}\varepsilon_{t-j_3}\varepsilon_{t-j_4}\right].$$

When all j_i are the different $j_1 \neq j_2 \neq j_3 \neq j_4$,

$$E\left[\varepsilon_{t-j_1}\varepsilon_{t-j_2}\varepsilon_{t-j_3}\varepsilon_{t-j_4}\right] = 0.$$

When one pair of j's are different with another pair, e.g. $j_1 = j_2 \neq j_3 = j_4$ or equivalent,

$$E\left[\varepsilon_{t-j_1}\varepsilon_{t-j_2}\varepsilon_{t-j_3}\varepsilon_{t-j_4}\right] = \sigma^4.$$

When one j_i are different with other equivalent ones, e.g. $j_1 = j_2 = j_3 \neq j_4$,

$$E\left[\varepsilon_{t-j_1}\varepsilon_{t-j_2}\varepsilon_{t-j_3}\varepsilon_{t-j_4}\right] = 0.$$

When all j_i are the same, $j_1 = j_2 = j_3 = j_4$,

$$E\left[\varepsilon_{t-j_1}\varepsilon_{t-j_2}\varepsilon_{t-j_3}\varepsilon_{t-j_4}\right] = E[\varepsilon_{t-j_1}^4] = \kappa_4.$$

All situations show that the x_t^4 is bounded and therefore

$$\begin{aligned}
E\left[\frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^4\varepsilon_t^4 v_t^l x_t^4 |\Omega_{t-1}\right] &= \frac{1}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^4 x_t^4 v_t^l E\left[\varepsilon_t^4|\Omega_{t-1}\right] \\
&= \frac{\mu_4}{T^2h^2}\sum_{t=1}^{\top} K_h(\tau_t-\tau)^4 v_t^l x_t^4 \\
&\xrightarrow{p} 0. \qquad \text{(A.36)}
\end{aligned}$$

Therefore, we can show that

$$\sum_{t=1}^{\top} E\left[\boldsymbol{a}_t(\tau)\boldsymbol{a}_t^\top(\tau)\boldsymbol{a}_t(\tau)\boldsymbol{a}_t^\top(\tau)|\Omega_{t-1}\right] \xrightarrow{p} 0.$$

Condition (1) and condition (2) hold, we have

$$\sqrt{Th}\boldsymbol{\lambda}^\top \widetilde{J}_T \xrightarrow{d} N\Big(0, \quad \boldsymbol{\lambda}^\top \Pi(\tau)\boldsymbol{\lambda}\Big). \tag{A.37}$$

and it suffices to show by the Cramer-Wold device

$$\sqrt{Th}\widetilde{J}_T \xrightarrow{d} N\Big(0, \quad \Pi(\tau)\Big). \tag{A.38}$$

Therefore, we have

$$\begin{aligned}\sqrt{Th}\left(\widehat{\theta}(\tau)-\theta(\tau)-\bar{B}_T-o_p(h^2)\right) &= \sqrt{Th}\widetilde{\boldsymbol{t}}_T = [\mathrm{I}_2, 0_2]S_T^{-1}\sqrt{Th}\widetilde{J}_T \\ &\xrightarrow{d} N\Big(0, \quad [\mathrm{I}_2, 0_2]\Sigma_S^{-1}(\tau)\Pi(\tau)\Sigma_S^{-1}(\tau)^\top [\mathrm{I}_2, 0_2]^\top\Big).\end{aligned}$$

■

Proof of Lemma A.3. The uniform convergence properties are standard in semiparametric regressions, see, for example, Li et al. (2011b) and Chen et al. (2012) for the proof. ■

Proof of Lemma A.4. (1) $\frac{1}{T}\left(\tilde{X}^\top\tilde{X}\right)^{-1} \xrightarrow{p} 1/\sigma_\varepsilon^2$.
(2) $\frac{1}{\sqrt{T}}\tilde{X}^\top\left(\mathbf{I}-\tilde{s}\right)\left(Z_1^\top\theta(\tau_2)), ..., Z_1^\top\theta(\tau_T)\right)^\top = o_p(1)$.
(3) $\frac{1}{\sqrt{T}}\tilde{X}^\top\left(\mathbf{I}-\tilde{s}\right)\left(\lambda-\widehat{\lambda}\right)\varepsilon = o_p(1)$.
(4) $\frac{1}{\sqrt{T}}\tilde{X}^\top\left(\mathbf{I}-\tilde{s}\right)\widehat{\lambda}\varepsilon = o_p(1)$.

Proof of (1): $\frac{1}{T}\left(\tilde{X}^\top\tilde{X}\right)^{-1} \xrightarrow{p} 1/\sigma_\varepsilon^2$.

The inverse matrix in $\mathbb{N}_\lambda(1)$ and $\mathbb{N}_\lambda(2)$ can be expanded to

$$\frac{1}{T}\tilde{X}^\top\tilde{X} = \frac{1}{T}\sum_{t=2}^{T}\left(\widehat{\varepsilon}_t - Z_{t-1}^\top s(\tau_t)\widehat{\varepsilon}\right)^2 = \mathbb{N}_{XX}(1) + \mathbb{N}_{XX}(2) + \mathbb{N}_{XX}(3).$$

where $\mathbb{N}_{XX}(1) = \frac{1}{T}\sum_{t=2}^{T}\widehat{\varepsilon}_t^2$, $\mathbb{N}_{XX}(2) = \frac{2}{T}\sum_{t=2}^{T}\widehat{\varepsilon}_t Z_{t-1}^\top s(\tau_t)\widehat{\varepsilon}$, and
$\mathbb{N}_{XX}(3) = \frac{1}{T}\sum_{t=2}^{T}\left(Z_{t-1}^\top s(\tau_t)\widehat{\varepsilon}\right)^2$. The following results

$$\mathbb{N}_{XX}(1) \xrightarrow{p} \sigma_\varepsilon^2, \quad \mathbb{N}_{XX}(2) \xrightarrow{p} 0, \quad \mathbb{N}_{XX}(3) \xrightarrow{p} 0, \tag{A.39}$$

suffice to prove $\frac{1}{T}\left(\tilde{X}^\top \tilde{X}\right)^{-1} \xrightarrow{p} 1/\sigma_\varepsilon^2$.

Consider $\mathbb{N}_{XX}(1)$,

$$
\begin{aligned}
\mathbb{N}_{XX}(1) =& \frac{1}{T}\sum_{t=2}^{T}\left(\varepsilon_t + \gamma(\tau_t) - \widehat{\gamma}(\tau_t) + x_{t-1}\left(\rho(\tau_t) - \widehat{\rho}(\tau_t)\right)\right)^2 \\
=& \frac{1}{T}\sum_{t=2}^{T}\varepsilon_t^2 + \frac{1}{T}\sum_{t=2}^{T}\left(\gamma(\tau_t) - \widehat{\gamma}(\tau_t)\right)^2 + \frac{1}{T}\sum_{t=2}^{T}x_{t-1}^2\left(\rho(\tau_t) - \widehat{\rho}(\tau_t)\right)^2 \\
&+ \frac{1}{T}\sum_{t=2}^{T}2\varepsilon_t\left(\gamma(\tau_t) - \widehat{\gamma}(\tau_t)\right) + \frac{1}{T}\sum_{t=2}^{T}2\left(\gamma(\tau_t) - \widehat{\gamma}(\tau_t)\right)\left(\rho(\tau_t) - \widehat{\rho}(\tau_t)\right) \\
&+ \frac{1}{T}\sum_{t=2}^{T}2\varepsilon_t\left(\rho(\tau_t) - \widehat{\rho}(\tau_t)\right) \\
=& \mathbb{N}_{XX}(1,1) + \mathbb{N}_{XX}(1,2) + \mathbb{N}_{XX}(1,3) + \mathbb{N}_{XX}(1,4) + \mathbb{N}_{XX}(1,5) + \mathbb{N}_{XX}(1,6).
\end{aligned}
$$

Since η_t follows m.d.s.$(0,\sigma_\varepsilon^2)$, it is obvious to have $\mathbb{N}_{XX}(1,1) = \frac{1}{T}\sum_{t=2}^{T}\varepsilon_t^2 \xrightarrow{p} 0$. We consider $\mathbb{N}_{XX}(1,3)$. Under uniform convergence XX, we have

$$
\begin{aligned}
\mathbb{N}_{XX}(1,3) &= \frac{1}{T}\sum_{t=2}^{T}\left(\gamma(\tau_t) - \widehat{\gamma}(\tau_t)\right)\left(\rho(\tau_t) - \widehat{\rho}(\tau_t)\right) \leq \sup_{0\leq\tau\leq 1}|\rho(\tau_t) - \widehat{\rho}(\tau_t)| \sup_{0\leq\tau\leq 1}|\gamma(\tau_t) - \widehat{\gamma}(\tau_t)| \\
&= o_p(1)\cdot o_p(1) \xrightarrow{p} 0.
\end{aligned}
$$

Similarly, we

$$
\mathbb{N}_{XX}(1,6) = \frac{1}{T}\sum_{t=2}^{T}2\varepsilon_t\big(\rho(\tau_t) - \widehat{\rho}(\tau_t)\big) \leq \sup_{0\leq\tau\leq 1}|\rho(\tau_t) - \widehat{\rho}(\tau_t)|\frac{1}{T}\sum_{t=2}^{T}2\varepsilon_t = O_p(1)\cdot o_p(1) \xrightarrow{p} 0.
$$

Therefore, $\mathbb{N}_{XX}(1) \xrightarrow{p} \sigma_\varepsilon^2$ holds.

