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:

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

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, $f\left(x_{t-1}'\theta_0; \gamma_0\right)$, 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.

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

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

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 β, 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 $\tilde{\beta}$:

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

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

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

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

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

Let

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

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

Then we have an approximate model of the form:

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

Step 2: estimate (θ, γ) using nonlinear least square(NLS) method:

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

over $(\theta, \gamma) \in (\Theta, \Gamma)$. The $(\hat{\theta}, \hat{\gamma})$ 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. As $\hat{\theta}$ and $\hat{\gamma}$ have been estimated, we can substitute them back to (2.2) and get the estimate of β:

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

.

In the 2-step procedure above, we have obtained the NLS estimators $(\widehat{\theta}, \widehat{\gamma}, \widehat{\beta})$. 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.

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

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

And the CLS estimator of β can be written as:

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

.

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:

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

with

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

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

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 γ_0 varies depending on the nonlinear form $f\left(x'_{t-1}\theta_0, \gamma_0\right)$ we select.

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

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

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

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

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

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

where $\widehat{\alpha}^{(r)}$ denote the r-th replication of the estimate. Following the above definitions, we then calculate biases, standard deviations and covariance for θ, β and γ.

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

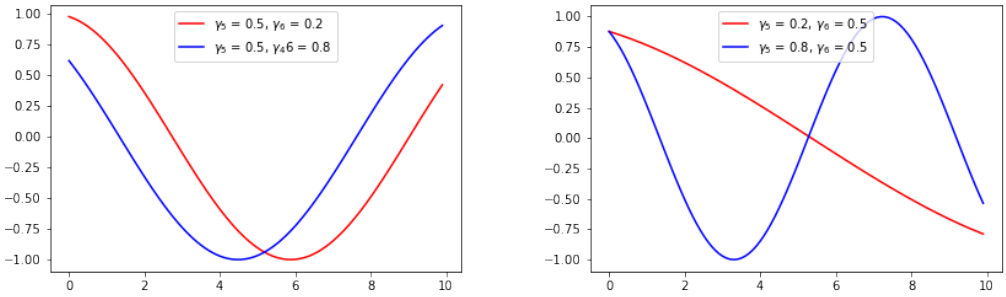
In the partially nonlinear model 1.1, we consider different forms of $f\left(x'_{t-1}\theta_0, \gamma_0\right)$ to include varies type of nonlinearities. The first group of nonlinear functions we consider is the trigonometric functions:

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

where $u_{t-1} = x'_{t-1}\theta_0$. Among the above 4 trigonometric functions, $f_1\left(x'_{t-1}\theta_0, \gamma_0\right)$ and $f_2\left(x'_{t-1}\theta_0, \gamma_0\right)$ can be regarded as special cases of $f_3\left(x'_{t-1}\theta_0, \gamma_0\right)$ and $f_4\left(x'_{t-1}\theta_0, \gamma_0\right)$ with γ_3 and γ_5 equals to 1.

Parameters γ_1, γ_2, γ_4 and γ_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 γ_3 and γ_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



(a) f_4 with different location parameter

(b) f_4 with different shape parameter

In addition, we also consider two exponential functions:

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

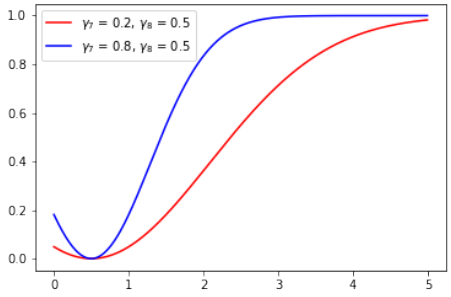
where $\gamma_7, \gamma_{10} \in (0, \infty)$. The two exponential functions are bounded: $f_5\left(x'_{t-1}\theta_0, \gamma_0\right)$ lies in (0,1) and $f_6\left(x'_{t-1}\theta_0, \gamma_0\right)$ lies in (0, γ_9).

Figure 2 and 3 present the plots of $f_5\left(x'_{t-1}\theta_0, \gamma_0\right)$ and $f_6\left(x'_{t-1}\theta_0, \gamma_0\right)$. 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 γ_7 changes the steepness of $f_5\left(x'_{t-1}\theta_0, \gamma_0\right)$ (figure 2a) and γ_8 changes its position (figure 2b). And for $f_6\left(x'_{t-1}\theta_0, \gamma_0\right)$, γ_9 acts as the upper bound for the function and changes its position (figure 3a), while γ_{10} changes its steepness (figure 3b).

