Parametric Single-Index Models: Simulation and Empirical Results

Ying Zhou

Supervised by:

Professor Jiti Gao

Dr Hsein Kew

A thesis presented for PhD progress review.

Department of Econometrics and Business Statistics

March, 2021

1 Overview

My PhD thesis encompasses three chapters. Chapter 1 considers a semi-parametric nonlinear model with a cointegrated system. This model extends the fully non-parametric time-varying coefficients models developed in Li et al. (2016) by adding non time-varying coefficients. This is a useful extension because it allows for the inclusion of dummy variables and deterministic time trends in the regression. We develop a multi-step estimation procedure that can be employed to estimate both the time-varying coefficients nonparametrically and the non-time-varying coefficients parametrically. We apply the proposed model to study the impact of tax incentives on fertility. The main findings in chapter 1 include:

1. The time-varying coefficients model suggests a nonlinear cointegrating relationship between tax benefits and general fertility rate.

2. The nonlinear cointegrating relationship between tax benefits and fertility rate has weakened considerably over time.

Chapter 2 extends the nonlinear model with a single integrated regressor developed in Park and Phillips (2001) to allow for multiple integrated regressors. To help ease the curse of dimensionality associated with estimating multivariate non-linear models with integrated time series, we present an econometric model in which the multiple integrated regressors can be reduced to a univariate single-index form via a known univariate nonlinear function. This single-index component allows for either cointegrated predictors or non-cointegrated predictors. We develop a new estimation procedure for the model. We apply the proposed model to study stock return predictability because the commonly used predictors typically contain a unit root. The main findings in chapter 2 include:

1. We show, via a Monte Carlo experiment, that the new estimator has better finite sample properties than the standard nonlinear least-squares estimator.

2. Exploiting nonlinearities in the data can lead to improved forecast accuracy relative to the historical average when predicting stock returns.

Chapter 3 extends the nonlinear single-index model developed in Chapter 2 by allowing for lagged dependent variables. This is a useful extension because autocorrelation is a pervasive feature of economic time-series data. We develop a novel two-step procedure that can be employed to estimate the parameters of both (i) the non-linear univariate function containing the single-index component with either the cointegrated predictors or the non-cointegrated predictors and (ii) the linear functions containing the lagged dependent variables.

2 Thesis Structure

Chapter 1

Whittington et al. (1990) adopt a linear model to investigate the effect that tax benefits have on general fertility rate. They find that every increase of $100 (in 2015 dollar) in tax benefits is associated with an increase of 2.1 to 4.2 births. However, since the variables in the regression are highly persistent, the finding may suffer from the problem of spurious regression.

Crump et al. (2011) revisit the analysis of Whittington et al. (1990) and perform a variety of cointegration tests. They find no evidence of a linear cointegrating relationship between fertility and personal exemption. This motivates us to relax the linearity assumption by allowing for instability in the cointegrating relationship.

This chapter contributes to the literature on the effect of tax benefits on fertility by estimating a nonlinear cointegrating regression model with time-varying coefficients.

Model and Methodology

Phillips et al. (2017) propose the following nonlinear cointegration model with time-varying coefficient functions:

$$y_t = x_t' \beta_t + \epsilon_t, \quad t = 1, ..., T,$$
$$x_t = x_{t-1} + u_t,$$

where $\beta_t = \beta(\tau_t), \tau_t = t/T$, and $\beta(\cdot)$ is a d-dimensional vector of time-varying coefficients, x_t is a d-dimensional $I(1)$ processes, u_t is a martingale difference sequence and ϵ_t is an error term. This is a fully nonparametric model because $\beta_t = (\beta_{1,t}, ..., \beta_{d,t})'$ is an unknown function of time.

We extend the above nonlinear cointegration model to allow for some coefficients to be constant over time instead of time-varying because the empirical model of fertility involves two dummy variables and a time trend variable. The coefficients of these variables should not be time-varying. We therefore consider a semi-parametric nonlinear cointegration model of the form:

$$y_t = x_t'\beta(\tau_t) + z_t'\theta + \epsilon_t \tag{2.1}$$

where $\tau_t = t/T$, and $\beta(\cdot)$ is a d-dimensional vector of time-varying coefficients, x_t is a d-dimensional $I(1)$ processes, z_t is a m-dimensional vector containing time trend and dummy variables, ϵ_t is an error term.

The estimation method developed in Phillips et al. (2017) cannot be used to estimate a semi-parametric model because it contains a non-parametric part $\beta(\cdot)$ and a parametric part θ. Therefore, we propose a three-step estimation procedure, in which the nonparametric part is estimated first and used to construct a parametric estimator $\hat{\theta}$ in a multi-step estimation approach.

In Step 1, we rewrite model (2.1) as follows:

$$y_t - z_t'\theta = x_t'\beta(\tau) + \varepsilon_t.$$

Let θ be given at this stage and this model becomes a fully non-parametric model. Hence, we can apply the usual non-parametric estimation method and get the estimation of β:

$$\bar{\beta}(\tau) = \bar{\beta}(\tau, \theta).$$

In step 2, we estimate θ by minimising:

$$\sum_{t=1}^{T} \left(y_t - z_t'\theta - x_t'\bar{\beta}(\tau_t, \theta)\right)^2$$

and get the estimation of θ and denote it by $\hat{\theta}$.

In step 3, we re-estimate $\beta(\tau)$ by plugging in $\widehat{\theta}$ to define $\widehat{\beta}(\tau) = \bar{\beta}\left(\tau, \widehat{\theta}\right)$.

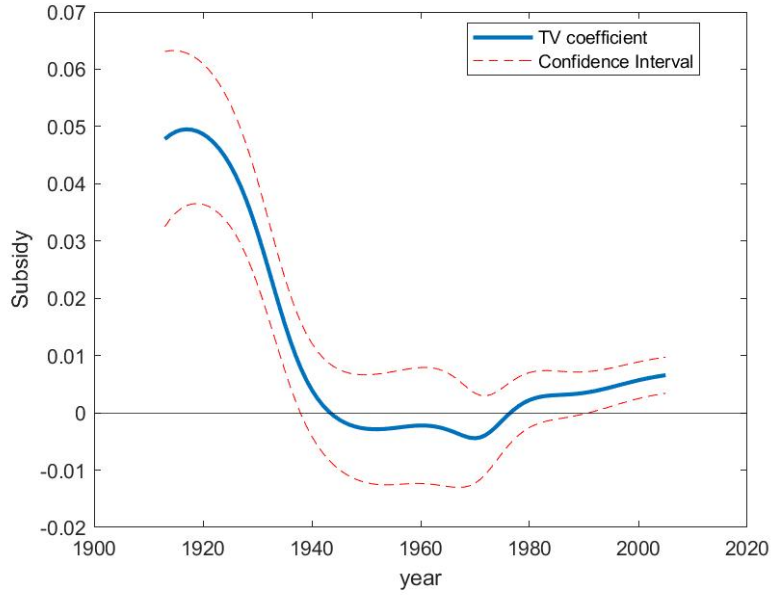
Since our model is newly proposed and we do not have the asymptotic distribution for the coefficients, we use a bootstrap method to construct the 95% confidence interval for the time-varying regression coefficients.

Empirical Study

In an empirical study, we use model (2.1) to investigate the time-varying nature of the cointegrating relationship between tax benefits and general fertility rate. Figure (1) present the impact of one of the tax benefits (Subsidy) have on general fertility rate (the time-varying coefficient, blue line) along with its point-wise 95% confidence interval (dotted lines) constructed using a wild bootstrap method.

In the early part of the period, the impact was significantly different from zero at the 5% level based on the 95% confidence interval. Beginning in the 1930s, the effect of Subsidy on the general fertility rate diminishes until the late 1940s after which they stabilize, though for most of the sample period the coefficient is statistically insignificant.

Figure 1: Time-varying Coefficient for Subsidy



To justify that our time-varying model specifies a nonlinear cointegrating relationship between tax benefits and general fertility rate, we apply the KPSS stationarity test statistic to the residual of the regression. As the residual is stationary at 10% level, the instability present in the cointegrating relation shown in Figure 1 is large enough to change the conclusion of no linear cointegration.

Chapter 2

Previous studies in the empirical literature have taken nonlinearities into account when modeling financial and macroeconomic series (see for example Lettau and Van Nieuwerburgh (2008), Qi (1999)). As financial and macroeconomics time-series data are usually non-stationary, subsequent studies have developed estimation methods for the nonlinear univariate econometric models with non-stationary regressors (Park and Phillips (1999, 2000, 2001)). However, when extending the univariate framework to include multivariate predictors, it suffers from the curse of dimensionality problem because the limiting distribution would naturally be a function of a d-dimensional vector Brownian motion process. As the dimension of multiple predictor increases, the vector Brownian motion process is known to exhibit increasing spatial behaviour (Revuz and Yor (2013)). To avoid this problem, we present an econometric model in which the multiple integrated regressors can be reduced to a univariate single-index form via a known univariate nonlinear function.

We then propose a nonstationary nonlinear single-index model. This model: (a) accounts for the nonlinearities in the data when forecasting next period's macroeconomic and financial variables; (b) helps ease the curse of dimensionality when it comes to parametric nonlinear function estimation involving multivariate integrated regressors; and (c) allows for the presence of cointegrated regressors or non-cointegrated regressors.

We introduce a new estimation procedure for the model and investigate its finite-sample properties via Monte Carlo simulations. This model is then used to examine stock return predictability via various combinations of integrated lagged economic and financial variables.

Model and Methodology

This chapter considers the estimation of a nonstationary nonlinear single-index predictive regression model of the form:

$$y_t = f\left(x'_{t-1}\theta_0, \gamma_0\right) + e_t, \quad t = 2, ..., T, \tag{2.2}$$

where $f(.,.)$ is a known univariate function, x_{t-1} is a d-dimensional integrated process of order one, θ_0 is a d-dimensional unknown true parameter vector that lies in the parameter set Θ, γ_0 is a m-dimensional unknown true parameter vector that lies in the parameter set Γ and e_t is a martingale difference process. The parameter sets Θ and Γ are assumed to be compact and convex subsets of $\mathbb{R}^d$ and $\mathbb{R}^m$ respectively.

The linear combination $x'_{t-1}\theta_0$ in (2.2) is called the single-index component. Since x_{t-1} is a vector of $I(1)$ time series, we impose the following two assumptions on the single-index component. The first assumption rules out cointegration among the predictors, x_{t-1}, and θ_0 the vector of single-index coefficients. By contrast, the second assumption permits cointegrated predictors via $x'_{t-1}\theta_0 \sim I(0)$ with θ_0 being the cointegrating vector.

To illustrate the role played by the parameter vector $\gamma_0 = \left(\gamma_{1,0}, ..., \gamma_{m,0}\right)'$ in (2.1), consider the case where y_t is related to the single-index component through a quadratic functional form

$$f\left(u_{t-1},\gamma_0\right)=\gamma_{1,0}+\gamma_{2,0}u_{t-1}+\gamma_{3,0}u_{t-1}^2,$$

where $u_{t-1}=x_{t-1}'\theta_0$. Here γ_0 is the vector of coefficients for the single-index components. Non-zero elements of γ_0 indicate that the single-index component is a useful predictor of y_t.

Model (2.2) can be estimated by minimizing sum-of-squared-errors:

$$\left(\widehat{\theta},\widehat{\gamma}\right)=\arg\min_{\theta\in\Theta,\gamma\in\Gamma}Q_T\left(\theta,\gamma\right).$$

where $\left(\widehat{\theta},\widehat{\gamma}\right)$ is the nonlinear least square (NLS) estimator.

In an attempt to improve the finite sample properties of the NLS estimator, we truncate the squared-errors $\left(y_t-f\left(x_{t-1}'\theta,\gamma\right)\right)^2$ and we impose a constraint on the coefficient vector θ in the estimation procedure. To this end, we define the modified sum-of-squared-errors by

$$Q_{T,M}\left(\theta,\gamma\right)=\sum_{t=1}^{T}\left(y_t-f\left(x_{t-1}'\theta,\gamma\right)\right)^2 I\left(\|x_{t-1}\|\leq M_T\right)+\lambda\left(\|\theta\|^2-1\right), \tag{2.3}$$

where $I\left(.\right)$ denotes the indicator function, $\|.\|$ is the Euclidean norm and $M_T=\sqrt{T}$. It is a positive and increasing sequence satisfying $M_T\to\infty$ as $T\to\infty$ and λ is a Lagrange multiplier. We then obtain the constrained least squares (denoted CLS) estimator $\tilde{\theta}$ and $\tilde{\gamma}$ by minimizing (2.3).

We investigate the finite sample properties of the NLS and the proposed CLS estimators in multivariate nonstationary settings. The simulation results show that both NLS and CLS estimators have a good finite sample performance. In addition, there is significant finite sample gains from imposing the constraint on the estimation procedure when comparing the NLS estimator with the CLS estimator.

Empirical Study

When predicting stock returns, Welch and Goyal (2008) find that the linear predictive models predict poorly out-of-sample than a simple model with constant means. We aim to examine whether the single-index predictive model, which explicitly allows for nonlinearities, can produce better out-of-sample fits than the historical average benchmark.We find that several combinations of the predictors used in prior studies deliver out-of-sample forecasting gains relative to the standard historical average benchmark when using the single-index predictive model that accounts for the nonlinearities in the time-series data.

Chapter 3

In this chapter, we are interested in a partially nonlinear single-index models of the form:

$$y_t = \beta_0' z_t + g\left(x_{t-1}'\theta_0; \gamma_0\right) + e_t, \quad t = 2, ..., T,$$

where $z_t = (y_{t-1}, \cdots, y_{t-p}, w_{t-1}')'$, in which w_{t-1} is a vector of stationary predictors, $g\left(.,.\right)$ is a known univariate nonlinear function, x_{t-1} is a d-dimensional integrated process of order one, θ_0, γ_0 are unknown d-dimensional and m-dimensional true parameter vectors, and e_t is a martingale difference process.

Our model allows for lagged dependent variables because key macroeconomic/financial variables, such as the growth rate of GDP, the rate of unemployment and interest rates, are typically autocorrelated. Our model may also be useful in cases where there are additional stationary predictors, w_{t-1}, for which the linear specification fits the data better than the nonlinear specification.

To reduce computational burden, we propose a novel 2-step estimation method in which β will have a closed form solution while θ and γ can be estimated by the method of nonlinear least squares or constrained nonlinear least squares. In the coming year, I am going to investigate the finite sample properties of the proposed estimator via a Monte Carlo experiment. In the empirical analysis, we will apply the partially nonlinear single-index predictive model to predict stock returns.

Contents

Chapter 2

Non-stationary Parametric Single-Index Predictive Models

1 Introduction

There is a considerable practical and theoretical effort being channelled into predicting stock returns. Early studies (Campbell and Shiller (1988), Fama (1990), Pesaran and Timmermann (1995)) claim that stock returns can be predicted by lagged macroeconomic and financial variables using linear models, while the out-of-sample (OOS) forecasts fail to outperform the historical average (Welch and Goyal (2008)). Many different models have since been proposed to find a better OOS forecast; for example, Lee et al. (2015) and Chen and Hong (2016)) have proposed nonparametric predictive models. Also Dangl and Halling (2012) and Johannes et al. (2014) have developed time-varying coefficients predictive models.

Previous studies in the empirical literature have taken nonlinearities into account when forecasting financial and macroeconomic series (Lettau and Van Nieuwerburgh (2008), Qi (1999)). To account for this nonlinearity feature, a wide variety of nonlinear econometric models has been developed under the assumption that the regressors are stationary. However, financial and macroeconomics time-series data are usually non-stationary and cointegrated. Subsequent studies have developed estimation methods for the nonlinear econometric models with non-stationary regressors. For instance, Wooldridge (1994) and Andrews and McDermott (1995) develop an asymptotic theory of estimation for stationary time series around a deterministic (but not stochastic) trend function. In a series of papers, Park and Phillips (1999), Park and Phillips (2000), Park and Phillips

(2001) develop an asymptotic theory for a class of nonlinear regressions with integrated scalar regressors. These papers continue the work of Park and Phillips (1988) and Park and Phillips (1989) on linear models in which the regressor is an integrated time series. Chang et al. (2001) and Chang and Park (2003) extend the non-linear models to allow for multiple integrated regressors. Also, Park (2002) and Chung and Park (2007) develop an asymptotic theory for nonlinear heteroskedastic models.