Next we prove $\mathbb{N}_{XX}(2) \xrightarrow{p} 0$.

$$
\begin{aligned}
\mathbb{N}_{XX}(2) &= \frac{2}{T}\sum_{t=2}^{T}\widehat{\varepsilon}_t Z_{t-1}^\top s(\tau_t)\widehat{\varepsilon} \\
&= \frac{2}{T}\sum_{t=2}^{T}\left[\left(\varepsilon_t + \phi(\tau_t) - \widehat{\phi}(\tau_t)\right)Z_{t-1}^\top s(\tau_t)\widehat{\varepsilon}\right] \\
&= \frac{2}{T}\sum_{t=2}^{T}\varepsilon_t Z_{t-1}^\top s(\tau_t)\widehat{\varepsilon} + \frac{2}{T}\sum_{t=2}^{T}\left(\phi(\tau_t) - \widehat{\phi}(\tau_t)\right),
\end{aligned}
$$

where we have the following term shown in the both parts

$$
\begin{aligned}
Z_{t-1}^{\top} s(\tau_t)\widehat{\varepsilon} =& Z_{t-1}^{\top} [\mathbf{I}, \mathbf{0}] \left(D(\tau)^{\top} K(\tau) D(\tau) \right)^{-1} D(\tau)^{\top} K(\tau)\widehat{\varepsilon} \\
=& Z_{t-1}^{\top} [\mathbf{I}, \mathbf{0}] \left(\frac{1}{T} D(\tau)^{\top} K(\tau) D(\tau) \right)^{-1} \frac{1}{T} D(\tau)^{\top} K(\tau)\widehat{\varepsilon}.
\end{aligned}
$$

From the previous results, we have

$$
\frac{1}{T} D(\tau)^{\top} K(\tau) D(\tau) = \frac{1}{T} \sum_{t=2}^{T} K\left(\frac{\tau_t - \tau}{h} \right) \begin{bmatrix} Z_{t-1} \\ Z_{t-1} \frac{\tau_t - \tau}{h} \end{bmatrix} \begin{bmatrix} Z_{t-1}^{\top} & Z_{t-1}^{\top} \frac{\tau_t - \tau}{h} \end{bmatrix} \xrightarrow{p} \mathbf{\Sigma}_S(\tau).
$$

It can be shown that

$$
\begin{aligned}
\frac{1}{T} D(\tau)^{\top} K(\tau)\widehat{\varepsilon} =& \frac{1}{T} \sum_{t=2}^{T} K\left(\frac{\tau_t - \tau}{h} \right) \begin{bmatrix} Z_{t-1} \\ Z_{t-1} \frac{\tau_t - \tau}{h} \end{bmatrix} \Big(\varepsilon_t + \gamma(\tau_t) - \widehat{\gamma}(\tau_t) + x_{t-1} \left(\rho(\tau_t) - \widehat{\rho}(\tau_t) \right) \Big) \\
=& \frac{1}{T} \sum_{t=2}^{T} K_h(\tau_t - \tau) \begin{bmatrix} Z_{t-1} \\ Z_{t-1} \frac{\tau_t - \tau}{h} \end{bmatrix} \varepsilon_t \\
&+ \frac{1}{T} \sum_{t=2}^{T} K_h(\tau_t - \tau) \begin{bmatrix} Z_{t-1} \\ Z_{t-1} \frac{\tau_t - \tau}{h} \end{bmatrix} \left(\gamma(\tau_t) - \widehat{\gamma}(\tau_t) \right) \\
&+ \frac{1}{T} \sum_{t=2}^{T} K_h(\tau_t - \tau) \begin{bmatrix} Z_{t-1} \\ Z_{t-1} \frac{\tau_t - \tau}{h} \end{bmatrix} x_{t-1} \left(\rho(\tau_t) - \widehat{\rho}(\tau_t) \right) \\
&= \mathbb{N}_{XX}(2,1) + \mathbb{N}_{XX}(2,2) + \mathbb{N}_{XX}(2,3) \xrightarrow{p} 0,
\end{aligned}
$$

where it suffices for us to show that

$$
\mathbb{N}_{XX}(2,1) \xrightarrow{p} 0, \quad \mathbb{N}_{XX}(2,2) \xrightarrow{p} 0 \quad \mathbb{N}_{XX}(2,3) \xrightarrow{p} 0. \tag{A.40}
$$

Consider $\mathbb{N}_{XX}(2,1)$,

$$\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)\begin{bmatrix} Z_{t-1} \\ Z_{t-1}\frac{\tau_t-\tau}{h} \end{bmatrix}\varepsilon_t = \begin{bmatrix} \frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)\varepsilon_t \\ \frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)x_{t-1}\varepsilon_t \\ \frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t\varepsilon_t \\ \frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)x_{t-1}v_t\varepsilon_t \end{bmatrix}.$$

We have $\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)\varepsilon_t \xrightarrow{p} 0$ due to

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)\varepsilon_t\right] =& 0 \\ E\left[\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)\varepsilon_t\right]^2 =& \frac{1}{T^2}\sum_{t=2}^{T} K_h(\tau_t-\tau)^2\sigma_\varepsilon^2 \approx \frac{1}{T}\sigma_{l=0}^2\sigma_\varepsilon^2 \longrightarrow 0, \end{aligned}$$

$\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)x_{t-1}\varepsilon_t \xrightarrow{p} 0$ due to

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)x_{t-1}\varepsilon_t\right] =& 0, \\ E\left[\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)x_{t-1}\varepsilon_t\right]^2 =& \frac{1}{T^2}\sum_{t=2}^{T} K_h(\tau_t-\tau)^2x_{t-1}^2\sigma_\varepsilon^2 \approx \frac{1}{T}A_{T\tau}^2\sigma_l^2\sigma_\varepsilon^2 \longrightarrow 0, \end{aligned}$$

$\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t\varepsilon_t \xrightarrow{p} 0$ due to

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t\varepsilon_t\right] =& 0, \\ E\left[\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t\varepsilon_t\right]^2 =& \frac{1}{T^2}\sum_{t=2}^{T} K_h(\tau_t-\tau)^2v_t^2\sigma_\varepsilon^2 \approx \frac{1}{T}\sigma_{l=1}^2\sigma_\varepsilon^2 \longrightarrow 0, \end{aligned}$$

$\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)x_{t-1}v_t\varepsilon_t \xrightarrow{p} 0$ due to

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)x_{t-1}\varepsilon_t\right] =& 0, \\ E\left[\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)x_{t-1}\varepsilon_t\right]^2 =& \frac{1}{T^2}\sum_{t=2}^{T} K_h(\tau_t-\tau)^2x_{t-1}^2\sigma_\varepsilon^2 \approx \frac{1}{T}A_{T\tau}^2\sigma_l^2\sigma_\varepsilon^2 \longrightarrow 0. \end{aligned}$$

Consider $\mathbb{N}_{XX}(2,2)$,

$$\frac{1}{T}\sum_{t=2}^{T}K_h(\tau_t-\tau)\begin{bmatrix} Z_{t-1} \\ Z_{t-1}\frac{\tau_t-\tau}{h}\end{bmatrix}(\gamma(\tau_t)-\widehat{\gamma}(\tau_t))=\begin{bmatrix}\frac{1}{T}\sum_{t=2}^{T}K_h(\tau_t-\tau)\,(\gamma(\tau_t)-\widehat{\gamma}(\tau_t)) \\ \frac{1}{T}\sum_{t=2}^{T}K_h(\tau_t-\tau)x_{t-1}\,(\gamma(\tau_t)-\widehat{\gamma}(\tau_t)) \\ \frac{1}{T}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t\,(\gamma(\tau_t)-\widehat{\gamma}(\tau_t)) \\ \frac{1}{T}\sum_{t=2}^{T}K_h(\tau_t-\tau)x_{t-1}v_t\,(\gamma(\tau_t)-\widehat{\gamma}(\tau_t))\end{bmatrix}.$$

Since $\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^l\longrightarrow\tilde{\sigma}_l<\infty$ and $\frac{1}{Th}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t^lx_t\xrightarrow{p}A_{T\tau}\sigma_l<\infty$, we have

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T}K_h(\tau_t-\tau)\,(\gamma(\tau_t)-\widehat{\gamma}(\tau_t))&\leq\sup_{0\leq\tau\leq1}|\gamma(\tau)-\widehat{\gamma}(\tau)|\frac{1}{T}\sum_{t=2}^{T}|K_h(\tau_t-\tau)|\\&=o_p(1)\cdot O_p(1)\\&=o_p(1).\end{aligned}$$

Similarly, by the uniform convergence lemma, it is straightforward to have the following results,

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T}K_h(\tau_t-\tau)x_{t-1}\,(\gamma(\tau_t)-\widehat{\gamma}(\tau_t))&=o_p(1)\\\frac{1}{T}\sum_{t=2}^{T}K_h(\tau_t-\tau)v_t\,(\gamma(\tau_t)-\widehat{\gamma}(\tau_t))&=o_p(1)\\\frac{1}{T}\sum_{t=2}^{T}K_h(\tau_t-\tau)x_{t-1}v_t\,(\gamma(\tau_t)-\widehat{\gamma}(\tau_t))&=o_p(1).\end{aligned}$$

Consider $\mathbb{N}_{XX}(2,3)$,

$$\begin{aligned}\mathbb{N}_{XX}(2,3)&=\frac{1}{T}\sum_{t=2}^{T}K_h(\tau_t-\tau)\begin{bmatrix} Z_{t-1} \\ Z_{t-1}\frac{\tau_t-\tau}{h}\end{bmatrix}x_{t-1}\,(\rho(\tau_t)-\widehat{\rho}(\tau_t))\\&=\begin{bmatrix}\frac{1}{T}\sum_{t=2}^{T}K_h(\tau_t-\tau)x_{t-1}\,(\rho(\tau_t)-\widehat{\rho}(\tau_t)) \\ \frac{1}{T}\sum_{t=2}^{T}K_h(\tau_t-\tau)x_{t-1}^2\,(\rho(\tau_t)-\widehat{\rho}(\tau_t)) \\ \frac{1}{T}\sum_{t=2}^{T}K_h(\tau_t-\tau)x_{t-1}v_t\,(\rho(\tau_t)-\widehat{\rho}(\tau_t)) \\ \frac{1}{T}\sum_{t=2}^{T}K_h(\tau_t-\tau)x_{t-1}^2v_t\,(\rho(\tau_t)-\widehat{\rho}(\tau_t))\end{bmatrix}.\end{aligned}$$

