

# Chapter 3: Partially Nonlinear Single-Index Predictive Model

## 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:

$$y_t = \beta'_0 z_t + f(x'_{t-1} \theta_0; \gamma_0) + e_t, \quad t = 2, \dots, T, \quad (1.1)$$

where  $z_t = (y_{t-1}, \dots, y_{t-p}, w'_{t-1})'$ , in which  $w_{t-1}$  is a vector of stationary predictors,  $g(\cdot, \cdot)$  is a known univariate nonlinear function,  $x_{t-1}$  is a  $d$ -dimensional integrated process of order one,  $\theta_0$  is a  $d$ -dimensional unknown true parameter vector that lies in the parameter set  $\Theta$ ,  $\gamma_0$  is a  $m$ -dimensional unknown true parameter vector that lies in the parameter set  $\Gamma$  and  $e_t$  is a martingale difference process. The parameter sets  $\Theta$  and  $\Gamma$  are assumed to be compact and convex subsets of  $\mathbb{R}^d$  and  $\mathbb{R}^m$  respectively. In order to ensure that  $\theta_0$  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,  $f(x'_{t-1} \theta_0; \gamma_0)$ , 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  $\beta$  will have a closed form solution while  $\theta$  and  $\gamma$  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.

This chapter is organized as follows. Section 2 introduces the partially nonlinear single-index predictive model and proposes a 3-step estimation method. To improve the performance of the model, this section also introduces constraint and truncation conditions in addition to the normal nonlinear least square. Section 3 examines the finite-sample properties of the normal nonlinear least square estimators (NLS estimators hereafter) and the constrained nonlinear least square estimators (CLS estimators hereafter) using Monte Carlo evaluation. The discussion includes models with different nonlinear functional part and considers both co-integrated and non-cointegrated cases. Section 4 applies the partially nonlinear models to stock return predictability and investigate both in-sample and out-of-sample performances of the models. Section 5 concludes.

## 2 Model and Methodology

### 2.1 Estimation Method

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

$$\begin{aligned} \frac{\partial L(\beta, \theta, \gamma)}{\partial \beta} &= -2 \sum_{t=1}^T z'_t \left( y_t - \beta' z_t - f(x'_{t-1}\theta; \gamma) \right) \\ \frac{\partial L(\beta, \theta, \gamma)}{\partial \theta} &= -2 \sum_{t=1}^T \left( y_t - \beta' z_t - f(x'_{t-1}\theta; \gamma) \right) \frac{\partial g(x'_{t-1}\theta; \gamma)}{\partial \theta} \\ \frac{\partial L(\beta, \theta, \gamma)}{\partial \gamma} &= -2 \sum_{t=1}^T \left( y_t - \beta' z_t - f(x'_{t-1}\theta; \gamma) \right) \frac{\partial g(x'_{t-1}\theta; \gamma)}{\partial \gamma} \end{aligned} \quad (2.1)$$

The minimum value of  $L(\beta, \theta, \gamma)$  occurs when the above gradient 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  $\beta$ , we introduce a novel two-step approach for estimation in order to reduce the computational burden.

**Step 1:** set  $\frac{\partial L(\beta, \theta, \gamma)}{\partial \beta} = 0$  and solve the first equation in 2.1 to obtain  $\tilde{\beta}$ :

$$\tilde{\beta} = \left( \sum_{t=1}^T z_t z'_t \right)^{-1} \sum_{t=1}^T \left( y_t - f(x'_{t-1}\theta; \gamma) \right) z_t \quad (2.2)$$

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

$$y_t = \tilde{\beta}' z_t + f(x'_{t-1}\theta; \gamma) + e_t,$$

Substitute  $\tilde{\beta}$  in (2.2) and rearrange the equation above, we can get:

$$y_t - \left( \left( \sum_{t=1}^T z_t z'_t \right)^{-1} \sum_{t=1}^T y_t z_t \right)' z_t = f(x'_{t-1}\theta; \gamma) - z'_t \left( \sum_{t=1}^T z_t z'_t \right)^{-1} \sum_{t=1}^T f(x'_{t-1}\theta; \gamma) z_t + e_t.$$

Let

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

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

Then we have an approximate model of the form:

$$\tilde{y} = \tilde{g}(x_{t-1}'\theta; \gamma) + e_t. \quad (2.3)$$

**Step 2:** estimate  $(\theta, \gamma)$  using nonlinear least square(NLS) method:

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

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

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

which can be solved using an iterative procedure since there is no closed form solution. As  $\hat{\theta}$  and  $\hat{\gamma}$  have been estimated, we can substitute them back to (2.2) and get the estimate of  $\beta$ :

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

.

In the 2-step procedure above, we have obtained the NLS estimators  $(\hat{\theta}, \hat{\gamma}, \hat{\beta})$ . In order to improve finite sample properties of the estimators, we impose a truncation condition  $I(\|x_{t-1}\| \leq M_T)$  on  $x_{t-1}$  and an identification condition on coefficient vector  $\theta$ . We then define the modified sum-of-squared errors by:

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

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

Then in step 1, we can estimate the constrained least squares (denoted CLS) estimators  $\bar{\theta}$  and  $\bar{\gamma}$  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

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

And the CLS estimator of  $\beta$  can be written as:

$$\bar{\beta} = \left( \sum_{t=1}^T z_t z_t' \right)^{-1} \sum_{t=1}^T \left( y_t - f(x_{t-1}' \bar{\theta}; \bar{\gamma}) \right) z_t.$$

.

### 3 Monte Carlo Simulation

#### 3.1 Data Generation Processes

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

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

with

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

$$x_t = x_{t-1} + v_t \tag{3.1}$$

In terms of  $x_t$ , we consider both cointegrated  $x_t$  and non-cointegrated  $x_t$ . When  $x_t$  is not cointegrated,  $v_t$  in (3.1) is given by:

$$v_t = \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right)$$

To generate the co-integrated  $x_t$ , we follow a vector integrated process driven by an MA(1) innovations and construct  $v_t$  in (3.1) 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  $C = \begin{pmatrix} -1 & 4/3 \\ 0 & 0 \end{pmatrix}$ .

In the data generation process, we consider true parameter values  $\hat{\theta}_0 = (0.8, -0.6)'$ ,  $\beta_{1,0} = 0.5$ ,  $\beta_{2,0} = 1.0$ , and  $\gamma_0 = (0.2, 0.3, 0.5)$ . The vector length of  $\gamma_0$  varies depending on the nonlinear form  $f(x'_{t-1}\theta_0, \gamma_0)$  we select.

In our simulation study, we consider sample sizes  $T = 100, 500, 1000$ , replication time  $M = 5000$  and calculate the following statistics.

$$\text{bias}(\hat{\alpha}_i) = \bar{\bar{\alpha}}_i - \theta_{i,0},$$

where  $\bar{\bar{\alpha}}_i = M^{-1} \sum_{r=1}^M \hat{\alpha}_i^{(r)}$ ; and the standard deviation:

$$\text{std}(\hat{\alpha}_i) = \sqrt{M^{-1} \sum_{r=1}^M \left( \hat{\alpha}_i^{(r)} - \bar{\bar{\alpha}}_i \right)^2}.$$

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

$$\sigma_{ij} = \frac{1}{M} \sum_{r=1}^M \left( \hat{\alpha}_i^{(r)} - \bar{\bar{\alpha}}_i \right) \left( \hat{\alpha}_j^{(r)} - \bar{\bar{\alpha}}_j \right), \quad \text{cov}(\hat{\alpha}) = \sqrt{\sum_{i=1}^2 \sum_{j=1}^2 \sigma_{ij}^2}.$$

where  $\hat{\alpha}^{(r)}$  denote the  $r$ -th replication of the estimate. Following the above definitions, we then calculate biases, standard deviations and covariance for  $\theta$ ,  $\beta$  and  $\gamma$ .

### 3.2 Nonlinear Function: $f(x'_{t-1}\theta_0, \gamma_0)$

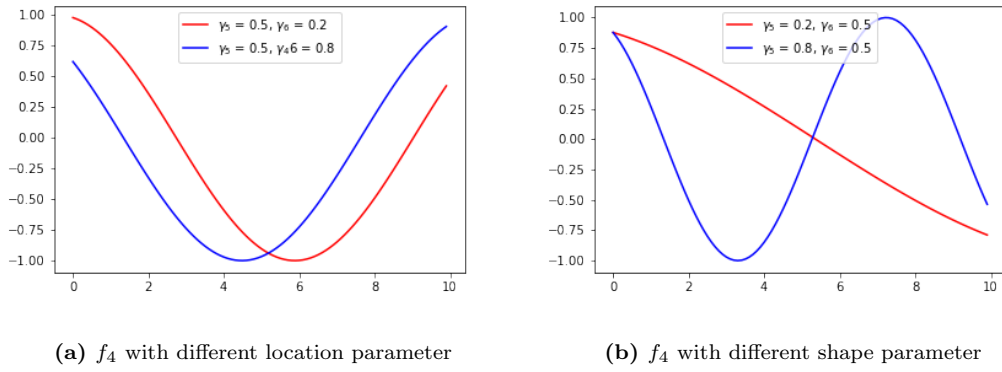
In the partially nonlinear model 1.1, we consider different forms of  $f(x'_{t-1}\theta_0, \gamma_0)$  to include varies type of nonlinearities. The first group of nonlinear functions we consider is the trigonometric functions:

$$\begin{aligned} \sin : f_1(u_{t-1}, \gamma_1) &= \sin(u_{t-1} + \gamma_1), \\ \cos : f_2(u_{t-1}, \gamma_2) &= \cos(u_{t-1} + \gamma_2), \\ \sin\_scaled : f_3(u_{t-1}, \gamma_3, \gamma_4) &= \sin(\gamma_3 u_{t-1} + \gamma_4), \\ \cos\_scaled : f_4(u_{t-1}, \gamma_5, \gamma_6) &= \cos(\gamma_5 u_{t-1} + \gamma_6), \end{aligned}$$

where  $u_{t-1} = x'_{t-1}\theta_0$ . Among the above 4 trigonometric functions,  $f_1(x'_{t-1}\theta_0, \gamma_0)$  and  $f_2(x'_{t-1}\theta_0, \gamma_0)$  can be regarded as special cases of  $f_3(x'_{t-1}\theta_0, \gamma_0)$  and  $f_4(x'_{t-1}\theta_0, \gamma_0)$  with  $\gamma_3$  and  $\gamma_5$  equals to 1.

Parameters  $\gamma_1, \gamma_2, \gamma_4$  and  $\gamma_6$  decide the position of the functions (the location parameter). As shown in picture (1a), the two curves have the same shape but the right curve is to the right of the blue one. Parameters  $\gamma_3$  and  $\gamma_5$  decide the shape of the functions, a bigger value will make the shape compressed while a smaller value will make it stretched (the shape parameter). The effect of the shape parameters can be seen in picture (1b). The blue curve shows almost one period of a cosine function but the red curve shows less than a quarter.