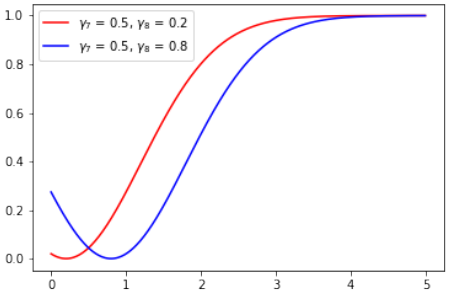
In addition, we also consider a quadratic polynomial function:

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

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

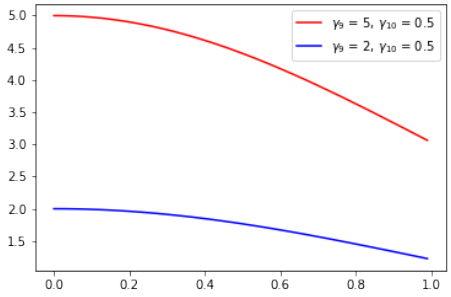


(a) f_5 with different γ_7

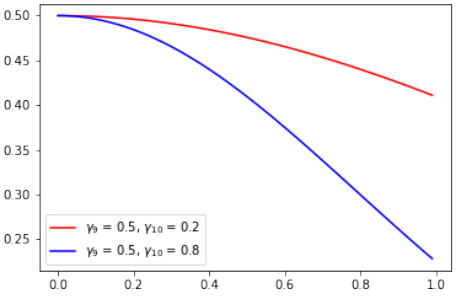


(b) f_5 with different γ_8

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



(a) f_6 with different γ_7



(b) f_6 with different γ_8

It is a parabola where γ_{11} decides the y-intercept and γ_{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 known 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 $(1, -\hat{\eta})$:

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

Then θ_0 is given by normalizing the cointegration vector $(1, -\hat{\eta})$:

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

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 nonlinear part $f\left(\hat{u}_{t-1}, \gamma_0\right)$. For the two sine functions: $f_1\left(u_{t-1}, \gamma_0\right) = \sin\left(u_{t-1} + \gamma_1\right)$ and $f_3\left(\gamma_3 u_{t-1}, \gamma_4\right) = \sin\left(\gamma_3 u_{t-1} + \gamma_4\right)$, 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 (β, γ) for different functional forms can be calculated by estimating the following linear regressions:

$$\begin{aligned}
y_t &= \gamma_1 + \hat{u}_{t-1} + \beta_1 y_{t-1} + \beta_2 w_{t-1} + e_t \\
y_t &= \gamma_2 + \hat{u}_{t-1} + \beta_1 y_{t-1} + \beta_2 w_{t-1} + e_t \\
y_t &= \gamma_4 + \gamma_3 \hat{u}_{t-1} + \beta_1 y_{t-1} + \beta_2 w_{t-1} + e_t \\
y_t &= \gamma_6 + \gamma_5 \hat{u}_{t-1} + \beta_1 y_{t-1} + \beta_2 w_{t-1} + e_t \\
y_t &= \gamma_7 \hat{u}_{t-1}^2 - 2\gamma_7 \gamma_8 \hat{u}_{t-1} + \gamma_7 \gamma_8^2 + \beta_1 y_{t-1} + \beta_2 w_{t-1} + e_t \\
y_t &= \gamma_9 (1 - \gamma_{10} \hat{u}_{t-1}^2 + \frac{\gamma_{10}^2 \hat{u}_{t-1}^4}{2}) + \beta_1 y_{t-1} + \beta_2 w_{t-1} + e_t \\
y_t &= \gamma_{11} + \gamma_{12} \hat{u}_{t-1} + \gamma_{13} \hat{u}_{t-1}^2 + \beta_1 y_{t-1} + \beta_2 w_{t-1} + e_t
\end{aligned}$$

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

Together with θ_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 $e_t \sim i.i.d.N(0,1), \quad t = 2, ..., T..$ 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 θ_1 as an example, when using true values, the standard deviation of θ_1 is 0.00029, while when initial values equal to 0, standard deviation of θ_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 θ_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 γ: $\text{cov}(\gamma)$ 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

		θ_1	θ_2	β_1	β_2	γ_1	γ_2	γ_3
		Initial Values = Taylor Initials						
	Bias	0.00088	0.00026	-0.00079	0.00029	0.00326	-0.00064	0.00121
T=100	std	0.00139	0.00105	0.00504	0.00504	0.00914	0.00451	0.00141
		0.00244		0.00461			0.01029	
	Bias	-0.00013	-0.00008	-0.00040	-0.00058	0.00063	-0.00032	0.00146
T=500	std	0.00041	0.00031	0.00068	0.00206	0.00436	0.00175	0.00043
		0.00167		0.00167			0.00472	
	Bias	-0.00005	-0.00003	-0.00025	-0.00012	0.00035	-0.00021	0.00087
T=1000	std	0.00026	0.00019	0.00043	0.00136	0.00319	0.00110	0.00026
		0.00045		0.00110			0.00339	
		Initial Values = 0						
	Bias	0.12214	0.35638	0.09401	-0.27546	2.40099	0.19899	-0.10567
T=100	std	0.28186	0.45009	0.31824	1.10049	2.95226	0.67484	0.36090
		0.51793		1.02441			0.17287	
	Bias	0.13182	0.47163	0.19141	-0.33712	2.25798	0.19733	-0.18554
T=500	std	0.28545	0.52487	0.38507	1.13323	2.50154	0.78597	0.38406
		0.39734		0.57196			0.12342	
	Bias	0.15770	0.47870	0.21117	-0.25275	2.34776	0.30953	-0.20737
T=1000	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						
	Bias	0.00008	0.00007	-0.00170	0.00082	0.00368	-0.00028	0.00080
T=100	std	0.00127	0.00095	0.00196	0.00572	0.00735	0.00350	0.00174
		0.00222		0.00261			0.00832	
	Bias	0.00003	0.00002	-0.00098	0.00085	0.00233	-0.00003	0.00068
T=500	std	0.00029	0.00022	0.00069	0.00192	0.00330	0.00135	0.00060
		0.00050		0.00098			0.00362	
	Bias	0.00000	0.00001	-0.00091	0.00084	0.00023	-0.00003	0.00068
T=1000	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 5 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 NLSand 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 β_1 in $f_1(u_{t-1}, \gamma_1)$ is -0.00007 and 0.01574 respectively, but the values for the NLS estimator is 0.00154 and 0.02500.