Building on this growing literature, our study focuses on a parametric single-index predictive model with integrated regressors. This model: (a) accounts for the nonlinearities in the data when forecasting next period's macroeconomic and financial variables; (b) helps ease the curse of dimensionality when it comes to parametric nonlinear function estimation involving multivariate integrated regressors; and (c) allows for the presence of cointegrated regressors or non-cointegrated regressors. Our work is related to the semi-parametric single-index framework used by Dong et al. (2016) in which the integrated predictors are not cointegrated, and by Zhou et al. (2018) in which the integrated predictors are cointegrated (see also Xu (2016), Lee et al. (2018) and Koo et al. (2020)).

In our parametric specification, we develop a new estimation procedure for the single-index model and show, via a Monte Carlo experiment, that it has better finite sample properties than the usual nonlinear least-squares estimator. In our empirical analysis, we use the single-index model containing multivariate integrated regressors to predict future stock returns. Our results suggest that forecasts generated from this model perform better than the usual historical average benchmark in an out-of-sample forecasting exercise when we account for the nonlinear relationship between stock returns and the multivariate integrated predictors.

This chapter is organised as follows. Section 2 introduces the parametric single-index predictive model and presents a parametric estimator for this model. Section 3 examines the small-sample properties of the estimators by a Monte Carlo evaluation. The application to stock market return predictability is given in Section 4. Section 5 concludes.

2 Parametric Single-Index Predictive Model

In this chapter, we study the estimation for a parametric single-index predictive model with nonstationary predictors

$$y_t = f\left(x'_{t-1}\theta_0, \gamma_0\right) + e_t, \quad t = 2, ..., T, \tag{2.1}$$

where $f(.,.)$ is a known univariate function, x_{t-1} is a d-dimensional integrated process of order one, θ_0 is a d-dimensional unknown true parameter vector that lies in the parameter set Θ, γ_0 is a m-dimensional unknown true parameter vector that lies in the parameter set Γ and e_t is a martingale difference process. The parameter sets Θ and Γ are assumed to be compact and convex subsets of $\mathbb{R}^d$ and $\mathbb{R}^m$ respectively. In order to ensure that θ_0 is uniquely identifiable, we will need to impose $\theta'_0\theta_0 = 1$.

The linear combination $x'_{t-1}\theta_0$ in (2.1) is called the single-index component. Since x_{t-1} is a vector of $I(1)$ time series, we impose the following two assumptions on the single-index component. The first assumption rules out cointegration among the predictors, x_{t-1}, and θ_0 the vector of single-index coefficients. By contrast, the second assumption permits cointegrated predictors via $x'_{t-1}\theta_0 \sim I(0)$ with θ_0 being the cointegrating vector.

Our single-index predictive model can be used to predict stock market returns without imposing linearity. The most commonly used predictors, such as dividend-price ratio, earning-price ratio, book-to-market ratio, term spread and dividend-payout ratio, have been found to be non-stationary and often contain an autoregressive unit root; see Table 7 of Kostakis et al. (2015) and Table 4 of Campbell and Yogo (2006). Campbell et al. (2004) estimate a linear trivariate predictive model containing earning-price ratio, term spread and book-to-market ratio as predictors. In the absence of a cointegrating relationship among these predictors, we could utilize this single-index model to re-examine the out-of-sample forecasting performance of these predictors. In the cointegrated predictors case, Lettau and Ludvigson (2001) consider $x_{t-1} = (c_{t-1}, a_{t-1}, y_{t-1})'$ where c_t is log consumption, a_t is log asset wealth and y_t is log labor income. They show that, under the general household budget constraint framework, the predictors c_t, a_t and y_t should move together over the long run and are hence cointegrated. They suggest that the cointegrating residual (termed the 'cay' variable) from regressing c_{t-1} on a_{t-1} and y_{t-1} is a relevant predictor of stock returns when using a linear predictive model. Again, we could utilize this single-index model to re-assess the forecasting ability of the cointegrated cay predictor.

To illustrate the role played by the parameter vector $\gamma_0 = \left(\gamma_{1,0}, ..., \gamma_{m,0}\right)'$ in (2.1), consider the case where y_t is related to the single-index component through a quadratic functional form

$$f\left(u_{t-1}, \gamma_0\right) = \gamma_{1,0} + \gamma_{2,0} u_{t-1} + \gamma_{3,0} u_{t-1}^2,$$

where $u_{t-1} = x_{t-1}'\theta_0$. Here γ_0 is the vector of coefficients for the single-index components. Non-zero elements of γ_0 indicate that the single-index component is a useful predictor of y_t. The single-index model (2.1) includes as a special case the simple neural network model – with $f\left(u_{t-1}, \gamma_0\right) = \gamma_{1,0} + \gamma_{2,0} G\left(\gamma_{3,0} + u_{t-1}\right)$ for a known function $G\left(.\right)$ – considered by Chang and Park (2003).

In the case where x_{t-1} is a univariate integrated predictor, Park and Phillips (2001) consider a parametric nonlinear nonstationary regression model: $y_t = f\left(x_{t-1}, \gamma_0\right) + e_t$. They use a nonlinear least squares (NLS) estimator to estimate their model and show that the limiting distribution of this estimator is a functional of a univariate Brownian motion. An extension of their univariate framework to include multivariate predictors suffers from the curse of dimensionality problem because the limiting distribution would naturally be a function of a d-dimensional vector Brownian motion. As the dimension of x_t increases, the vector Brownian motion process is known to exhibit increasing spatial behaviour; see Revuz and Yor (2013). The use of the single-index method avoids this spatial problem because it reduces the dimension of multivariate predictors to a univariate linear combination $x_{t-1}'\theta_0$, and thereby eliminates the necessity for vector Brownian motion in the development of the limit theory.

Following Chang and Park (2003), the single-index model (2.1) can be estimated by a NLS estimator. Define the sum-of-squared-errors by

$$Q_T\left(\theta, \gamma\right) = \sum_{t=1}^{T} \left(y_t - f\left(x_{t-1}'\theta, \gamma\right)\right)^2.$$

The NLS estimator $\hat{\theta}$ and $\hat{\gamma}$ is given by minimizing $Q_T\left(\theta, \gamma\right)$ over $\theta \in \Theta$ and $\gamma \in \Gamma$, that is

$$\left(\widehat{\theta}, \widehat{\gamma}\right) = \arg \min_{\theta \in \Theta, \gamma \in \Gamma} Q_T\left(\theta, \gamma\right). \tag{2.2}$$

The solutions to (2.2) must be found numerically because there is no closed-form solution. The '*nloptr*' package in R with Augmented Lagrangian Algorithm can be used as an optimization routine to numerically solve (2.2).

In an attempt to improve the finite sample properties of the NLS estimator, we truncate the squared-errors $\left(y_t - f\left(x_{t-1}'\theta, \gamma\right)\right)^2$ and we impose a constraint on the coefficient vector θ in the estimation procedure. To this end, we define the modified sum-of-squared-errors by

$$Q_{T,M}\left(\theta,\gamma\right) = \sum_{t=1}^{T} \left(y_t - f\left(x_{t-1}'\theta,\gamma\right)\right)^2 I\left(\|x_{t-1}\| \leq M_T\right) + \lambda\left(\|\theta\|^2 - 1\right), \tag{2.3}$$

where $I\left(.\right)$ denotes the indicator function, $\|.\|$ is the Euclidean norm and $M_T = \sqrt{T}$. It is a positive and increasing sequence satisfying $M_T \to \infty$ as $T \to \infty$ and λ is a Lagrange multiplier.

The reason for truncating the squared-errors $\left(y_t - f\left(x_{t-1}'\theta, \gamma\right)\right)^2$ in (2.3) is that the presence of integrated predictors will tend to produce far too few observations at distinct spatial locations. These observations may cause a standard optimisation routine to fail to converge when solving (2.2). This truncation method was originally used by Li et al. (2016) for the case when the univariate regressor x_{t-1} follows the null recurrent Markov process, which is known to exhibit spatial structure.

The constrained least squares (denoted CLS) estimator $\widetilde{\theta}$ and $\widetilde{\gamma}$ is given by minimizing $Q_{T,M}\left(\theta,\gamma\right)$ over $\theta \in \Theta$ and $\gamma \in \Gamma$ such that the restriction $\|\theta\|^2 = 1$ holds; that is

$$\left(\widetilde{\theta}, \widetilde{\gamma}\right) = \arg \min_{\theta\in\Theta,\gamma\in\Gamma,\|\theta\|^2=1} Q_{T,M}\left(\theta,\gamma\right). \tag{2.4}$$

A constrained optimization method is required to find the solutions to (2.4) and the '*nloptr*' package in R can still be used to numerically solve (2.4) subject to the constraint that $\|\theta\|^2 = 1$.

The constraint $\|\theta\|^2 = 1$ in (2.3) scales the CLS estimator to the surface of the unit ball and Zhou et al. (2018) demonstrate that, in their semi-parametric single-index model, this constraint on the estimation procedure causes the CLS estimator of θ_0 to converge to their true values at a faster rate than that for the case without contraints. In our parametric single-index model, the simulation results in the next section show significant finite sample gains from imposing this constraint on the estimation procedure when comparing the NLS estimator with the CLS estimator.

3 Monte Carlo Simulation

3.1 Data Generation Process

In this section, we investigate the finite sample properties of the NLS and the proposed CLS estimators in multivariate nonstationary settings. The predictors x_{t-1} is a 2-vector integrated time series. Data were generated on the following models:

$$\begin{aligned} y_t &= f\left(x_{t-1}'\theta_0, \gamma_0\right) + e_t, \quad e_t \sim i.i.d.N\left(0,1\right), \\ x_t &= x_{t-1} + v_t, \quad x_0 = \left(0,0\right)', \\ v_t &= \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix} \sim i.i.N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right). \end{aligned}$$

We consider the following nonlinear functions:

$$\begin{aligned} f_1\left(u,\gamma_0\right) &= \sin\left(u+\gamma_{1,0}\right), \\ f_2\left(u,\gamma_0\right) &= \cos\left(u+\gamma_{1,0}\right), \\ f_3\left(u,\gamma_0\right) &= 1-e^{-\gamma_{1,0}\left(u-\gamma_{2,0}\right)^2}, \\ f_4\left(u,\gamma_0\right) &= \gamma_{1,0}e^{-\gamma_{2,0}u^2}, \\ f_5\left(u,\gamma_0\right) &= \gamma_{1,0}+\gamma_{2,0}u+\gamma_{3,0}u^2. \end{aligned} \tag{3.1}$$

with $\gamma_{1,0} = 0.2, \gamma_{2,0} = 0.3$ and $\gamma_{3,0} = 0.3$. The first four functions are bounded on $\mathbb{R}$ and the last one is unbounded on $\mathbb{R}$.

We consider sample sizes $T = 100, 500, 1000$ and we use $M = 1000$ simulation replications. In the CLS estimation method, we follow Li et al. (2016) by choosing $M_T = \sqrt{T}$. Let $\theta_0 = \left(\theta_{1,0}, \theta_{2,0}\right)'$. To evaluate the finite sample performance of the NLS and proposed CLS estimators, we compute the bias and standard deviation of each element of $\hat{\theta}$ and $\tilde{\theta}$ defined in the previous section. For example, let $\widehat{\theta} = \left(\widehat{\theta}_1, \widehat{\theta}_2\right)'$ and $\widehat{\theta}_i^{(j)}$ denote the j-th replication of the estimate $\widehat{\theta}_i$ for $i = 1, 2$. Then, for the NLS estimator, we have

$$\text{bias} = \overline{\overline{\theta}}_i - \theta_{i,0},$$

where $\overline{\overline{\theta}}_i = M^{-1}\sum_{j=1}^{M}\widehat{\theta}_i^{(j)}$; and

$$\text{standard deviation (s.d.)} = \sqrt{M^{-1}\sum_{j=1}^{M}\left(\widehat{\theta}_i^{(j)} - \overline{\overline{\theta}}_i\right)^2}.$$

Since $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are correlated, we also calculate a type of estimated covariance of the form:

$$\sigma_\theta = \frac{1}{M}\sum_{j=1}^{M}\left(\widehat{\theta}_1^{(j)} - \bar{\widehat{\theta}}_1\right)\left(\widehat{\theta}_2^{(j)} - \bar{\widehat{\theta}}_2\right), \qquad \mathrm{std}_\theta = |\sigma_\theta|. \tag{3.2}$$

3.2 Computational Details

In this section, we provide an instruction on how to estimate the parameters using CLS estimation. In equation (2.3), $\left(\|\theta\|^2 - 1\right)$ is an equality constraint that only applies to the parameter θ and we use nonlinear least square method with Lagrange multiplier to estimate both θ and γ. To simplify, we first denote the Lagrangian multiplier as:

$$L_{n,m}(\theta, \gamma, \lambda) = \sum_{t=1}^{T} F_t(\theta, \gamma) + \lambda h(\theta)$$

where $F_t(\theta, \gamma) = \left(y_t - f\left(x'_{t-1}\theta, \gamma\right)\right)^2 I\left(\|x_{t-1}\| \le M_T\right)$ and $h(\theta) = \left(\|\theta\|^2 - 1\right)$.

The necessary conditions for an extremum of F_t with the equality constraints $h(\theta) = 0$ are that:

$$\nabla L((\theta^*, \gamma^*), \lambda^*) = \begin{bmatrix} \nabla_\lambda L(\lambda^*, \theta, \gamma) \\ \nabla_\theta L(\theta^*, \gamma, \lambda) \\ \nabla_\gamma L(\gamma^*, \theta, \lambda) \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial \lambda} \\ \frac{\partial L}{\partial \theta} \\ \frac{\partial L}{\partial \gamma} \end{bmatrix} = 0 \tag{3.3}$$

where $\frac{\partial L}{\partial \lambda} = \frac{\partial h}{\partial \lambda}$, $\frac{\partial L}{\partial \theta} = \left(\frac{\partial F_t}{\partial \theta_1} \frac{\partial F_t}{\partial \theta_2} \cdots \frac{\partial F_t}{\partial \theta_d}\right)^T$, and $\frac{\partial L}{\partial \gamma} = \left(\frac{\partial F_t}{\partial \gamma_1} \frac{\partial F_t}{\partial \gamma_2} \cdots \frac{\partial F_t}{\partial \gamma_m}\right)^T$.

Given that condition (3.3) is satisfied, we then move on to check whether the Hessian matrix with a Lagrangian multiplier (or bordered Hessian) is negative definite. Bordered hessian is given below:

$$\mathbf{H}(L) = \begin{bmatrix} \frac{\partial^2 L}{\partial \lambda^2} & \frac{\partial^2 L}{\partial \lambda \partial \theta} & \frac{\partial^2 L}{\partial \lambda \partial \gamma} \\ \left(\frac{\partial^2 L}{\partial \lambda \partial \theta}\right)^\top & \frac{\partial^2 L}{\partial \theta^2} & \frac{\partial^2 L}{\partial \theta \partial \gamma} \\ \left(\frac{\partial^2 L}{\partial \lambda \partial \gamma}\right)^\top & \left(\frac{\partial^2 L}{\partial \theta \partial \gamma}\right)^\top & \frac{\partial^2 L}{\partial \gamma^2} \end{bmatrix} = \begin{bmatrix} 0 & \frac{\partial^2 L}{\partial \lambda \partial \theta} & \frac{\partial^2 L}{\partial \lambda \partial \gamma} \\ \left(\frac{\partial^2 L}{\partial \lambda \partial \theta}\right)^\top & \frac{\partial^2 L}{\partial \theta^2} & \frac{\partial^2 L}{\partial \theta \partial \gamma} \\ \left(\frac{\partial^2 L}{\partial \lambda \partial \gamma}\right)^\top & \left(\frac{\partial^2 L}{\partial \theta \partial \gamma}\right)^\top & \frac{\partial^2 L}{\partial \gamma^2} \end{bmatrix} \tag{3.4}$$

So if $(-1)det(H(L)) > 0$, then L has a local minimum. As an example, consider the functional form:

$$f_1(u, \gamma) = \sin(u + \gamma), u = (\theta_1 x_1 + \theta_2 x_2).$$

The first order optimality condition is given by:

$$\nabla L((\theta^*,\gamma^*),\lambda^*) = \begin{bmatrix} \nabla_\lambda L(\lambda^*,\theta,\gamma) \\ \nabla_\theta L(\theta^*,\gamma,\lambda) \\ \nabla_\gamma L(\gamma^*,\theta,\lambda) \end{bmatrix} = \begin{bmatrix} \theta_1^2+\theta_2^2-1 \\ -\sum_{t=1}^T (y-f_1(u,\gamma))\frac{\partial f_1(u,\gamma)}{\partial\theta_1}+2\lambda\theta_1 \\ -\sum_{t=1}^T (y-f_1(u,\gamma))\frac{\partial f_1(u,\gamma)}{\partial\theta_2}+2\lambda\theta_2 \\ -\sum_{t=1}^T (y-f_1(u,\gamma))\frac{\partial f_1(u,\gamma)}{\partial\gamma} \end{bmatrix} = 0$$

where I denote the truncation condition: $I\left(\|x_{t-1}\| \le M_T\right)$.