By the uniform convergence lemma, we have
$\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)x_{t-1}\left(\rho(\tau_t)-\widehat{\rho}(\tau_t)\right) \leq \sup_{0\leq\tau\leq 1}|\rho(\tau)-\widehat{\rho}(\tau)|\frac{1}{T}\sum_{t=2}^{T}|K_h(\tau_t-\tau)x_{t-1}| = o_p(1)$
$\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)x_{t-1}^2\left(\rho(\tau_t)-\widehat{\rho}(\tau_t)\right) \leq \sup_{0\leq\tau\leq 1}|\rho(\tau)-\widehat{\rho}(\tau)|\frac{1}{T}\sum_{t=2}^{T}|K_h(\tau_t-\tau)x_{t-1}^2| = o_p(1)$,
$\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)v_t x_{t-1}\left(\rho(\tau_t)-\widehat{\rho}(\tau_t)\right) \leq \sup_{0\leq\tau\leq 1}|\rho(\tau)-\widehat{\rho}(\tau)|\frac{1}{T}\sum_{t=2}^{T}|K_h(\tau_t-\tau)v_t x_{t-1}| = o_p(1)$,
$\frac{1}{T}\sum_{t=2}^{T} K_h(\tau_t-\tau)x_{t-1}^2 v_t\left(\rho(\tau_t)-\widehat{\rho}(\tau_t)\right) \leq \sup_{0\leq\tau\leq 1}|\rho(\tau)-\widehat{\rho}(\tau)|\frac{1}{T}\sum_{t=2}^{T}|K_h(\tau_t-\tau)x_{t-1}^2 v_t| = o_p(1)$.
According to the above results,

$$\frac{1}{T}\left(D(\tau)^\top K(\tau)D(\tau)\right)^{-1} \xrightarrow{p} \mathbf{\Sigma}_S(\tau)^{-1} \quad \text{and} \quad \frac{1}{T}D(\tau)^\top K(\tau)\widehat{\varepsilon} \xrightarrow{p} 0,$$

we then obtain

$$S(\tau_t)\widehat{\varepsilon} = \left(D(\tau)^\top K(\tau)D(\tau)\right)^{-1} D(\tau)^\top K(\tau)\widehat{\varepsilon} \xrightarrow{p} 0. \tag{A.41}$$

Recall that

$$\frac{2}{T}\sum_{t=2}^{T}\widehat{\varepsilon}_t Z_{t-1}^\top S(\tau_t)\widehat{\varepsilon} = \frac{2}{T}\sum_{t=2}^{T}\left[\varepsilon_t Z_{t-1}^\top S(\tau_t)\widehat{\varepsilon}\right] + \frac{2}{T}\sum_{t=2}^{T}\left[\left(\phi(\tau_t)-\widehat{\phi}(\tau_t)\right) Z_{t-1}^\top s(\tau_t)\widehat{\varepsilon}\right], \tag{A.42}$$

we have

$$\begin{aligned}\frac{2}{T}\sum_{t=2}^{T}\varepsilon_t Z_{t-1}^\top s(\tau_t)\widehat{\varepsilon} \leq \frac{2}{T}\sum_{t=2}^{T}|\varepsilon_t Z_{t-1}^\top s(\tau_t)\widehat{\varepsilon}| &\leq \frac{2}{T}\sum_{t=2}^{T}|\varepsilon_t Z_{t-1}^\top| \sup_{0\leq\tau\leq 1}|s(\tau)\widehat{\varepsilon}| \\ &= O_p(1)\cdot o_p(1).\end{aligned}$$

and

$$
\begin{aligned}
\frac{2}{T}\sum_{t=2}^{T}\left[\left(\phi(\tau_t)-\widehat{\phi}(\tau_t)\right)Z_{t-1}^{\top}S(\tau_t)\widehat{\varepsilon}\right] &\leq \frac{2}{T}\sum_{t=2}^{T}|Z_{t-1}^{\top}|\sup_{0\leq\tau\leq 1}|\phi(\tau_t)-\widehat{\phi}(\tau_t)|\sup_{0\leq\tau\leq 1}|S(\tau)\widehat{\varepsilon}| \\
&= O_p(1)\cdot o_p(1)\cdot o_p(1) \\
&\xrightarrow{p} 0.
\end{aligned}
$$

since obviously $\frac{2}{T}\sum_{t=2}^{T}|\varepsilon_t Z_{t-1}^{\top}| \xrightarrow{p} 0$ and $\frac{2}{T}\sum_{t=2}^{T}|Z_{t-1}^{\top}| \xrightarrow{p} 0$. Therefore, we have the following result

$$
\mathbb{N}_{XX}(2) = \frac{2}{T}\sum_{t=2}^{T}\varepsilon_t Z_{t-1}^{\top}S(\tau_t)\widehat{\varepsilon} \xrightarrow{p} 0. \tag{A.43}
$$

Now we prove $\mathbb{N}_{XX}(3) = \frac{1}{T}\sum_{t=2}^{T}\left(Z_{t-1}^{\top}S(\tau_t)\widehat{\varepsilon}\right)^2 \xrightarrow{p} 0$.

$$
\begin{aligned}
\mathbb{N}_{XX}(3) &= \frac{1}{T}\sum_{t=2}^{T}\left(Z_{t-1}^{\top}S(\tau_t)\widehat{\varepsilon}\right)^2 \\
&= \frac{1}{T}\sum_{t=2}^{T}\begin{bmatrix} 1 & x_{t-1} \\ x_{t-1} & x_{t-1}^2 \end{bmatrix}\cdot o_p(1)\cdot o_p(1) \\
&\xrightarrow{p} 0.
\end{aligned} \tag{A.44}
$$

Therefore, the lemma $\left(\frac{1}{T}\tilde{X}^{\top}\tilde{X}\right)^{-1} \xrightarrow{p} \frac{1}{\sigma_{\varepsilon}^2}$ holds.

We now prove: (2) $\frac{1}{\sqrt{T}}\tilde{X}^{\top}\left(\mathbf{I}-\tilde{s}\right)\left(Z_1^{\top}\theta(\tau_2)\right),...,Z_1^{\top}\theta(\tau_T)$.

$$
\begin{aligned}
&\frac{1}{T}\tilde{X}^{\top}\left(\mathbf{I}-\tilde{s}\right)\left(Z_1^{\top}\theta(\tau_2)),...,Z_1^{\top}\theta(\tau_T)\right)^{\top} \\
&\qquad = \frac{1}{T}\sum_{t=2}^{T}\left(\widehat{\varepsilon}_t - Z_{t-1}^{\top}S(\tau_t)\widehat{\varepsilon}\right)\left(Z_{t-1}^{\top}\theta(\tau_t) - Z_{t-1}^{\top}S(\tau_t)\left(Z_1^{\top}\theta(\tau_2)),...,Z_{T-1}^{\top}\theta(\tau_T)\right)^{\top}\right).
\end{aligned}
$$

Expanding each element in $\left(Z_1^\top\theta(\tau_2)),...,Z_1^\top\theta(\tau_T)\right)^\top$ at τ_t using Taylor series, then

$$\begin{aligned}
\begin{bmatrix} Z_1^\top\theta(\tau_2)) \\ \vdots \\ Z_{T-1}^\top\theta(\tau_T)) \end{bmatrix} &= \begin{bmatrix} Z_1' & Z_2'\left(\frac{\tau_1-\tau_t}{h}\right) \\ \vdots & \vdots \\ Z_t' & Z_t'\left(\frac{\tau_T-\tau_t}{h}\right) \end{bmatrix} \begin{bmatrix} \theta(\tau_t) \\ h\theta(\tau_t)' \end{bmatrix} + \begin{bmatrix} Z_1'\left(\frac{\tau_1-\tau_t}{h}\right)^2 \\ \vdots \\ Z_t'\left(\frac{\tau_T-\tau_t}{h}\right)^2 \end{bmatrix} \left[\frac{h^2}{2}\theta(\tau_t)'' + o(h^2)\right] \\
&= D(\tau_t) \begin{bmatrix} \theta(\tau_t) \\ h\theta(\tau_t)' \end{bmatrix} + \begin{bmatrix} Z_1'\left(\frac{\tau_2-\tau_t}{h}\right)^2 \\ \vdots \\ Z_t'\left(\frac{\tau_T-\tau_t}{h}\right)^2 \end{bmatrix} \left[\frac{h^2}{2}\theta(\tau_t)'' + o(h^2)\right].
\end{aligned}$$

Therefore

$$\begin{bmatrix} Z_1^\top\theta(\tau_2)) \\ \vdots \\ Z_{T-1}^\top\theta(\tau_T)) \end{bmatrix} = D(\tau_t) \begin{bmatrix} \theta(\tau_t) \\ h\theta(\tau_t)' \end{bmatrix} + \widetilde{B}(\tau_t) + o(h^2).$$

So,

$$\begin{aligned}
&Z_{t-1}^\top\theta(\tau_t) - Z_{t-1}^\top S(\tau_t)\left(Z_1^\top\theta(\tau_2)),...,Z_{T-1}^\top\theta(\tau_T)\right)^\top \\
=&Z_{t-1}^\top\theta(\tau_t) - Z_{t-1}^\top S(\tau_t)D(\tau_t) \begin{bmatrix} \theta(\tau_t) \\ h\theta(\tau_t)' \end{bmatrix} - Z_{t-1}^\top S(\tau_t)\widetilde{B}(\tau_t) - o(h^2) \\
=&Z_{t-1}^\top\theta(\tau_t) - Z_{t-1}^\top\theta(\tau_t) - \widetilde{B}_\theta(\tau_t) - o(h^2) \\
=&-\widetilde{B}_\theta(\tau_t) - o(h^2),
\end{aligned}$$

where the new bias term is define as $\widetilde{B}_\theta(\tau_t) = Z_{t-1}^\top\theta(\tau_t) \begin{bmatrix} Z_1'\left(\frac{\tau_2-\tau}{h}\right)^2 \\ \vdots \\ Z_t'\left(\frac{\tau_T-\tau}{h}\right)^2 \end{bmatrix} \left[\frac{h^2}{2}\theta(\tau)'' + o(h^2)\right].$

Then we can show

$$
\begin{aligned}
&\frac{\sqrt{T}}{T}\tilde{X}^\top\left(\mathbf{I}-\tilde{s}\right)\left(Z_1^\top\theta(\tau_2)),...,Z_1^\top\theta(\tau_T)\right)^\top \\
&\qquad=\frac{1}{\sqrt{T}}\tilde{X}^\top\left(\mathbf{I}-\tilde{s}\right)\left(Z_1^\top\theta(\tau_2)),...,Z_1^\top\theta(\tau_T)\right)^\top \\
&\qquad=\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\left(\widehat{\varepsilon}_t-Z_{t-1}^\top S(\tau_t)\widehat{\varepsilon}\right)\left(\widetilde{B}_\theta(\tau_t)-o(h^2)\right) \\
&\qquad=\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\widehat{\varepsilon}_t\widetilde{B}_\theta(\tau_t)-\frac{1}{\sqrt{T}}\sum_{t=2}^{T}Z_{t-1}^\top S(\tau_t)\widehat{\varepsilon}\widetilde{B}_\theta(\tau_t)-o(h^2) \\
&\qquad=o_p(1).
\end{aligned}
$$

■

Proof of Lemma A.5. Since we assume $\eta_t \sim i.i.d.(0,\sigma_\eta^2)$, the CLT will be produced from the following term