The covariance of θ, β and γ are also smaller when the constraint is applied. Take the covariance of θ using $f_2(u_{t-1}, \gamma_2)$ as an example, $\text{cov}(\hat{\theta})$ dropped from 0.09124 and 0.09918, 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				
		θ_1	θ_2	β_1	β_2	γ_1	θ_1	θ_2	β_1	β_2	γ_1
						$f_1(u_{t-1}, \gamma_1)$					
	Bias	-0.00550	-0.00190	-0.00295	0.00334	-0.03311	-0.00145	-0.00116	-0.00038	0.00089	0.04358
T = 100	std	0.03268	0.03688	0.00467	0.02733	0.12446	0.01523	0.01683	0.00122	0.01254	0.08301
		0.04608		0.02783			0.02050		0.01191		
	Bias	0.00933	0.00062	-0.00297	-0.00311	-0.06790	-0.00408	-0.00063	-0.00012	0.00147	0.00911
T = 500	std	0.00444	0.00625	0.00128	0.00672	0.01740	0.00369	0.00432	0.00021	0.00364	0.03815
		0.07470		0.06914			0.00574		0.00345		
	Bias	0.00154	0.00065	0.00130	-0.00212	-0.00327	-0.00147	-0.00026	-0.00007	0.00100	0.00286
T = 1000	std	0.02500	0.01999	0.04538	0.05794	0.06140	0.01892	0.01693	0.01574	0.02946	0.01798
		0.03173		0.05844			0.00249		0.00279		
						$f_2(u_{t-1}, \gamma_2)$					
	Bias	-0.00887	-0.00870	0.06036	0.01303	-0.02714	-0.00424	-0.00120	-0.00100	0.00334	0.04116
T = 100	std	0.04654	0.06049	0.00889	0.05785	0.02916	0.01781	0.01584	0.00222	0.01433	0.08129
		0.07372		0.05954			0.02291		0.01280		
	Bias	-0.00112	-0.00230	-0.00154	0.00344	-0.01280	-0.00147	-0.00052	-0.00036	0.00208	0.00771
T = 500	std	0.01444	0.01375	0.00311	0.01694	0.01270	0.00396	0.00411	0.00065	0.00553	0.03370
		0.09124		0.01770			0.00536		0.00493		
	Bias	-0.00036	0.00869	-0.00193	0.00172	-0.00788	-0.00074	-0.00032	-0.00017	0.00096	0.00287
T = 1000	std	0.00713	0.00670	0.00168	0.01557	0.00863	0.00222	0.00174	0.00040	0.00367	0.01557
		0.09918		0.01590			0.00283		0.00328		

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

		NLS						Constrained-NLS					
		theta	theta	beta	beta	gamma	gamma	theta	theta	beta	beta	gamma	gamma
							f3						
	Bias	0.25080	0.11643	0.09701	0.09621	-0.02764	0.04677	-0.005119	-0.009574	0.003672	0.040889	0.000233	0.000843
T = 100	std	0.42010	0.27514	0.10589	0.14919	0.02840	0.04952	0.122508	0.091186	0.024306	0.111065	0.010231	0.021262
		0.49218		0.17846		0.02405		0.148196		0.115617		0.023531	
	Bias	0.18672	0.09059	0.07042	0.11776	-0.01412	0.02610	-0.006816	-0.004300	0.000619	-0.000600	-0.000829	0.000783
T = 500	std	0.37435	0.24955	0.09681	0.14953	0.01354	0.02486	0.025858	0.022944	0.006150	0.023337	0.002762	0.004753
		0.44802		0.17453		0.01163		0.037832		0.023827		0.005613	
	Bias	0.14667	0.08110	0.05778	0.11680	-0.00923	0.01717	-0.001771	-0.001621	0.000129	-0.001485	-0.000329	0.000380
T = 1000	std	0.34103	0.23593	0.09100	0.14950	0.00941	0.01754	0.014331	0.013881	0.004064	0.015287	0.001655	0.002476
		0.41141		0.17074		0.00825		0.022410		0.015683		0.002679	
							f4						
	Bias	0.26070	0.13566	0.09572	0.10944	-0.00129	0.00539	-0.008391	-0.007001	0.000848	0.010916	-0.026530	0.042717
T = 100	std	0.41629	0.28334	0.10132	0.14225	0.00287	0.01948	0.050088	0.044224	0.013495	0.053716	0.027326	0.047434
		0.47705		0.17498		0.01708		0.064981		0.056411		0.025406	
	Bias	0.18304	0.09114	0.07063	0.12854	-0.00039	0.00245	-0.001530	-0.000129	-0.000441	0.004487	-0.012395	0.022335
T = 500	std	0.37174	0.24185	0.09417	0.14590	0.00071	0.00598	0.013671	0.013996	0.003518	0.017634	0.012163	0.022159
		0.43332		0.17804		0.00530		0.020391		0.018473		0.010363	
	Bias	0.13966	0.06411	0.05101	0.12171	-0.00023	0.00163	-0.000149	0.000492	-0.000255	0.002375	-0.007772	0.014324
T = 1000	std	0.33860	0.21123	0.08496	0.14666	0.00042	0.00383	0.006464	0.006840	0.001630	0.015591	0.008064	0.014897
		0.40072		0.17760		0.00342		0.010085		0.015883		0.006995	