The second order condition, that is the Hessian matrix is given by:

$$\mathbf{H}(L) = \begin{bmatrix} 0 & -2\theta_1 & -2\theta_2 & 0 \\ -2\theta_1 & D_{\theta_1\theta_1}+2\lambda & D_{\theta_1\theta_2} & D_{\theta_1\gamma} \\ -2\theta_2 & D_{\theta_2\theta_1} & D_{\theta_2\theta_2}+2\lambda & D_{\theta_2\gamma} \\ 0 & D_{\gamma\theta_1} & D_{\gamma\theta_2} & D_{\gamma\gamma} \end{bmatrix}$$

where $D_{\alpha\beta} = \sum_{t=1}^T \frac{\partial f_1(u,\gamma)}{\partial\alpha}\left(\frac{\partial f_1(u,\gamma)}{\partial\beta}\right)' - \sum_{t=1}^T \left(y_t - f_1(u,\gamma)\right)\frac{\partial^2 f_1(u,\gamma)}{\partial\alpha^2}$.

Then we can check the sign of det(H(L)) to decide whether a local minimum exists. If there is a local minimum, we can find it by solving (3.2).

3.3 Simulation Results

We first consider the case in which the predictors are not cointegrated, so that the single-index component $x_{t-1}'\theta_0$ is purely non-stationary by setting $\theta_0 = (0.8, 0.6)'$. The simulation results are reported in Table 2.1. As we can see, the biases, standard deviations and σ_θ for the NLS and CLS estimators decrease as the sample size increases. These results are promising, and indicate that both NLS and CLS are consistent estimators of θ_0. Table 2.1 also shows that our proposed CLS estimator exhibits a much better finite sample performance than the NLS estimator, especially much smaller finite sample biases and standard deviations across all experiments. Take the function $f_1(u,\gamma_0)$ in Table (2.1) as an example, when T = 100, the bias of $\theta_{1,0}$ decreases from 0.01497 to 0.00422 when comparing the NLS and CLS estimators. It is thus useful to incorporate the truncation method and impose a constraint on the coefficient vector θ in the estimation procedure.

We now allow for the possibility of cointegration among the predictors by setting $\theta_0 = (0.6, -0.8)'$ and hence the single-index component $x_{t-1}'\theta_0$ becomes $I(0)$. We also allow the predictors x_{t-1} to follow a vector integrated process driven by an MA(1) innovations; that is

$$v_t = \epsilon_t + C\epsilon_{t-1},$$

where $\epsilon_t \sim i.i.d.N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$ and $C = \begin{pmatrix} -1 & 4/3 \\ 0 & 0 \end{pmatrix}$.

We report simulation results in Table 2.2 only for the CLS estimator since in the non-cointegrated case we find that there are finite sample gains when resorting to a constrained estimation procedure. As can be seen, the performance of the CLS estimator tends to improve as the sample size increases across the different regression functions. Deriving the theoretical properties of the CLS estimator under both non-cointegrated and cointegrated predictors should be a good topic for future research.

4 Empirical illustration: Stock market return predictability

To illustrate the use of the single-index model, we provide an application to predictability of U.S. stock market returns. We focus on out-of-sample forecasts because it involves forecasters making predictions in real time (Stock and Watson (2007)). There is a large and growing empirical literature that attempts to predict stock returns using lagged macroeconomic and financial variables. Most empirical studies in this literature have focused on univariate linear predictive regression models, including, for example, the papers by Campbell (1987), Fama and French (1989), Pesaran and Timmermann (1995) and Choi et al. (2016). However, several subsequent papers argue that multiple predictors can explain the variation in stock returns better than a single predictor because multivariate predictive models are less prone to suffer from omitted predictors problem (Cochrane (2011)).

A linear specification for the bivariate/multivariate predictive regression is used in many empirical applications. However, Welch and Goyal (2008) find that the linear predictive models predict poorly out-of-sample than a simple model with constant means. Meanwhile Qi (1999) and Zhou et al. (2018) find evidence of nonlinearities in the stock return and lagged financial variables relationship. The main objective of this section is to examine whether the parametric single-index model, which explicitly allows for

Table 2.1: Finite sample properties of NLS and CLS estimators: Non-cointegrated predictors

			CLS			NLS		
			T=100	**T=500**	**T=1000**	**T=100**	**T=500**	**T=1000**
$f_1(u,\gamma_0)$	$\theta_{1,0}$	Bias	0.00422	0.00112	0.00038	0.01497	0.00532	0.00369
		s.d.	0.02191	0.00600	0.00365	0.08951	0.04345	0.03039
	$\theta_{2,0}$	Bias	-0.00068	-0.00013	-4.5E-05	-0.00227	-0.00250	0.00200
		s.d.	0.00271	0.00061	0.00037	0.15105	0.06067	0.04121
		std_θ	0.00845	0.00689	0.00137	0.10982	0.07392	0.03368
	$\gamma_{1,0}$	Bias	-0.00779	-0.00032	-0.00152	0.00582	0.00326	0.00266
		s.d.	0.17540	0.07832	0.05733	0.13945	0.05851	0.04129
$f_2(u,\gamma_0)$	$\theta_{1,0}$	Bias	0.00447	0.00139	0.00064	-0.00128	-0.00016	-0.00011
		s.d.	0.023040	0.00642	0.00336	0.03094	0.00755	0.00390
	$\theta_{2,0}$	Bias	-0.00073	-0.00016	-7E-05	-0.00058	-0.00003	-0.00010
		s.d.	0.00292	0.00066	0.00034	0.02912	0.00722	0.00386
		std_θ	0.00845	0.00689	0.00137	0.11256	0.29867	0.01399
	$\gamma_{1,0}$	Bias	-0.00211	-0.0057	-0.00161	0.04223	0.00609	-0.00229
		s.d.	0.16254	0.08333	0.05583	0.09154	0.04721	0.02063
$f_3(u,\gamma_0)$	$\theta_{1,0}$	Bias	-0.00011	-0.00005	0.00007	0.00634	0.00409	0.00380
		s.d.	0.00371	0.00312	0.00254	0.02708	0.01484	0.01076
	$\theta_{2,0}$	Bias	0.00007	0.00003	-0.00006	0.06275	0.01113	0.00459
		s.d.	0.00267	0.00225	0.00195	0.19590	0.07324	0.05620
		std_θ	0.00345	0.00279	0.00011	0.03922	0.00981	0.00136
	$\gamma_{1,0}$	Bias	-0.00083	0.00460	0.00861	0.01342	0.00349	0.00185
		s.d.	0.02793	0.14505	0.16913	0.03588	0.0.00877	0.00576
	$\gamma_{2,0}$	Bias	0.09702	0.07387	0.03148	0.03655	0.01197	0.00696
		s.d.	1.21163	0.90087	0.51446	0.05355	0.02329	0.01733

Table 2.1: Continued

			CLS			NLS		
			T=100	**T=500**	**T=1000**	**T=100**	**T=500**	**T=1000**
$f_4(u,\gamma_0)$	$\theta_{1,0}$	Bias	-0.00047	-0.00054	-0.00028	0.03116	0.03489	0.02974
		s.d.	0.03926	0.00974	0.00485	0.41504	0.15343	0.10855
	$\theta_{2,0}$	Bias	-0.00049	0.00012	4E-05	-0.00729	0.01939	0.01315
		s.d.	0.02934	0.00853	0.00444	0.37102	0.14480	0.08879
		std_θ	0.00552	0.00398	0.00275	0.00992	0.00584	0.00412
	$\gamma_{1,0}$	Bias	0.00288	-0.00023	-0.00011	0.01621	-0.00864	-0.00925
		s.d.	0.02392	0.00738	0.00336	0.09407	0.05596	0.03377
	$\gamma_{2,0}$	Bias	0.00171	-0.00087	-0.00009	-0.00210	0.00759	0.01689
		s.d.	0.03928	0.00664	0.00874	0.37198	0.16559	0.11484
$f_5(u,\gamma_0)$	$\theta_{1,0}$	Bias	-0.00077	-0.00001	0.00000	0.00237	0.00167	0.00094
		s.d.	0.00571	0.00090	0.00041	0.01033	0.00494	0.00326
	$\theta_{2,0}$	Bias	0.00006	0.00000	0.00000	0.00220	0.00139	0.00068
		s.d.	0.00053	0.00009	0.00004	0.01132	0.00483	0.00329
		std_θ	0.00053	0.00019	0.00008	0.00444	0.00192	0.00107
	$\gamma_{1,0}$	Bias	-0.00071	0.00167	0.00168	-0.00003	-0.00049	-0.00013
		s.d.	0.01609	0.00513	0.00322	0.00809	0.00296	0.00214
	$\gamma_{2,0}$	Bias	0.00008	0.00000	0.00000	-0.00174	-0.00111	-0.00062
		s.d.	0.00472	0.00060	0.00028	0.00446	0.00238	0.00148
	$\gamma_{3,0}$	Bias	0.00014	-0.00003	-0.00001	0.00270	0.00107	0.00045
		s.d.	0.00595	0.00084	0.00040	0.00581	0.00358	0.00233

Table 2.2: Finite sample properties of CLS estimator: Cointegrated predictors

		$f_1(u,\gamma_0)$			$f_2(u,\gamma_0)$			$f_3(u,\gamma_0)$			
		$\gamma_{1,0}$	$\theta_{1,0}$	$\theta_{2,0}$	$\gamma_{1,0}$	$\theta_{1,0}$	$\theta_{2,0}$	$\gamma_{1,0}$	$\gamma_{2,0}$	$\theta_{1,0}$	$\theta_{2,0}$
	bias	-0.00707	0.00814	-0.03127	-0.00511	0.00815	-0.02885	-0.01325	0.03478	0.02894	0.56838
T = 100	std	0.05976	0.08368	0.22101	0.05561	0.07984	0.19458	0.18462	0.25850	0 .39893	0.86347
	std_θ		0.21981			0.19987				0.05635	
	bias	-0.00049	0.00065	-0.00102	-0.00115	0.00124	-0.00589	-0.01193	-0.03216	0.02289	0.11934
T = 500	std	0.01789	0.02420	0.07311	0.01135	0.01536	0.05785	0.06733	0.08842	0.0.07056	0.21906
	std_θ		0.07272			0.06180				0.01955	
	bias	-0.00007	0.00083	-0.00120	-0.00066	0.00075	-0.00257	-0.00434	0.00793	0.00173	0.09899
T = 1000	std	0.00812	0.01115	0.04286	0.00764	0.01072	0.03726	0.01899	0.03194	0.03956	0.09878
	std_θ		0.04471			0.03731				0.01266	

		$f_4(u,\gamma_0)$				$f_5(u,\gamma_0)$				
		$\gamma_{1,0}$	$\gamma_{2,0}$	$\theta_{1,0}$	$\theta_{2,0}$	$\gamma_{1,0}$	$\gamma_{2,0}$	$\gamma_{3,0}$	$\theta_{1,0}$	$\theta_{2,0}$
	bias	0.04778	-0.05835	0.42574	0.44352	0.00244	-0.00290	0.00176	-0.00725	-0.02559
T = 100	std	0.22502	30026	79131	1.35994	0.00982	0.01532	0.02586	0.09341	0.05667
	std_θ			0.09709					0.10301	
	bias	-0.00359	0.00408	0.09030	17635	0.00041	-0.00047	0.00077	-0.00025	-0.00352
T = 500	std	0.07905	0.10786	0.32777	0.69730	0.00386	0.00609	0.00896	0.03894	0.02578
	std_θ			0.04742					0.05841	
	bias	-0.00266	0.00241	0.05001	0.07098	4.37E-05	-0.00027	0.00008	-0.00049	-0.00103
T = 1000	std	0.05233	0.07447	0.23882	0.37198	0.00266	0.00418	0.00532	0.02730	0.03349
	std_θ			0.008577438					0.03156	

nonlinearities in the predictive relations, can produce better out-of-sample fits than the historical average benchmark.

The dataset used for this study is available from Amit Goyal's website. The examined period is quarterly data from 1956:Q1 to 2018:Q4. The dependent variable, y_t, is the equity premium defined as the S&P500 value-weighted log excess returns. We select a combination of predictors x_{t-1} that has been found to be important, both theoretically and empirically, in prior studies of stock return predictability. Table 7 of Kostakis et al. (2015) provides evidence suggesting that the commonly used predictors contain a unit root. Using the Engle-Granger test for cointegration, we fail to reject the null hypothesis of no cointegration for the following combinations of predictors: (a) dividend-price ratio and T-bill rate, see Ang and Bekaert (2007); (b) dividend-price ratio, T-bill rate, default yield spread and term spread, see Ferson and Schadt (1996); (c) dividend-price ratio and book-to-market ratio, see Kothari and Shanken (1997); (d) dividend-price ratio and dividend-payout ratio, see Lamont (1998), and (e) earning-price ratio, term spread and book-to-market ratio, see Campbell et al. (2004). The following four combinations of cointegrated predictors are considered in Zhou et al. (2018): (f) dividend-price ratio and dividend yield; (g) T-bill rate and long-term yield; (h) dividend-price ratio and earning-price ratio; (i) baa- and aaa-rated corporate bond yields; and (j) c_t, a_t and y_t, see Lettau and Ludvigson (2001).

We use the same regression functions (namely $f_1(u,\gamma), f_2(u,\gamma), f_3(u,\gamma), f_4(u,\gamma)$ and $f_5(u,\gamma)$) as in the Monte Carlo section. The single-index model is estimated using the CLS estimator with $M_T = \sqrt{T}$. We consider one-step ahead forecasting of the equity premium using all the 10 combinations of predictors and the 5 nonlinear functions, for a total of 50 single-index predictive models. We generate pseudo out-of-sample forecasts from a sequence of recursive predictions.

We evaluate the out-of-sample performance using a recursive window scheme. The details are described as follows. We divide the total observations T into in-sample part containing T_1 observations from Q1 of 1956 to Q4 of 1987, while the remaining $T - T_1$ observations from Q1 of 1988 to Q4 of 2018 are regarded as out-of-sample period. To generate the first out-of-sample stock return forecast, we use the first T_1-1 pairs of observations $\{(x_1,\ y_2),\ (x_2,\ y_3),\ \cdots,\ (x_{T_1-1},\ y_{T_1})\}$ to estimate the nonlinear model and predict $\hat{y}_{T_1+1}$. We then include the information in the T_1+1 period and predict $\hat{y}_{T_1+2}$ using $\{(x_1,\ y_2),\ (x_2,\ y_3),\ \cdots,\ (x_{T_1},\ y_{T_1+1})\}$. The procedure continues until we obtain $\hat{y}_T$

and the predicted values are denoted by:

$$\hat{y}_{T_1+1},\, \hat{y}_{T_1+2},\, \cdots,\, \hat{y}_T$$

In order to compare the forecast accuracy, we calculate the out-of-sample R^2 following the work of Campbell and Thompson (2008) and compute the out-of-sample R^2 statistic:

$$R^2_{OOS} = 1 - \frac{T_2^{-1} \sum_{s=1}^{T_2} \left(y_{T_1+s} - \hat{y}_{T_1+s}\right)^2}{T_2^{-1} \sum_{s=1}^{T_2} \left(y_{T_1+s} - \bar{y}_{T_1+s}\right)^2},$$

where $T_2 = T - T_1$ with T_1 is the number of observations in the (in-sample) estimation period, $\hat{y}_{t+1} = f\left(x_t'\tilde{\theta}, \tilde{\gamma}\right)$ is the forecast in period $t+1$, and $\bar{y}_{t+1} = t^{-1}\sum_{s=1}^{t} y_s$ is the historical average benchmark forecast. The single-index model has a forecasting performance that beats the benchmark if $R^2_{OOS} > 0$. We report out-of-sample results in a set of figures that show the R^2_{OOS} statistic on the vertical axis and the beginning of the various out-of-sample evaluation periods on the horizontal axis. This will provide us with an assessment of the robustness of the out-of-sample forecasting results.

Figures 1-4 shows some of the OOS results we have. The red horizontal line represents 0. If the blue line (R^2_{OOS}) lies above the red line, it means the OOS forecasts generated by our nonlinear models outperform the historical average benchmark. Here, we present only the combinations of predictors and nonlinear functions that are found to produce positive R^2_{OOS} values over a substantial number of out-of-sample periods.

