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:

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

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

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

fits the data better than the nonlinear specification.

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

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

2 Model and Methodology

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

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

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

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

The minimum value of $L(\beta, \theta, \gamma)$ occurs when the above score functions equals to 0. Notice that these score functions $\frac{\partial L(\beta,\theta,\gamma)}{\partial \theta}$ and $\frac{\partial L(\beta,\theta,\gamma)}{\partial \gamma}$ are nonlinear functions of both the variables and the parameters, and so they do not have closed form solutions. To estimate the parameters, we need to include an iterative procedure and obtain optimal values by using gradient descent algorithms. However, recognising that this score function $\frac{\partial L(\beta,\theta,\gamma)}{\partial \beta}$

is linear in the parameter β, we introduce a novel two-step approach for estimation in order to reduce the computational burden.

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

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

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

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

which is equivalent to:

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

Let

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

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

Then we have an approximate model of the form:

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

We can estimate (θ, γ) by minimizing:

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

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

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

which can be solved using an iterative procedure since there is no closed form solution. In order to improve finite sample properties of the estimators, we impose a truncation

condition $I\left(\|x_{t-1}\| \leq M_T\right)$ on x_{t-1} and an identification condition on coefficient vector θ. We then define the modified sum-of-squared errors by:

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

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

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

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

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

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

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

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

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

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 $\hat{\theta}_0 = (0.8, -0.6)'$, $\beta_{1,0} = 0.5$ and $\beta_{2,0} = 1.0$. And γ_0 will vary depending on the nonlinear form $f\left(x'_{t-1}\theta_0, \gamma_0\right)$ we select.

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.

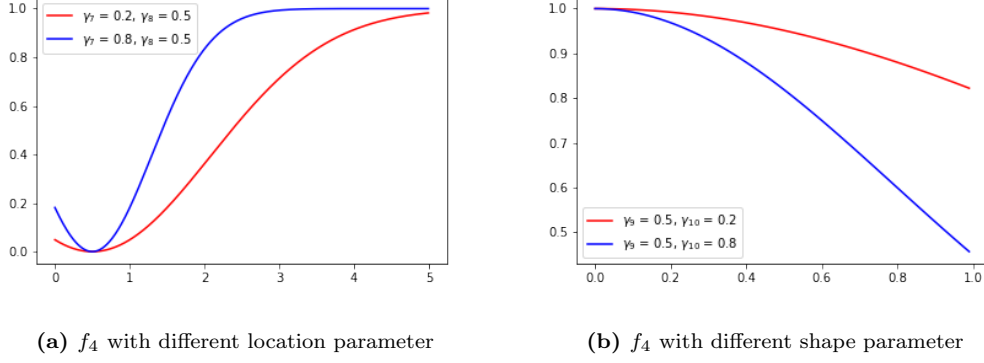**Figure 1:** Plots for Trigonometric Functions

(a) f_4 with different location parameter

(b) f_4 with different shape parameter
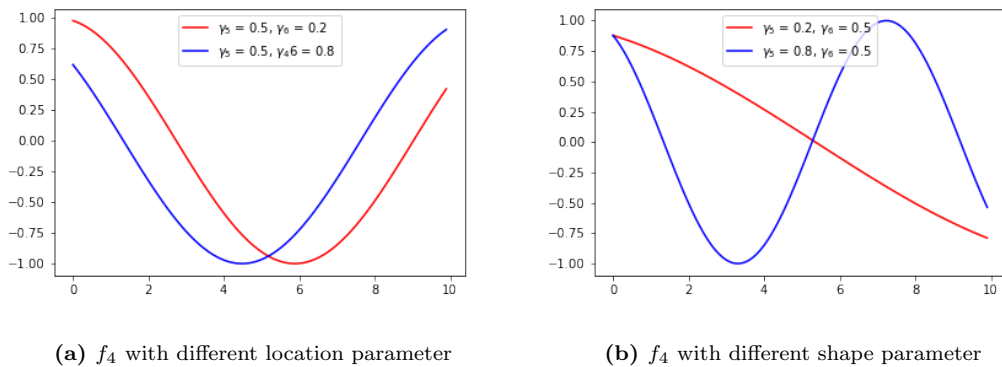
In addition, we also consider two exponential functions:

$$
\text{exp\_shift} : f_5\left(u_{t-1}, \gamma_7, \gamma_8\right) = 1 - e^{-\gamma_7\left(u_{t-1} - \gamma_8\right)^2}
$$

$$\exp : f_6\left(u_{t-1}, \gamma_9, \gamma_{10}\right) \quad = \quad \gamma_9 e^{-\gamma_{10} u_{t-1}^2},$$

where $\gamma_7, \gamma_{10} \in (0, \infty)$.