**Figure 1:** Plots for Trigonometric Functions



In addition, we also consider two exponential functions:

$$\begin{aligned} \text{exp\_shift} : f_5(u_{t-1}, \gamma_7, \gamma_8) &= 1 - e^{-\gamma_7(u_{t-1} - \gamma_8)^2} \\ \text{exp} : f_6(u_{t-1}, \gamma_9, \gamma_{10}) &= \gamma_9 e^{-\gamma_{10} u_{t-1}^2}, \end{aligned}$$

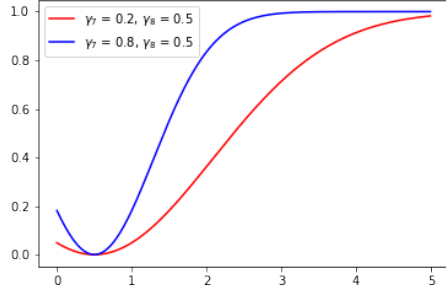
where  $\gamma_7, \gamma_{10} \in (0, \infty)$ . The two exponential functions are bounded:  $f_5(x'_{t-1}\theta_0, \gamma_0)$  lies in  $(0, 1)$  and  $f_6(x'_{t-1}\theta_0, \gamma_0)$  lies in  $(0, \gamma_9)$ .

Figure 2 and 3 present the plots of  $f_5(x'_{t-1}\theta_0, \gamma_0)$  and  $f_6(x'_{t-1}\theta_0, \gamma_0)$ . Even though both  $f_5$  and  $f_6$  are exponential functions, they have very different shapes.  $f_5$  is an increasing function while  $f_6$  is a decreasing function. We can see that  $\gamma_7$  changes the steepness of  $f_5(x'_{t-1}\theta_0, \gamma_0)$  (figure 2a) and  $\gamma_8$  changes its position (figure 2b). And for  $f_6(x'_{t-1}\theta_0, \gamma_0)$ ,  $\gamma_9$  acts as the upper bound for the function and changes its position (figure 3a), while  $\gamma_{10}$  changes its steepness (figure 3b).

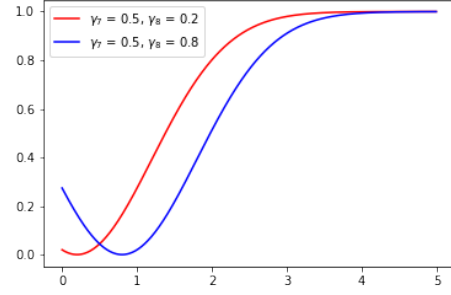
In addition, we also consider a quadratic polynomial function:

$$\text{Polynomial} : f_7(u_{t-1}, \gamma_{11}, \gamma_{12}, \gamma_{13}) = \gamma_{11} + \gamma_{12}u_{t-1} + \gamma_{13}u_{t-1}^2$$

**Figure 2:**  $f_5(x'_{t-1}\theta_0, \gamma_0)$  with different  $\gamma$

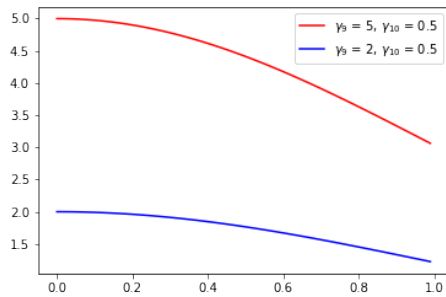


(a)  $f_5$  with different  $\gamma_7$

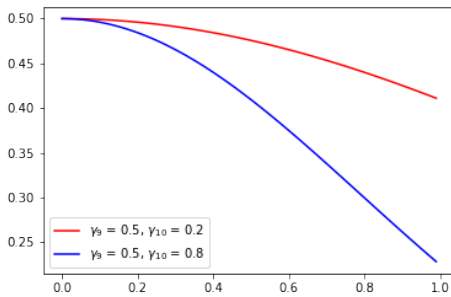


(b)  $f_5$  with different  $\gamma_8$

**Figure 3:**  $f_6(x'_{t-1}\theta_0, \gamma_0)$  with different  $\gamma$



(a)  $f_6$  with different  $\gamma_9$



(b)  $f_6$  with different  $\gamma_{10}$



It is a parabola where  $\gamma_{11}$  decides the y-intercept and  $\gamma_{13}$  changes the parabola's wideness.

The nonlinear functional forms we consider have included the bounded functions (the exponential functions), bounded periodic functions (trigonometric functions), and the unbounded function (polynomial function).

### 3.3 Initial Values (re-write)

As the gradient functions (2.1) do not have a closed form solution, we will use an iterative procedure to estimate the model, which requires initial values. Most studies use 0 or true values as the initial values to start the procedure(LR here). Normally, true values provide the quickest and best convergence, but they are not known in empirical studies. And values that are close to the optimal values will tend to give a better convergence.

Even though we do not know the optimal values beforehand, we can linearize our partially nonlinear model using Taylor expansion and estimate the linearized model to obtain better initial values.

Before applying Taylor expansion, we need to calculate the single-index  $u_{t-1} = x'_{t-1}\theta$ . Since  $x_1$  and  $x_2$  are cointegrated, we can first estimate the following regression to get the cointegration vector  $(1, -\hat{\eta})$ :

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

Then  $\theta_0$  is given by normalizing the cointegration vector  $(1, -\hat{\eta})$ :

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

which satisfies the constrain that  $\|\theta_0\| = 1$ . Then the single-index can be calculated as  $\hat{u}_{t-1} = x'_{t-1}\theta_0$ .

As  $\hat{u}_{t-1}$  has been estimated, we can apply Taylor expansion to approximate the non-linear part  $f(\hat{u}_{t-1}, \gamma_0)$ . For the two sine functions:  $f_1(u_{t-1}, \gamma_0) = \sin(u_{t-1} + \gamma_1)$  and  $f_3(\gamma_3 u_{t-1}, \gamma_4) = \sin(\gamma_3 u_{t-1} + \gamma_4)$ , we can use their first order Taylor expansions and substitute  $f_1$  and  $f_3$  by:

$$\tilde{f}_1 = (\hat{u}_{t-1} + \gamma_1)$$

$$\tilde{f}_3 = (\gamma_3 \hat{u}_{t-1} + \gamma_4)$$

Then for the two cosine functions  $f_2(u_{t-1}, \gamma_2) = \cos(u_{t-1} + \gamma_2)$  and  $f_4(\gamma_5 u_{t-1} + \gamma_6) = \cos(\gamma_5 u_{t-1} + \gamma_6)$ , we approximate them using their second-order Taylor expansions:

$$\tilde{f}_2 = 1 - (\hat{u}_{t-1} + \gamma_2)^2$$

$$\tilde{f}_4 = 1 - (\gamma_5 \hat{u}_{t-1} + \gamma_6)^2$$

Similarly, the two exponential functions  $f_5(u_{t-1}, \gamma_7, \gamma_8) = 1 - e^{-\gamma_7(u_{t-1} - \gamma_8)^2}$  and  $f_6(u_{t-1}, \gamma_9, \gamma_{10}) = \gamma_9 e^{-\gamma_{10} u_{t-1}^2}$  can also be approximated by their first and second order Taylor expansions:

$$\tilde{f}_5 = \gamma_7(\hat{u}_{t-1} - \gamma_8)^2$$

$$\tilde{f}_6 = \gamma_9(1 - \gamma_{10} \hat{u}_{t-1}^2 + \frac{\gamma_{10}^2 \hat{u}_{t-1}^4}{2})$$

And for functional 7, the polynomial function, it is already a linear function when the single-index is known. Then we can substitute the nonlinear part  $f_1$  to  $f_6$  in model (1.1) by  $\tilde{f}_1$  to  $\tilde{f}_6$  above. Therefore, the initial values of  $(\beta, \gamma)$  for different functional forms can be calculated by estimating the following linear regressions:

$$y_t = \gamma_1 + \hat{u}_{t-1} + \beta_1 y_{t-1} + \beta_2 w_{t-1} + e_t$$

$$y_t = \gamma_2 + \hat{u}_{t-1} + \beta_1 y_{t-1} + \beta_2 w_{t-1} + e_t$$

$$y_t = \gamma_4 + \gamma_3 \hat{u}_{t-1} + \beta_1 y_{t-1} + \beta_2 w_{t-1} + e_t$$

$$y_t = \gamma_6 + \gamma_5 \hat{u}_{t-1} + \beta_1 y_{t-1} + \beta_2 w_{t-1} + e_t$$

$$y_t = \gamma_7 \hat{u}_{t-1}^2 - 2\gamma_7 \gamma_8 \hat{u}_{t-1} + \gamma_7 \gamma_8^2 + \beta_1 y_{t-1} + \beta_2 w_{t-1} + e_t$$

$$y_t = \gamma_9(1 - \gamma_{10} \hat{u}_{t-1}^2 + \frac{\gamma_{10}^2 \hat{u}_{t-1}^4}{2}) + \beta_1 y_{t-1} + \beta_2 w_{t-1} + e_t$$

$$y_t = \gamma_{11} + \gamma_{12} \hat{u}_{t-1} + \gamma_{13} \hat{u}_{t-1}^2 + \beta_1 y_{t-1} + \beta_2 w_{t-1} + e_t$$

where  $\hat{u}_{t-1} = x'_{t-1} \theta_0$  and  $e_t \sim i.i.d.N(0, 1)$ ,  $t = 2, \dots, T$ .

Together with  $\theta_0$ , we have calculated the initial values  $(\theta_0, \beta_0, \gamma_0)$  for the iterative procedure. In the section below part, we can see that these initial values are better than random ones (for example, 0).

### 3.4 Simulation Results

#### Compare Different Initial Values

Table 1 shows the simulation results for the partially nonlinear model with polynomial  $f(u_{t-1}, \gamma)$ :

$$y_t = \gamma_{11} + \gamma_{12}u_{t-1} + \gamma_{13}u_{t-1}^2 + \beta_1 y_{t-1} + \beta_2 w_{t-1} + e_t$$

where  $u_{t-1} = x'_{t-1}\theta$  and  $e_t \sim i.i.d.N(0, 1)$ ,  $t = 2, \dots, T$ . From the table, we can see that the CLS estimators achieve the best convergence when using true values to start the iterative procedure. As shown in the bottom panel, the biases and standard deviations of the estimators are very low even when sample size  $T = 100$ . And it continues to decrease when sample size increases.

But when using 0 as initial values, the simulation performance is much worse. Take sample size  $T=500$  and  $\theta_1$  as an example, when using true values, the standard deviation of  $\theta_1$  is 0.00029, while when initial values equal to 0, standard deviation of  $\theta_1$  is 0.2855. It also shows a worse convergence when sample size increase.

Even though the true values provide much better results, it is impossible for us to use true values to do the estimation in real world. Therefore, we adopt the Taylor initial values following the procedure in previous section to improve the finite sample performance. As we have approximated the model using Taylor expansions, the Taylor initial values we obtain are relatively close to the true values. Therefore we expect the results to be better than using random initial values.

We can see from the first panel in table 1 that when using the Taylor initial values, the standard deviation of  $\theta_1$  has decreased from 0.2855 (when initial values equals to 0) to 0.0004. It also shows a good convergence, where the biases and standard deviations both decreases when then sample size increases. Take the covariance of  $\gamma$ :  $\text{cov}(\gamma)$  as an example, the covariance drops from 0.01029 when  $T = 100$ , to 0.00339 when  $T = 1000$ .