Figure 2.1 and Figure 2.2 show that for some non-cointegrated predictors studied previously, when using $f_3(u, \gamma)$ as a nonlinear function, the following combinations of predictors generate out-of-sample gains: (a) earning-price ratio, book-to-market ratio and term spread (see Figure 2.1); (b) dividend-price ratio and dividend-payout ratio (see Figure 2.2). In particular, Figure 2.1 shows that the multivariate predictors generate consecutive out-of-sample gains after 2014.

The result presented in Figure 2.3 is obtained using the 3 cointegrated predictors c_{t-1}, a_{t-1} and y_{t-1} and using $f_3(u, \gamma)$. These predictors have been studied in Lettau and Ludvigson (2001) and we find that, using our model, this combination of cointegrated predicotrs generates better forecasts than the historical average benchmark around the year 2000 and at the end of the forecasting period.

Figure 2.1: Predictors: earning-price ratio, book-to-market ratio, term spread and $f_3(u, \gamma)$

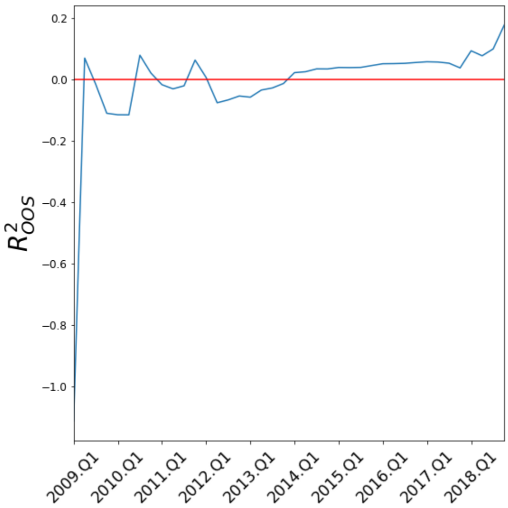


Figure 2.2: Predictors: dividend price, dividend-payout and $f_3(u, \gamma)$

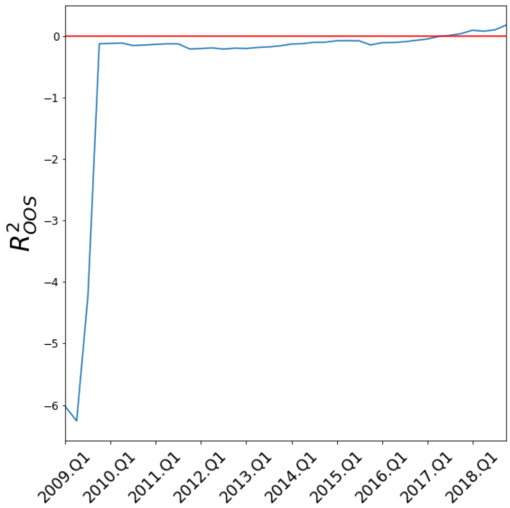


Figure 2.3: Predictors: log consumption, log asset wealth and log labor income and $f_3(u, \gamma)$

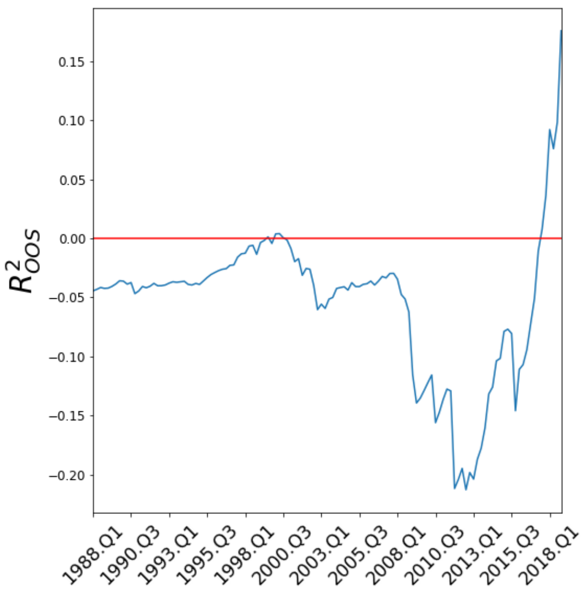
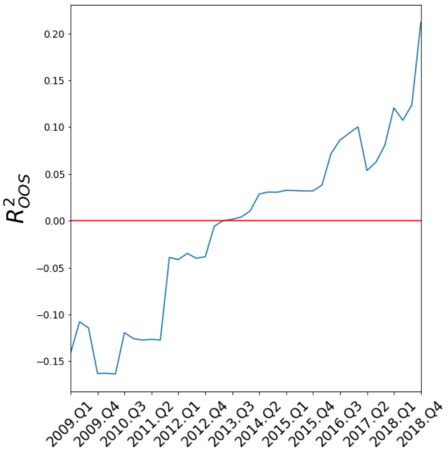
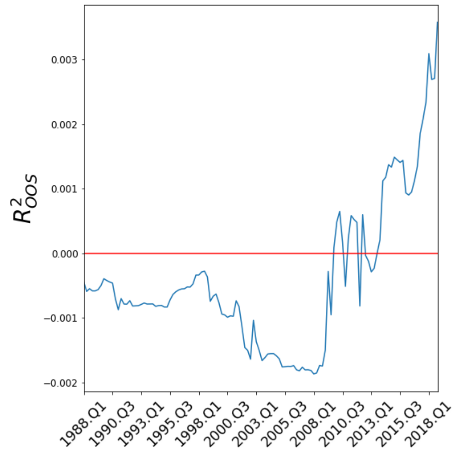


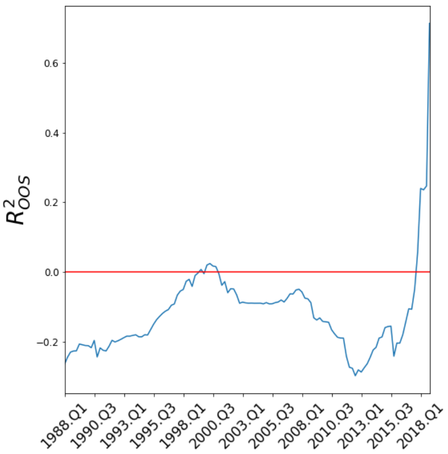
Figure 2.4: Forecasting results using cointegrated predictors



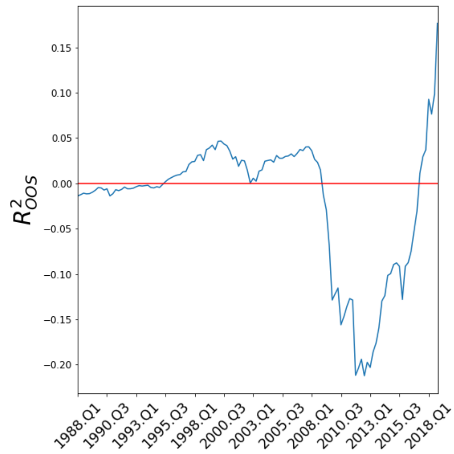
(a) Predictors: long-term yield and T-bill rate and $f_1(u, \gamma)$



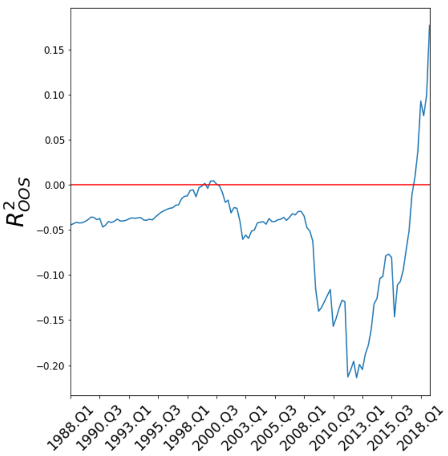
(b) Predictors: long-term yield and T-bill rate and $f_4(u, \gamma)$



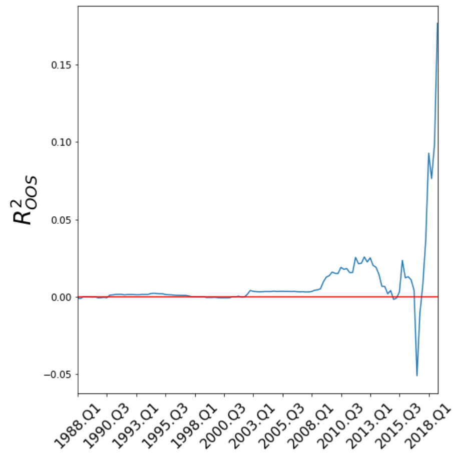
(c) Predictors: long-term yield and T-bill rate and $f_5(u, \gamma)$



(d) Predictors: dividend-price and earning-price and $f_3(u, \gamma)$



(e) Predictors: dividend-price and dividend yield and $f_3(u, \gamma)$



(f) Predictors: BAA and AAA and $f_3(u, \gamma)$

Figure 2.4 displays the results when we use pairs of cointegrated predictors considered in Zhou et al. (2018). As shown in Figure 2.4 (a), (b) and (c), the pair of long-term yield and 3-month T-bill rate produces $R^2_{OOS} > 0$ when using $f_1(u, \gamma)$, $f_4(u, \gamma)$ and $f_5(u, \gamma)$ as nonlinear functions. The function $f_1(u, \gamma)$ produces positive $R^2_{OOS} > 0$ after 2013, while the function $f_4(u, \gamma)$ gives consecutive positive results after 2009. We also find that when using $f_3(u, \gamma)$ as a nonlinear function, the following three pairs of predictors generate positive R^2_{OOS} values: (a) dividend-price ratio and earning-price ratio; (b) dividend-price ratio and dividend yield; and (c) baa- and aaa- rated corporate bond yields (see Figures 2.4 (c), (d) and (e)). From Figure 2.4 (d), we can see that the dividend-price ratio and earning-price ratio not only produce positive $R^2_{OOS} > 0$ at the end of the forecasting period, but generates consecutive better forecasts from around 1995 to 2008. In Figure 2.4 (e), the combination of aaa- and baa- rated corporate bond rates produces long and continuous positive $R^2_{OOS} > 0$ over the out-of-sample forecasting period.

These findings indicate that exploiting nonlinearities in the data can lead to improved forecast accuracy relative to the historical average benchmark. Overall, the single-index predictive model shows that the largest out-of-sample forecast gains in all the figures come from using the combination of aaa- and baa- rated corporate bond rates, and using the nonlinear function $f_3(u, \gamma)$.

5 Conclusion

This chapter considers a parametric single-index predictive model with integrated predictors. We propose a new constrained estimation procedure to estimate this model and show that it has better finite sample properties than the usual NLS estimator. We apply the model to examine the predictability of stock returns using cointegrated and non-cointegrated predictors. We find that several combinations of the predictors used in prior studies deliver out-of-sample forecasting gains relative to the standard historical average benchmark over a large number of evaluation periods when using the single-index predictive model that accounts for the nonlinearities in the time-series data.

Chapter 3

Partially Nonlinear Single-Index Predictive Models

1 Introduction

Conventional partially linear models involving both parametric and nonparametric components have been widely studied in the literature, such as by Gao (2007). In the time series literature, partially linear single-index models have also attracted attention in recent years, see for example Dong et al. (2016). We are now interested in a new class of nonlinear time series models - partially nonlinear single-index models of the form:

$$y_t = \beta_0' z_t + g\left(x_{t-1}'\theta_0; \gamma_0\right) + e_t, \quad t = 2, ..., T, \tag{1.1}$$

where $z_t = (y_{t-1}, \cdots, y_{t-p}, w_{t-1}')'$, in which w_{t-1} is a vector of stationary predictors, $g(.,.)$ is a known univariate nonlinear function, x_{t-1} is a d-dimensional integrated process of order one, θ_0 is a d-dimensional unknown true parameter vector that lies in the parameter set Θ, γ_0 is a m-dimensional unknown true parameter vector that lies in the parameter set Γ and e_t is a martingale difference process. The parameter sets Θ and Γ are assumed to be compact and convex subsets of $\mathbb{R}^d$ and $\mathbb{R}^m$ respectively. In order to ensure that θ_0 is uniquely identifiable, we will need to impose $\theta_0'\theta_0 = 1$.

Our model allows for lagged dependent variables because key macroeconomic/financial variables, such as the growth rate of GDP, the rate of unemployment and interest rates are typically autocorrelated. Failing to account for this autocorrelation will lead to serially correlated residuals. We are thus interested in using this model to assess whether including lagged dependent variables would, in fact, improve forecasts of y_t relative to using only the nonlinear single-index component, $g\left(x_{t-1}'\theta_0;\gamma_0\right)$, containing either the cointegrated predictors or non-cointegrated predictors. Our model may also be useful in cases where there are additional stationary predictors, w_{t-1}, for which the linear specification fits the data better than the nonlinear specification.

There is a considerable theoretical effort being put into developing new estimation method of the partial linear model (see for example, Dong et al. (2016)) and nonlinear models with single-index (see for example Chang and Park (2003)). In our study, we propose a novel 2-step estimation method in which β will have a closed form solution while θ and γ can be estimated by the method of nonlinear least squares or constrained nonlinear least squares.

This study aims to use the Monte Carlo simulation method to investigate the finite sample properties of the estimators. There will also be an empirical analysis to study the predictability of stock returns. We will use the dataset from Welch and Goyal (2008) and investigate the out-of-sample forecast ability of model (1.1). Compared with chapter 2, we will include lagged dependent variables in the model. We will investigate whether the partially nonlinear single-index model will help further improve the stock return predictability in my future research.

2 Model and Methodology

Since in our case, $g\left(x_{t-1}'\theta_0;\gamma_0\right)$ is known, model (1.1) can be estimated by using a nonlinear least square method. Let $L(\beta,\theta,\gamma)=\sum_{t=1}^{T}\left(y_t-\beta' z_t-g\left(x_{t-1}'\theta;\gamma\right)\right)^2$, and hence we have the following score functions:

$$\frac{\partial L(\beta,\theta,\gamma)}{\partial\beta}=-2\sum_{t=1}^{T}z_t'\left(y_t-\beta' z_t-g\left(x_{t-1}'\theta;\gamma\right)\right)$$

$$\frac{\partial L(\beta,\theta,\gamma)}{\partial\theta}=-2\sum_{t=1}^{T}\left(y_t-\beta' z_t-g\left(x_{t-1}'\theta;\gamma\right)\right)\frac{\partial g(x_{t-1}'\theta;\gamma)}{\partial\theta}$$

$$\frac{\partial L(\beta,\theta,\gamma)}{\partial\gamma}=-2\sum_{t=1}^{T}\left(y_t-\beta' z_t-g\left(x_{t-1}'\theta;\gamma\right)\right)\frac{\partial g(x_{t-1}'\theta;\gamma)}{\partial\gamma}$$

The minimum value of $L(\beta,\theta,\gamma)$ occurs when the above score functions equals to 0. Notice that these score functions $\frac{\partial L(\beta,\theta,\gamma)}{\partial\theta}$ and $\frac{\partial L(\beta,\theta,\gamma)}{\partial\gamma}$ are nonlinear functions of both the variables and the parameters, and so they do not have closed form solutions. To estimate the parameters, we need to include an iterative procedure and obtain optimal values by using gradient descent algorithms. However, recognising that this score function $\frac{\partial L(\beta,\theta,\gamma)}{\partial \beta}$ is linear in the parameter β, we introduce a novel two-step approach for estimation in order to reduce the computational burden.