Simulation Results for Cointegrated x_t

Table 6 to 9 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 θ_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 γ in the second panel as an example, the covariance of γ decrease from 0.01916 to 0.00242 when sample size T rises from 100 to 1000.

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,

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

		NLS						Constrained-NLS					
		theta	theta	beta	beta	gamma	gamma	theta	theta	beta	beta	gamma	gamma
							f5						
T = 100	Bias	0.12540	0.08285	0.03502	0.10013	-0.00125	0.00465	-0.015121	-0.008498	0.002983	0.035998	0.000354	0.000099
	std	0.32644	0.22965	0.07947	0.14053	0.00291	0.01768	0.089389	0.097611	0.016707	0.107387	0.010282	0.020406
		0.41166		0.16238		0.01553		0.136240		0.108824		0.022402	
T = 500	Bias	0.07635	0.03312	0.02314	0.07846	-0.00040	0.00256	-0.006261	-0.005330	0.000978	0.000203	-0.000791	0.000692
	std	0.26124	0.15768	0.06617	0.13341	0.00061	0.00513	0.024477	0.023667	0.006174	0.029384	0.002705	0.004183
		0.33234		0.15343		0.00456		0.038780		0.029935		0.005000	
T = 1000	Bias	0.04853	0.02568	0.01525	0.06429	-0.00021	0.00148	-0.001450	-0.001384	0.000064	-0.001541	-0.000334	0.000268
	std	0.21269	0.13821	0.05665	0.12720	0.00038	0.00335	0.013480	0.013180	0.003425	0.015259	0.001756	0.002307
		0.27633		0.13912		0.00298		0.020501		0.015664		0.002670	
							f6						
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
	Bias	0.002111	-0.0016	3.087179	-0.09573	-0.00189	-0.06007	0.072155	0.00029	0.00008	0.00633	-0.00099	0.00037	-0.00109	0.00073
T = 100	std	0.36946	0.331152	7.401177	0.74312	0.294723	0.178366	1.414562	0.00360	0.00271	0.03609	0.00582	0.00242	0.00399	0.01136
		0.048890075		0.074443156		0.01423			0.00514		0.03664		0.01101		
	Bias	0.002983	0.00214	0.307132	-0.06586	0.001327	-0.00554	0.016146	0.00007	0.00003	0.00223	-0.00009	0.00003	-0.00013	0.00002
T = 500	std	0.04705	0.043007	1.972141	0.285891	0.041959	0.047356	0.338921	0.00036	0.00035	0.01185	0.00218	0.00030	0.00042	0.00210
		0.00704		0.01993		0.00342			0.00058		0.01206		0.00209		
	Bias	0.000261	-0.00032	0.248519	-0.0176	0.000291	-0.0011	-0.00152	0.00001	0.00002	0.00074	-0.00001	0.00001	-0.00005	-0.00001
T = 1000	std	0.020383	0.017837	1.113794	0.190168	0.020954	0.016493	0.19888	0.00016	0.00014	0.01082	0.00105	0.00014	0.00022	0.00093
		0.00030		0.01130		0.00198			0.00026		0.01087		0.00088		

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

		NLS					Constrained-NLS				
		θ_1	θ_2	β_1	β_2	γ_1	θ_1	θ_2	β_1	β_2	γ_1
						$f_1(u_{t-1}, \gamma_1)$					
	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		
						$f_2(u_{t-1}, \gamma_2)$					
	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		

in the first panel of table 7, the absolute value of bias of θ_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 γ ($\text{cov}(\hat{\gamma})$) 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					
		θ_1	θ_2	β_1	β_2	γ_1	γ_2	θ_1	θ_2	β_1	β_2	γ_1	γ_2
$f_3(u_{t-1}, \gamma_3, \gamma_4)$													
	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	
$f_4(u_{t-1}, \gamma_5, \gamma_6)$													
	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	

For the models containing two exponential functions $f_5\left(u_{t-1}, \gamma_7, \gamma_8\right) = 1 - e^{-\gamma_7(u_{t-1}-\gamma_8)^2}$ and $f_6\left(u_{t-1}, \gamma_9, \gamma_{10}\right) = \gamma_9 e^{-\gamma_{10} u_{t-1}^2}$, 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 β_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 β_1 drop from 0.08371

to 0.00105.