Therefore, compare with previous studies, our Taylor initial values significantly improved the finite sample performance. We will then apply this method to our empirical studies later.

**Table 1:** Simulation Results for Polynomial  $f(u_{t-1}, \gamma)$  Using Different Initial Values

		$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$	$\gamma_3$
Initial Values = Taylor Initials								
T=100	Bias	0.00088	0.00026	-0.00079	0.00029	0.00326	-0.00064	0.00121
	std	0.00139	0.00105	0.00504	0.00504	0.00914	0.00451	0.00141
		0.00244		0.00461		0.01029		
T=500	Bias	-0.00013	-0.00008	-0.00040	-0.00058	0.00063	-0.00032	0.00146
	std	0.00041	0.00031	0.00068	0.00206	0.00436	0.00175	0.00043
		0.00167		0.00167		0.00472		
T=1000	Bias	-0.00005	-0.00003	-0.00025	-0.00012	0.00035	-0.00021	0.00087
	std	0.00026	0.00019	0.00043	0.00136	0.00319	0.00110	0.00026
		0.00045		0.00110		0.00339		
Initial Values = 0								
T=100	Bias	0.12214	0.35638	0.09401	-0.27546	2.40099	0.19899	-0.10567
	std	0.28186	0.45009	0.31824	1.10049	2.95226	0.67484	0.36090
		0.51793		1.02441		0.17287		
T=500	Bias	0.13182	0.47163	0.19141	-0.33712	2.25798	0.19733	-0.18554
	std	0.28545	0.52487	0.38507	1.13323	2.50154	0.78597	0.38406
		0.39734		0.57196		0.12342		
T=1000	Bias	0.15770	0.47870	0.21117	-0.25275	2.34776	0.30953	-0.20737
	std	0.26417	0.50324	0.04013	0.324009	0.21639	0.09167	0.06618
		0.15924		0.10229		0.07233		
Initial Values = True Values								
T=100	Bias	0.00008	0.00007	-0.00170	0.00082	0.00368	-0.00028	0.00080
	std	0.00127	0.00095	0.00196	0.00572	0.00735	0.00350	0.00174
		0.00222		0.00261		0.00832		
T=500	Bias	0.00003	0.00002	-0.00098	0.00085	0.00233	-0.00003	0.00068
	std	0.00029	0.00022	0.00069	0.00192	0.00330	0.00135	0.00060
		0.00050		0.00098		0.00362		
T=1000	Bias	0.00000	0.00001	-0.00091	0.00084	0.00023	-0.00003	0.00068
	std	0.00019	0.00014	0.00046	0.00133	0.00226	0.00090	0.00041
		0.00033		0.00065		0.00247		

## Results for Non-cointegrated $x_t$

Table 2 presents the simulation results for non-cointegrated  $x_t$  using  $f_1(u_{t-1}, \gamma_1)$  and  $f_2(u_{t-1}, \gamma_2)$ . We can see from the table that both the NLS and CLS estimators converge when the sample size increases. And the CLS estimators perform better than the NLS estimators. For example, when  $T = 1000$ , the bias and standard deviation of CLS estimator in  $\beta_1$  in  $f_1(u_{t-1}, \gamma_1)$  is -0.00007 and 0.01574 respectively, but the values for the NLS estimator is 0.00835 and 0.01148.

The covariance of  $\theta$ ,  $\beta$  and  $\gamma$  are also smaller when the constraint is applied. Take the covariance of  $\theta$  using  $f_2(u_{t-1}, \gamma_2)$  as an example,  $\text{cov}(\hat{\theta})$  dropped from 0.70877 and 0.06903, to 0.00536 and 0.00283 when  $T = 500$  and  $T = 1000$ .

**Table 2:** Simulation Results for Non-cointegrated  $x_t$  Using  $f_1$  and  $f_2$

		NLS					Constrained-NLS				
		$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$
$f_1(u_{t-1}, \gamma_1)$											
T = 100	Bias	0.38379	0.18386	0.08270	0.48694	-0.24803	-0.00145	-0.00116	-0.00038	0.00089	0.04358
	std	0.50727	0.32935	0.13342	0.04772	0.46857	0.01523	0.01683	0.00122	0.01254	0.08301
		0.63646		0.47723		0.02050		0.01191			
T = 500	Bias	0.38567	0.19677	0.07986	0.49866	-0.23763	-0.00408	-0.00063	-0.00012	0.00147	0.00911
	std	0.47487	0.30993	0.11830	0.01579	0.26452	0.00369	0.00432	0.00021	0.00364	0.03815
		0.60420		0.26785		0.00574		0.00345			
T = 1000	Bias	0.03802	0.01811	0.00835	0.04999	-0.01991	-0.00147	-0.00026	-0.00007	0.00100	0.00286
	std	0.04750	0.02966	0.01148	0.00124	0.02029	0.01892	0.01693	0.01574	0.02946	0.01798
		0.05939		0.02050		0.00249		0.00279			
$f_1(u_{t-1}, \gamma_2)$											
T = 100	Bias	0.44222	0.18241	0.08787	0.05589	0.47601	-0.00424	-0.00120	-0.00100	0.00334	0.04116
	std	0.57804	0.33512	0.12559	0.02178	0.58834	0.01781	0.01584	0.00222	0.01433	0.08129
		0.69034		0.58181		0.02291		0.01280			
T = 500	Bias	0.42377	0.16441	0.08132	0.05121	0.48900	-0.00147	-0.00052	-0.00036	0.00208	0.00771
	std	0.59056	0.32840	0.13094	0.00719	0.55783	0.00396	0.00411	0.00065	0.00553	0.03370
		0.70877		0.55450		0.00536		0.00493			
T = 1000	Bias	0.04422	0.01824	0.00559	0.04760	0.00879	-0.00074	-0.00032	-0.00017	0.00096	0.00287
	std	0.05780	0.03351	0.00218	0.05883	0.01256	0.00222	0.00174	0.00040	0.00367	0.01557
		0.06903		0.05818		0.00283		0.00328			

Table 2 presents the simulation results for non-cointegrated  $x_t$  using  $f_3(u_{t-1}, \gamma_3, \gamma_4)$

and  $f_4(u_{t-1}, \gamma_5, \gamma_6)$ . When compared with the NLS estimators, CLS estimators have smaller bias and deviations are smaller. The biases of  $\theta_1$  in the case of CLS is -0.00177 while the value is 0.01467 in the case of NLS. And the CLS estimators also show a better convergence. As shown in the second panel, when sample size increases, the bias of  $\beta_2$  decreases from 0.01092 to 0.00237.

**Table 3:** Simulation Results for Non-cointegrated  $x_t$  Using  $f_3$  and  $f_4$

		NLS						Constrained-NLS					
		$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$	$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$
$f_3(u_{t-1}, \gamma_3, \gamma_4)$													
T = 100	Bias	0.25080	0.11643	0.09701	0.09621	-0.02764	0.04677	-0.00512	-0.00957	0.00367	0.04089	0.00023	0.00084
	std	0.42010	0.27514	0.10589	0.14919	0.02840	0.04952	0.12251	0.09119	0.02431	0.11107	0.01023	0.02126
		0.49218		0.17846		0.02405		0.14820		0.11562		0.02353	
T = 500	Bias	0.18672	0.09059	0.07042	0.11776	-0.01412	0.02610	-0.00682	-0.00430	0.00062	-0.00060	-0.00083	0.00078
	std	0.37435	0.24955	0.09681	0.14953	0.01354	0.02486	0.02586	0.02294	0.00615	0.02334	0.00276	0.00475
		0.44802		0.17453		0.01163		0.03783		0.02383		0.00561	
T = 1000	Bias	0.01467	0.08110	0.05778	0.11680	-0.00923	0.01717	-0.00177	-0.00162	0.00013	-0.00149	-0.00033	0.00038
	std	0.03410	0.23593	0.09100	0.14950	0.00941	0.01754	0.01433	0.01388	0.00406	0.01529	0.00166	0.00248
		0.41141		0.17074		0.00825		0.02241		0.01568		0.00268	
$f_4(u_{t-1}, \gamma_5, \gamma_6)$													
T = 100	Bias	0.26070	0.13566	0.09572	0.10944	-0.00129	0.00539	-0.00839	-0.00700	0.00085	0.01092	-0.02653	0.04272
	std	0.41629	0.28334	0.10132	0.14225	0.00287	0.01948	0.05009	0.04422	0.01350	0.05372	0.02733	0.04743
		0.47705		0.17498		0.01708		0.06498		0.05641		0.02541	
T = 500	Bias	0.18304	0.09114	0.07063	0.12854	-0.00039	0.00245	-0.00153	-0.00013	-0.00044	0.00449	-0.01240	0.02233
	std	0.37174	0.24185	0.09417	0.14590	0.00071	0.00598	0.01367	0.01400	0.00352	0.01763	0.01216	0.02216
		0.43332		0.17804		0.00530		0.02039		0.01847		0.01036	
T = 1000	Bias	0.01397	0.06411	0.05101	0.12171	-0.00023	0.00163	-0.00015	0.00049	-0.00025	0.00237	-0.00777	0.01432
	std	0.03386	0.21123	0.08496	0.14666	0.00042	0.00383	0.00646	0.00684	0.00163	0.01559	0.00806	0.01490
		0.40072		0.17760		0.00342		0.01009		0.01588		0.00699	

Models with the two exponential functions (see table 4) also show similar performance as other models. The NLS and CLS estimators converge when sample size increases and the CLS estimators perform better.

In table 5, we can also see a good convergence and an improvement in finite sample performance when constraint and truncation have been applied.

To conclude, for the non-cointegrated  $x_t$ , the proposed CLS estimators show a good convergence and a better performance than the normal NLS estimators. In the next section, we will investigate the performance of the estimators when  $x_t$  is cointegrated.