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\tilde{X}^\top\left(\mathbf{I}-\tilde{s}\right)\eta &= \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\left(\widehat{\varepsilon}_t-Z_{t-1}^\top S(\tau_t)\widehat{\varepsilon}\right)\left(\eta_t-Z_{t-1}^\top S(\tau_t)\eta\right) \\
&= \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\widehat{\varepsilon}_t\eta_t-\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\widehat{\varepsilon}_t Z_{t-1}^\top S(\tau_t)\eta \\
&\quad-\frac{1}{\sqrt{T}}\sum_{t=2}^{T}Z_{t-1}^\top\eta_t S(\tau_t)\widehat{\varepsilon}+\frac{1}{\sqrt{T}}\sum_{t=2}^{T}Z_{t-1}^\top S(\tau_t)\widehat{\varepsilon}Z_{t-1}^\top S(\tau_t)\eta \\
&= \mathbb{N}_\lambda(2,1)-\mathbb{N}_\lambda(2,2)-\mathbb{N}_\lambda(2,3)-\mathbb{N}_\lambda(2,4).
\end{aligned}
$$

We will show that $\mathbb{N}_\lambda(2,1)$ is the leading term that produces the CLT and all remaining terms converge to zero in probability.

$$
\begin{aligned}
\mathbb{N}_\lambda(2,2) &= \frac{1}{T}\sum_{t=2}^{T}\widehat{\varepsilon}_t Z_{t-1}^\top S(\tau_t)\eta \\
&= \frac{1}{T}\sum_{t=2}^{T}\left(\varepsilon_t+Z_{t-1}^\top\left(\phi(\tau_t)-\widehat{\phi}(\tau_t)\right)\right)Z_{t-1}^\top S(\tau_t)\eta \\
&= \frac{1}{T}\sum_{t=2}^{T}\varepsilon_t Z_{t-1}^\top S(\tau_t)\eta+\frac{1}{T}\sum_{t=2}^{T}Z_{t-1}^\top\left(\phi(\tau_t)-\widehat{\phi}(\tau_t)\right)Z_{t-1}^\top S(\tau_t)\eta \\
&\xrightarrow{p} 0.
\end{aligned}
$$

By Lemma A.3, $\mathbb{N}_\lambda(2,3) \xrightarrow{p} 0$ and $\mathbb{N}_\lambda(2,4) \xrightarrow{p} 0$.

$$\begin{aligned}\mathbb{N}_\lambda(2,1) = \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\widehat{\varepsilon}_t\eta_t =& \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\eta_t\left(\varepsilon_t + Z_{t-1}^\top\left(\phi(\tau_t) - \widehat{\phi}(\tau_t)\right)\right)\\ =& \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\eta_t\varepsilon_t + \frac{1}{\sqrt{T}}\sum_{t=2}^{T}\eta_t Z_{t-1}^\top\left(\phi(\tau_t) - \widehat{\phi}(\tau_t)\right).\end{aligned}$$

Multiplying $\sqrt{T}$ on the both sides and we obtain

$$\begin{aligned}\sqrt{T}\frac{1}{T}\mathbb{N}_\lambda(2,1) =& \sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\widehat{\varepsilon}_t\eta_t\\ =& \sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\eta_t\varepsilon_t + \sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\eta_t Z_{t-1}^\top\left(\phi(\tau_t) - \widehat{\phi}(\tau_t)\right)\\ =& \mathbb{N}_\lambda(2,1,1) + \mathbb{N}_\lambda(2,1,2).\end{aligned}$$

We can quickly prove, in the same manner with the previous proof, that

$$\begin{aligned}\mathbb{N}_\lambda(2,1,2) &= \sqrt{T}\frac{1}{T}\sum_{t=2}^{T}\eta_t Z_{t-1}^\top\left(\phi(\tau_t) - \widehat{\phi}(\tau_t)\right)\\ &\leq \frac{1}{\sqrt{T}}\sum_{t=2}^{T}|\eta_t Z_{t-1}^\top| \sup_{0\leq\tau\leq 1}|\phi(\tau_t) - \widehat{\phi}(\tau_t)|\\ &= O_p(1)\cdot o_p(1) \xrightarrow{p} 0.\end{aligned}$$

We denote $e_t = \eta_t\varepsilon_t$, where ε is a m.d.s. and η_t is i.i.d.

$$E\left[\mathbb{N}_\lambda(2,1,1)\right] = E\left[\frac{1}{\sqrt{T}}\sum_{t=2}^{T}e_t\right] = \frac{1}{\sqrt{T}}\sum_{t=2}^{T}E\left[\eta_t\right]E\left[\varepsilon_t\right] = 0.$$

The variance is

$$\begin{aligned}
\operatorname{var}\left(\sqrt{T}\frac{1}{T}\sum_{t=2}^{T} e_t\right) &= E\left[\left(\sqrt{T}\frac{1}{T}\sum_{t=2}^{T} e_t\right)^2\right] - \left[E\left(\sqrt{T}\frac{1}{T}\sum_{t=2}^{T} e_t\right)\right]^2 \\
&= \frac{1}{T}\left[\sum_{t=2}^{T} E\left(\eta_t^2\varepsilon_t^2\right) + 2\sum_{s=2}^{T}\sum_{t\neq s}^{T} E\left(\eta_s\varepsilon_s\eta_t\varepsilon_t\right)\right] - 0 \\
&= \frac{1}{T}\sum_{t=2}^{T}\sigma_\eta^2\sigma_\varepsilon^2 \\
&= \sigma_\eta^2\sigma_\varepsilon^2,
\end{aligned}$$

since $\frac{T-1}{T} \to 1$ as T goes to infinity.

Thus we have the asymptotic distribution of $e_t = \eta_t\varepsilon_t$

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=2}^{T}\eta_t\varepsilon_t\right) \xrightarrow{d} N\left(0, \quad \sigma_\eta^2\sigma_\varepsilon^2\right). \tag{A.45}$$

■

Appendix B

Additional Results

B.1 Rolling window estimation

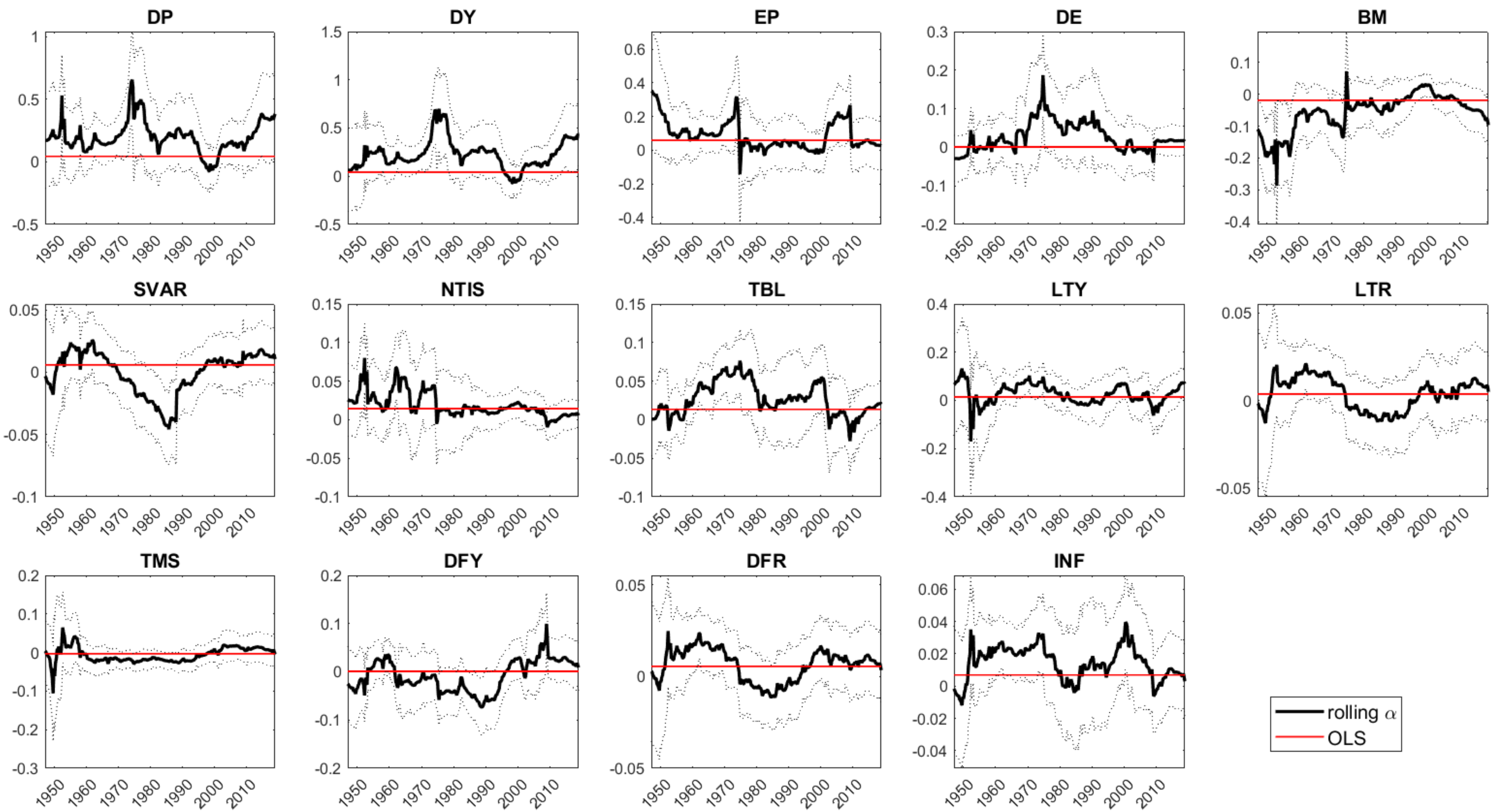
The full empirical rolling estimates of the intercept α and slope β in the predictive regression model for all 14 predictors are reported in Figure B.1) and Figure B.2, respectively. The gradual or structural changes of coefficients mainly happen in the oil shock period of 1973–1974, roughly from 1995 to 2000 near the dot.com bubble and the 2007-2009 global financial crisis. Not surprisingly, there are strong structural instabilities in both intercept and the forecasting relationships between the equity premium and 14 economic variables, complementing recent empirical evidence of structural breaks in individual equity premium predictive regression models (Paye and Timmermann, 2006; Rapach et al., 2010). The results, in general, indicate time-varying forecasting relationship between the expected equity premium and the commonly used predictors, and, therefore, suggest the failure of using the traditional models with constant coefficients.