The CLS estimators also shows a significant decrease in the covariance. Take β 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					
		θ_1	θ_2	β_1	β_2	γ_1	γ_2	θ_1	θ_2	β_1	β_2	γ_1	γ_2
$f_5(u_{t-1}, \gamma_7, \gamma_8)$													
	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
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	-0.00105	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	
$f_6(u_{t-1}, \gamma_9, \gamma_{10})$													
	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	-0.00107	-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 9 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 β and γ.

As for the convergence, take β_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 performance for all the estimators, but the effect is more significant for the estimator γ. For example, when T = 1000, the covariance of β in the case of NLS is 0.00190 and the value

is 0.00110 in the case of CLS. And for the covariance of γ, 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						
		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	

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 9, the standard deviation of θ 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 $f_2(u_{t-1}, \gamma_0)$).

In next section, we will investigate how models with different nonlinear component perform empirical. As we have shown in simulation that the CLS estimator consistently provides better performance, we will focus on the constrained least square method in the empirical study.

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.

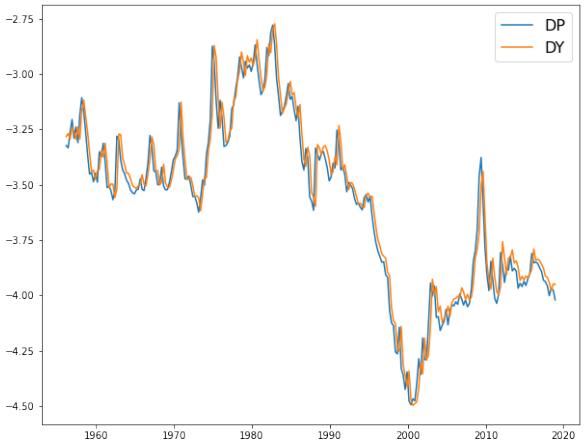
  The earning-price ratio is the difference between the log of earnings on the S&P 500 index and the log of stock prices.

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

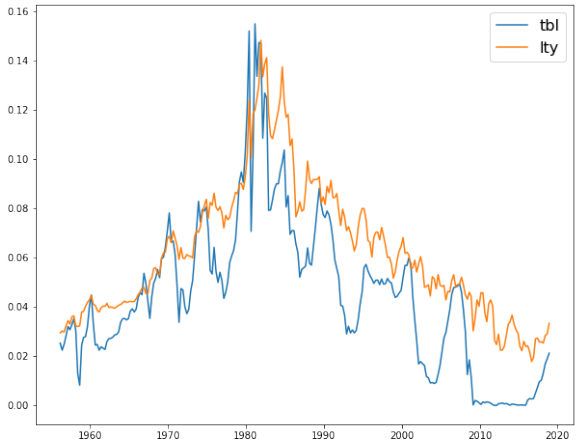
Figure 4 plots the time series of the co-integrated variable pairs. It is clear that the variables in each plot are not stationary and share a similar trend.

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

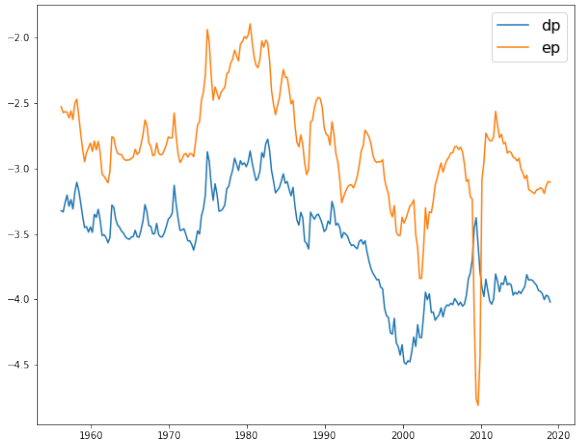
Figure 4: Time Series Plots of Co-integrated Variables



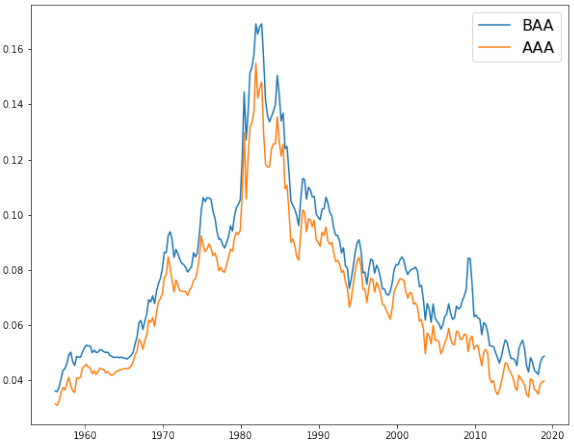
(a) co1



(b) co2



(c) co3



(d) co4

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 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 non-linear 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.