**Table 4:** Simulation Results for Non-cointegrated  $x_t$  Using  $f_5$  and  $f_6$ 

		NLS						Constrained-NLS					
		$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$	$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$
$f_5(u_{t-1}, \gamma_7, \gamma_8)$													
T = 100	Bias	0.12540	0.08285	0.03502	0.10013	-0.00125	0.00465	-0.01512	-0.00850	0.00298	0.03600	0.00035	0.00010
	std	0.32644	0.22965	0.07947	0.14053	0.00291	0.01768	0.08939	0.09761	0.01671	0.10739	0.01028	0.02041
		0.41166		0.16238		0.01553		0.13624		0.10882		0.02240	
T = 500	Bias	0.07635	0.03312	0.02314	0.07846	-0.00040	0.00256	-0.00626	-0.00533	0.00098	0.00020	-0.00079	0.00069
	std	0.26124	0.15768	0.06617	0.13341	0.00061	0.00513	0.02448	0.02367	0.00617	0.02938	0.00270	0.00418
		0.33234		0.15343		0.00456		0.03878		0.02994		0.00500	
T = 1000	Bias	0.04853	0.02568	0.01525	0.06429	-0.00021	0.00148	-0.00145	-0.00138	0.00006	-0.00154	-0.00033	0.00027
	std	0.21269	0.13821	0.05665	0.12720	0.00038	0.00335	0.01348	0.01318	0.00342	0.01526	0.00176	0.00231
		0.27633		0.13912		0.00298		0.02050		0.01566		0.00267	
$f_6(u_{t-1}, \gamma_9, \gamma_{10})$													
T = 100	Bias	0.16972	0.09560	0.03762	0.09567	-0.00140	0.00279	-0.00253	-0.00113	0.05286	0.11522	-0.00797	0.02183
	std	0.36696	0.24840	0.08178	0.13985	0.00590	0.02946	0.02819	0.02091	0.17887	0.31808	0.00821	0.05086
		0.45088		0.15968		0.02947		0.04905		0.36180		0.04597	
T = 500	Bias	0.13362	0.08659	0.03314	0.08628	-0.00043	0.00218	-0.00254	-0.00179	0.01126	0.01520	-0.00432	0.02071
	std	0.33483	0.23097	0.07623	0.13596	0.00139	0.01472	0.01036	0.00771	0.02667	0.13611	0.00234	0.02118
		0.42320		0.15627		0.01379		0.01806		0.13714		0.01923	
T = 1000	Bias	0.08091	0.04687	0.02416	0.08304	-0.00009	0.00082	-0.00177	-0.00128	0.00945	-0.00417	-0.00394	0.02014
	std	0.26874	0.18120	0.06819	0.13751	0.00020	0.00334	0.00698	0.00520	0.01713	0.08853	0.00161	0.01523
		0.34376		0.15229		0.00317		0.01218		0.08865		0.01382	

**Table 5:** Simulation Results for Non-cointegrated  $x_t$  Using  $f_7$ 

		NLS							Constrained-NLS						
		theta	theta	beta	beta	gamma	gamma	gamma	theta	theta	beta	beta	gamma	gamma	gamma
T = 100	Bias	0.02111	-0.01598	30.87179	-0.95731	-0.01892	-0.60074	0.72155	0.00286	0.00084	0.06327	-0.00989	0.00370	-0.01088	0.00725
	std	0.36946	0.31152	0.74012	0.74312	0.29472	0.17836	0.14145	0.03604	0.02714	0.36089	0.05816	0.02418	0.03986	0.11360
		0.48890		0.74443			0.14232		0.05138		0.36636			0.11006	
T = 500	Bias	0.02983	0.02140	3.07132	-0.65861	0.01327	-0.05536	0.16146	0.00066	0.00026	0.02231	-0.00085	0.00025	-0.00126	0.00019
	std	0.47050	0.43007	0.19241	0.285891	0.41959	0.47356	0.33892	0.00355	0.00347	0.11853	0.02182	0.00301	0.00422	0.02105
		0.07039		0.19932			0.03417		0.00581		0.12056			0.02087	
T = 1000	Bias	0.00261	-0.00316	2.48519	-0.17601	0.00291	-0.01104	-0.01521	0.00014	0.00019	0.00743	-0.00007	0.00014	-0.00050	-0.00008
	std	0.20383	0.17837	0.1113794	0.19016	0.20954	0.16493	0.19888	0.00155	0.00138	0.10818	0.01050	0.00141	0.00219	0.00929
		0.00300		0.11301			0.01985		0.00263		0.10870			0.00876	

## Simulation Results for Cointegrated $x_t$

Table 6 to 10 show the simulation results on co-integrated  $x_t$  using Taylor initial values. Table 6 presents the results for model containing nonlinear component  $f_1(u_{t-1}, \gamma_1) = \sin(u_t + \gamma_1)$  and  $f_2(u_{t-1}, \gamma_2) = \cos(u_t + \gamma_2)$ . From the table we can see that the constrained CLS estimator performs better than the NLS estimator. Bias, standard deviation and covariance of the estimators are lower in the constrained least square case than in the nonlinear least square case. For example, for model containing  $f_1$ , the absolute value of bias of  $\theta_1$  is 0.00818 when using NLS and 0.00129 when using CLS estimator (sample size  $T = 1000$ ).

The estimators also show a good convergence. Take the covariance of  $\gamma$  in the second panel as an example, the covariance of  $\gamma$  decrease from 0.01916 to 0.00242 when sample size  $T$  rises from 100 to 1000.

**Table 6:** Simulation Results for Cointegrated  $x_t$  Using Models with  $f_1$  and  $f_2$

		NLS					Constrained-NLS				
		$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$
$f_1(u_{t-1}, \gamma_1)$											
T = 100	Bias	0.00221	0.01180	-0.01443	0.01320	0.00447	0.01287	0.01004	-0.01604	-0.00325	-0.02298
	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		
T = 500	Bias	-0.00776	0.01514	-0.01165	0.01131	-0.00464	0.00247	0.00186	-0.01223	0.00879	-0.01146
	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		
T = 1000	Bias	-0.00818	0.01622	-0.01119	0.01126	-0.00584	0.00129	0.00097	-0.01176	0.01021	-0.02066
	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		
$f_2(u_{t-1}, \gamma_2)$											
T = 100	Bias	0.00069	0.00276	-0.00646	0.00180	0.03309	-0.00189	-0.00192	0.00104	-0.00143	0.04245
	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		
T = 500	Bias	0.00068	0.00075	-0.01248	0.01028	-0.00206	-0.00010	-0.00026	-0.00049	0.00061	0.00770
	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		
T = 1000	Bias	-0.00012	0.00005	-0.01275	0.01034	0.00029	-0.00015	0.00087	-0.00041	0.00051	0.00100
	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		



Similarly for  $f_3(u_{t-1}, \gamma_3, \gamma_4) = \sin(\gamma_3 u_t + \gamma_4)$  and  $f_4(\gamma_5 u_{t-1}, \gamma_6) = \cos(\gamma_5 u_t + \gamma_6)$ , both NLS and CLS estimators converge and the CLS estimators perform better. For example, in the first panel of table 7, the absolute value of bias of  $\theta_1$  using  $f_3(u_{t-1}, \gamma_3, \gamma_4)$  decreases from 0.00219 to 0.00062 in the case of CLS and drops from 0.01336 to 0.00201 in the case of NLS.

And in most cases, the CLS estimators have smaller bias, standard deviation and covariance. As shown in the second panel of table 7, the covariance of CLS estimator  $\gamma$  ( $\text{cov}(\hat{\gamma})$ ) when sample size  $T = 100$  is slightly higher than the bias of the NLS estimator (0.19193 for CLS and 0.17287 for NLS), but when  $T$  increases to 500 and 1000,  $\text{cov}(\hat{\gamma})$  of the CLS estimator become smaller (0.05015 for CLS when  $T = 1000$ , and 0.09848 for NLS).

**Table 7:** Simulation Results for Cointegrated  $x_t$  Using Models with  $f_3$  and  $f_4$

		NLS						Constrained-NLS					
		$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$	$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$
$f_3(u_{t-1}, \gamma_3, \gamma_4)$													
T = 100	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
	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	
T = 500	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
	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	
T = 1000	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
	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	
$f_4(u_{t-1}, \gamma_5, \gamma_6)$													
T = 100	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
	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	
T = 500	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
	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	
T = 1000	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
	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	

For the models containing two exponential functions  $f_5(u_{t-1}, \gamma_7, \gamma_8) = 1 - e^{-\gamma_7(u_{t-1} - \gamma_8)^2}$  and  $f_6(u_{t-1}, \gamma_9, \gamma_{10}) = \gamma_9 e^{-\gamma_{10} u_{t-1}^2}$ , we can see that the NLS estimators does not have good convergence, but after applying constraint, the CLS perform much better. For example,

in the case of NLS, the biases of  $\beta_1$  is -0.00060 and is -0.00051 when  $T = 100$  and 1000 respectively. But in the case of CLS, the absolute values of bias of  $\beta_1$  drop from 0.08371 to 0.00105.

The CLS estimators also shows a significant decrease in the covariance. Take  $\beta$  in  $f_6(u_{t-1}, \gamma_9, \gamma_{10})$  as an example, the covariance of the CLS estimator is 0.00263 and that of the NLS estimator is 0.01184.

**Table 8:** Simulation Results for Cointegrated  $x_t$  Using Models with  $f_5$  and  $f_6$

		NLS						Constrained-NLS					
		$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$	$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$
$f_5(u_{t-1}, \gamma_7, \gamma_8)$													
T = 100	Bias	0.14852	0.07716	-0.00060	-0.00069	0.03979	0.09704	0.00049	0.00707	-0.08371	0.00270	0.05869	0.11357
	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	
T = 500	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
	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	
T = 1000	Bias	0.04737	0.02568	-0.00051	0.00050	0.02912	0.08033	-0.00377	-0.00197	-0.00105	0.00843	0.00041	-0.00583
	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	
$f_6(u_{t-1}, \gamma_9, \gamma_{10})$													
T = 100	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
	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	
T = 500	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
	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	
T = 1000	Bias	0.18891	0.09882	-0.01164	0.01017	0.00430	0.17839	-0.00107	-0.00681	-0.00047	0.00036	0.07620	0.11676
	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 10 presents the simulation results for the model containing polynomial function  $f_7(u_{t-1}, \gamma_{11}, \gamma_{12}, \gamma_{13}) = \gamma_{11} + \gamma_{12}u_{t-1} + \gamma_{13}u_{t-1}^2$ . Both NLS and CLS estimators have good convergence and the CLS shows a better performance for  $\beta$  and  $\gamma$ .

As for the convergence, take  $\beta_1$  as an example, the absolute value of bias for the NLS estimator  $\hat{\beta}$  decreases from 0.00056 to 0.00001 as sample size  $T$  increases from 100 to 1000. At the same time, the absolute value of bias for the CLS estimator  $\bar{\beta}$  decreases from 0.00079 to 0.00025.

The constraint and truncation conditions we applied improve the finite sample per-

formance for all the estimators, but the effect is more significant for the estimator  $\gamma$ . For example, when  $T = 1000$ , the covariance of  $\beta$  in the case of NLS is 0.00190 and the value is 0.00110 in the case of CLS. And for the covariance of  $\gamma$ , the value in the case of CLS is 0.00339, which is much lower than 0.01231 in the case of NLS.

**Table 9:** Simulation Results for Cointegrated  $x_t$  Using Models with  $f_7$

		NLS							Constrained-NLS						
		$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$	$\gamma_3$
T = 100	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
	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	
T = 500	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
	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	
T = 1000	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
	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	

From the simulation results above, we can conclude that:

1. Both the NLS and CLS estimators converge when sample size increases. But the constrained least square method tend to give better convergence in certain cases.
2. The CLS consistently perform better than the normal NLS for most functional forms. This implies that the constraint and truncation condition we proposed in section 2 do help improve the finite sample performances.

If we compare models with the 7 different functional forms, we find that the polynomial functional form tend to provide the best results but models with the two exponential functions do not perform as well as other ones.