Some existing studies have documented the instability of forecasting, but a few attention has been paid to the instability of autoregressive processes for the predictors. We plot the 20-years rolling estimation results with quarterly data from 1927:I to 2018:IV using the traditional constant-parameter AR model for all 14 predictors in Figure B.3 and Figure B.4, respectively. Again, we find that the parameter instability exists not only in the predictive models but also in the AR(1) models. The time-variation in persistence for these predictors largely concentrate in the economic recessions: the oil shock of 1973–1974, the dot.com bubble of the late 1990s, and the 2007-2009 global financial crisis.

B.2 Sensitivity analysis of bandwidth selection

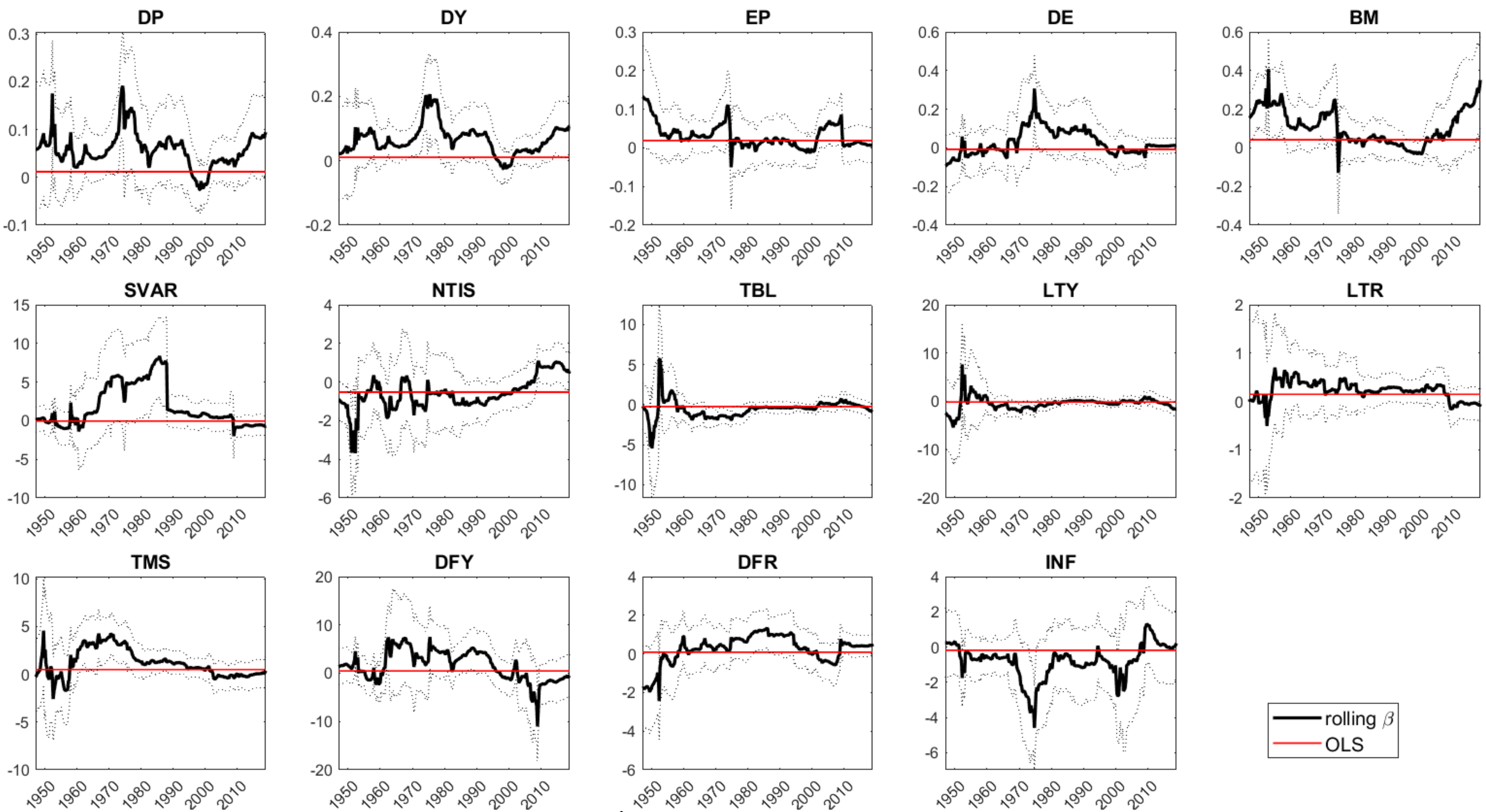
Rule-of-thumb bandwidth will give us reasonable benchmark choices. For example, the Silverman's rule-of-thumb bandwidth is given by, $h_{rot} \approx 1.06 \cdot \widehat{\sigma} T^{-1/5}$ in which $\widehat{\sigma}$ is the residual from regressions. Since minimizing the approximated mean integrated squared error (AMISE) with respect to bandwidth parameter h, the first-order-condition for h yield the AMISE optimal bandwidth $h_{op} \sim T^{-1/5}$ where T is the sample size. We use $T^{-1/5}$ as a benchmark bandwidth or starting point and also use some neighborhoods of the benchmark choice: $0.1 \cdot T^{-1/5}$, $0.3 \cdot T^{-1/5}$, $0.5 \cdot T^{-1/5}$, $0.8 \cdot T^{-1/5}$, $T^{-1/5}$, $1.2 \cdot T^{-1/5}$, and $1.5 \cdot T^{-1/5}$. Figure B.5 and Figure B.7 plot the local linear estimates of the time-varying functions γ_t and α_t, in the time-varying AR model and the time-varying predictive regression model respectively over the full sample length. As we can see clearly from the two graphs, the local linear estimates are very close to each other when we choose the bandwidths bigger than $h = 0.5 \cdot T^{-1/5}$. So we can pick a bandwidth $h = \cdot T^{-1/5}$, since the estimates are insensitive to this bandwidth choice.

Figure B.1: 20-years Rolling Estimates of α on Predictive Model $r_t = \alpha + \beta x_{t-1} + e_t$ for the 14 Predictors, 1947:I to 2018:IV.



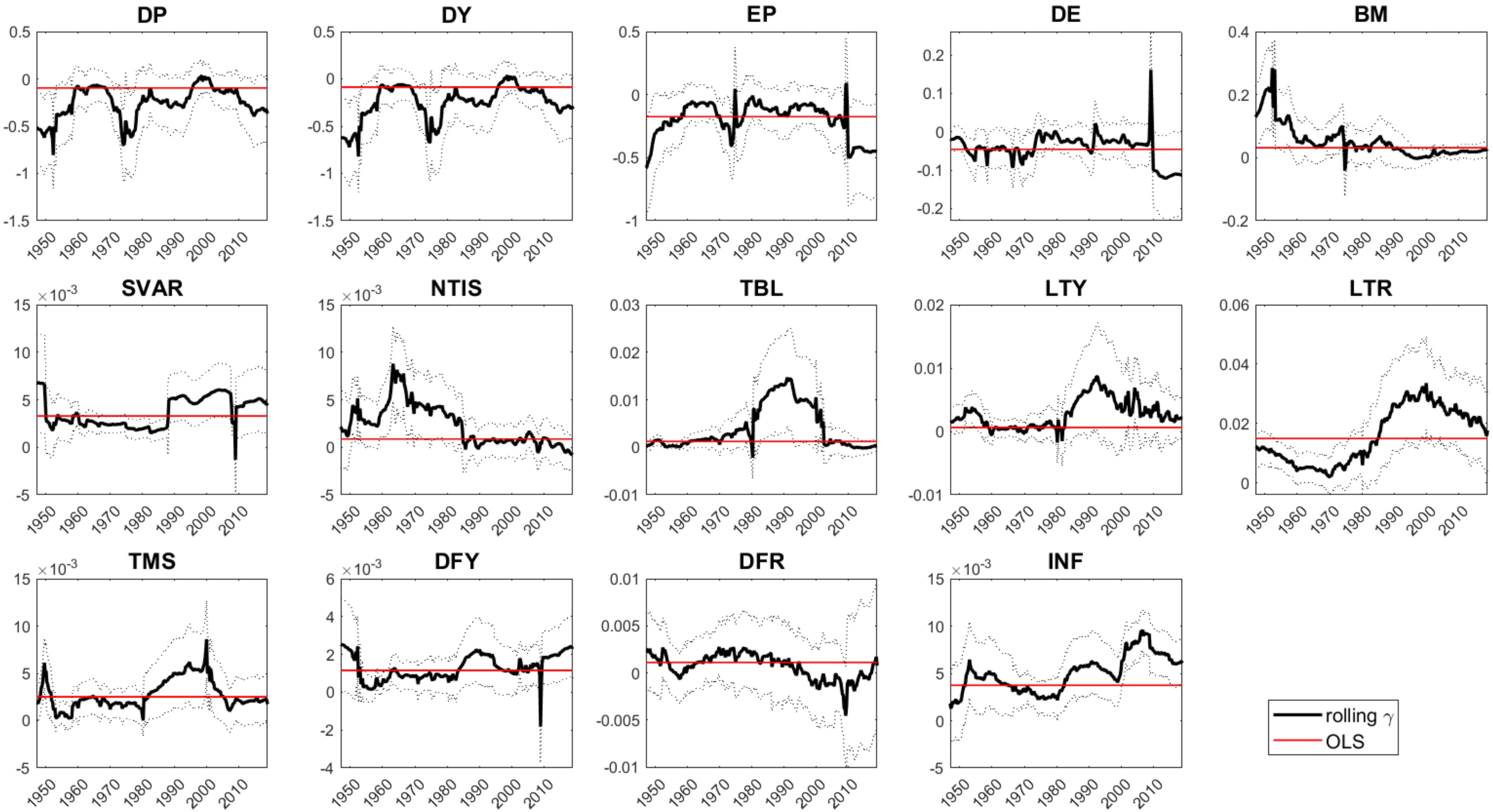
Note. This figure shows the time-varying local linear estimates $\widehat{\alpha}(\tau_t)$ in the univariate predictive regression model $r_t = \alpha + \beta x_{t-1} + e_t$ with 95% confidence bands assuming that x_t is stationary compared with the OLS estimates using the full sample for all predictors. Data are quarterly from 1927:I to 2018:IV.