In step 1, we set $\frac{\partial L(\beta,\theta,\gamma)}{\partial \beta} = 0$. Solve the equation and we have:

$$\hat{\beta} = \left(\sum_{t=1}^{T} z_t z_t'\right)^{-1} \sum_{t=1}^{T} \Big(y_t - g\left(x_{t-1}'\theta;\gamma\right)\Big) z_t \tag{2.1}$$

In other words, $\hat{\beta}$ is of a linear from by OLS expression. Thus model (1.1) can be approximated by:

$$y_t = \hat{\beta}' z_t + g\left(x_{t-1}'\theta;\gamma\right) + e_t,$$

which is equivalent to:

$$y_t - \left(\left(\sum_{t=1}^{T} z_t z_t'\right)^{-1} \sum_{t=1}^{T} y_t z_t\right)' z_t = g\left(x_{t-1}'\theta;\gamma\right) - z_t'\left(\sum_{t=1}^{T} z_t z_t'\right)^{-1} \sum_{t=1}^{T} g\left(x_{t-1}'\theta;\gamma\right) z_t + e_t.$$

Let

$$\tilde{y} = y_t - z_t'\left(\sum_{t=1}^{T} z_t z_t'\right)^{-1} \sum_{t=1}^{T} y_t z_t,$$

$$\tilde{g}(x_{t-1}'\theta;\gamma) = g\left(x_{t-1}'\theta;\gamma\right) - z_t'\left(\sum_{t=1}^{T} z_t z_t'\right)^{-1} \sum_{t=1}^{T} g\left(x_{t-1}'\theta;\gamma\right) z_t$$

Then we have an approximate model of the form:

$$\tilde{y} = \tilde{g}\left(x_{t-1}'\theta;\gamma\right) + e_t. \tag{2.2}$$

We can estimate (θ,γ) by minimizing:

$$Q_T(\theta,\gamma) = \sum_{t=1}^{T} \Big(\tilde{y}_t - \tilde{g}\left(x_{t-1}'\theta,\gamma\right)\Big)^2$$

over $(\theta,\gamma) \in (\Theta,\Gamma)$. The NLS estimator $(\hat{\theta},\hat{\gamma})$ is given by:

$$\left(\widehat{\theta},\widehat{\gamma}\right) = \arg\min_{\theta\in\Theta,\gamma\in\Gamma} Q_T\left(\theta,\gamma\right),$$

which can be solved using an iterative procedure since there is no closed form solution. In order to improve finite sample properties of the estimators, we impose a truncation condition $I\left(\|x_{t-1}\| \leq M_T\right)$ on x_{t-1} and an identification condition on coefficient vector θ. We then define the modified sum-of-squared errors by:

$$Q_{T,M}(\theta,\gamma) = \sum_{t=1}^{T}\left(y_t - f\left(x'_{t-1}\theta,\gamma\right)\right)^2 I\left(\|x_{t-1}\| \leq M_T\right) + \lambda\left(\|\theta\|^2 - 1\right),$$

where $I(.)$ denotes the indicator function, $\|.\|$ is the Euclidean norm, M_T is a positive and increasing sequence satisfying $M_T \to \infty$ as $T \to \infty$ and λ is a Lagrange multiplier.

The constrained least squares (denoted CLS) estimator $\widetilde{\theta}$ and $\widetilde{\gamma}$ is given by minimizing $Q_{T,M}(\theta,\gamma)$ over $\theta \in \Theta$ and $\gamma \in \Gamma$ such that the restriction $\|\theta\|^2 = 1$ holds; that is

$$\left(\widetilde{\theta},\widetilde{\gamma}\right) = \arg\min_{\theta\in\Theta,\gamma\in\Gamma,\|\theta\|^2=1} Q_{T,M}(\theta,\gamma).$$

In step 2, using equation (2.1), β may be re-estimated by:

$$\hat{\beta} = \left(\sum_{t=1}^{T} z_t z'_t\right)^{-1} \sum_{t=1}^{T}\left(y_t - g\left(x'_{t-1}\tilde{\theta};\tilde{\gamma}\right)\right) z_t.$$

3 Monte Carlo Simulation

3.1 Data Generation Process

We investigate the finite sample properties of the NLS and the proposed CLS estimators for partially nonlinear model in multivariate nonstationary settings. The predictors x_{t-1} is a 2-vector integrated time series. Data were generated on the following models:

$$y_t = \beta_{1,0}y_{t-1} + \beta_{2,0}w_{t-1} + f\left(x'_{t-1}\theta_0,\gamma_0\right) + e_t, \quad e_t \sim i.i.d.N(0,1), \quad t = 2,...,T,$$

with

$$w_t = 0.8 * w_{t-1} + s_t, \quad s_t \sim i.i.d.N(0,1),$$

$$x_t = x_{t-1} + v_t$$

In the data generation process, we consider $\hat{\theta}_0 = (0.8,-0.6)'$, $\hat{\gamma}_0 = (0.2,0.3,0.3)'$ $\beta_{1,0} = 0.5$ and $\beta_{2,0} = 1.0$.

To generate co-integrated x_t, we follow a vector integrated process driven by an MA(1) innovations and construct v_t as:

$$v_t = \epsilon_t + C\epsilon_{t-1},$$

where $\epsilon_t \sim i.i.d.N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$ and $C = \begin{pmatrix} -1 & 4/3 \\ 0 & 0 \end{pmatrix}$.

As for $f\left(x'_{t-1}\theta_0, \gamma_0\right)$, we consider the following nonlinear functions:

$$\begin{aligned}
\sin : f_1\left(u_{t-1}, \gamma_0\right) &= \sin\left(u_{t-1} + \gamma_{1,0}\right), \\
\cos : f_2\left(u_{t-1}, \gamma_0\right) &= \cos\left(u_{t-1} + \gamma_{1,0}\right), \\
\text{sin_scaled} : f_3\left(u_{t-1}, \gamma_0\right) &= \sin\left(\gamma_{1,0}u_{t-1} + \gamma_{2,0}\right), \\
\text{cos_scaled} : f_4\left(u_{t-1}, \gamma_0\right) &= \cos\left(\gamma_{1,0}u_{t-1} + \gamma_{2,0}\right), \\
\text{exp_shift} : f_5\left(u_{t-1}, \gamma_0\right) &= 1 - e^{-\gamma_{1,0}\left(u_{t-1} - \gamma_{2,0}\right)^2} \\
\exp : f_6\left(u_{t-1}, \gamma_0\right) &= \gamma_{1,0}e^{-\gamma_{2,0}u_{t-1}^2} \\
\text{Polynomial} : f_7\left(u_{t-1}, \gamma_0\right) &= \gamma_{1,0} + \gamma_{2,0}u_{t-1} + \gamma_{3,0}u_{t-1}^2
\end{aligned}$$

where $u_{t-1} = x'_{t-1}\theta_0$.

In our simulation study, we consider sample sizes $T = 100, 500, 1000$, replication time M $= 5000$ and the following statistics. Take θ as an example:

$$\text{bias} = \overline{\widehat{\theta}}_i - \theta_{i,0},$$

where $\overline{\widehat{\theta}}_i = M^{-1}\sum_{r=1}^{M}\widehat{\theta}_i^{(r)}$; and

$$\text{standard deviation (std)} = \sqrt{M^{-1}\sum_{r=1}^{M}\left(\widehat{\theta}_i^{(r)} - \overline{\widehat{\theta}}_i\right)^2}.$$

Since $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are correlated, we also calculate a type of estimated covariance of the form:

$$\sigma_\theta = \frac{1}{M}\sum_{r=1}^{M}\left(\widehat{\theta}_i^{(r)} - \overline{\widehat{\theta}}_i\right)\left(\widehat{\theta}_j^{(r)} - \overline{\widehat{\theta}}_j\right), \qquad \text{std}_\theta = \sqrt{\sum_{i,j}\sigma_{ij}^2}.$$

where $\widehat{\theta}^{(r)}$ denote the r-th replication of the estimate.

Following the above definitions, we then calculate biases, standard deviations for β_i and γ_i, and also σ_β and σ_γ.

3.2 Initial Values

As our model is partially nonlinear, the gradient functions do not have a closed solution. Therefore, we use an iterative procedure to estimate the model. To start the iterative procedure, initial values are necessary. We can assign random values as the initial values, however, finding initial values that are close to the optimal values is crucial.

To find better initial values, we follow a 2-step procedure and consider Taylor expansions to approximate the partially nonlinear models. In the first step, we calculate initial values for θ by using the following two steps.

Step 1:

$$x_1 = \alpha x_2 + e_t, \quad e_t \sim i.i.d.N\left(0, 1\right), \quad t = 2, ..., T$$

$\hat{\alpha}$ can be obtained by linear regression.

Step 2: Then the initial values of θ is given by:

$$\theta_0 = (\frac{1}{\sqrt{1+\hat{\alpha}^2}}, -\frac{\hat{\alpha}}{\sqrt{1+\hat{\alpha}^2}})$$

It satisfies the constrain that $\|\theta\| = 1$. The true values of θ calculated are around (0.6, -0.8)

Then for different functional forms, we first substitute the nonlinear function f by its Taylor expansion and then calculate initial values for β and γ by estimating a linear model.

For the first case with trigonometric function $f_1\left(u_{t-1}, \gamma_0\right) = \sin\left(u_{t-1} + \gamma_{1,0}\right)$, we can use its first order Taylor expansion:

$$\tilde{f}_1 = \sin\left(u_{t-1} + \gamma_{1,0}\right) \sim \left(u_{t-1} + \gamma_{1,0}\right)$$

to approximate the nonlinear function. Then, β and γ are obtained by estimating the following regression:

$$y_t = (\gamma_1 + u_{t-1}) + \beta_1 y_{t-1} + \beta_2 w_{t-1}.$$

Similarly, for the second case $f_1\left(u_{t-1}, \gamma_0\right) = \cos\left(u_{t-1} + \gamma_{1,0}\right)$, we can use its first order Taylor expansion:

$$\tilde{f}_2 = \cos\left(u_{t-1} + \gamma_{1,0}\right) \sim 1 - \left(u_{t-1} + \gamma_{1,0}\right)^2$$

to approximate the nonlinear function. Then, β and γ are obtained by estimating the following regression:

$$y_t = 1 - \left(\gamma_1 + u_{t-1}\right)^2 + \beta_1 y_{t-1} + \beta_2 w_{t-1}.$$

Then for the two scaled trigonometric function (functional form 3 - 4), β and γ are calculated by the following model:

$$y_t = \gamma_2 + \gamma_1 u_{t-1} + \beta_1 y_{t-1} + \beta_2 w_{t-1}$$

and

$$y_t = 1 - (\gamma_2 + \gamma_1 u_{t-1})^2 + \beta_1 y_{t-1} + \beta_2 w_{t-1}.$$

For functional 5, to better approximate the function, we use the second order Taylor expansion and calculate the initial values using linear regression. Therefore, we first approximate the nonlinear function $1 - e^{-\gamma_{1,0}(u_{t-1}-\gamma_{2,0})^2}$ by:

$$\tilde{f}_5 = \gamma_{1,0}(u_{t-1} - \gamma_{2,0})^2 - \frac{\gamma_{1,0}^2(u_{t-1} - \gamma_{2,0})^4}{2}$$

and β and γ are obtained by estimating the following regression:

$$y_t = \gamma_{1,0} u_{t-1}^2 - 2\gamma_{1,0}\gamma_{2,0} u_{t-1} + \gamma_{1,0}\gamma_{2,0}^2 - \frac{\gamma_{1,0}^2(u_{t-1} - \gamma_{2,0})^4}{2} + \beta_1 y_{t-1} + \beta_2 w_{t-1}$$

and $\gamma_{2,0}$ can be recovered from the coefficient of u_{t-1} (denoted by η) by $\gamma_{2,0} = -\eta/2\gamma_{1,0}$.

For functional 6, we use third order Taylor expansion and approximate the nonlinear function $\gamma_{1,0} e^{-\gamma_{2,0} u_{t-1}^2}$ by:

$$\tilde{f}_6 = \gamma_{1,0}(1 - \gamma_{2,0} u_{t-1}^2 + \frac{\gamma_{2,0}^2 u_{t-1}^4}{2} - \frac{\gamma_{2,0}^3 u_{t-1}^6}{3})$$

and β and γ are obtained by estimating the following regression:

$$y_t = \tilde{f}_6 = \gamma_{1,0}(1 - \gamma_{2,0} u_{t-1}^2 + \frac{\gamma_{2,0}^2 u_{t-1}^4}{2} - \frac{\gamma_{2,0}^3 u_{t-1}^6}{3}) + \beta_1 y_{t-1} + \beta_2 w_{t-1}$$

where $\gamma_{2,0}$ can be recovered from the coefficient of u_{t-1}^2 (denoted by η) by $-\eta/\gamma_{1,0}$.

For functional 7, the polynomial function, we can then get the initial values for β and γ by doing the following linear regression:

$$y_t = \beta_1 y_{t-1} + \beta_2 w_{t-1} + \gamma_1 + \gamma_2 u_{t-1} + \gamma_3 u_{t-1}^2$$

where $u_{t-1} = x_{t-1}'\theta_0$.

3.3 Simulation Results for Co-integrated x_t

Table 3.1 to 3.4 show the simulation results on co-integrated x_t using partially nonlinear models. From the simulation results, we find that:

1. Both NLS and constrained NLS estimators converge when sample size increases.
2. The constrained NLS perform better than the normal NLS for most functional forms as it gives smaller biases and standard deviations.
3. Among the 7 different partially nonlinear, the polynomial function $(f_7(u_{t-1}, \gamma_0))$ gives the best performance.

As presented in the tables, both NLS and constrained NLS estimators converge for most functional forms when sample size increases. For example, in table 3.2, the absolute value of bias of θ_1 using $f_3(u_{t-1}, \gamma_0)$ decreases from 0.00219 to 0.00062 in the case of constrained NLS. And in the case of NLS, the number also decreases from 0.01336 to 0.00201.

In addition, the constrained NLS estimators have a better performance than the NLS estimators as the magnitude of bias and standard deviation is smaller. Take θ in table 3.1 as an example, for $f_1(u_{t-1}, \gamma_0)$, when sample size T=1000, the bias of θ_1 and θ_2 using constrained NLS is 0.00129 and 0.00097 respectively, while the corresponding values using normal NLS is -0.00818 and 0.01622. Similarly for $f_2(u_{t-1}, \gamma_0)$, when T = 1000, the standard deviation for θ is 0.00476 in the case of constrained NLS and 0.02110 in the case of NLS.

Among the 7 partially nonlinear models, we find that the polynomial functional form tend to provide the best results. As shown in table 3.4, the standard deviation of θ using constrained NLS is 0.00045 when T=1000. The corresponding value is 0.00262 in $f_1(u_{t-1}, \gamma_0)$ and 0.00476 in $f_2(u_{t-1}, \gamma_0)$.

Table 3.1: Simulation Results for models containing $f_1(u_{t-1}, \gamma_0)$ and $f_2(u_{t-1}, \gamma_0)$

		NLS					Constrained-NLS				
		θ_1	θ_2	β_1	β_2	γ_1	θ_1	θ_2	β_1	β_2	γ_1
$f_1(u_{t-1}, \gamma_0)$											
	Bias	0.00221	0.01180	-0.01443	0.01320	0.00447	0.01287	0.01004	-0.01604	-0.00325	-0.02298
T = 100	std	0.09413	0.08928	0.05052	0.06048	0.52541	0.01474	0.01180	0.04778	0.05835	0.46140
		0.08751		0.03648			0.02653		0.04417		
	Bias	-0.00776	0.01514	-0.01165	0.01131	-0.00464	0.00247	0.00186	-0.01223	0.00879	-0.01146
T = 500	std	0.05713	0.04579	0.02192	0.02447	0.30383	0.00289	0.00219	0.02094	0.02295	0.29016
		0.04800		0.01013			0.00508		0.00918		
	Bias	-0.00818	0.01622	-0.01119	0.01126	-0.00584	0.00129	0.00097	-0.01176	0.01021	-0.02066
T = 1000	std	0.06111	0.04695	0.01541	0.01690	0.29261	0.00149	0.00112	0.01526	0.01632	0.27447
		0.05325		0.00590			0.00262		0.00471		
$f_2(u_{t-1}, \gamma_0)$											
	Bias	0.00069	0.00276	-0.00646	0.00180	0.03309	-0.00189	-0.00192	0.00104	-0.00143	0.04245
T = 100	std	0.04767	0.03619	0.03458	0.04583	0.12469	0.02647	0.02972	0.01103	0.01648	0.08817
		0.08369		0.04306			0.03984		0.01916		
	Bias	0.00068	0.00075	-0.01248	0.01028	-0.00206	-0.00010	-0.00026	-0.00049	0.00061	0.00770
T = 500	std	0.01564	0.01175	0.01493	0.02015	0.05825	0.00622	0.00698	0.00272	0.00421	0.04282
		0.02738		0.01652			0.00927		0.00518		
	Bias	-0.00012	0.00005	-0.01275	0.01034	0.00029	-0.00015	0.00087	-0.00041	0.00051	0.00100
T = 1000	std	0.01206	0.00905	0.01188	0.01482	0.04136	0.00329	0.00330	0.00156	0.00191	0.02102
		0.02110		0.01129			0.00476		0.00242		

Table 3.2: Simulation Results for $f_3(u_{t-1}, \gamma_0)$ and $f_4(u_{t-1}, \gamma_0)$