$$\begin{aligned}
\sin : \tilde{g}_1\left(u_{t-1}, \gamma_0\right) &= \left(\gamma_1 + u_{t-1}\right) + \beta z'_{t-1}, \\
\cos : \tilde{g}_2\left(u_{t-1}, \gamma_0\right) &= 1 - \left(\gamma_1 + u_{t-1}\right)^2 + \beta z'_{t-1}, \\
\text{sin_scaled} : \tilde{g}_3\left(u_{t-1}, \gamma_0\right) &= \gamma_2 + \gamma_1 u_{t-1} + \beta z'_{t-1}, \\
\text{cos_scaled} : \tilde{g}_4\left(u_{t-1}, \gamma_0\right) &= 1 - \left(\gamma_2 + \gamma_1 u_{t-1}\right)^2 + \beta z'_{t-1}, \\
\text{exp_shift} : \tilde{g}_5\left(u_{t-1}, \gamma_0\right) &= \gamma_{1,0} u_{t-1}^2 - 2\gamma_{1,0}\gamma_{2,0} u_{t-1} + \gamma_{1,0}\gamma_{2,0}^2 + \beta z'_{t-1} \\
\exp : \tilde{g}_6\left(u_{t-1}, \gamma_0\right) &= \gamma_{1,0}(1 - \gamma_{2,0} u_{t-1}^2 + \frac{\gamma_{2,0}^2 u_{t-1}^4}{2}) + \beta z'_{t-1}, \\
\text{Polynomial} : \tilde{g}_7\left(u_{t-1}, \gamma_0\right) &= \gamma_1 + \gamma_2 u_{t-1} + \gamma_3 u_{t-1}^2 + \beta z'_{t-1}, \\
\text{constrained_linear} : \tilde{g}_8\left(u_{t-1}, \gamma_0\right) &= \gamma_1 + \gamma_2 u_{t-1} + \beta z'_{t-1}
\end{aligned}$$

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^2_{IS} as a measurement for the in-sample performance:

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

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

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

where MSE_{bm} is the mean squared error of benchmark model and $\text{MSE}_{CLS} = 1/n \sum_{t=1}^{n} (y_t - \hat{y}_t)^2$ 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^2_{IS} 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^2_{IS} is reported in table 10. Among the 8 functional forms, g_3, g_4, g_5, g_7 and g_8 provide positive R^2_{IS} 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.

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:

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

Table 10: Results of R^2_{IS} for all the models (benchmark: sample mean model)

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

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 (5) and (6) 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.

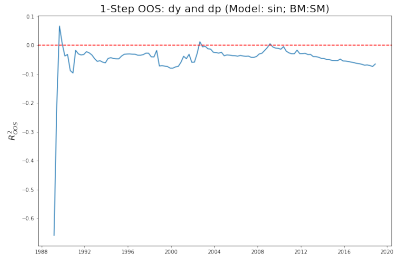
Figure (7) and (8) display results for the scaled trigonometric functions g_3 and g_4. These two functions are different from g_1 and g_2 in that they have a scale parameter in front of the single-index while g_1 and g_2 do not. Similar to the trigonometric functions, g_3 and g_4 only show positive R^2_{OOS} 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 (9) and (10). 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).

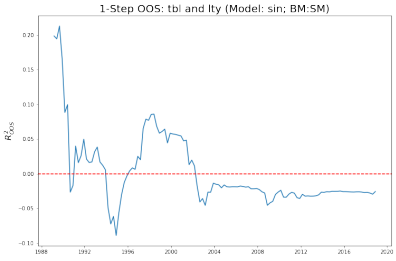
For functional g_6, the positive R^2_{OOS} can be found in the first half of the forecasting period. We can see from sub-figures (b) and (d) of figure (10) that g_6 gives consecutive positive results before 2014 when using combinations co2 and co4.

Figure (11) 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 (12) 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.

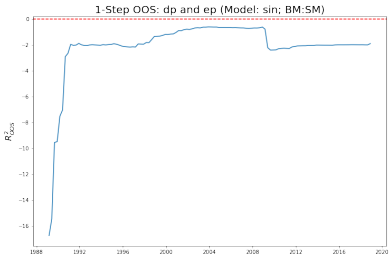
Figure 5: OOS Results for Model with f_1



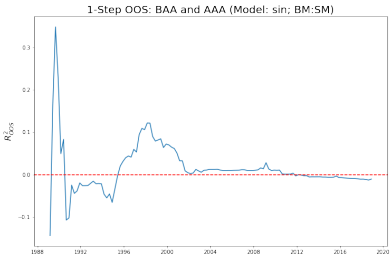
(a) co1



(b) co2

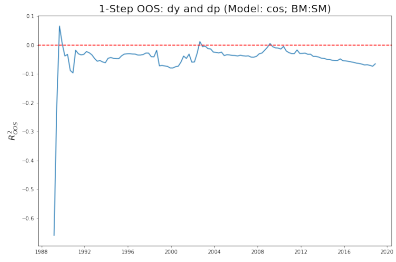


(c) co3

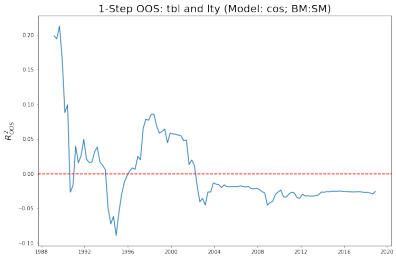


(d) co4

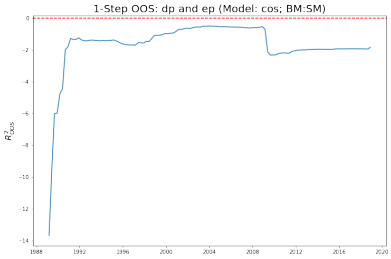
Figure 6: OOS Results for Model with f_2