Figure 2: Plots for Exponential Functions

(a) f_4 with different location parameter   (b) f_4 with different shape parameter

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

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

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

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

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

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

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

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

3.2 Initial Values

As our model is partially nonlinear, the gradient functions do not have a closed solution. Therefore, we use an iterative procedure to estimate the model. To start the iterative

procedure, initial values are necessary. We can assign random values as the initial values, however, finding initial values that are close to the optimal values is crucial.

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

Step 1:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.3   Simulation Results for Co-integrated x_t

Table 1 to 4 show the simulation results on co-integrated x_t using partially nonlinear models. From the simulation results, we find that:

1. Both NLS and constrained NLS estimators converge when sample size increases.

2. The constrained NLS perform better than the normal NLS for most functional forms as it gives smaller biases and standard deviations.

8

3. Among the 7 different partially nonlinear, the polynomial function $(f_7(u_{t-1}, \gamma_0))$ gives the best performance.

As presented in the tables, both NLS and constrained NLS estimators converge for most functional forms when sample size increases. For example, in table 2, the absolute value of bias of θ_1 using $f_3(u_{t-1}, \gamma_0)$ decreases from 0.00219 to 0.00062 in the case of constrained NLS. And in the case of NLS, the number also decreases from 0.01336 to 0.00201.

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

Among the 7 partially nonlinear models, we find that the polynomial functional form tend to provide the best results. As shown in table 4, the standard deviation of θ using constrained NLS is 0.00045 when T=1000. The corresponding value is 0.00262 in $f_1(u_{t-1}, \gamma_0)$ and 0.00476 in $f_2(u_{t-1}, \gamma_0)$.

9

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

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

Table 2: Simulation Results for $f_3(u_{t-1}, \gamma_0)$ and $f_4(u_{t-1}, \gamma_0)$

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

Table 3: Simulation Results for $f_5(u_{t-1}, \gamma_0)$ and $f_6(u_{t-1}, \gamma_0)$

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

Table 4: Simulation Results for $f_7(u_{t-1}, \gamma_0)$

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

4 Empirical Study

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

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

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

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

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

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

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

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

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

Figure 3 plots the time series of the co-integrated variable pairs. It is clear that the variables in each plot are not stationary and share a similar trend.

