

Parametric Single-Index Models: Simulation and Empirical Results

Ying Zhou

Supervised by:

Professor Jiti Gao

Dr Hsein Kew

A thesis presented for PhD progress review.



Department of Econometrics and Business Statistics

March, 2021

1 Overview

Accurate forecasting of economic and financial variables is a challenging task in time-series analysis because the real world data have different characteristics. Simply using a transformed version of the original data (for example, taking the first order difference) may ignore some key features involved in the original data.

Linear models are the most commonly used model in economics and finance as it is simple to implement and easy to interpret. However, when integrated time-series are included, the findings concluded from linear models may suffer from potential misbalancing or spurious problem ([Phillips \(2015\)](#)). Therefore, econometricians have created various models and estimation methodologies, and a vast number of studies have been published accordingly. The main objective of this thesis is to develop models and methods to deal with these problems caused by predictors with various features.

The spurious problem occurs when variables in the regression are highly persistent. As in the study of [Whittington et al. \(1990\)](#), which adopts a linear model to investigate the effect that tax benefits have on general fertility rate. However, since both tax benefits and general fertility rate are highly persistent, the model actually shows a misleading statistical evidence of linear relationship. [Crump et al. \(2011\)](#) revisits their study and they find no evidence of a linear co-integrating relationship between fertility and personal exemption. Thus, we consider a semi-parametric nonlinear model with a cointegrated system.

This model extends the fully non-parametric time-varying coefficients models developed in [Li et al. \(2016\)](#) by adding non time-varying coefficients. This is a useful extension because it allows for the inclusion of dummy variables and deterministic time trends in the regression. We develop a multi-step estimation procedure that can be employed to estimate both the time-varying coefficients non-parametrically and the non-time-varying coefficients parametrically. We apply the proposed model to study the impact of tax incentives on fertility. Our main findings include:

1. The time-varying coefficients model suggests a nonlinear co-integrating relationship between tax benefits and general fertility rate.
2. The nonlinear co-integrating relationship between tax benefits and fertility rate has weakened considerably over time.

3. Even though new tax incentives have been introduced, they are not effective in increasing fertility rate any more.

By using the semi-parametric time-varying model, we avoid the spurious problem and uncover the nonlinear co-integrating relationship between the dependent variables and independent variables.

Another challenge in using linear predictive model is the misbalancing problem, which happens when some of the predictors have long memory and the dependent variable has short memory. (Phillips (2015)). Stock return prediction is a typical case because stock return is stationary while most of the predictors used by previous studies are non-stationary (Campbell and Shiller (1988), Fama (1990) and Pesaran and Timmermann (1995)). In-sample, a large number of studies find evidence of predictability. Out-of-sample, little consensus exists on the question of whether stock return is predictable and which variables provide a better forecasting performance (Welch and Goyal (2008)).

To achieve a better out-of-sample performance, some researchers consider nonlinear models to capture the nonlinear relationship between stock return and its predictors (Park and Phillips (2001) and Park (2002)). Our study extends the nonlinear model with a single integrated regressor developed in Park and Phillips (2001) to allow for multiple integrated regressors. The nonlinear we propose is given below:

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

where $f(., .)$ is a known univariate function, x_{t-1} is a d -dimensional integrated process of order one, θ_0 and γ_0 are unknown parameters and e_t is a martingale difference process.

To help ease the curse of dimensionality associated with estimating multivariate nonlinear models with integrated time series, we present an econometric model in which the multiple integrated regressors can be reduced to a univariate single-index form via a known univariate nonlinear function. This single-index component allows for either cointegrated predictors or non-cointegrated predictors. We develop a new estimation procedure for the model. We apply the proposed model to study stock return predictability and the main findings in chapter 2 include:

1. We show, via a Monte Carlo experiment, that the new estimator has better finite sample properties than the standard nonlinear least-squares estimator.
2. Exploiting nonlinearities in the data can lead to improved forecast accuracy relative to the historical average when predicting stock returns.