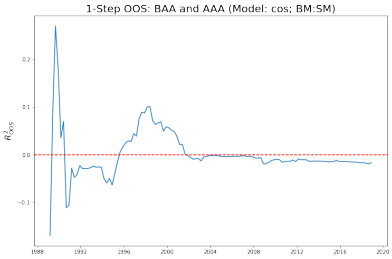
(a) co1



(b) co2

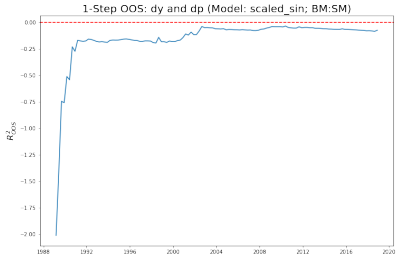


(c) co3

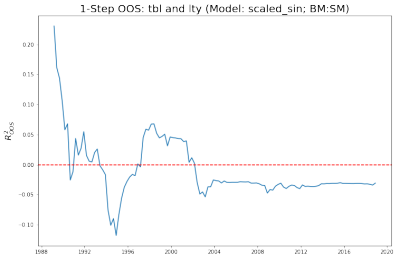


(d) co4

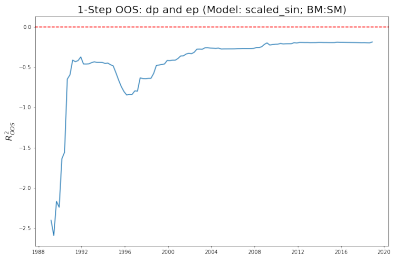
Figure 7: OOS Results for Model with f_3



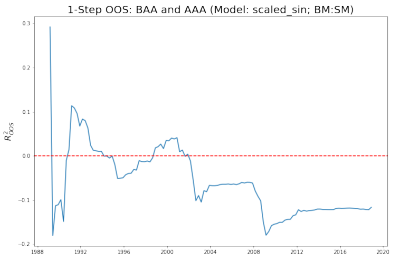
(a) co1



(b) co2

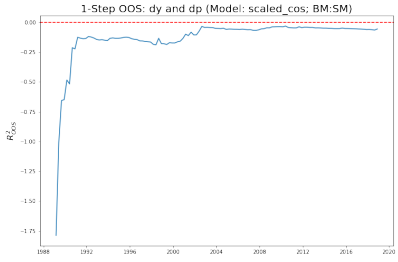


(c) co3

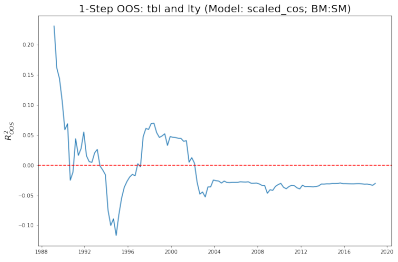


(d) co4

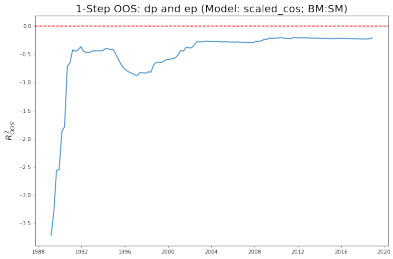
Figure 8: OOS Results for Model with f_4



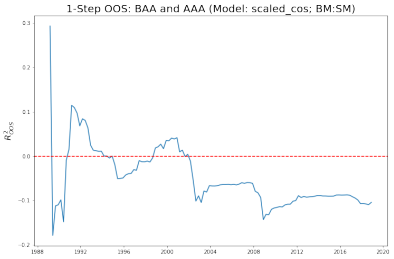
(a) co1



(b) co2

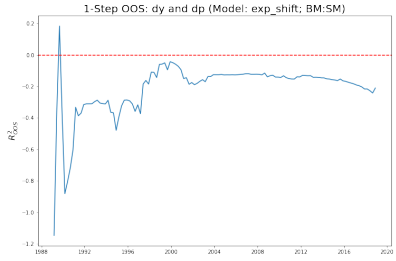


(c) co3

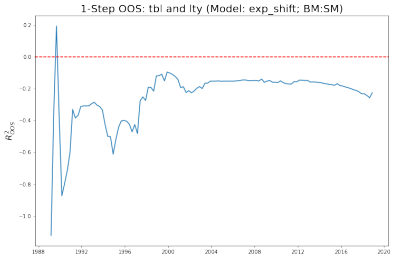


(d) co4

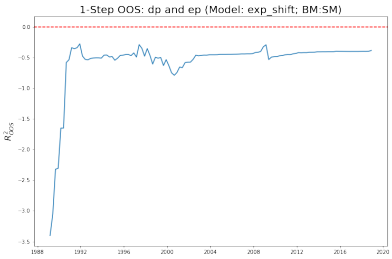
Figure 9: OOS Results for Model with f_5