As shown in table 10, the standard deviation of  $\theta$  using CLS is 0.00045 when  $T=1000$ , which is much lower than this value in other functional cases(for example, the corresponding value is 0.00262 in  $f_1(u_{t-1}, \gamma_0)$  and 0.00476 in  $f_2(u_{t-1}, \gamma_0)$ ).

When compare the simulation results for non-cointegrated  $x_t$  with the results for cointegrated  $x_t$  (see table below), we can also see that the models using cointegrated  $x_t$  have a better finite sample performance. Therefore, in next section, we will focus on models with cointegrated predictors. In addition, as the CLS estimators consistently provide better performance, we will focus on the constrained least square method in the empirical study.

**Table 10:** Simulation Results for Cointegrated  $x_t$  and Non-cointegrated  $x_t$  Using  $f_7$ 

		Cointegrated $x_t$							Non-cointegrated $x_t$						
		$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$	$\gamma_3$
T = 100	Bias	0.00088	0.00026	-0.00079	0.00029	0.00326	-0.00064	0.00121	0.00286	0.00084	0.06327	-0.00989	0.00370	-0.01088	0.00725
	std	0.00139	0.00105	0.00161	0.00504	0.00914	0.00451	0.00141	0.03604	0.02714	0.36089	0.05816	0.02418	0.03986	0.11360
		0.00244		0.00461			0.01029		0.05138		0.36636			0.11006	
T = 500	Bias	-0.00013	-0.00008	-0.0004	-0.00058	0.00063	-0.00032	0.00146	0.00066	0.00026	0.02231	-0.00085	0.00025	-0.00126	0.00019
	std	0.00041	0.00031	0.00068	0.00206	0.00436	0.00175	0.00043	0.00355	0.00347	0.11853	0.02182	0.00301	0.00422	0.02105
		0.00071		0.00167			0.00472		0.00581		0.12056			0.02087	
T = 1000	Bias	-0.00005	-0.00003	-0.00025	-0.00012	0.00035	-0.00021	0.00087	0.00014	0.00019	0.00743	-0.00007	0.00014	-0.00050	-0.00008
	std	0.00026	0.00043	0.00043	0.00136	0.00319	0.0011	0.00026	0.00155	0.00138	0.10818	0.01050	0.00141	0.00219	0.00929
		0.00045		0.0011			0.00339		0.00263		0.10870			0.00876	

## 4 Empirical Study

We have examined the finite sample properties in the previous section using Monte Carlo simulation study. To illustrate the use of our partially nonlinear model in the real world, 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 (as in [Zhou et al. \(2018\)](#)) and another set of variable from [Lettau and Ludvigson \(2001\)](#), which have all been found to be co-integrated :

- 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 4 plots the time series of the co-integrated variable pairs. It is clear that the variables in each plot are not stationary and have similar shape. Especially for the first pair, "co1", the d/p ratio and the dividend yield are very close to each other.

**Figure 4:** Time Series Plots of Co-integrated Variables

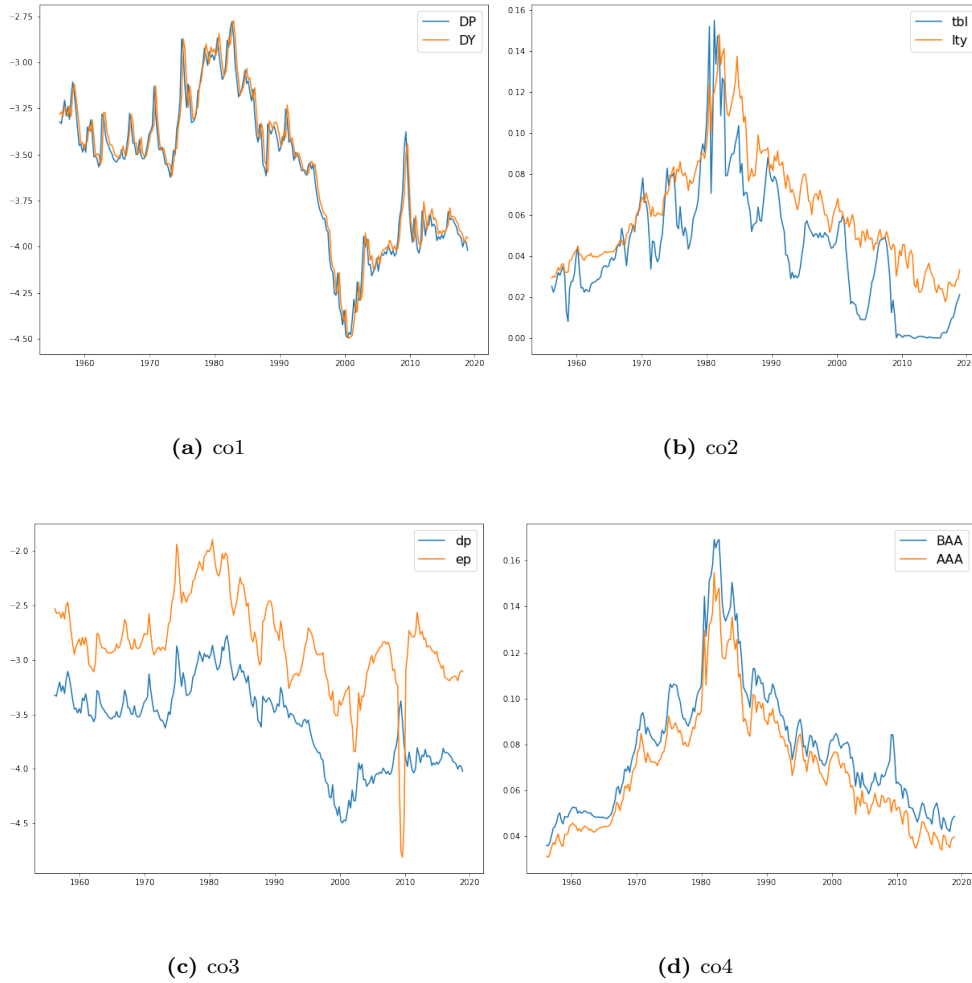
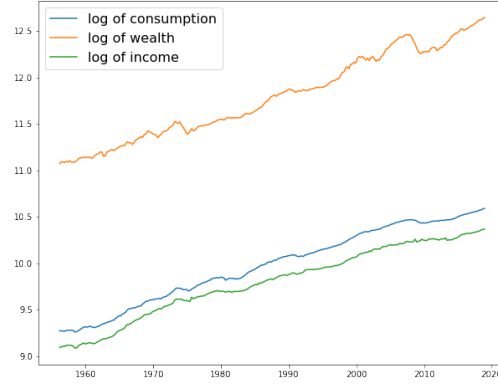


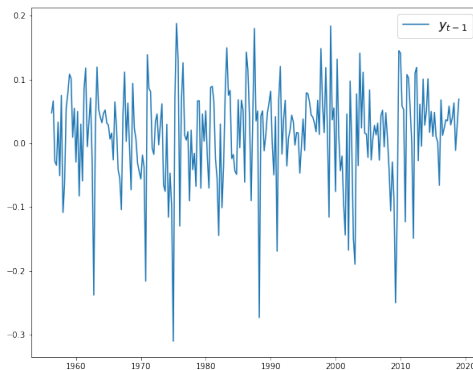
Figure 5 shows the three variables used in Lettau and Ludvigson (2001). It is clear that log consumption, log asset wealth and log labor income share a common stochastic trend.

**Figure 5:** "cay" variable

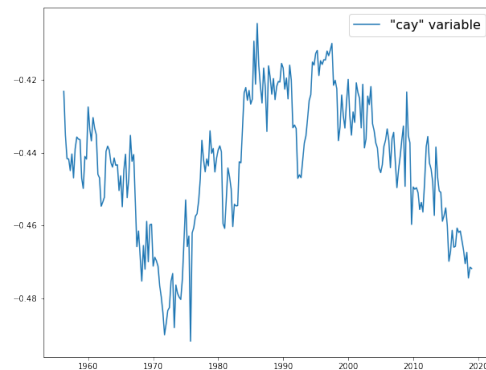


Apart from the non-stationary variables, the partially nonlinear models also include stationary predictors: the lagged stock return ( $y_{t-1}$ ) and the regression residual (the "cay" variable, hereafter) from the study of Lettau and Ludvigson (2001). In their study, they show that the three variables are cointegrated and share a common trend. The deviations from this shared trend is stationary and can be used as a predictor for stock return. Figure 6 below presents the two stationary variables (??).

**Figure 6:** Time Series Plots of Stationary Variables



**(a)** lag of dependent variable



**(b)** "cay" variable

Therefore in the partially nonlinear models we consider:

$$y_t = \beta_1 y_{t-1} + \beta_2 w_{t-1} + f(x'_{t-1} \theta, \gamma) + e_t, \quad e_t \sim i.i.d.N(0, 1), \quad t = 2, \dots, T,$$

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, the "cay" variable, from the [Lettau and Ludvigson \(2001\)](#).

In terms of the nonlinear part  $f(x'_{t-1} \theta, \gamma)$ , we consider the following 7 nonlinear functions that have been discussed in the simulation before, which includes trigonometric functions, exponential functions and polynomial functions. And these partially nonlinear models in the empirical study can be written as:

$$\begin{aligned} \sin : g_1(u_{t-1}, \gamma_1) &= \beta z'_{t-1} + \sin(u_{t-1} + \gamma_1), \\ \cos : g_2(u_{t-1}, \gamma_2) &= \beta z'_{t-1} + \cos(u_{t-1} + \gamma_2), \\ \text{scale\_sin} : g_3(u_{t-1}, \gamma_3, \gamma_4) &= \beta z'_{t-1} + \sin(\gamma_3 u_{t-1} + \gamma_4), \\ \text{scale\_cos} : g_4(u_{t-1}, \gamma_5, \gamma_6) &= \beta z'_{t-1} + \cos(\gamma_5 u_{t-1} + \gamma_6), \\ \text{exp\_shift} : g_5(u_{t-1}, \gamma_7, \gamma_8) &= \beta z'_{t-1} + 1 - e^{-\gamma_7(u_{t-1} - \gamma_8)^2} \\ \text{exp} : g_6(u_{t-1}, \gamma_9, \gamma_{10}) &= \beta z'_{t-1} + \gamma_9 e^{-\gamma_{10} u_{t-1}^2} \\ \text{Polynomial} : g_7(\hat{u}_{t-1}, \gamma_{11}, \gamma_{12}, \gamma_{13}) &= \beta z'_{t-1} + \gamma_{11} + \gamma_{12} u_{t-1} + \gamma_{13} u_{t-1}^2 \end{aligned}$$

where  $\beta = (\beta_1, \beta_2)$ ,  $z'_{t-1} = (y_{t-1}, w_{t-1})$ , and  $u_{t-1} = x'_{t-1} \theta$ .

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

$$\text{constrained\_linear} : g_8(u_{t-1}, \gamma_{14}, \gamma_{15}) = \gamma_{14} + \beta z'_{t-1} + \gamma_{15}(x_{1,t-1} \theta_1 + x_{2,t-1} \theta_2).$$

It is a special case of the polynomial function  $g_7$  where  $\gamma_{13} = 0$ .