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

Figure 3: Time Series Plots of Co-integrated Variables
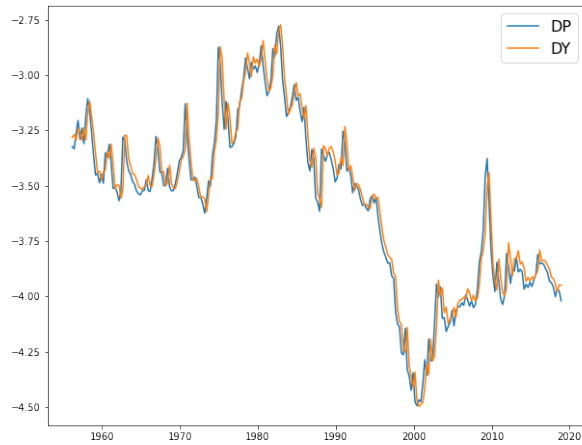
(a) co1
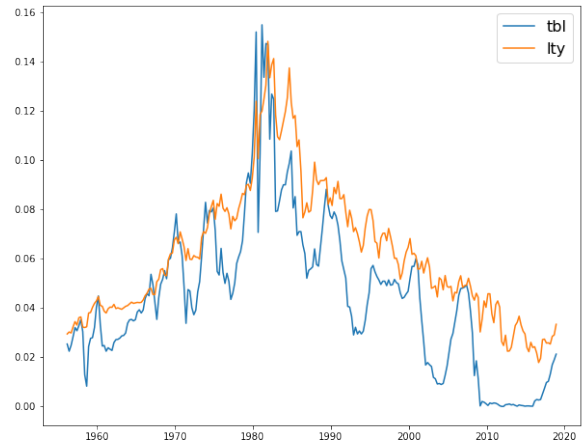
(b) co2
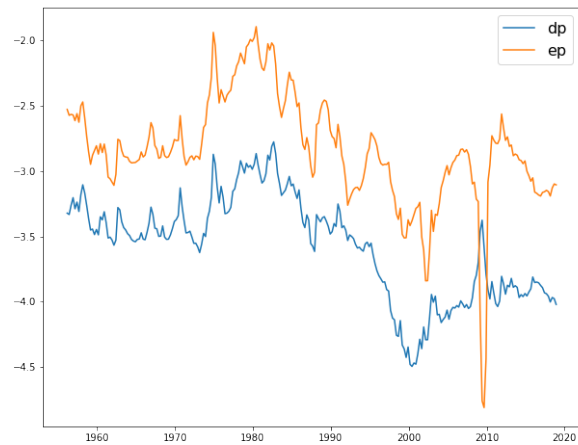
(c) co3
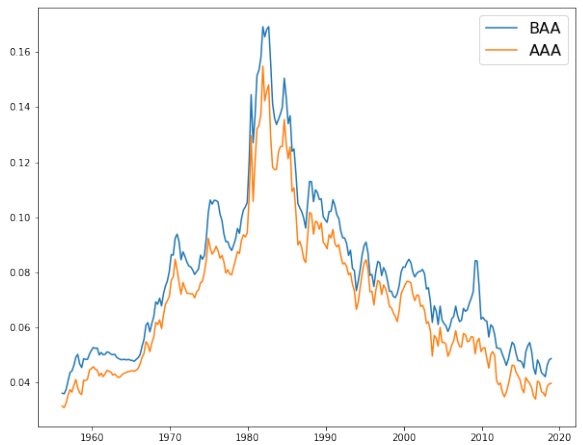
(d) co4

14

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

Therefore in the partially nonlinear models we consider:

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

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

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

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

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

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

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

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

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

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

15

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 $\hat{\alpha}$:

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

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 $f(u_{t-1}, \gamma)$ by its Taylor expansion and estimate the following linear regression to obtain the initial values for β and γ.

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

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:

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

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:

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

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

In equation (4.2), if R^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 5. 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.

Table 5: 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		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

4.2 OOS Results

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

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

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

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

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

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

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

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

The OOS results for different functional forms are shown in the following figures. We report out-of-sample results in 8 figures, each of which shows the prediction of different

functional forms using the 4 co-integrated combinations. We put the R^2_{OOS} statistics on the vertical axis and the beginning of the various out-of-sample evaluation periods on the horizontal axis.

Figure (4) and (5) 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 (6) and (7) 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 (8) and (9). 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 (9) that g_6 gives consecutive positive results before 2014 when using combinations co2 and co4.

Figure (10) 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 (11) 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 4: OOS Results for Model with f_1
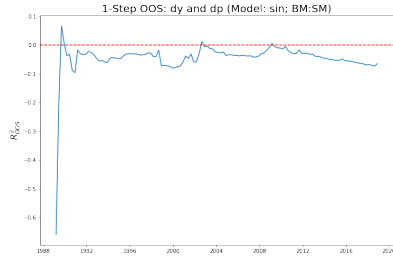
(a) co1
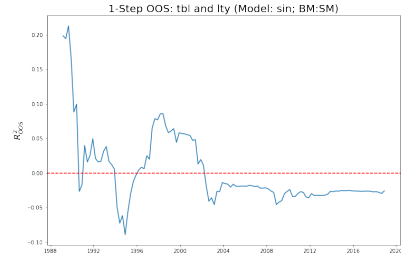
(b) co2
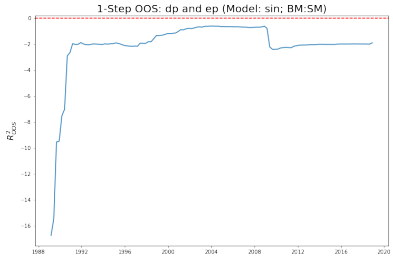
(c) co3
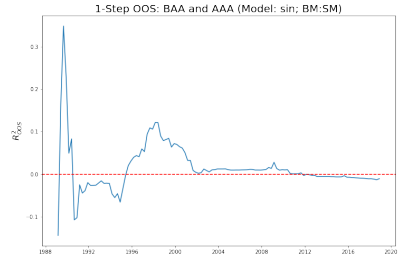
(d) co4