		NLS						Constrained-NLS					
		θ_1	θ_2	β_1	β_2	γ_1	γ_2	θ_1	θ_2	β_1	β_2	γ_1	γ_2
		$f_3(u_{t-1}, \gamma_0)$											
	Bias	-0.01336	-0.00146	0.00040	0.00012	0.00258	0.03875	-0.00219	0.00367	-0.01027	0.00355	0.00143	-0.00586
T = 100	std	0.09908	0.10823	0.01118	0.01976	0.02365	0.11074	0.07270	0.05655	0.03583	0.05240	0.06347	0.12251
		0.14788		0.02250		0.11450		0.12888		0.05136		0.13895	
	Bias	-0.00561	-0.00664	-0.00106	0.00100	0.00074	0.00084	0.00039	0.00178	-0.01354	0.00635	-0.00269	-0.00414
T = 500	std	0.02504	0.02534	0.00283	0.00462	0.00631	0.03351	0.03891	0.02937	0.01772	0.02257	0.02483	0.04649
		0.03833		0.00539		0.03384		0.06821		0.02083		0.05447	
	Bias	-0.00201	-0.00204	-0.00048	0.00044	0.00039	-0.00138	0.00062	0.00143	-0.01299	0.00738	-0.00171	-0.00021
T = 1000	std	0.01434	0.01451	0.00197	0.00230	0.00392	0.01576	0.03144	0.02371	0.01291	0.01631	0.01754	0.03414
		0.02283		0.00277		0.01611		0.05511		0.01493		0.04048	
		$f_4(u_{t-1}, \gamma_0)$											
	Bias	-0.00462	0.00898	-0.00485	0.00572	0.00210	0.03106	-0.00283	0.00192	0.00022	-0.00015	-0.01107	-0.00375
T = 100	std	0.10139	0.10403	0.03259	0.05225	0.01558	0.10063	0.06346	0.04934	0.01071	0.02085	0.05994	0.18408
		0.14118		0.04906		0.17287		0.11249		0.02418		0.19193	
	Bias	-0.00586	0.00356	-0.01139	0.01223	0.00094	0.00065	-0.01064	-0.00614	-0.00082	0.00065	-0.00232	-0.02078
T = 500	std	0.01965	0.02729	0.01540	0.02242	0.00549	0.03119	0.04250	0.03141	0.00315	0.00483	0.02115	0.07684
		0.03187		0.02015		0.12557		0.07384		0.00576		0.07694	
	Bias	-0.00286	0.00068	-0.00984	0.01005	0.00026	-0.00096	-0.00967	-0.00599	-0.00052	0.00048	-0.00040	-0.01090
T = 1000	std	0.01282	0.01323	0.01171	0.01616	0.00364	0.01763	0.03508	0.02570	0.00194	0.00233	0.01420	0.05158
		0.01564		0.01282		0.09848		0.06074		0.00272		0.05015	

Table 3.3: Simulation Results for $f_5(u_{t-1}, \gamma_0)$ and $f_6(u_{t-1}, \gamma_0)$

		NLS						Constrained-NLS					
		θ_1	θ_2	β_1	β_2	γ_1	γ_2	θ_1	θ_2	β_1	β_2	γ_1	γ_2
$f_5(u_{t-1}, \gamma_0)$													
	Bias	0.14852	0.07716	-0.00060	-0.00069	0.03979	0.09704	0.00049	0.00707	-0.00837	0.00270	0.05869	0.11357
T = 100	std	0.36721	0.25730	0.00983	0.01876	0.08946	0.14248	0.08064	0.06468	0.03280	0.04974	0.22359	0.36681
		0.47514		0.02122		0.16524		0.14459		0.04685		0.44678	
	Bias	0.06033	0.02538	-0.00080	0.00085	0.03530	0.09690	-0.00189	0.00011	-0.01111	0.00823	0.00059	0.02411
T = 500	std	0.25251	0.16061	0.00290	0.00427	0.08116	0.14316	0.03955	0.02972	0.01524	0.01987	0.04029	0.17323
		0.30811		0.00490		0.16249		0.06920		0.01641		0.17454	
	Bias	0.04737	0.02568	-0.00051	0.00050	0.02912	0.08033	-0.00377	-0.00197	-0.01053	0.00843	0.00041	-0.00583
T = 1000	std	0.21925	0.14210	0.00204	0.00215	0.07485	0.13657	0.02952	0.02203	0.01198	0.01481	0.02361	0.11419
		0.28125		0.00249		0.15389		0.05151		0.01177		0.11498	
$f_6(u_{t-1}, \gamma_0)$													
	Bias	0.29939	0.12435	-0.00836	0.00577	0.09685	0.86918	0.00239	0.00565	-0.00015	0.00038	0.10008	0.10486
T = 100	std	0.44703	0.31855	0.03431	0.04555	0.26677	1.16532	0.06176	0.04835	0.01144	0.01696	0.10519	0.14436
		0.52722		0.04150		1.23387		0.10976		0.01892		0.16633	
	Bias	0.20686	0.12096	-0.01268	0.01065	0.01278	0.33971	-0.00633	-0.00314	-0.00108	0.00100	0.08506	0.11410
T = 500	std	0.39309	0.27335	0.01578	0.02035	0.09200	0.81052	0.04018	0.03030	0.00304	0.00424	0.09984	0.14511
		0.47572		0.01677		0.82640		0.07041		0.00456		0.16128	
	Bias	0.18891	0.09882	-0.01164	0.01017	0.00430	0.17839	-0.01067	-0.00681	-0.00047	0.00036	0.07620	0.11676
T = 1000	std	0.37535	0.25110	0.01210	0.01538	0.06249	0.54192	0.03382	0.02468	0.00180	0.00239	0.09744	0.14438
		0.45836		0.01184		0.56098		0.05847		0.00263		0.16354	

Table 3.4: Simulation Results for $f_7(u_{t-1}, \gamma_0)$

		NLS							Constrained-NLS						
		θ_1	θ_2	β_1	β_2	γ_1	γ_2	γ_3	θ_1	θ_2	β_1	β_2	γ_1	γ_2	γ_3
	Bias	0.00018	-0.00018	-0.00056	0.00037	0.02711	-0.00147	0.00020	0.00088	0.00026	-0.00079	0.00029	0.00326	-0.00064	0.00121
T = 100	std	0.00407	0.00335	0.00171	0.01482	0.07001	0.00886	0.00308	0.00139	0.00105	0.00161	0.00504	0.00914	0.00451	0.00141
		0.00528		0.01502			0.07063		0.00244		0.00461			0.01029	
	Bias	0.00006	0.00003	-0.00003	0.00024	0.00261	-0.00072	0.00026	-0.00013	-0.00008	-0.00040	-0.00058	0.00063	-0.00032	0.00146
T = 500	std	0.00051	0.00040	0.00039	0.00354	0.01664	0.00315	0.00047	0.00041	0.00031	0.00068	0.00206	0.00436	0.00175	0.00043
		0.00068		0.00357			0.01695		0.00071		0.00167			0.00472	
	Bias	0.00008	0.00004	-0.00001	-0.00014	0.00218	-0.00008	-0.00006	-0.00005	-0.00003	-0.00025	-0.00012	0.00035	-0.00021	0.00087
T = 1000	std	0.00026	0.00015	0.00015	0.00194	0.01219	0.00173	0.00020	0.00026	0.00019	0.00043	0.00136	0.00319	0.00110	0.00026
		0.00032		0.00190			0.01231		0.00045		0.00110			0.00339	

4 Empirical Study

To illustrate the use of our partially nonlinear model, we conduct the in-sample and out-of-sample prediction exercises using U.S. stock market returns data. The datasets is available from Amit Goyal's website and it is quarterly data ranging from 1956 Q1 to 2018 Q4 . The dependent variable, stock returns, are measured as continuously compound returns on the S&P 500 index. Predictors used in Welch and Goyal (2008) include dividend-price ratio (log), dividend yield (log), earnings-price ratio (log), dividend-payout ratio (log), and other 11 predictors.

In our study, we want to evaluate the performance of the partially nonlinear models when the non-stationary predictors are co-integrated, therefore we choose the following 4 variable pairs from the Welch and Goyal (2008) datasets, which has been found to be co-integrated (as in Zhou et al. (2018)):

- co1: dividend-price ratio (dp) and dividend yield (dy).

  Dividend-price ratio is the difference between the log of dividends and the log of stock prices; dividend yield is the difference between the log of dividends and the log of lagged stock prices.

- co2: T-bill rate (tbl) and long-term yield (lty).

  T-bill rate is the interest rate on a three-month Treasury bill; long-term yield stands for the long-term government bond yield.

- co3: dividend-price ratio and earning-price ratio.

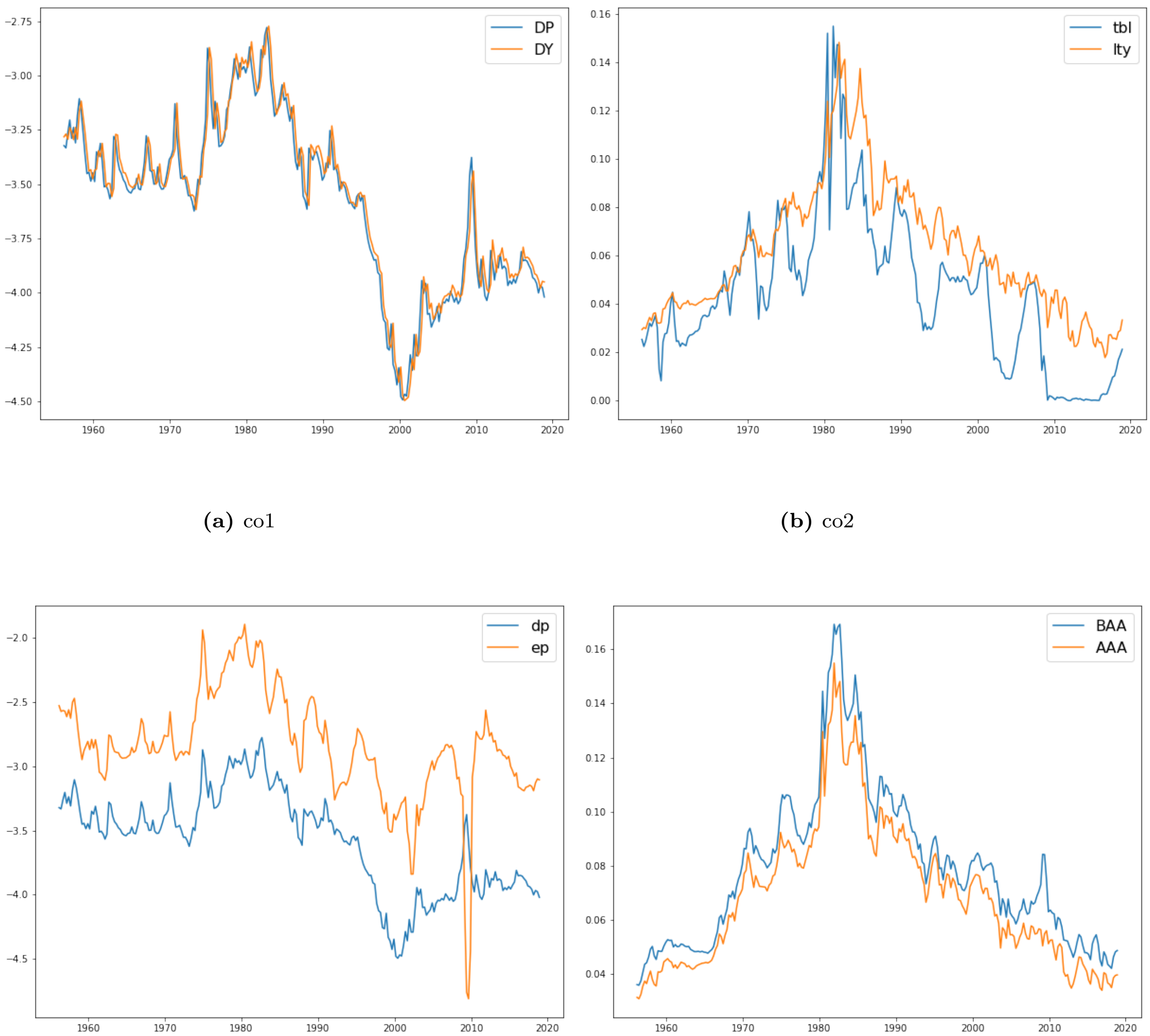
  The earning-price ratio is the difference between the log of earnings on the S&P 500 index and the log of stock prices.

- co4: baa- and aaa-rated corporate bonds yields.

Figure 3.1 plots the time series of the co-integrated variable pairs. It is clear that the variables in each plot are not stationary and share a similar trend.

Apart from the non-stationary variables, the partially nonlinear models also consider two stationary variables: the lagged stock return (y_{t-1}) and the regression residual from the study of Lettau and Ludvigson (2001). In their study, they show that log consumption, log asset wealth and log labor income share a common stochastic trend and are

Figure 3.1: Time Series Plots of Co-integrated Variables



(a) co1 **(b)** co2

(c) co3 **(d)** co4

co-integrated. The deviations from this shared trend is stationary and can be used as a predictor for stock return.

Therefore in the partially nonlinear models we consider:

$$y_t = \beta_{1,0} y_{t-1} + \beta_{2,0} w_{t-1} + f\left(x'_{t-1}\theta_0, \gamma_0\right) + e_t, \quad e_t \sim i.i.d.N\left(0,1\right), \quad t = 2, ..., T,$$

the dependent variable, y_t, is the equity premium defined as the S&P500 value-weighted log excess returns. y_{t-1} is the lagged equity premium, x_{t-1} are the co-integrated variables from Zhou et al. (2018) and w_{t-1} is the regression residual from Lettau and Ludvigson (2001).

In terms of the nonlinear part $f\left(x'_{t-1}\theta_0, \gamma_0\right)$, we consider the following 7 commonly used nonlinear functions including trigonometric functions, exponential functions and polynomial functions. Then the partially nonlinear models in the empirical study can be written as:

$$\begin{aligned}
\text{sin} : g_1\left(u_{t-1}, \gamma_0\right) &= \beta_0 z'_{t-1} + \sin\left(u_{t-1} + \gamma_{1,0}\right), \\
\text{cos} : g_2\left(u_{t-1}, \gamma_0\right) &= \beta_0 z'_{t-1} + \cos\left(u_{t-1} + \gamma_{1,0}\right), \\
\text{scale_dsin} : g_3\left(u_{t-1}, \gamma_0\right) &= \beta_0 z'_{t-1} + \sin\left(\gamma_{1,0} u_{t-1} + \gamma_{2,0}\right), \\
\text{scale_cos} : g_4\left(u_{t-1}, \gamma_0\right) &= \beta_0 z'_{t-1} + \cos\left(\gamma_{1,0} u_{t-1} + \gamma_{2,0}\right), \\
\text{exp_shift} : g_5\left(u_{t-1}, \gamma_0\right) &= \beta_0 z'_{t-1} + 1 - e^{-\gamma_{1,0}\left(u_{t-1} - \gamma_{2,0}\right)^2} \\
\text{exp} : g_6\left(u_{t-1}, \gamma_0\right) &= \beta_0 z'_{t-1} + \gamma_{1,0} e^{-\gamma_{2,0} u_{t-1}^2} \\
\text{Polynomial} : g_7\left(u_{t-1}, \gamma_0\right) &= \beta_0 z'_{t-1} + \gamma_{1,0} + \gamma_{2,0} u_{t-1} + \gamma_{3,0} u_{t-1}^2
\end{aligned}$$

where $\beta_0 = (\beta_{1,0}, \beta_{2,0})$, $z'_{t-1} = (y_{t-1}, w_{t-1})$, and $u_{t-1} = x'_{t-1}\theta_0$.

In addition, we also include a linear functional form with single-index:

$$\text{constrained\_linear} : g_8\left(u_{t-1}\right) = \gamma_{1,0} + \beta_0 z'_{t-1} + \gamma_{2,0}(x_{1,t-1}\theta_{1,0} + x_{2,t-1}\theta_{2,0}).$$

As for other functional forms, in this constrained linear function, we set $\theta_{1,0}^2 + \theta_{2,0}^2 = 1$ and use the same iterative procedure to estimate the model.

To estimate model g_1 to g_7, we adopt the constrained nonlinear least square method described in section (to be filled). But as mentioned before, when minimizing the nonlinear least squares:

$$S_T(\beta, \theta, \gamma) = \sum_{t=1}^{T} \left(y_t - g\left(u_{t-1}, \gamma\right)\right)^2$$

The gradient equations do not have a closed solution, so we use an iterative algorithm. To implement the iterative algorithm, we must choose a vector of initial values to start. Initial values can be specified by random number generation. One would generate many sets of initial values and then choose the one that leads to a better result. In our case, to better fit the data, we choose the initial values using linear regressions. The detailed steps are shown below:

Step 1: estimate the following linear regression to get $\hat{\alpha}$:

$$x_1 = \alpha x_2 + e_t, \quad e_t \sim i.i.d.N\left(0,1\right), \quad t = 2, ..., T$$

where x_1 and x_2 are the non-stationary co-integrated variables. Then, the initial values for θ is $\theta_0 = (\frac{1}{\sqrt{1+\alpha^2}}, \frac{-\alpha}{\sqrt{1+\alpha^2}})$. As θ_0 is known, we can calculate u_{t-1}.