Figure B.2: 20-years Rolling Estimates of β on Predictive Model $r_t = \alpha + \beta x_{t-1} + e_t$ for all 14 predictors



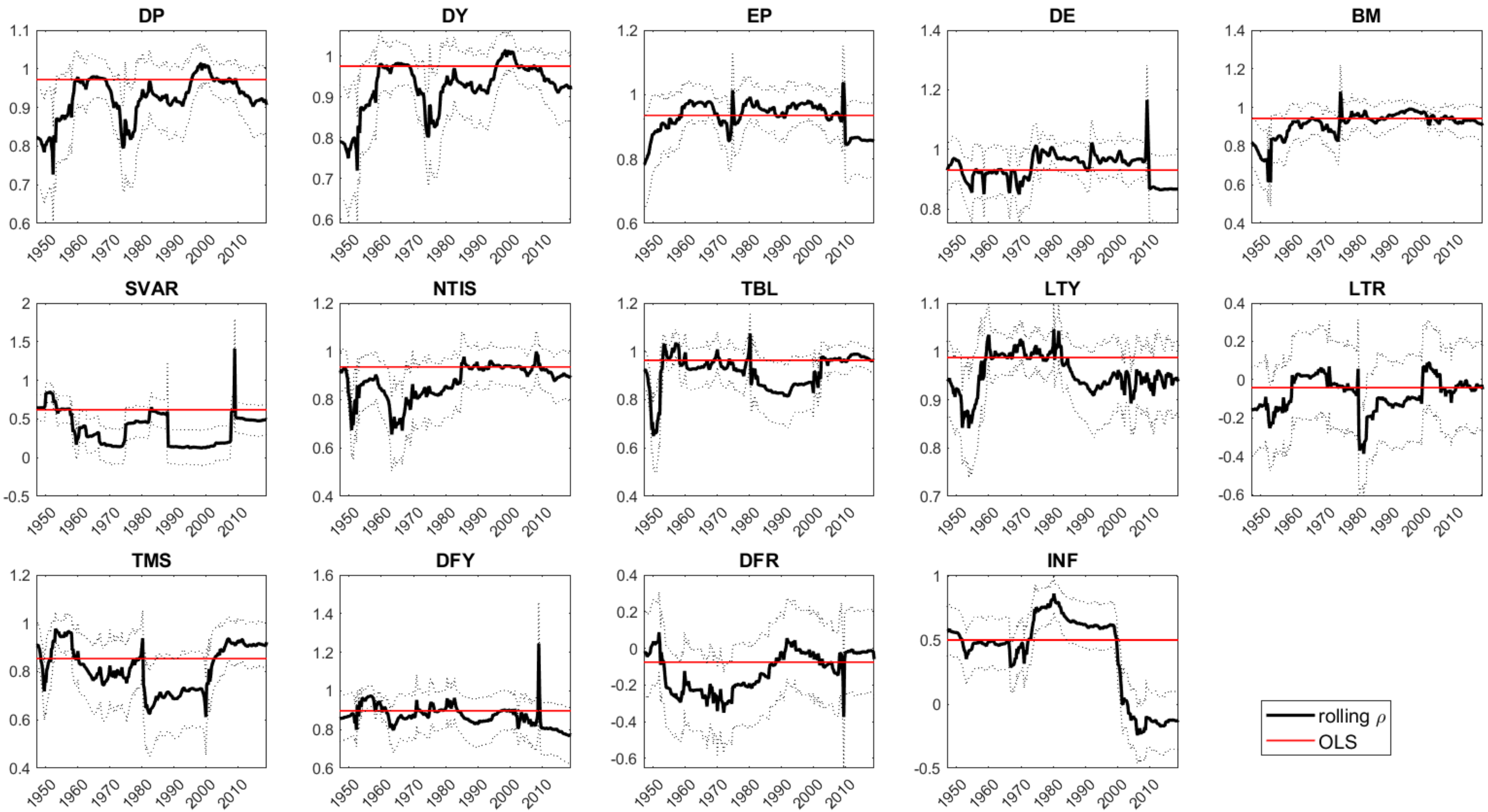
Note. This figure the 20-years rolling OLS estimate $\widehat{\beta}$ in the univariate predictive regression model $r_t = \alpha + \beta x_{t-1} + e_t$ with 95% confidence bands assuming that x_t is stationary compared and the OLS estimates using the full sample for all predictors. Data are quarterly from 1927:I to 2018:IV.

Figure B.3: 20-years rolling estimates of γ on predictive model $x_t = \gamma + \rho x_{t-1} + e_t$ for all predictors, 1947:I to 2018:IV.



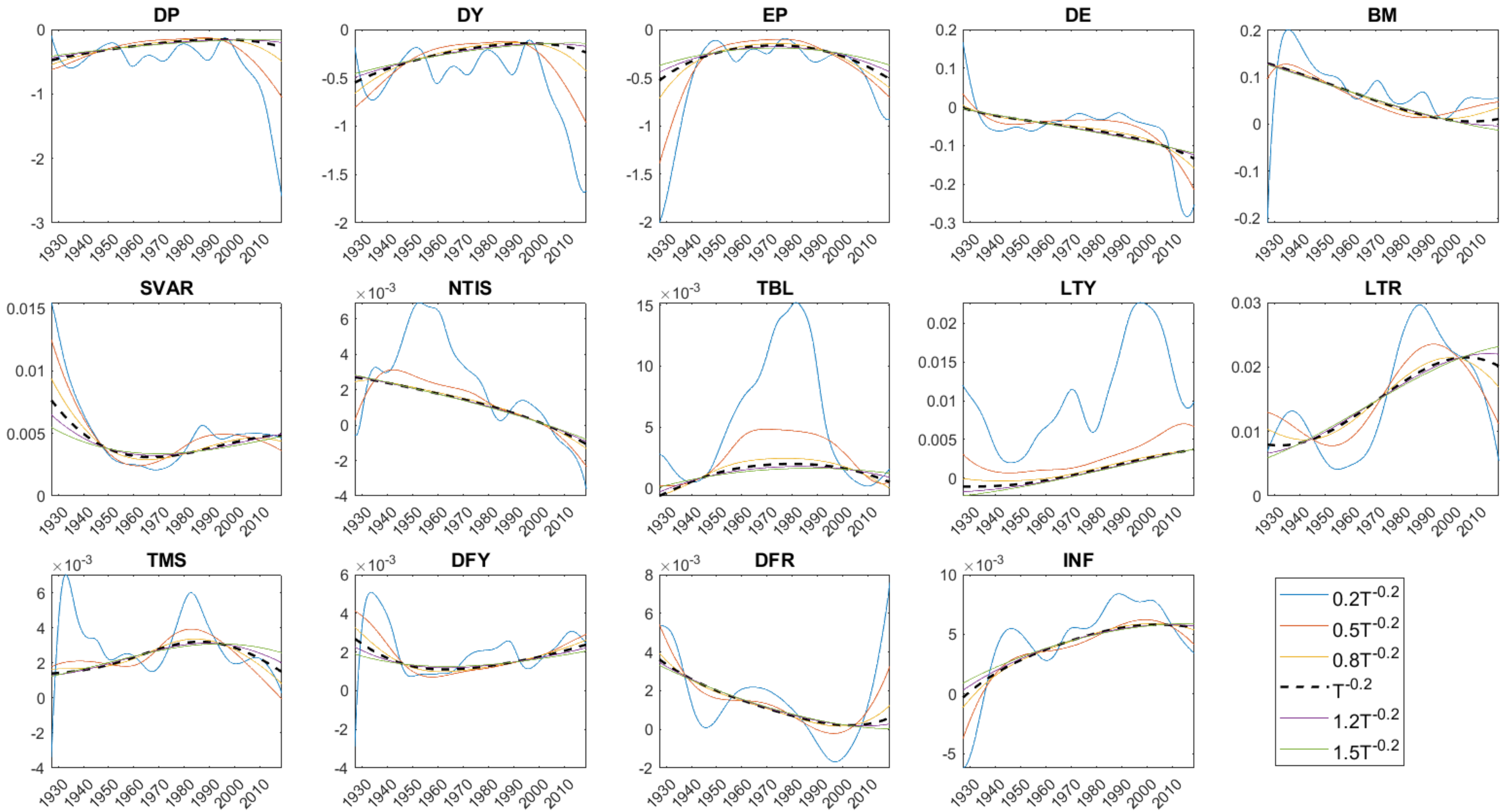
The figure the 20-years rolling OLS estimate $\widehat{\gamma}$ in the univariate AR regression model $x_t = \gamma + \rho x_{t-1} + e_t$ with 95% confidence bands assuming that x_t is stationary compared and the OLS estimates using the full sample for all predictors. Data are quarterly from 1927:I to 2018:IV.

Figure B.4: 20-years Rolling Estimates of ρ on Predictive Model $x_t = \gamma + \rho x_{t-1} + e_t$ for the 14 Predictors, 1947:I to 2018:IV.