[expand findings; put models in overview](#)

The empirical results of nonlinear models proved that by considering the nonlinear relationship among variables, the out-of-sample performance can be improved significantly. However, the nonlinear model we propose above only include integrated regressors, while in economic and financial studies, the regressors are a mixture of stationary and non-stationary variables. For example, some stationary predictors are also proved to be important, such as the regression residuals in the study of [Lettau and Ludvigson \(2001\)](#).

In addition, the nonlinear model fail to consider the lag variables, $y_{t-1}, y_{t-2}, \dots, y_{t-p}$, which will lead to serially correlated residuals. Therefore, we further extend model [1.1](#) and propose the partially nonlinear model:

$$y_t = \beta_0' z_t + f(x_{t-1}' \theta_0; \gamma_0) + e_t, \quad t = 2, \dots, T, \quad (1.2)$$

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

To estimate model [\(2.4\)](#), we propose a novel 3-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. As we need to use an iterative procedure to apply the nonlinear least square estimation, we choose initial values calculated from Taylor expansion to start the procedure.

The findings of this chapter are summarized as follows:

1. From the Monte Carlo simulation, we find that the constrained nonlinear least square estimators have good finite sample performances
2. The Taylor initial values we calculate from linear regression help further improve the performance.
3. Model [\(2.4\)](#) outperforms pure linear or nonlinear models, including autoregressive models and model [\(1.1\)](#).
4. The partially nonlinear model we proposed also provide a better out-of-sample performance than historical average.

2 Overview of the Thesis

The thesis contains 4 chapters. The contents of each chapter are described below.

Chapter 1: Introduction

The introductory chapter for the thesis reviews the relevant studies about non-stationary time series. The first part of this chapter provides an introduction to the semi-parametric nonlinear cointegration model with time-varying coefficients. The model is used to study the relationship between tax benefit and fertility rate. As the two series are non-stationary and there is a lack of cointegrating relationship ([Crump et al. \(2011\)](#)), we propose the nonlinear cointegration model to examine the instability in the time-series relationship.

The model we propose is an extension from [Phillips et al. \(2017\)](#). It also allows for some coefficients to be constant over time instead of being time varying because the empirical model of fertility involves two dummy variables. We then introduce the 3-step estimation procedure used to estimate the semi-parametric model. It starts by estimating the non-parametric part and then estimates the parametric part by minimizing the sum of least squares.

In the first part of this chapter, we will also present the empirical study results using the time-varying model. First, we uncover the nonlinear cointegrating relationship between tax benefits and fertility rate. In addition, the impact that tax benefits have on fertility rate diminishes over time. And even when new tax benefits are introduced, the diminishing impact remains the same.

The second part of this chapter introduces a parametric nonlinear model with a single-index component. This model follows the studies of [Dong et al. \(2016\)](#) and can deal with multiple non-stationary time-series. It accounts for the nonlinearities in the data and also allows for the presence of cointegrated and non-cointegrated regressors. To estimate the model, we introduce a new estimation procedure and include constraints to achieve a better finite sample performance.

We will also present the empirical results using stock return dataset. It shows that the parametric single-index model outperforms the historical average in out-of-sample prediction. To make sure the results are significant, we apply a hypothesis test following [Clark and West \(2007\)](#) to test the significant of the out-of-sample performances. In addition, we also follow [Neely et al. \(2014\)](#) to produce economic utility gains relative to

the historical average model.

In the third part of the introduction, we will discuss a partially nonlinear single-index model. This model extends the parametric single-index model in chapter 2 to include lagged dependent variables and other stationary variables. We develop a 3-step procedure to estimate the model and apply truncation condition and constraint on parameters. To achieve a better convergence, we design a new method to select the initial values. We will use Monte Carlo simulation to show that the new initial values we select can significantly improve the convergence.