Step 2: substitute the nonlinear function $f\left(u_{t-1}, \gamma\right)$ by its Taylor expansion and estimate the following linear regression to obtain the initial values for β and γ.

$$\begin{aligned}
\text{sin} : \tilde{g_1}\left(u_{t-1}, \gamma_0\right) &= \left(\gamma_1 + u_{t-1}\right) + \beta z'_{t-1}, \\
\text{cos} : \tilde{g_2}\left(u_{t-1}, \gamma_0\right) &= 1 - \left(\gamma_1 + u_{t-1}\right)^2 + \beta z'_{t-1}, \\
\text{sin_scaled} : \tilde{g_3}\left(u_{t-1}, \gamma_0\right) &= \gamma_2 + \gamma_1 u_{t-1} + \beta z'_{t-1}, \\
\text{cos_scaled} : \tilde{g_4}\left(u_{t-1}, \gamma_0\right) &= 1 - \left(\gamma_2 + \gamma_1 u_{t-1}\right)^2 + \beta z'_{t-1}, \\
\text{exp_shift} : \tilde{g_5}\left(u_{t-1}, \gamma_0\right) &= \gamma_{1,0} u_{t-1}^2 - 2\gamma_{1,0}\gamma_{2,0} u_{t-1} + \gamma_{1,0}\gamma_{2,0}^2 + \beta z'_{t-1} \\
\text{exp} : \tilde{g_6}\left(u_{t-1}, \gamma_0\right) &= \gamma_{1,0}(1 - \gamma_{2,0} u_{t-1}^2 + \frac{\gamma_{2,0}^2 u_{t-1}^4}{2}) + \beta z'_{t-1}, \\
\text{Polynomial} : \tilde{g_7}\left(u_{t-1}, \gamma_0\right) &= \gamma_1 + \gamma_2 u_{t-1} + \gamma_3 u_{t-1}^2 + \beta z'_{t-1}, \\
\text{constrained_linear} : \tilde{g_8}\left(u_{t-1}, \gamma_0\right) &= \gamma_1 + \gamma_2 u_{t-1} + \beta z'_{t-1}
\end{aligned}$$

where $\beta = (\beta_1, \beta_2)$, $z'_{t-1} = (y_{t-1}, w_{t-1})$, and u_{t-1} has been calculated in step 1 above.

In the simulation section, we have shown that by using Taylor-initials, the convergence of the estimators have improved significantly. In this section, we will investigate the empirical performances of our partially nonlinear models using the Taylor-initials.

4.1 In-sample Results

In this section, we use the complete sample from 1956 Q1 to 2018 Q4 to investigate the in-sample performances of our partially nonlinear models. We define the in-sample R_{IS}^2

as a measurement for the in-sample performance:

$$R_{IS}^2 = 1 - \frac{\sum_{t=1}^{n} (y_t - \widehat{y}_t)^2}{\sum_{t=1}^{n} (y_t - \bar{y})^2} \tag{4.1}$$

where y_t is the observed stock return in time t, $\bar{y}$ is the predicted return from the benchmark model, and $\hat{y}_t$ is the corresponding predicted stock return.

R_{IS}^2 can also be rewritten as:

$$R_{IS}^2 = 1 - \frac{MSE_{CLS}}{MSE_{bm}} \tag{4.2}$$

where MSE_{bm} is the mean squared error of benchmark model and $\mathrm{MSE}_{CLS} = 1/n \sum_{t=1}^{n} (y_t - \hat{y}_t)^2$ is the mean squared error of our partially nonlinear models. In the study of Welch and Goyal (2008), they found that sample mean is a competitive model in stock return prediction. Therefore, we use the sample mean model as the benchmark.

In equation (4.2), if R_{IS}^2 for a given model is positive, it indicates that the model outperforms the benchmark model, and the bigger the value is, the better the corresponding model performs.

The results of R_{IS}^2 is reported in table 3.5. Among the 8 functional forms, g_3, g_4, g_5, g_7 and g_8 provide positive R_{IS}^2 for all 4 variable combinations, which means the 5 functional forms have better in-sample performances than historical benchmark. And for g_1, g_2 and g_6, they can outperform sample mean model for some of the variable combinations.

Table 3.5: Results of R_{IS}^2 for all the models (benchmark: sample mean model)

variables	**function**	R_{IS}^2	**function**	R_{IS}^2	**function**	R_{IS}^2	**function**	R_{IS}^2
co1	g_1	-0.37483	g_3	0.00656	g_5	0.02287	g_7	0.01445
co2		0.01603		0.01633		0.00000		0.01626
co3		-5.82166		0.00115		0.02376		0.01814
co4		-0.00622		0.01026		0.00000		0.01493
co1	g_2	-0.57967	g_4	0.01183	g_6	-0.07239	g_8	0.00780
co2		-0.03409		0.01633		-0.02301		0.01633
co3		-5.50946		0.00618		0.01751		0.00077
co4		0.00615		0.01026		-0.02565		0.01025

4.2 OOS Results

Since the existing literatures show that the evidence for stock return predictability only hold for in-sample, in this section, we investigate how our models perform out-of-sample. Following Campbell and Thompson (2008), we use the OOS R^2 to measure the forecasting performance. The R^2_{OOS} is defined as:

$$R^2_{OOS,j,n,R} = 1 - \frac{\sum_{r=1}^{R} \left(y_{n+r,j} - \widehat{y}_{n+r,j}\right)^2}{\sum_{r=1}^{R} \left(y_{n+r,j} - \bar{y}_{n+r,j}\right)^2} \tag{4.3}$$

where n is the sample size of initial data to get a regression estimate at the start of evaluation period, R is the total number of expansive windows. In our case, n = 128 (from 1956 Q1 to 1987 Q4) and the maxmum of R is 124 (from 1988 Q1 to 2018 Q4). We also set j = 1 because we only consider 1-step forecast. To make the notation simpler, we will ignore the subscript j in the rest of the chapter.

In the above definition, $\hat{y}_{n+r}$ is the 1-step predicted return in the r-th window. $\bar{y}_{n+r}$ is the sample mean of observations using the information up to $n + r - 1$, y_{n+r} is the observed return in period $n + r$.

To generate the first out-of-sample stock return forecast, we use the first n-1 pairs of observations $\{(x_1, y_2), (x_2, y_3), \cdots, (x_{n-1}, y_n)\}$ to estimate the nonlinear model and predict $\hat{y}_{n+1}$ and $\bar{y}_{n+1}$. We then include the information in the $n + 1$ period and predict $\hat{y}_{n+2}$ and $\bar{y}_{n+2}$ using $\{(x_1, y_2), (x_2, y_3), \cdots, (x_n, y_{n+1})\}$. The procedure continues until we obtain $\hat{y}_{n+R}$ and $\bar{y}_{n+R}$. The predicted values are denoted by:

$$\hat{y}_{n+1}, \hat{y}_{n+2}, \cdots, \hat{y}_{n+R}$$
$$\bar{y}_{n+1}, \bar{y}_{n+2}, \cdots, \bar{y}_{n+R}$$

Using these predicted values, we can calculate the $R^2_{OOS,j,n,R}$ using the definition (4.3). To show how the out-of-sample forecasting performance when the forecasting window is expanding, we look at the cumulative out-of-sample R^2.

Cumulative $R^2_{OOS,n,R}$ can be obtained with R ranging from 1 to 124 (in our previous chapter). In Cheng et al. (2019), they use R^2_{OOS} starting from R = 12 (1 year, 12 months). Therefore in this chapter, I will follow their choice of R^2_{OOS} and start from R = 4 (1 year, 4 quarters).

The OOS results for different functional forms are shown in the following figures. We report out-of-sample results in 8 figures, each of which shows the prediction of different functional forms using the 4 co-integrated combinations. We put the R^2_{OOS} statistics on the vertical axis and the beginning of the various out-of-sample evaluation periods on the horizontal axis.

Figure (3.2) and (3.3) show the performances of the two trigonometric functions compared with sample mean. They have similar OOS performance for each of the 4 variable pairs and we can only find positive results for combinations co2 (tbl and lty) and co4 (baa or aaa-rated bonds). But from the sub-figure (d), we can see that function g_1 is better than g_2 since it provide consecutive positive R^2_{OOS} from 1996 to 2012.

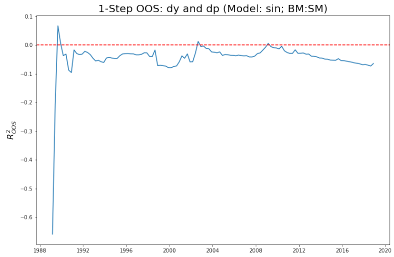
Figure (3.4) and (3.5) display results for the scaled trigonometric functions g_3 and g_4. These two functions are different from g_1 and g_2 in that they have a scale parameter in front of the single-index while g_1 and g_2 do not. Similar to the trigonometric functions, g_3 and g_4 only show positive R^2_{OOS} for combinations co2 (lty and tbl) and co4 (baa or aaa-rated bonds), and the results are almost identical for the 4 variable combinations except for co4 , where g_3 have a better forecast.

The results for the two exponential functions are presented in figure (3.6) and (3.7). For functional g_5, it cannot outperform sample mean model for most of the out-of-sample forecasting period. We can only find a positive spike around 1992 for co1 (dp and dy), (lty and tbl) and co4 (baa or aaa-rated bonds).

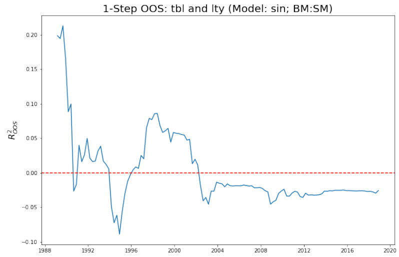
For functional g_6, the positive R^2_{OOS} can be found in the first half of the forecasting period. We can see from sub-figures (b) and (d) of figure (3.7) that g_6 gives consecutive positive results before 2014 when using combinations co2 and co4.

Figure (3.8) presents the forecasting results of the polynomial function g_7. Positive R^2_{OOS} can be found for combinations co2 and co4 before 2014. Forecasting results of g_8 in figure (3.9) are similar to g_7 for combinations co1, co2 and co3. Although the results of co4 is different, the positive values are only present in the first half of the forecasting period.

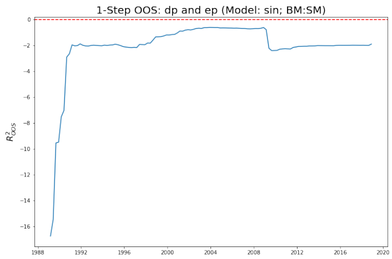
Figure 3.2: OOS Results for Model with f_1



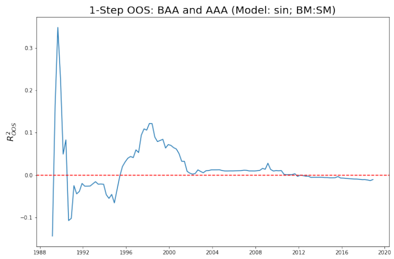
(a) co1



(b) co2

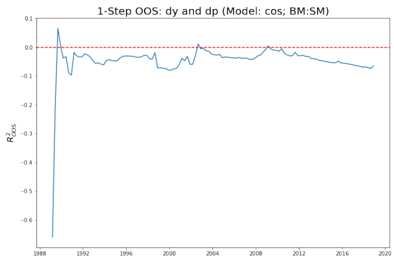


(c) co3

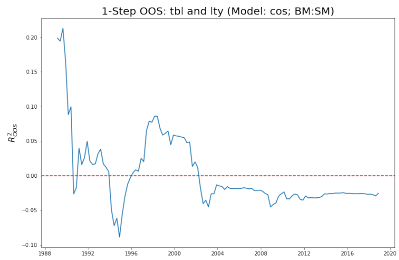


(d) co4

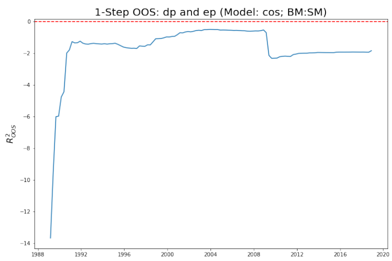
Figure 3.3: OOS Results for Model with f_2



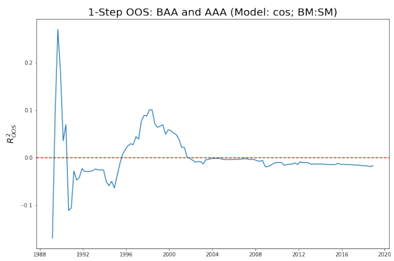
(a) co1



(b) co2

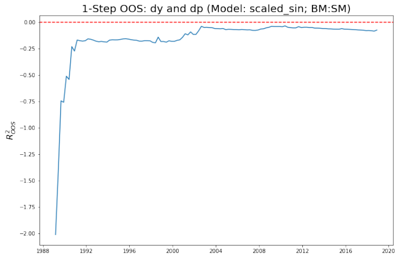


(c) co3

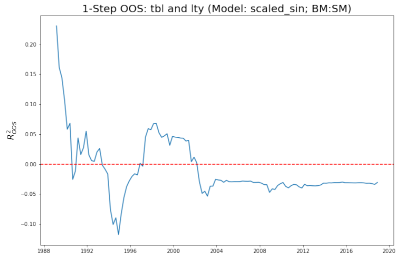


(d) co4

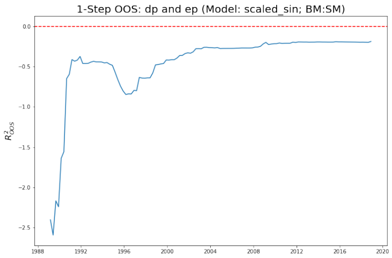
Figure 3.4: OOS Results for Model with f_3



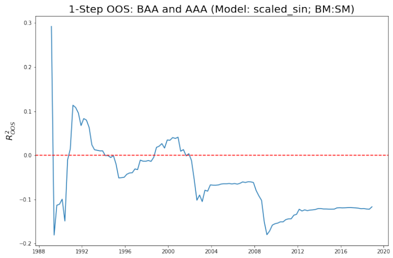
(a) co1



(b) co2

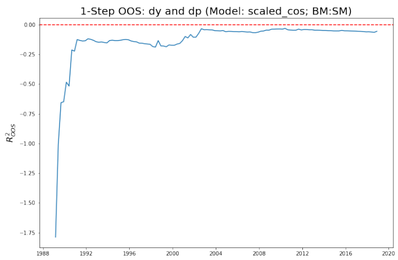


(c) co3

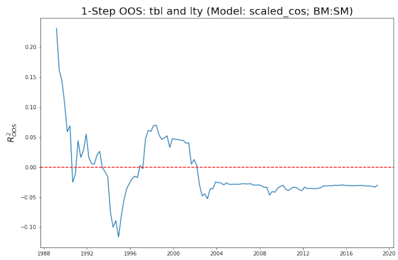


(d) co4

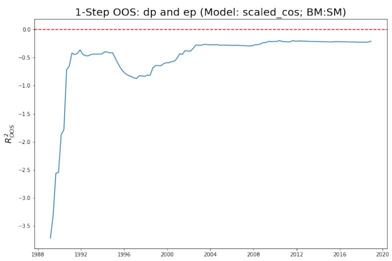
Figure 3.5: OOS Results for Model with f_4