For all the 8 functional forms, we set  $\theta_1^2 + \theta_2^2 = 1$  and will adopt the constrained nonlinear least square method described in section 2 to estimate. But as mentioned before, when minimizing the nonlinear least squares:

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

The gradient equation:

$$\frac{\partial S_T(\alpha)}{\partial \alpha}$$

does not have a closed solution, so we need to use an iterative algorithm. To implement the iterative algorithm, we must choose a vector of initial values to start. As discussed in the simulation study, we could use Taylor expansion to approximate the partially nonlinear models and choose initial values by estimating the approximated linear models. We first need to estimate the following linear regression to get  $\hat{\eta}$ :

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

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

Then following what we have done in section 3, we bring  $u_{t-1}$  back into  $f(u_{t-1}, \gamma)$  and substitute the nonlinear function  $f(u_{t-1}, \gamma)$  by its first or second order Taylor expansion and estimate the following linear regression to obtain the initial values for  $\beta$  and  $\gamma$ .

$$\begin{aligned} \sin : \tilde{g}_1(\hat{u}_{t-1}, \gamma_1) &= (\gamma_1 + \hat{u}_{t-1}) + \beta z'_{t-1}, \\ \cos : \tilde{g}_2(\hat{u}_{t-1}, \gamma_2) &= 1 - (\gamma_2 + \hat{u}_{t-1})^2 + \beta z'_{t-1}, \\ \sin\_scaled : \tilde{g}_3(\hat{u}_{t-1}, \gamma_3, \gamma_4) &= \gamma_4 + \gamma_3 \hat{u}_{t-1} + \beta z'_{t-1}, \\ \cos\_scaled : \tilde{g}_4(\hat{u}_{t-1}, \gamma_5, \gamma_6) &= 1 - (\gamma_6 + \gamma_5 \hat{u}_{t-1})^2 + \beta z'_{t-1}, \\ \exp\_shift : \tilde{g}_5(\hat{u}_{t-1}, \gamma_7, \gamma_8) &= \gamma_7 \hat{u}_{t-1}^2 - 2\gamma_8 \gamma_7 \hat{u}_{t-1} + \gamma_7 \gamma_8^2 + \beta z'_{t-1} \\ \exp : \tilde{g}_6(\hat{u}_{t-1}, \gamma_9, \gamma_{10}) &= \gamma_9 (1 - \gamma_{10} \hat{u}_{t-1}^2 + \frac{\gamma_{10}^2 \hat{u}_{t-1}^4}{2}) + \beta z'_{t-1}, \\ \text{Polynomial} : \tilde{g}_7(\hat{u}_{t-1}, \gamma_{11}, \gamma_{12}, \gamma_{13}) &= \gamma_{11} + \gamma_{12} \hat{u}_{t-1} + \gamma_{13} \hat{u}_{t-1}^2 + \beta z'_{t-1}, \\ \text{constrained\_linear} : \tilde{g}_8(\hat{u}_{t-1}, \gamma_{14}, \gamma_{15}) &= \gamma_{14} + \gamma_{15} \hat{u}_{t-1} + \beta z'_{t-1} \end{aligned}$$

where  $\beta = (\beta_1, \beta_2)$ ,  $z'_{t-1} = (y_{t-1}, w_{t-1})$ .

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 these 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. The full sample results can



be found in the appendix, table 14. In this section, we use in-sample  $r^2$  to examine the in-sample performance of our models. The in-sample  $R_{IS}^2$  is defined as:

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

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:

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

where  $MSE_{bm}$  is the mean squared error of benchmark model and  $MSE_{CLS} = 1/n \sum_{t=1}^n (y_t - \hat{y}_t)^2$  is the mean squared error of our partially nonlinear models. In 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.

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. To compare our models with some commonly used time series models and the nonlinear model we used in chapter 2:

- Sample mean model:

$$y_t = \frac{1}{t-1} \sum_{i=1}^{t-1} y_i$$

- AR1:

$$y_t = \beta_1 y_{t-1} + \epsilon_t$$

- AR2:

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \epsilon_t$$

- AR1 model with "cay" variable:

$$y_t = \beta_1 y_{t-1} + \beta_2 cay_{t-1} + \epsilon_t$$

- Nonlinear single-index model:

$$y_t = f(x'_{t-1} \theta, \beta) + \epsilon_t,$$

where  $\epsilon_t \sim i.i.d.N(0, 1)$ .

The results of  $R_{IS}^2$  against sample mean model is reported in table 11. 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 11:** 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		-0.37483		0.00656		0.02287		0.01445
co2		0.01603		0.01633		0.00000		0.01626
co3	$g_1$	-5.82166	$g_3$	0.00115	$g_5$	0.02376	$g_7$	0.01814
co4		-0.00622		0.01026		0.00000		0.01493
co1		-0.57967		0.01183		-0.07239		0.00780
co2		-0.03409		0.01633		-0.02301		0.01633
co3	$g_2$	-5.50946	$g_4$	0.00618	$g_6$	0.01751	$g_8$	0.00077
co4		0.00615		0.01026		-0.02565		0.01025

Table 12 presents the the in-sample  $R_{IS}^2$  against the nonlinear models we used in chapter 2. We can also see that our partially nonlinear models perform better for some variables and models. For example, for the variable combination co4 (baa and aaa rated bonds), models  $g_4, g_6, g_8$  can provide a better in-sample performance. It indicates that after including stationary variables, we can improve the in-sample performance in some cases.

The in-sample  $R_{IS}^2$  results against other benchmark models can be found in the appendix (put later).

## 4.2 OOS Results

### OOS Performance v.s. Sample Mean Model

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

**Table 12:** Results of  $R_{IS}^2$  for all the models (benchmark: Nonlinear Model)

variables	function	$R_{IS}^2$	function	$R_{IS}^2$	function	$R_{IS}^2$	function	$R_{IS}^2$
co1		0.31283		-0.00717		-0.08370		0.02620
co2		-0.00712		0.00240		-0.01162		-0.02511
co3	$g_1$	0.78506	$g_3$	0.01185	$g_5$	-1.20501	$g_7$	-0.02166
co4		-0.01189		-0.00214		-0.81296		-0.40862
co1		0.31278		-0.00985		-0.13675		-0.01857
co2		-0.00711		0.00312		0.09928		0.00240
co3	$g_2$	0.78993	$g_4$	0.09388	$g_6$	-0.17578	$g_8$	-0.00073
co4		-0.02822		0.01147		0.09928		-0.00181

performance. The  $R_{OOS}^2$  is defined as:

$$R_{OOS,j,n,R}^2 = 1 - \frac{\sum_{r=1}^R (y_{n+r,j} - \hat{y}_{n+r,j})^2}{\sum_{r=1}^R (y_{n+r,j} - \bar{y}_{n+r,j})^2} \quad (4.3)$$

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 maximum 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  $\{(x_1, y_2), (x_2, y_3), \dots, (x_{n-1}, y_n)\}$  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  $\{(x_1, y_2), (x_2, y_3), \dots, (x_n, y_{n+1})\}$ . The procedure continues until we obtain  $\hat{y}_{n+R}$  and  $\bar{y}_{n+R}$ . The predicted values are denoted by:

$$\begin{aligned} &\hat{y}_{n+1}, \hat{y}_{n+2}, \dots, \hat{y}_{n+R} \\ &\bar{y}_{n+1}, \bar{y}_{n+2}, \dots, \bar{y}_{n+R} \end{aligned}$$

Using these predicted values, we can calculate the  $R_{OOS,j,n,R}^2$  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  $R_{OOS,n,R}^2$  can be obtained with  $R$  ranging from 1 to 124 (in our previous chapter). In [Cheng et al. \(2019\)](#), they use  $R_{OOS}^2$  starting from  $R = 12$  (1 year, 12 months). Therefore in this chapter, I will follow their choice of  $R_{OOS}^2$  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  $R_{OOS}^2$  statistics on the vertical axis and the beginning of the various out-of-sample evaluation periods on the horizontal axis.

Figure (7) and (8) 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  $R_{OOS}^2$  from 1996 to 2012.

Figure (9) and (10) 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  $R_{OOS}^2$  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.

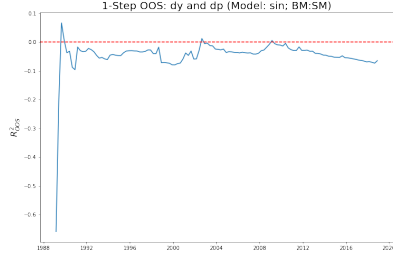
The results for the two exponential functions are presented in figure (11) and (12). For functional  $g_5$ , it cannot outperform sample mean model for most of the out-of-sample forecasting period. We can only find a positive spike around 1992 for co1 (dp and dy), (lty and tbl) and co4 (baa or aaa-rated bonds).

For functional  $g_6$ , the positive  $R_{OOS}^2$  can be found in the first half of the forecasting period. We can see from sub-figures (b) and (d) of figure (12) that  $g_6$  gives consecutive positive results before 2014 when using combinations co2 and co4.

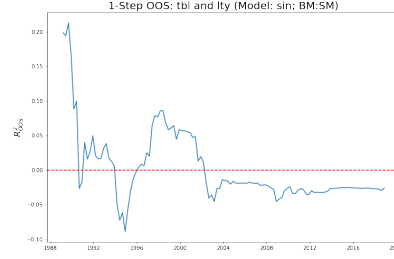
Figure (13) presents the forecasting results of the polynomial function  $g_7$ . Positive  $R_{OOS}^2$  can be found for combinations co2 and co4 before 2014. Forecasting results of  $g_8$

in figure (14) 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.