Figure 5: OOS Results for Model with f_2
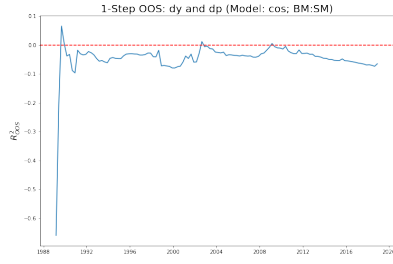
(a) co1
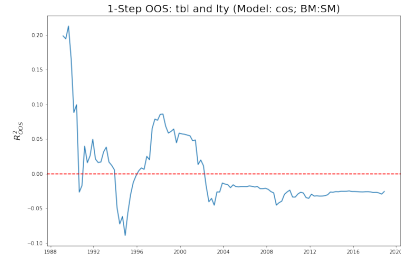
(b) co2
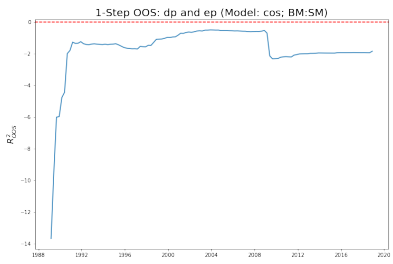
(c) co3
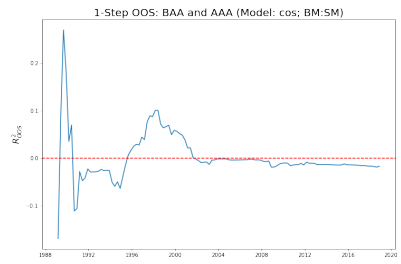
(d) co4

Figure 6: OOS Results for Model with f_3
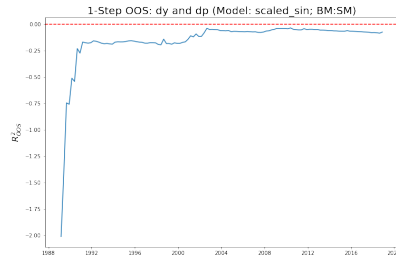
(a) co1
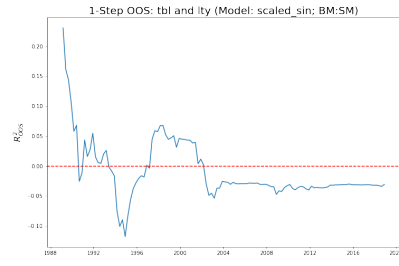
(b) co2
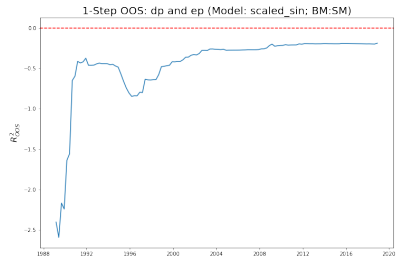
(c) co3
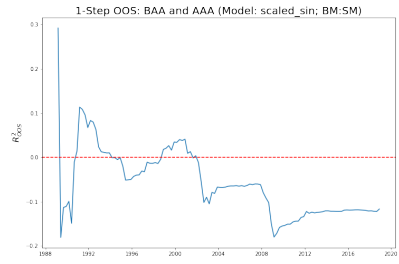
(d) co4