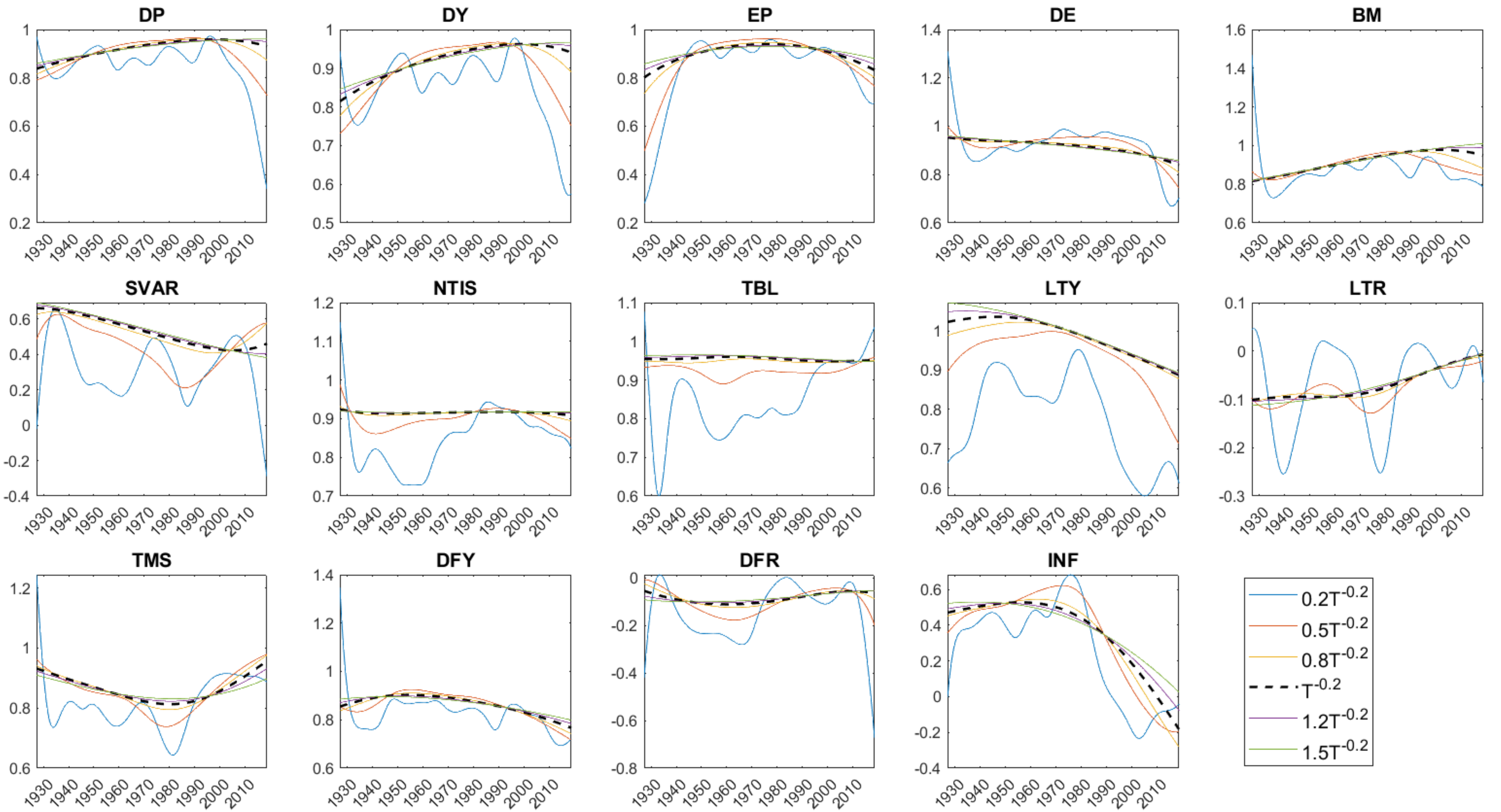
Note. This figure plots the 20-years rolling OLS estimates $\widehat{\rho}$ in the univariate AR regression model, $x_t = \gamma + \rho x_{t-1} + e_t$, with 95% confidence bands assuming that x_t is stationary and the OLS estimates using the full sample, for all predictors. Data are quarterly from 1927:I to 2018:IV.

Figure B.5: Local Linear Estimates $\widehat{\gamma}_t$ on Different Bandwidth Choices



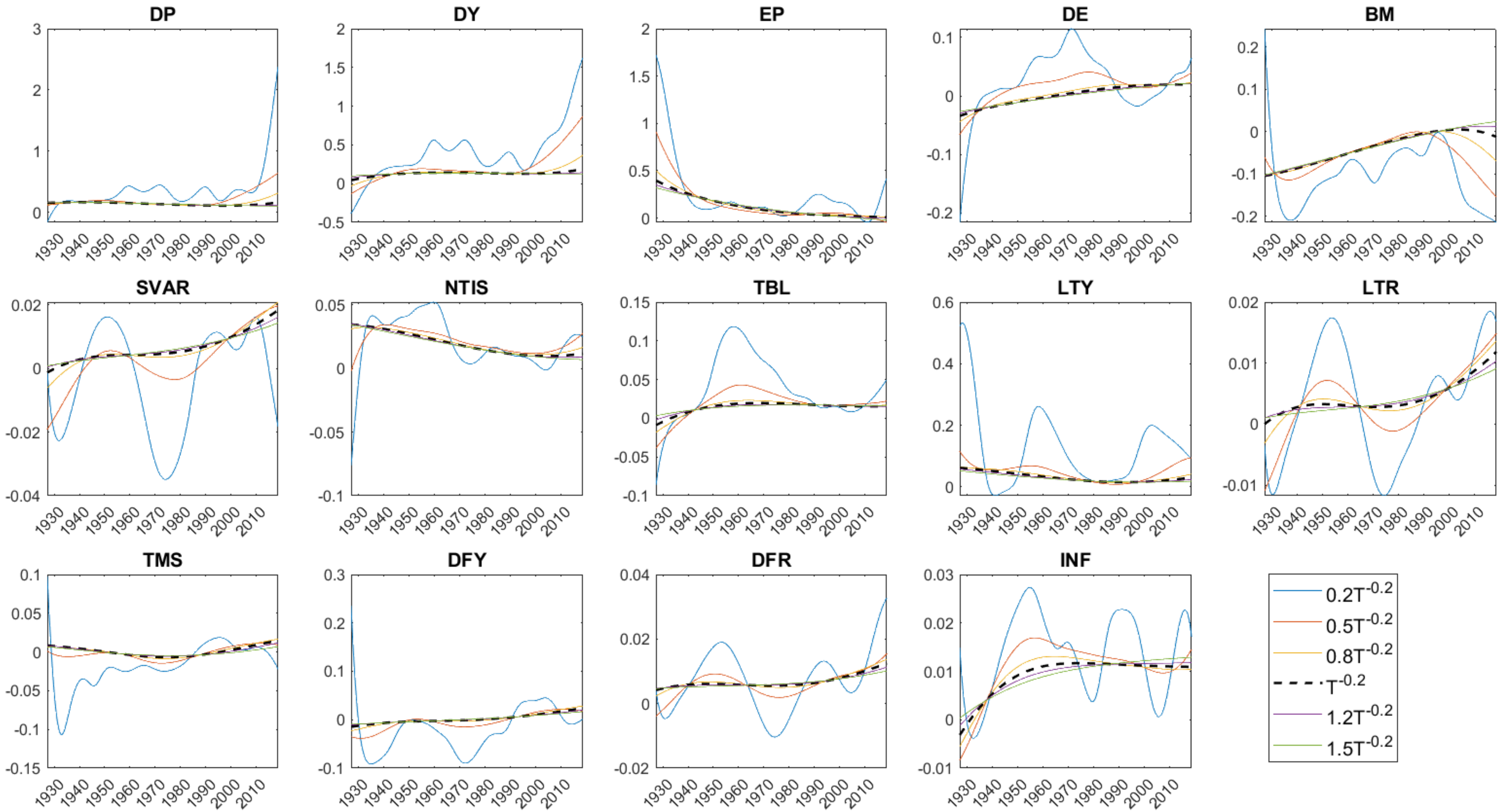
Note. This figure plots the local linear estimates $\widehat{\gamma}_t$ using a set of bandwidth choices in the time-varying AR model $x_t = \gamma(\tau_t) + \rho(\tau_t)x_{t-1} + \varepsilon_t$. The sample period is from 1927:I to 2018:IV.

Figure B.6: Local Linear Estimates $\widehat{\rho}_t$ on Different Bandwidth Choices



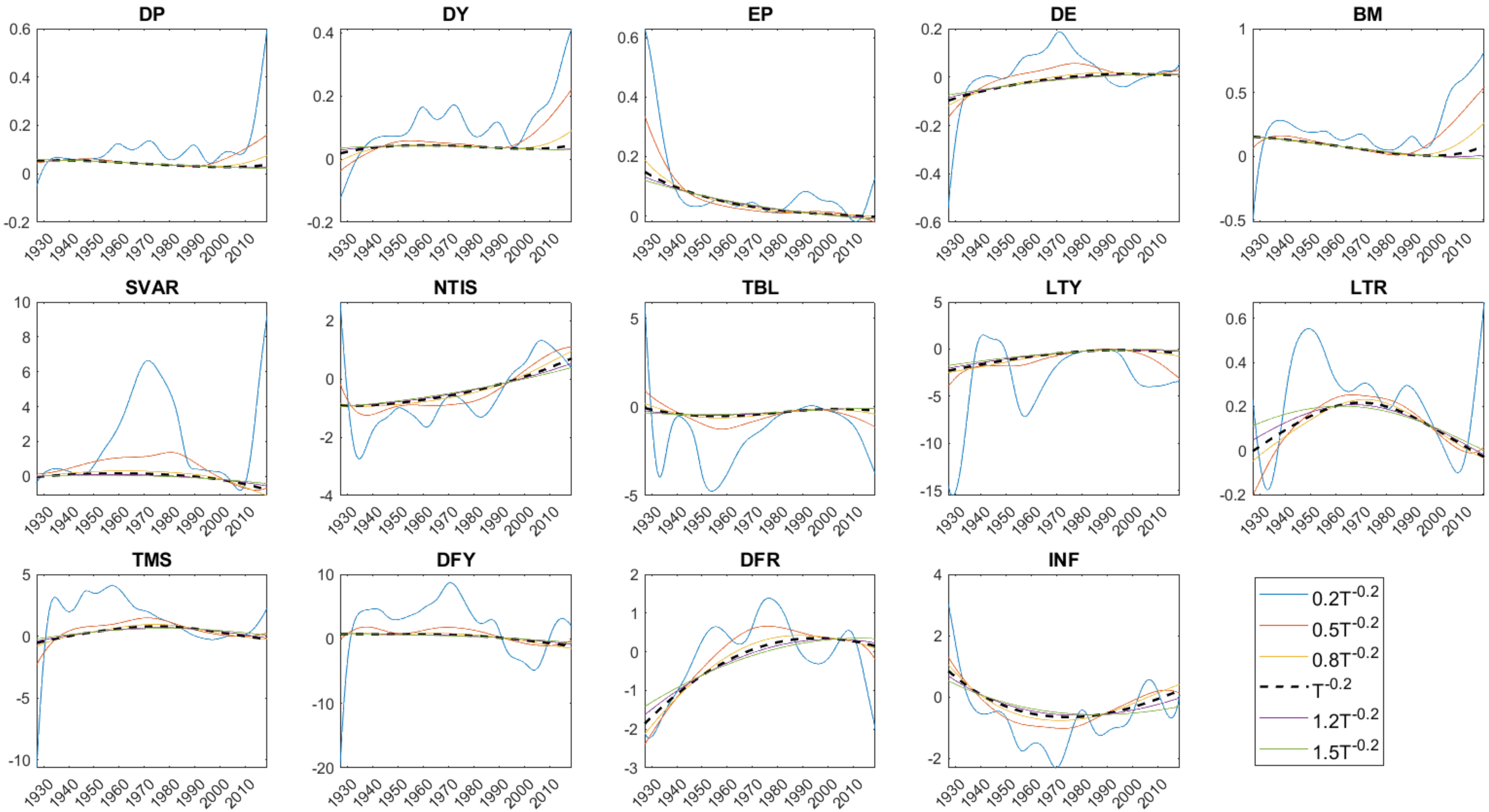
Note. This figure plots the local linear estimates $\widehat{\rho}_t$ using a set of bandwidth choices in the time-varying AR model $x_t = \gamma(\tau_t) + \rho(\tau_t)x_{t-1} + \varepsilon_t$. The sample period is from 1927:I to 2018:IV.