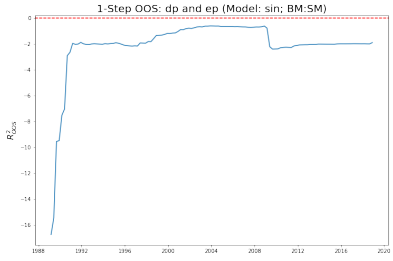
**Figure 7: OOS Results for Model with  $f_1$**



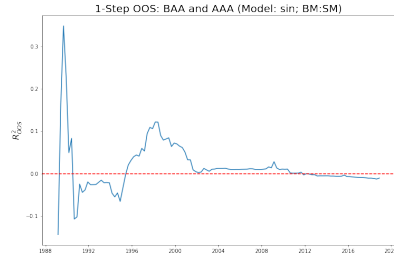
(a) co1



(b) co2

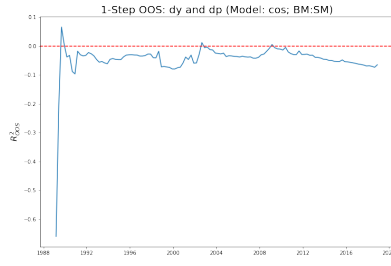


(c) co3

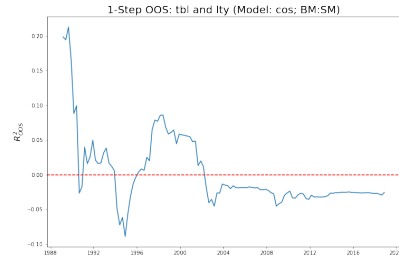


(d) co4

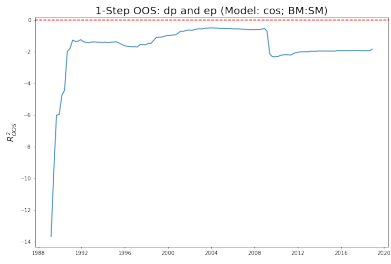
**Figure 8: OOS Results for Model with  $f_2$**



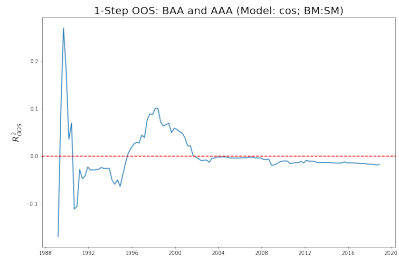
**(a) co1**



**(b) co2**

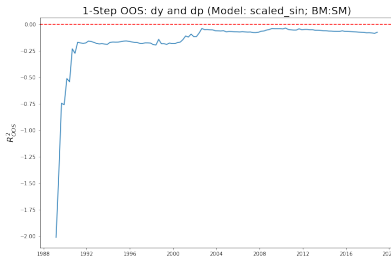


**(c) co3**

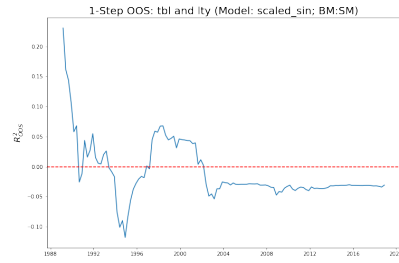


**(d) co4**

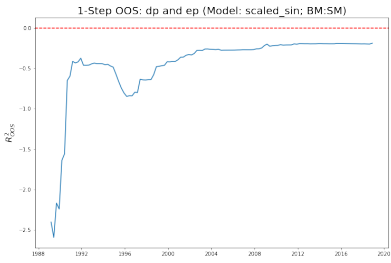
**Figure 9: OOS Results for Model with  $f_3$**



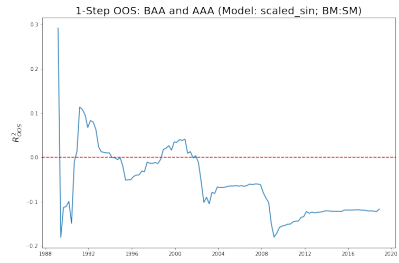
**(a) co1**



**(b) co2**

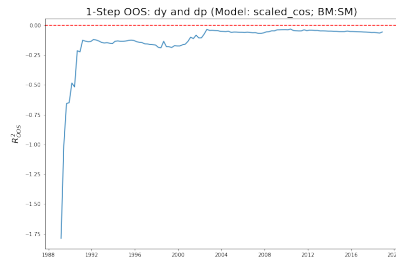


**(c) co3**

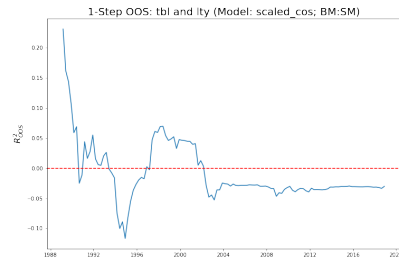


**(d) co4**

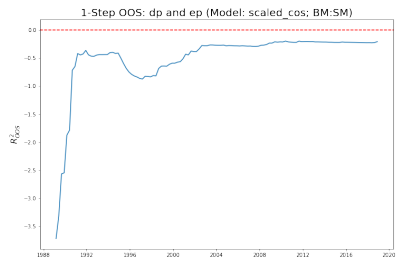
**Figure 10: OOS Results for Model with  $f_4$**



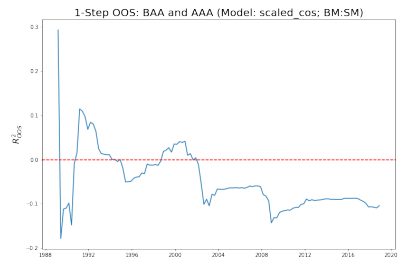
**(a) co1**



**(b) co2**

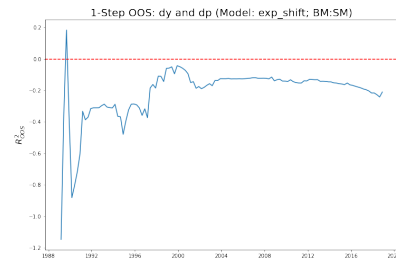


**(c) co3**

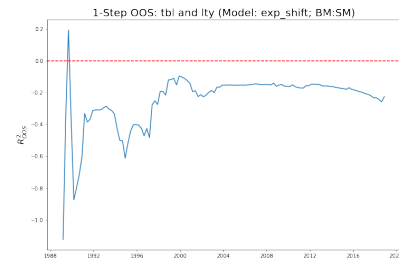


**(d) co4**

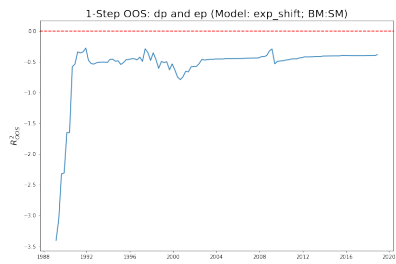
**Figure 11: OOS Results for Model with  $f_5$**



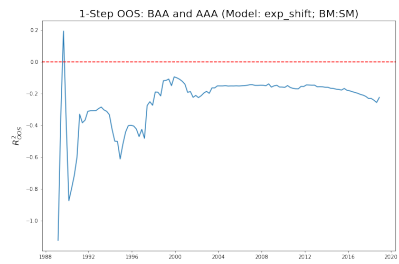
**(a) co1**



**(b) co2**

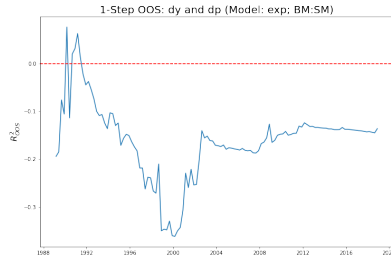


**(c) co3**

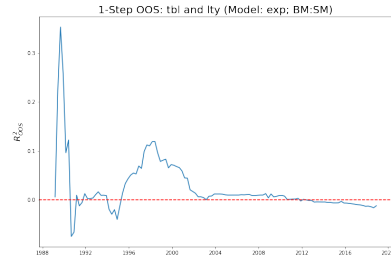


**(d) co4**

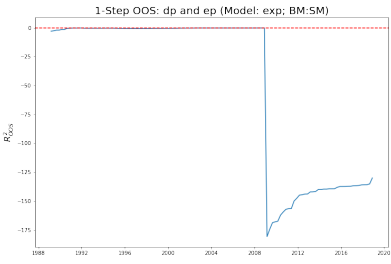
**Figure 12:** OOS Results for Model with  $f_6$



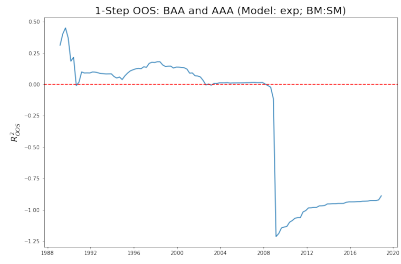
**(a)** co1



**(b)** co2

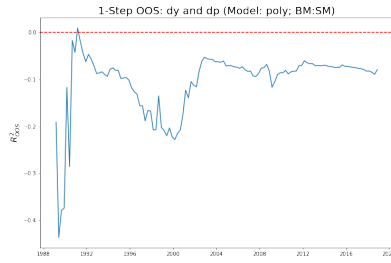


**(c)** co3

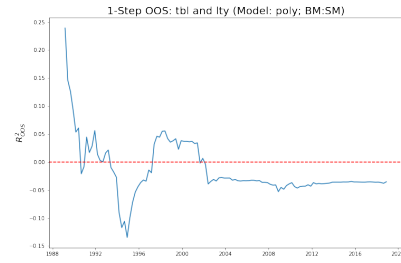


**(d)** co4

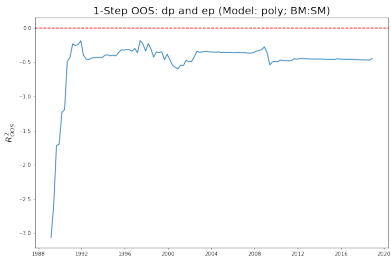
**Figure 13:** OOS Results for Model with  $f_7$



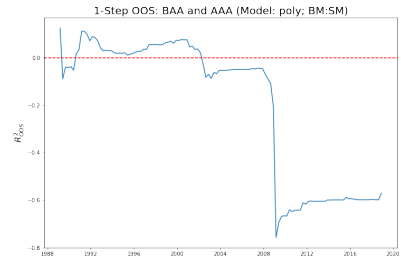
**(a)** co1



**(b)** co2

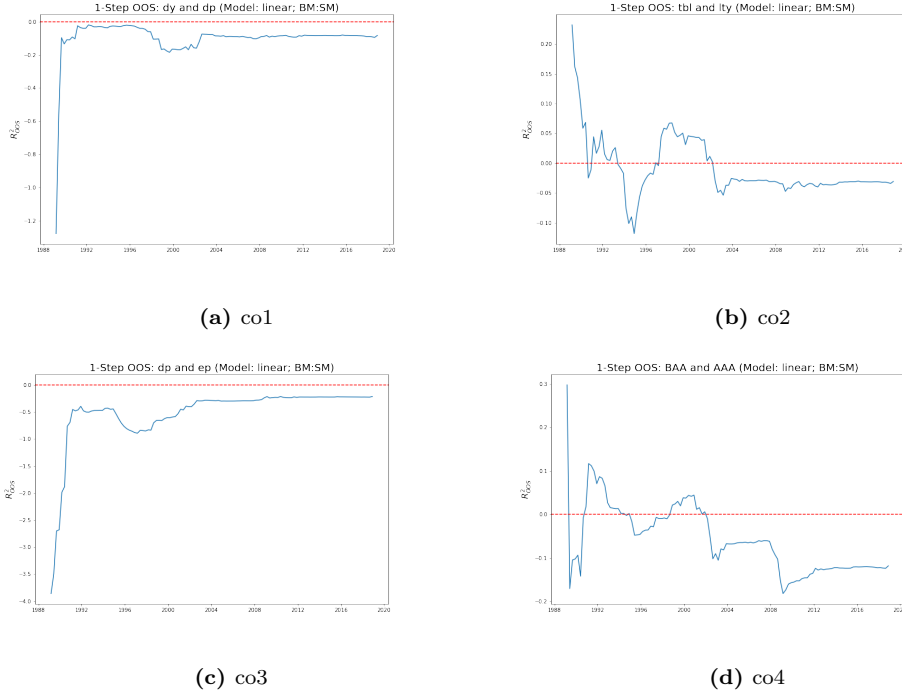


**(c)** co3



**(d)** co4



**Figure 14:** OOS Results for Model with  $g_8$ 

To compare the performances of the 8 functional forms with the 4 different variable combinations, we calculate the percentage of positive  $R_{OOS}^2$  in the forecasting period. The results are shown in table (13). 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 13:** Percentage of Positive  $R_{OOS}^2$ 

	$g_1$	$g_2$	$g_3$	$g_4$	$g_5$	$g_6$	$g_7$	$g_8$
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  $R_{OOS}^2$ . 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.