Figure 7: OOS Results for Model with f_4

(a) co1

(b) co2

(c) co3

(d) co4

21
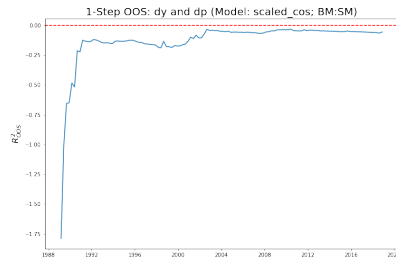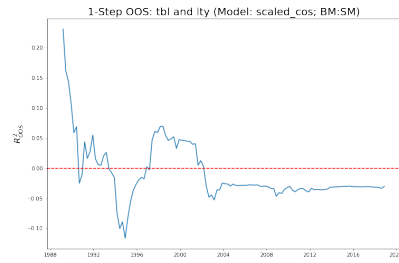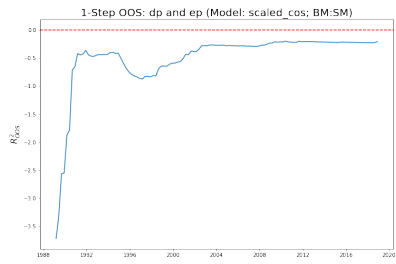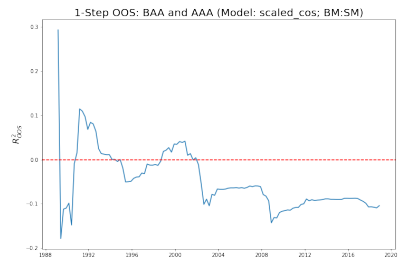
Figure 8: OOS Results for Model with f_5
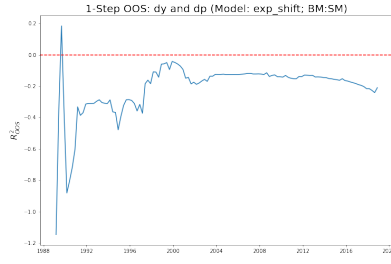
(a) co1
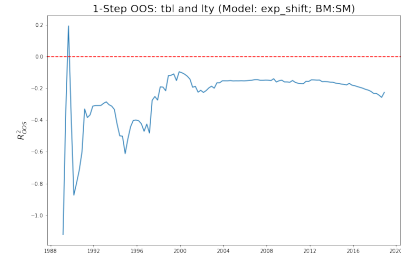
(b) co2
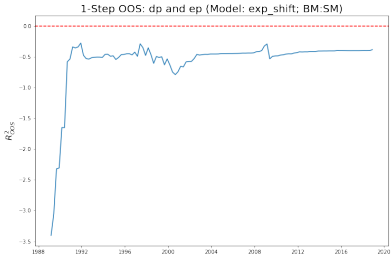
(c) co3
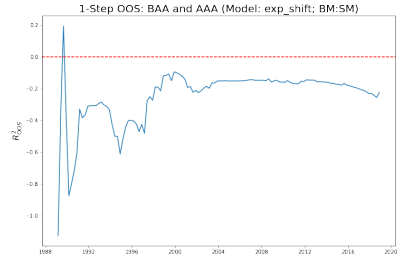
(d) co4