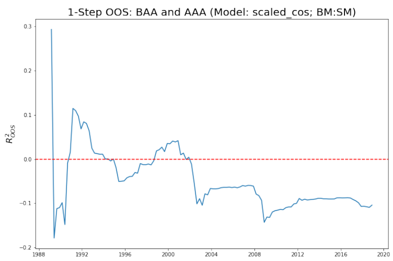
(a) co1



(b) co2

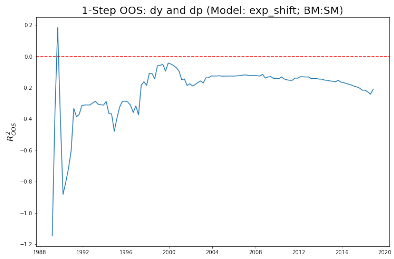


(c) co3

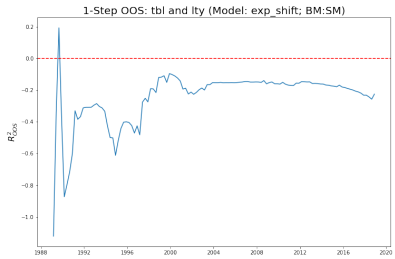


(d) co4

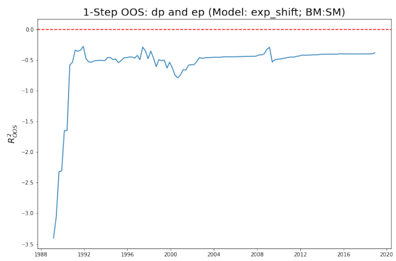
Figure 3.6: OOS Results for Model with f_5



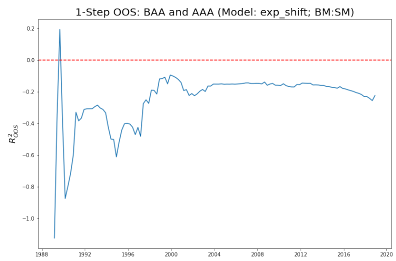
(a) co1



(b) co2

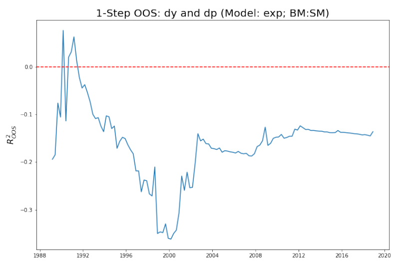


(c) co3

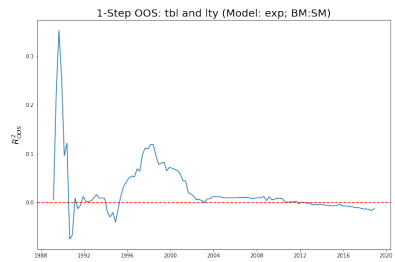


(d) co4

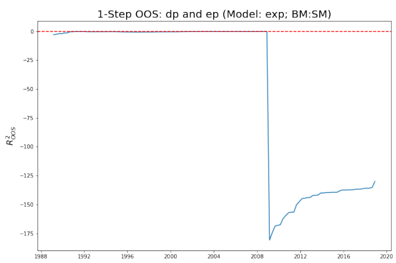
Figure 3.7: OOS Results for Model with f_6



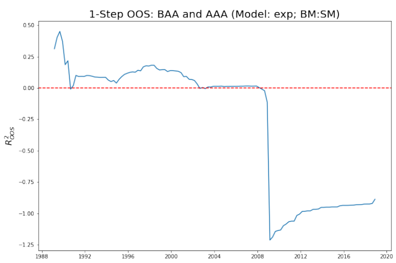
(a) co1



(b) co2

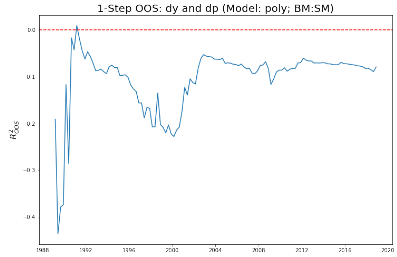


(c) co3

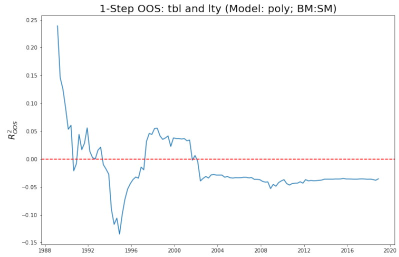


(d) co4

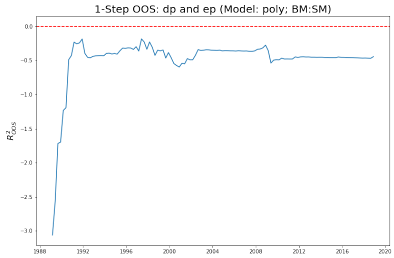
Figure 3.8: OOS Results for Model with f_7



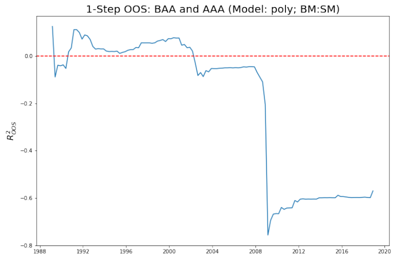
(a) co1



(b) co2

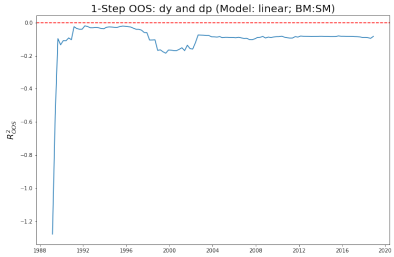


(c) co3

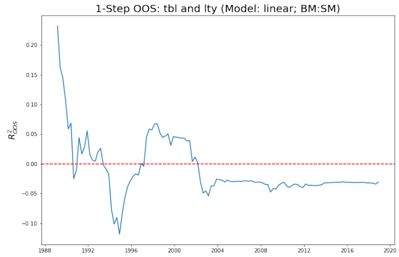


(d) co4

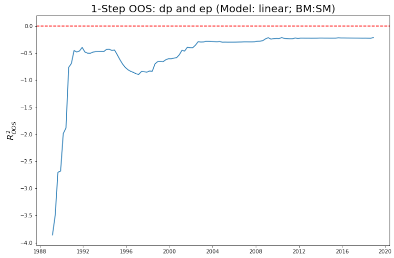
Figure 3.9: OOS Results for Model with g_8



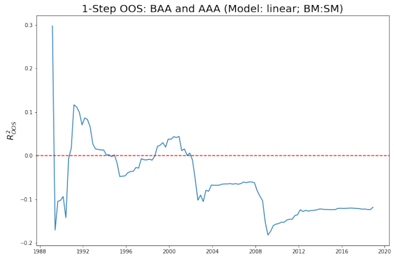
(a) co1



(b) co2



(c) co3



(d) co4

To compare the performances of the 8 functional forms with the 4 different variable combinations, we calculate the percentage of positive R^2_{OOS} in the forecasting period. The results are shown in table (3.6). It is clear that for combination co3 (dp and ep), none of the functions can perform better than sample mean prediction. However, for combinations co2 and co4, for more than half of out-of-sample period, g_1 and g_6 can outperform sample mean. And other functional forms also can provide positive results.

Table 3.6: Percentage of Positive R^2_{OOS}

	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8
co1	3.23%	3.23%	0.00%	0.00%	2.42%	7.26%	4.03%	0.00%
co2	37.90%	37.90%	32.26%	32.26%	2.42%	68.55%	29.84%	32.26%
co3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
co4	57.26%	24.19%	24.19%	26.61%	2.42%	62.90%	41.94%	27.42%

To conclude, no matter which function we use, combinations co2 and co4 tend to have better results than other variable pairs. And among the 8 functions, g_6 is the best for all the 4 combinations in terms of the percentage of positive R^2_{OOS}. The two trigonometric functions also provide good OOS results, especially for g_1 and co4. But considering the scale parameter in g_3 and g_4 does not improve the forecast. For polynomial function, it can outperform sample mean except for combination co3, although it is not as good as other functional forms.

5 conclusion

In this chapter, we consider a partially nonlinear single-index model, which allows for lagged dependent variables, stationary variables, cointegrated and non-cointegrated variables. We propose a two-step estimation method to estimate the model and includes a constraint on θ (the coefficient for the non-stationary variables). From the simulation results, we can see that the estimators have good finite sample properties and the constraint provides finite sample gains.

We apply the model to the Welch and Goyal (2008) dataset and investigate the predictability using co-integrated variable combinations. We find that by including lagged dependent variable and stationary variable, the partially nonlinear single-index models obtain a better out-of-sample performance than the nonlinear models. When using the partially nonlinear model, some of the variable combinations in previous studies give better out-of-sample forecast than the sample mean prediction over a consecutive period.

Timetable

Timetable for 2020

After my confirmation report, I followed the research schedule proposed for chapter 2 and finished the Monte Carlo simulation and empirical study.

- April 2020 - Jul 2020: Extended the simulation I have done in first year.

  In my confirmation report, I considered three functional forms (two trigonometric functions and one polynomial function) and two cases (stationary and noncointegrated predictors).

  I then included 2 exponential functions and also considered cointegrated predictor case in the simulation.

- Aug 2020 - mid Dec 2020: I finished the empirical study using the five nonlinear functions and find that they can provide a better performance than the benchmark model.

- Nov - Dec I started to write chapter 2.

- Jan 2020 - Feb 2021: Started to extend the nonstationary nonlinear single-index model to a partially nonlinear single-index model. Introduced a new estimation method for the model in chapter 3.

Timetable for 2021

In the coming year of my PhD program, I am going to keep working on chapter 3 - the partially nonlinear single-index model. Here is my timetable:

- Mar 2021 - Jul 2021: I will conduct simulation experiment for chapter 3 and compare the finite sample properties of NLS and CLS estimator under different cases (noncointegrated, cointegrated, serially correlated error.)

- Aug 2021 - Nov 2021: I will conduct the empirical study to illustrate the usefulness of the partially nonlinear single-index model. Then I will write chapter 3.
- Nov 2021 - Feb 2022: I will establish the asymptotic properties of the estimators.

Bibliography

Andrews, D. W. and McDermott, C. J. (1995), 'Nonlinear econometric models with deterministically trending variables', *The Review of Economic Studies* **62**(3), 343–360.

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

Campbell, J. Y. (1987), 'Stock returns and the term structure', *Journal of financial economics* **18**(2), 373–399.

Campbell, J. Y. and Shiller, R. J. (1988), 'The dividend-price ratio and expectations of future dividends and discount factors', *The 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?', *The Review of Financial Studies* **21**(4), 1509–1531.

Campbell, J. Y., Vuolteenaho, T. and Ramadorai, T. (2004), 'Caught on tape: Predicting institutional ownership with order flow', *Harvard Institute of Economic Research Discussion Paper* (2046).

Campbell, J. Y. and Yogo, M. (2006), 'Efficient tests of stock return predictability', *Journal of financial economics* **81**(1), 27–60.

Chang, Y. and Park, J. Y. (2003), 'Index models with integrated time series', *Journal of Econometrics* **114**(1), 73–106.

Chang, Y., Park, J. Y. and Phillips, P. C. (2001), 'Nonlinear econometric models with cointegrated and deterministically trending regressors', *The Econometrics Journal* **4**(1), 1–36.

Chen, Q. and Hong, Y. (2016), 'Predictability of equity returns over different time horizons: a nonparametric approach', *Available at SSRN 3390982* .

Cheng, T., Gao, J. and Linton, O. (2019), 'Nonparametric predictive regressions for stock return prediction'.

Choi, Y., Jacewitz, S. and Park, J. Y. (2016), 'A reexamination of stock return predictability', *Journal of Econometrics* **192**(1), 168–189.

Chung, H. and Park, J. Y. (2007), 'Nonstationary nonlinear heteroskedasticity in regression', *Journal of Econometrics* **137**(1), 230–259.

Cochrane, J. H. (2011), 'Presidential address: Discount rates', *The Journal of finance* **66**(4), 1047–1108.

Crump, R., Goda, G. S. and Mumford, K. J. (2011), 'Fertility and the personal exemption: comment', *American Economic Review* **101**(4), 1616–28.

Dangl, T. and Halling, M. (2012), 'Predictive regressions with time-varying coefficients', *Journal of Financial Economics* **106**(1), 157–181.

Dong, C., Gao, J., Tjøstheim, D. et al. (2016), 'Estimation for single-index and partially linear single-index integrated models', *The Annals of Statistics* **44**(1), 425–453.

Fama, E. F. (1990), 'Stock returns, expected returns, and real activity', *The journal of finance* **45**(4), 1089–1108.

Fama, E. F. and French, K. R. (1989), 'Business conditions and expected returns on stocks and bonds', *Journal of financial economics* **25**(1), 23–49.

Ferson, W. E. and Schadt, R. W. (1996), 'Measuring fund strategy and performance in changing economic conditions', *The Journal of finance* **51**(2), 425–461.

Gao, J. (2007), *Nonlinear time series: semiparametric and nonparametric methods*, CRC Press.

Johannes, M., Korteweg, A. and Polson, N. (2014), 'Sequential learning, predictability, and optimal portfolio returns', *The Journal of Finance* **69**(2), 611–644.

Koo, B., Anderson, H. M., Seo, M. H. and Yao, W. (2020), 'High-dimensional predictive regression in the presence of cointegration', *Journal of Econometrics* .

Kostakis, A., Magdalinos, T. and Stamatogiannis, M. P. (2015), 'Robust econometric inference for stock return predictability', *The Review of Financial Studies* **28**(5), 1506–1553.

Kothari, S. P. and Shanken, J. (1997), 'Book-to-market, dividend yield, and expected market returns: A time-series analysis', *Journal of Financial economics* **44**(2), 169–203.

Lamont, O. (1998), 'Earnings and expected returns', *The journal of Finance* **53**(5), 1563–1587.

Lee, J. H., Shi, Z. and Gao, Z. (2018), 'On lasso for predictive regression', *arXiv preprint arXiv:1810.03140* .

Lee, T.-H., Tu, Y. and Ullah, A. (2015), 'Forecasting equity premium: Global historical average versus local historical average and constraints', *Journal of Business & Economic Statistics* **33**(3), 393–402.

Lettau, M. and Ludvigson, S. (2001), 'Consumption, aggregate wealth, and expected stock returns', *the Journal of Finance* **56**(3), 815–849.

Lettau, M. and Van Nieuwerburgh, S. (2008), 'Reconciling the return predictability evidence: The review of financial studies: Reconciling the return predictability evidence', *The Review of Financial Studies* **21**(4), 1607–1652.

Li, D., Tjøstheim, D., Gao, J. et al. (2016), 'Estimation in nonlinear regression with harris recurrent markov chains', *The Annals of Statistics* **44**(5), 1957–1987.

Park, J. Y. (2002), 'Nonstationary nonlinear heteroskedasticity', *Journal of econometrics* **110**(2), 383–415.

Park, J. Y. and Phillips, P. C. (1988), 'Statistical inference in regressions with integrated processes: Part 1', *Econometric Theory* **4**(3), 468–497.

Park, J. Y. and Phillips, P. C. (1989), 'Statistical inference in regressions with integrated processes: Part 2', *Econometric Theory* pp. 95–131.

Park, J. Y. and Phillips, P. C. (1999), 'Asymptotics for nonlinear transformations of integrated time series', *Econometric Theory* **15**(3), 269–298.

Park, J. Y. and Phillips, P. C. (2000), 'Nonstationary binary choice', *Econometrica* **68**(5), 1249–1280.

Park, J. Y. and Phillips, P. C. (2001), 'Nonlinear regressions with integrated time series', *Econometrica* **69**(1), 117–161.

Pesaran, M. H. and Timmermann, A. (1995), 'Predictability of stock returns: Robustness and economic significance', *The Journal of Finance* **50**(4), 1201–1228.

Phillips, P. C., Li, D. and Gao, J. (2017), 'Estimating smooth structural change in cointegration models', *Journal of Econometrics* **196**(1), 180–195.

Qi, M. (1999), 'Nonlinear predictability of stock returns using financial and economic variables', *Journal of Business & Economic Statistics* **17**(4), 419–429.

Revuz, D. and Yor, M. (2013), *Continuous martingales and Brownian motion*, Vol. 293, Springer Science & Business Media.

Stock, J. H. and Watson, M. W. (2007), 'Why has us inflation become harder to forecast?', *Journal of Money, Credit and banking* **39**, 3–33.

Welch, I. and Goyal, A. (2008), 'A comprehensive look at the empirical performance of equity premium prediction', *The Review of Financial Studies* **21**(4), 1455–1508.

Whittington, L. A., Alm, J. and Peters, H. E. (1990), 'Fertility and the personal exemption: implicit pronatalist policy in the united states', *The American Economic Review* **80**(3), 545–556.

Wooldridge, J. M. (1994), 'Estimation and inference for dependent processes', *Handbook of econometrics* **4**, 2639–2738.

Xu, K.-L. (2016), 'Testing for return predictability with co-moving predictors of unknown form', *Available at SSRN 3177313* .

Zhou, W., Gao, J., Kew, H. and Harris, D. (2018), 'Semiparametric single-index predictive regression', *Available at SSRN 3214042* .