The model will be applied to the stock return dataset. We can see that it has both good in-sample and out-of-sample performances. In addition, we will also discuss the out-of-sample performance of the partially nonlinear model against multiple benchmark models, such as auto-regressive model and pure nonlinear models.

Chapter 2: Literature Review

This chapter provides a review of the relevant literatures on the semi-parametric time-varying model, parametric single-index predictive models and empirical studies about fertility rate and stock return. From the review, we can see the potential challenges when use non-stationary time series. We can then discuss the importance of:

1. considering a time-varying model to describe the instability in time-series relationship;
2. including single-index in multi-variate time-series models;
3. considering nonlinearities in financial and economic predictions;
4. combining linearities and nonlinearities in empirical studies.

Chapter 3: Time-varying Modelling of Fertility and the Tax Benefits

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

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

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

Model and Methodology

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

$$\begin{aligned} y_t &= x_t' \beta_t + \epsilon_t, \quad t = 1, \dots, T, \\ x_t &= x_{t-1} + u_t, \end{aligned}$$

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

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

$$y_t = x_t' \beta(\tau_t) + z_t' \theta + \epsilon_t \quad (2.1)$$

where $\tau_t = t/T$, and $\beta(\cdot)$ is a d -dimensional vector of time-varying coefficients, x_t is a

d -dimensional $I(1)$ processes, z_t is a m -dimensional vector containing time trend and dummy variables, ϵ_t is an error term.

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

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

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

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

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

In step 2, we estimate θ by minimising:

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

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

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

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

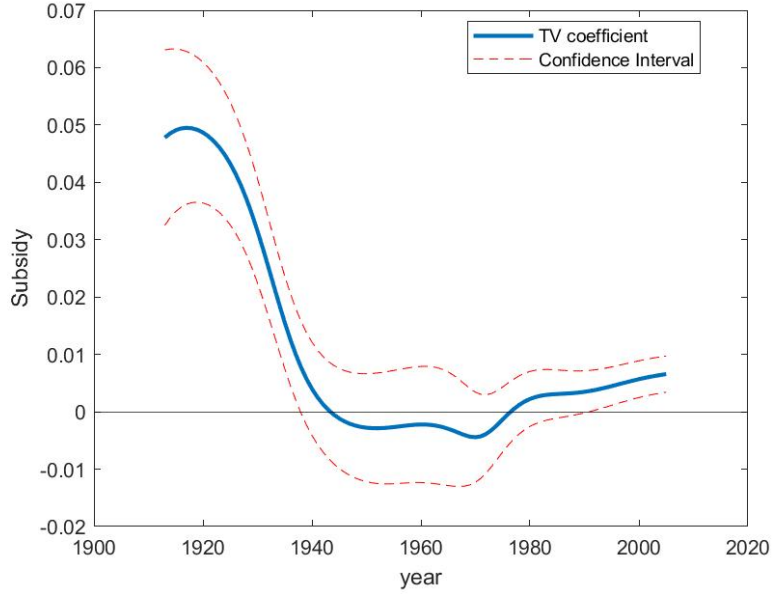
Empirical Study

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

In the early part of the period, the impact was significantly different from zero at the 5% level based on the 95% confidence interval. Beginning in the 1930s, the effect

of Subsidy on the general fertility rate diminishes until the late 1940s after which they stabilize, though for most of the sample period the coefficient is statistically insignificant.

Figure 1: Time-varying Coefficient for Subsidy



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

Table 1: KPSS Test Results

	Subsidy
Test Statistics	0.0971
Conclusion	Stationary

Note: The 10% critical value is 0.347.

Chapter 4: Non-stationary Parametric Single-Index Predictive Models

Since linear predictive models fail to outperform the historical average when predicting stock returns (Welch and Goyal (2008)), previous studies in the empirical literature have taken nonlinearities into account when forecasting financial and macroeconomic series (see for example Wooldridge (1994)). Following their studies, we consider a parametric single-index predictive model to forecast the stock return and investigate whether our model can achieve a better out-of-sample prediction.