Figure 9: OOS Results for Model with f_6

(a) co1

(b) co2

(c) co3

(d) co4

22
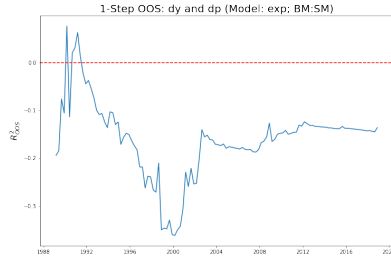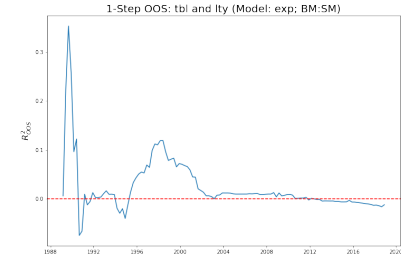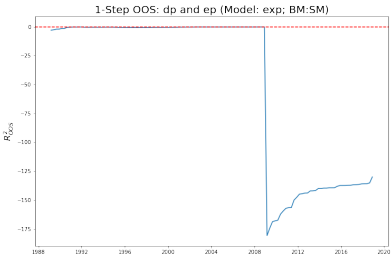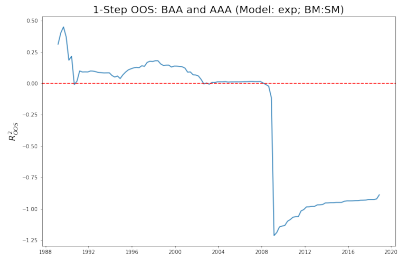
Figure 10: OOS Results for Model with f_7
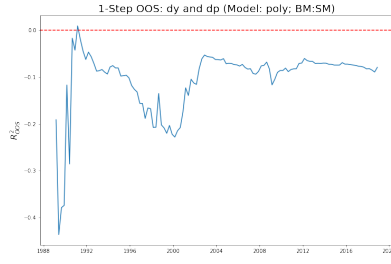
(a) co1
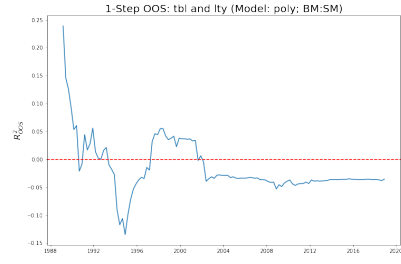
(b) co2
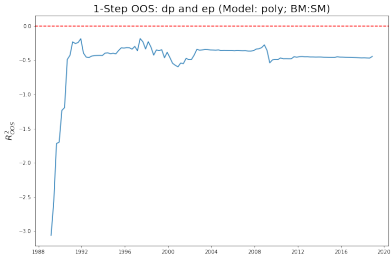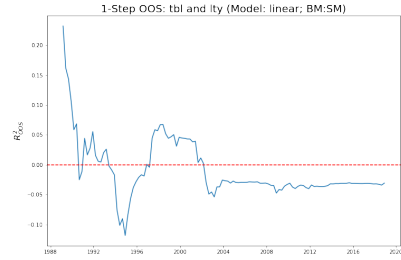
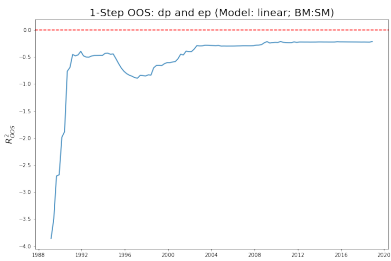**(c)** co3
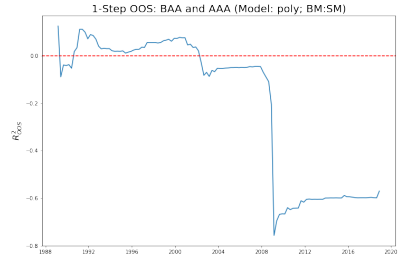
(d) co4

Figure 11: OOS Results for Model with g_8
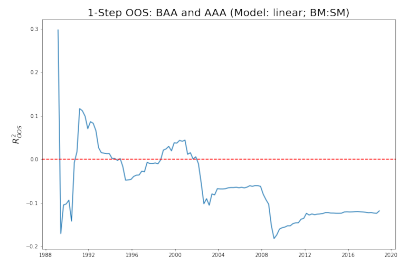

(a) co1

(b) co2

(c) co3

(d) co4

23
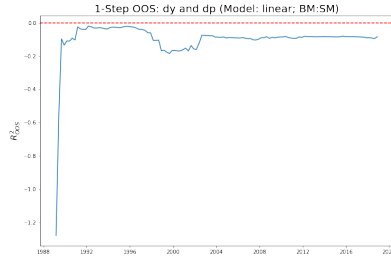
To compare the performances of the 8 functional forms with the 4 different variable combinations, we calculate the percentage of positive R^2_{OOS} in the forecasting period. The results are shown in table (6). It is clear that for combination co3 (dp and ep), none of the functions can perform better than sample mean prediction. However, for combinations co2 and co4, for more than half of out-of-sample period, g_1 and g_6 can outperform sample mean. And other functional forms also can provide positive results.

Table 6: Percentage of Positive R^2_{OOS}

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

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

5 conclusion

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

We apply the model to the Welch and Goyal (2008) dataset and investigate the predictability using co-integrated variable combinations. We find that by including lagged

dependent variable and stationary variable, the partially nonlinear single-index models obtain a better out-of-sample performance than the nonlinear models. When using the partially nonlinear model, some of the variable combinations in previous studies give better out-of-sample forecast than the sample mean prediction over a consecutive period.

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* .