Figure B.7: Local Linear Estimates $\widehat{\alpha}_t$ on Different Bandwidth Choices



Note. This figure plots the local linear estimates $\widehat{\alpha}_t$ using a set of bandwidth choices for the time-varying predictive regression model $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + e_t$. The sample period is from 1927:I to 2018:IV.

Figure B.8: Local Linear Estimates $\widehat{\beta}_t$ on Different Bandwidth Choices



Note. This figure plots the local linear estimates $\widehat{\beta}_t$ using a set of bandwidth choices for the time-varying predictive regression model $r_t = \alpha(\tau_t) + \beta(\tau_t)x_{t-1} + e_t$. The sample period is from 1927:I to 2018:IV.

Bibliography

Amihud, Y. and Hurvich, C. M. (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis*, 39(4):813–841.

Amihud, Y., Hurvich, C. M., and Wang, Y. (2008). Multiple-predictor regressions: Hypothesis testing. *Review of Financial Studies*, 22(1):413–434.

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies*, 20(3):651–707.

Ang, A. and Kristensen, D. (2012). Testing conditional factor models. *Journal of Financial Economics*, 106(1):132–156.

Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics*, 64(3):423–458.

Beveridge, S. and Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary Economics*, 7(2):151–174.

Boons, M., Duarte, F., de Roon, F., and Szymanowska, M. (2020). Time-varying inflation risk and stock returns. *Journal of Financial Economics*, 136(2):444–470.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188.

Cai, Z. and Wang, Y. (2014). Testing predictive regression models with nonstationary regressors. *Journal of Econometrics*, 178:4–14.

Cai, Z., Wang, Y., and Wang, Y. (2015). Testing instability in a predictive regression model with nonstationary regressors. *Econometric Theory*, 31(5):953–980.

Campbell, J. Y. (2008). Estimating the equity premium. *Canadian Journal of Economics*, 41(1):1–21.

Campbell, J. Y. and Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107(2):205–251.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21(4):1509–1531.

Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1):27–60.

Chan, N. H. and Wei, C.-Z. (1987). Asymptotic inference for nearly nonstationary ar (1) processes. *Annals of Statistics*, 15(3):1050–1063.

Chen, B. and Hong, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80(3):1157–1183.

Chen, J., Gao, J., and Li, D. (2012). Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics*, 171(1):71–85.

Chen, L., Da, Z., and Priestley, R. (2012). Dividend smoothing and predictability. *Management Science*, 58(10):1834–1853.

Chen, X. B., Gao, J., Li, D., and Silvapulle, P. (2018). Nonparametric estimation and forecasting for time-varying coefficient realized volatility models. *Journal of Business and Economic Statistics*, 36(1):88–100.

Clark, T. E. and McCracken, M. W. (2009). Tests of equal predictive ability with real-time data. *Journal of Business and Economic Statistics*, 27(4):441–454.

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291–311.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21(4):1533–1575.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance*, 66(4):1047–1108.

Cochrane, J. H. (2017). Macro-finance. *Review of Finance*, 21(3):945–985.

Corsi, F. (2008). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Cujean, J. and Hasler, M. (2017). Why does return predictability concentrate in bad times. *Journal of Finance*, 72(6):2717–2758.

Dahlhaus (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 25(1):1–37.

Dahlhaus, R. (1996). On the kullback-leibler information divergence of locally stationary processes. *Stochastic Processes and Their Applications*, 62(1):139–168.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1):157–181.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263.

Dong, C., Gao, J., Tjøstheim, D., and Yin, J. (2017). Specification testing for nonlinear multivariate cointegrating regressions. *Journal of Econometrics*, 200(1):104–117.

Elliott, G. and Stock, J. H. (1994). Inference in time series regression when the order of integration of a regressor is unknown. *Econometric Theory*, 10(3-4):672–700.

Fama, E. F. (1990). Stock returns, expected returns, and real activity. *Journal of Finance*, 45(4):1089–1108.

Fama, E. F. and French, K. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25(1):23–49.

Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3–25.

Fama, E. F. and French, K. R. (1988). Permanent and temporary components of stock prices. *Journal of Political Economy*, 96(2):246–273.

Fan, R. and Lee, J. H. (2019). Predictive quantile regressions under persistence and conditional heteroskedasticity. *Journal of Econometrics*, 213(1):261–280.

Favero, C. A., Gozluklu, A. E., and Tamoni, A. (2011). Demographic trends, the dividend/price ratio and the predictability of long-run stock market returns. *Journal of Financial and Quantitative Analysis*, 46(5):1493–1520.

Gao, J. and Anh, V. (1999). Semiparametric regression under long-range dependent errors. *Journal of Statistical Planning and Inference*, 80(1-2):37–57.

Gao, J. and Anh, V. (2000). A central limit theorem for a random quadratic form of strictly stationary processes. *Statistics and Probability Letters*, 49(1):69–79.

Gao, J. and Gijbels, I. (2008). Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association*, 103(484):1584–1594.

Gao, J. and Hawthorne, K. (2006). Semiparametric estimation and testing of the trend of temperature series. *Econometrics Journal*, 9(2):332–355.

Giglio, S. W., Maggiori, M., and Stroebel, J. (2016). No-bubble condition: Model-free tests in housing markets. *Econometrica*, 84(3):1047–1091.

Goyal, A. and Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management Science*, 49(5):639–654.

Hamilton, J. D. (2018). Why you should never use the hodrick-prescott filter. *Review of Economics and Statistics*, 100(5):831–843.

Henkel, S. J., Martin, J. S., and Nardari, F. (2011). Time-varying short-horizon predictability. *Journal of Financial Economics*, 99(3):560–580.

Hodrick, R. J. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies*, 5(3):357–386.

Jiang, F., Lee, J. A., Martin, X., and Zhou, G. (2019). Manager sentiment and stock returns. *Journal of Financial Economics*, 132(1):126–149.

Kasparis, I., Andreou, E., and Phillips, P. C. (2015). Nonparametric predictive regression. *Journal of Econometrics*, 185(2):468–494.

Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *Review of Financial Studies*, 28(5):1506–1553.

Kothari, S. and Shanken, J. A. (1997). Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics*, 44(2):169–203.

Künsch, H. R. (1995). A note on causal solutions for locally stationary ar-processes. *ETH Zürich*.

Lamont, O. (1998). Earnings and expected returns. *Journal of Finance*, 53(5):1563–1587.

Lee, J. H. (2016). Predictive quantile regression with persistent covariates: Ivx-qr approach. *Journal of Econometrics*, 192(1):105–118.

Lee, T.-H., Tu, Y., and Ullah, A. (2014). Nonparametric and semiparametric regressions subject to monotonicity constraints: estimation and forecasting. *Journal of Econometrics*, 182(1):196–210.

Lee, T.-H., Tu, Y., and Ullah, A. (2015). Forecasting equity premium: Global historical average versus local historical average and constraints. *Journal of Business and Economic Statistics*, 33(3):393–402.

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, 56(3):815–849.

Lettau, M., Ludvigson, S. C., and Wachter, J. A. (2008). The declining equity premium: What role does macroeconomic risk play? *Review of Financial Studies*, 21(4):1653–1687.

Lettau, M. and Nieuwerburgh, S. V. (2008). Reconciling the return predictability evidence. *Review of Financial Studies*, 21(4):1607–1652.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–235.

Li, D., Chen, J., and Gao, J. (2011a). Non-parametric time-varying coefficient panel data models with fixed effects. *Econometrics Journal*, 14(3):387–408.

Li, D., Chen, J., and Lin, Z. (2011b). Statistical inference in partially time-varying coefficient models. *Journal of Statistical Planning and Inference*, 141(2):995–1013.

Lintner, J. (1956). Distribution of incomes of corporations among dividends, retained earnings, and taxes. *American Economic Review*, 46(2).

McCracken, M. W. (2007). Asymptotics for out of sample tests of granger causality. *Journal of Econometrics*, 140(2):719–752.

Møller, S. V. and Sander, M. (2017). Dividends, earnings, and predictability. *Journal of Banking and Finance*, 78(5):153–163.

Neely, C. J., Rapach, D. E., Tu, J., and Zhou, G. (2014). Forecasting the equity risk premium: the role of technical indicators. *Management Science*, 60(7):1772–1791.

Nelson, C. R. and Kim, M. J. (1993). Predictable stock returns: The role of small sample bias. *Journal of Finance*, 48(2):641–661.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Paye, B. S. and Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, 13(3):274–315.

Pesaran, M. H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1):134–161.

Phillips, P. C. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica*, 82(3):1177–1195.

Phillips, P. C. (2015). Halbert white jr. memorial jfec lecture: Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics*, 13(3):521–555.

Phillips, P. C. and Lee, J. H. (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics*, 177(2):250–264.

Phillips, P. C., Li, D., and Gao, J. (2017). Estimating smooth structural change in cointegration models. *Journal of Econometrics*, 196(1):180–195.

Phillips, P. C. B. and Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics*, 20(2):971–1001.

Rapach, D. and Zhou, G. (2013). Forecasting stock returns. *Handbook of Economic Forecasting*, 2:328–383.

Rapach, D. E., Ringgenberg, M. C., and Zhou, G. (2016). Short interest and aggregate stock returns. *Journal of Financial Economics*, 121(1):46–65.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, 23(2):821–862.

Ren, Y., Tu, Y., and Yi, Y. (2019). Balanced predictive regressions. *Journal of Empirical Finance*, 54:118–142.

Robinson, P. M. (1989). Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change*, pages 253–264. Springer.

Robinson, P. M. (1991). Time-varying nonlinear regression. In *Economic Structural Change*, pages 179–190. Springer.

Spiegel, M. (2008). Forecasting the equity premium: Where we stand today. *Review of Financial Studies*, 21(4):1453–1454.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3):375–421.

Stock, J. H. and Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 41(3):788–829.

Torous, W., Valkanov, R., and Yan, S. (2004). On predicting stock returns with nearly integrated explanatory variables. *Journal of Business*, 77(4):937–966.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.

Xu, K.-L. (2020). Testing for multiple-horizon predictability: Direct regression based versus implication based. *Review of Financial Studies*, 33(9):4403–4443.

Zhang, T., Wu, W. B., et al. (2012). Inference of time-varying regression models. *Annals of Statistics*, 40(3):1376–1402.

Zhang, W. and Peng, H. (2010). Simultaneous confidence band and hypothesis test in generalised varying-coefficient models. *Journal of Multivariate Analysis*, 101(7):1656–1680.