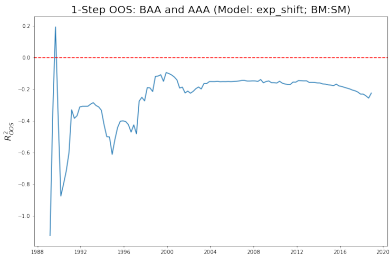
(a) co1



(b) co2

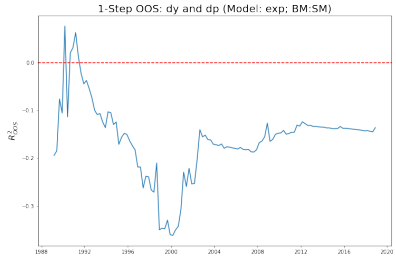


(c) co3

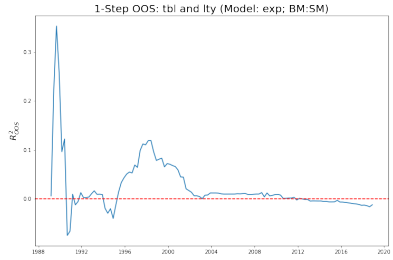


(d) co4

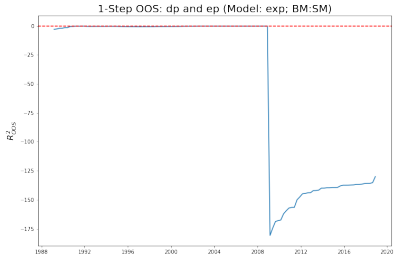
Figure 10: OOS Results for Model with f_6



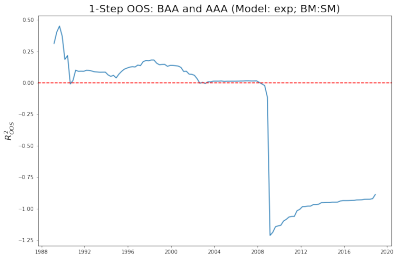
(a) co1



(b) co2

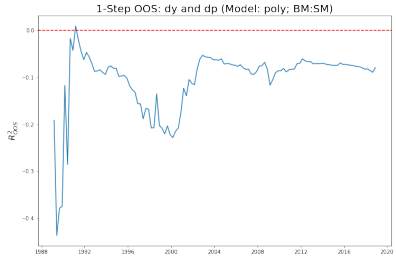


(c) co3

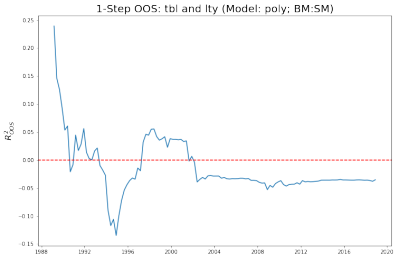


(d) co4

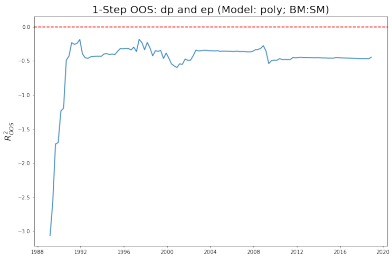
Figure 11: OOS Results for Model with f_7



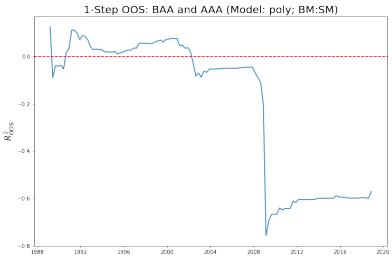
(a) co1



(b) co2

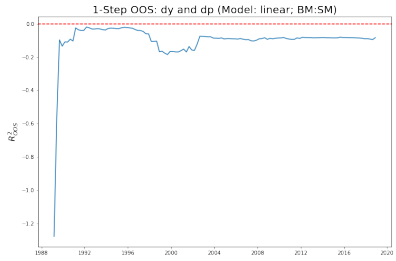


(c) co3

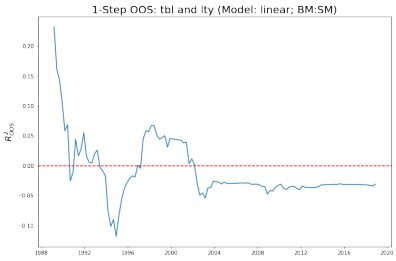


(d) co4

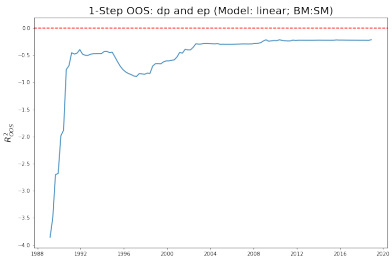
Figure 12: OOS Results for Model with g_8



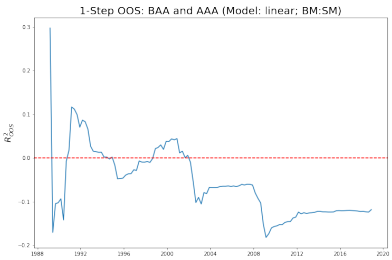
(a) co1



(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 (11). 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 11: 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 pre-

dictability 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.

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* .