### OOS Performance v.s. Other Benchmark Models

For another commonly used time series model, the AR(1) model, we can also see from the figure 15 that the partially nonlinear models can perform better. When using model containing the nonlinear functional form  $f_1$  and variable combination  $co2$ , our partially nonlinear model can outperform AR(1) model for most of the period before 2006. Similar for the model with functional form  $f_5$ , our model performs better when using  $co2$  before 2012.

**Figure 15:** OOS Performance (Benchmark: AR(1) Model)

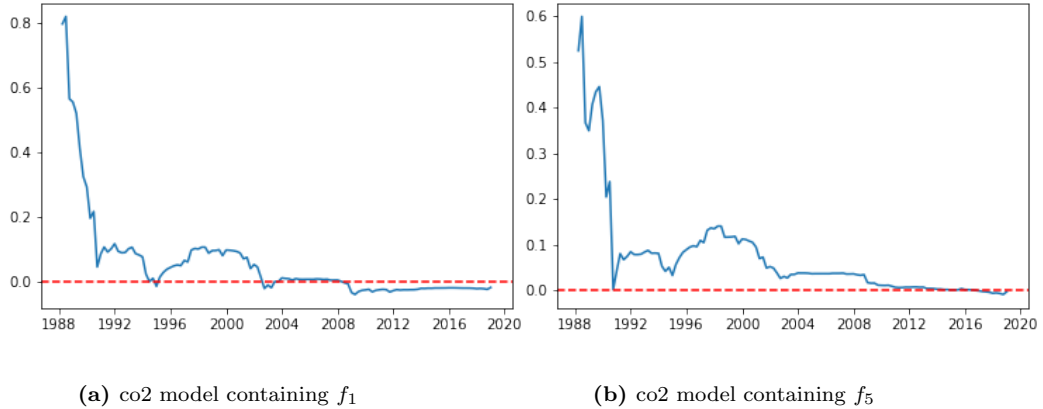
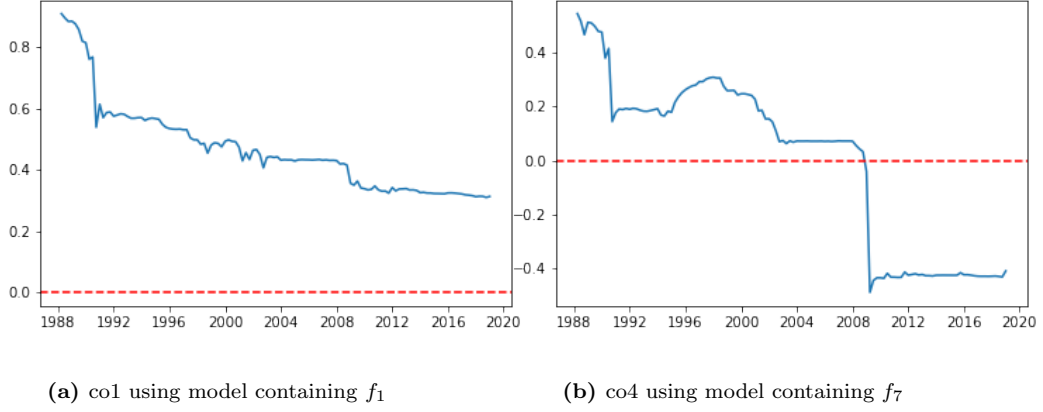


Figure 17 presents the  $R_{OOS}^2$  when using the nonlinear models in chapter 2 as the benchmark model. In the nonlinear model, we did not include the stationary variables but we considered the same nonlinear functional forms.

From the figure, we can see that when using variable combination  $co2$  and function with nonlinear form  $f_1$ , the partially nonlinear model can outperform the nonlinear model during the whole out-of-sample period. And for the variable combination  $co4$ , when using function containing nonlinear part  $f_7$ , the partially nonlinear model outperforms the nonlinear model before 2009.

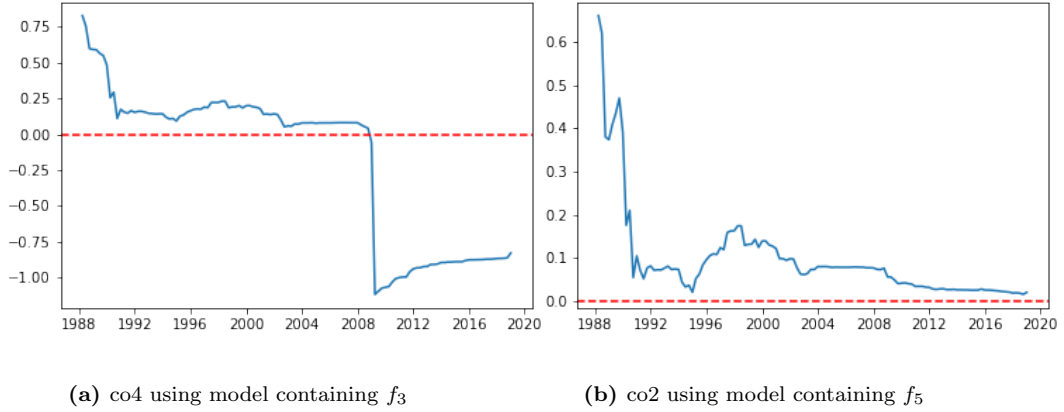
**Figure 16:** OOS Performance (Benchmark: Nonlinear Model)



When compared with the AR(2) models, the partially nonlinear models also provide a good out-of-sample performance. The model with nonlinear part  $f_3$  outperforms the AR(2) model before 2009 when using variable combination co4.

And for the model with nonlinear component  $f_5$ , it outperforms the AR(2) model during the whole out-of-sample period (for variable combination co2).

**Figure 17:** OOS Performance (Benchmark: AR(2) Models)



Therefore, we can conclude that the partially nonlinear models can outperform most of the commonly used time series models. And by including the stationary variables  $y_{t-1}$  and the "cay" variable, our partially nonlinear models tend to show a much better out-of-sample performance.

## 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 three-step estimation method to estimate the model and includes a constraint on  $\theta$  (the coefficient for the non-stationary variables) and a truncation condition on  $x_t$ .

From the simulation results, we can see that the estimators have good finite sample properties and good convergence. We also find that the constraint provides finite sample gains for the CLS estimators. In addition, we also find that when using cointegrated  $x_t$ , both the NLS and CLS estimators have a better performance. And among the 7 functional forms we consider, the model with polynomial functional form gives the best simulation performance.

Based on the simulation results, we focus on cointegrated variables and the constrained nonlinear least square method in the empirical study. We apply the model to the [Welch and Goyal \(2008\)](#) dataset and investigate the predictability using co-integrated variable combinations. We find that the partially nonlinear models can outperform sample mean model both in-sample and out-of-sample. When using other benchmark models, such as AR(1) model and AR(2) model, our models also provides better in-sample performance for certain variable combinations and nonlinear functional forms. In terms of out-of-sample, our models outperforms the two time series models in a consecutive period.

And by including lagged dependent variable and the stationary "cay" variable, the partially nonlinear models give better out-of-sample forecast than the nonlinear models over a consecutive period. Therefore, we can conclude that by considering nonlinearities and auto-correlation in the dependent variable, we can achieve out-of-sample forecasting gains.

# Bibliography

- 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.
- Chang, Y. and Park, J. Y. (2003), ‘Index models with integrated time series’, *Journal of Econometrics* **114**(1), 73–106.
- Cheng, T., Gao, J. and Linton, O. (2019), ‘Nonparametric predictive regressions for stock return prediction’.
- 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.
- Gao, J. (2007), *Nonlinear time series: semiparametric and nonparametric methods*, CRC Press.
- Lettau, M. and Ludvigson, S. (2001), ‘Consumption, aggregate wealth, and expected stock returns’, *the Journal of Finance* **56**(3), 815–849.
- 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.
- Zhou, W., Gao, J., Kew, H. and Harris, D. (2018), ‘Semiparametric single-index predictive regression’, *Available at SSRN 3214042*.

# Appendix 1 In-sample Estimation Results

**Table 14:** Full-sample Estimation Results Using Taylor Initials

		$\theta_1$	$\theta_2$	$\beta_1$	$\beta_2$	$\gamma$		
sin_func	co1	-0.699	0.715	-0.600	0.424	0.257		
	co2	-0.775	0.632	0.064	0.399	0.185		
	co3	-0.953	0.302	0.048	2.130	-1.011		
	co4	-0.742	0.670	0.081	0.474	0.236		
cos_func	co1	-0.699	0.715	-0.600	0.424	-1.314		
	co2	-0.775	0.632	0.063	0.398	30.030		
	co3	-0.953	0.302	0.048	2.130	-2.582		
	co4	-0.743	0.670	0.081	0.474	187.161		
scaled_sin_func	co1	0.678	-0.735	-0.139	0.476	9.132	0.333	
	co2	-0.839	0.544	0.069	0.442	0.212	0.615	
	co3	-0.947	0.320	0.100	0.508	0.313	-0.030	
	co4	0.654	-0.756	0.080	0.624	0.295	4.164	
scaled_cos_func	co1	-0.785	0.619	0.182	0.515	1.247	0.137	
	co2	-0.843	0.538	0.069	0.442	-1.359	0.610	
	co3	-0.932	0.361	0.099	0.507	1.261	0.031	
	co4	-0.654	0.756	0.080	0.624	5.007	-4.178	
exp_func	co1	-0.703	0.711	0.221	0.488	0.219	-8.759	
	co2	-0.662	0.750	0.076	0.127	0.062	-248.540	
	co3	-0.977	0.212	0.099	0.509	0.276	0.018	
	co4	-0.690	0.724	0.092	0.473	0.218	-518.196	
exp_shift_func	co1	-0.577	-0.817	0.101	2.228	-5.264	-0.166	
	co2	-0.835	-0.550	0.101	2.228	-8.194E+09	-1.162E+11	
	co3	-0.913	0.408	0.101	-0.003	2.293	-0.097	
	co4	-0.678	0.735	0.101	2.228	19.311	-0.418	
poly_func	co1	-0.712	0.702	2.089	0.622	0.388	-2.940	2.204
	co2	-0.696	0.718	0.068	0.382	0.168	2.277	-61.876
	co3	-0.910	0.415	0.105	0.471	0.646	-0.379	0.082
	co4	-0.661	0.751	0.056	0.708	0.326	-6.473	-372.950
linear_func	co1	-0.713	0.702	2.049	0.643	0.412	-2.792	
	co2	-0.841	0.542	0.069	0.443	0.210	0.600	
	co3	-0.950	0.313	0.099	0.508	0.308	-0.028	
	co4	-0.654	0.757	0.080	0.619	0.288	-3.938	