Model and Methodology

The nonlinear single-index predictive regression model we consider can be written as:

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

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

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

This model: (a) accounts for the nonlinearities in the data when forecasting next period's macroeconomic and financial variables; (b) helps ease the curse of dimensionality when it comes to parametric nonlinear function estimation involving multivariate integrated regressors; and (c) allows for the presence of cointegrated regressors or non-cointegrated regressors. This model is then used to examine stock return predictability via various combinations of integrated lagged economic and financial variables.

We introduce a new estimation procedure for the model and investigate its finite-sample properties via Monte Carlo simulations. Model (2.2) can be estimated by minimizing sum-of-squared-errors:

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

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

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

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

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

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

Empirical Study

To illustrate the use of the single-index model, we provide an application to predictability of U.S. stock market returns using updated [Welch and Goyal \(2008\)](#) dataset. We find that several combinations of the predictors used in prior studies deliver out-of-sample forecasting gains relative to the standard historical average benchmark when using the single-index predictive model that accounts for the nonlinearities in the time-series data. And the hypothesis suggests that the out-of-sample predictive gain is significant at 10% level.

In addition, we consider the realised utility gain for a mean-variance optimising investor over the out-of-sample period. Our results suggest that a risk-adverse investor who optimally allocates their resources between equities and risk-free bills can receive sizeable utility gains by using the single-index versus the historical-average forecasts.

Chapter 5: Partially Nonlinear Single-Index Predictive Models

In this chapter, we extend the model in chapter 2 to include stationary predictors. The new model is called partially nonlinear model. It combines linear components with a nonlinear single-index component, which can capture both linear and nonlinear relationship between dependent variable and predictors.

Model and Methodology

In this chapter, we propose a new class of nonlinear time series models - partially nonlinear single-index models of the form:

$$y_t = \beta'_0 z_t + f(x'_{t-1} \theta_0; \gamma_0) + e_t, \quad t = 2, \dots, T, \quad (2.4)$$

where $z_t = (y_{t-1}, \dots, y_{t-p}, w'_{t-1})'$, in which w_{t-1} is a vector of stationary predictors, $g(\cdot, \cdot)$ is a known univariate nonlinear function, x_{t-1} is a d -dimensional integrated process of order one, 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 $\theta'_0 \theta_0 = 1$.

To estimate the model, we introduce a novel 3-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 (??) to obtain $\tilde{\beta}$:

$$\tilde{\beta} = \left(\sum_{t=1}^T z_t z'_t \right)^{-1} \sum_{t=1}^T \left(y_t - f(x'_{t-1} \theta; \gamma) \right) z_t \quad (2.5)$$

In other words, $\tilde{\beta}$ is of a linear form by OLS expression. Thus model (2.4) can be approximated by:

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

Substitute $\tilde{\beta}$ in (2.5) and rearrange the equation above, we can get:

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

Let

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

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

Then we have an approximate model of the form:

$$\tilde{y} = \tilde{g}(x_{t-1}'\theta; \gamma) + e_t. \quad (2.6)$$

Step 2: estimate $(\hat{\theta}, \hat{\gamma})$ using nonlinear least square(NLS) method. Define the sum of least squares $Q_T(\theta, \gamma)$ as:

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

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

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

which can be solved using an iterative procedure since there is no closed form solutions.

Step 3: As $\hat{\theta}$ and $\hat{\gamma}$ have been estimated, we can substitute them back to (2.5) and get the estimate of β :

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

In the 3-step procedure above, we have obtained the NLS estimators $(\hat{\theta}, \hat{\gamma}, \hat{\beta})$. In order to improve finite sample properties of the estimators, we impose a truncation condition $I(\|x_{t-1}\| \leq M_T)$ on x_{t-1} and an identification condition on coefficient vector θ . We then define the modified sum-of-squared errors by:

$$Q_{T,M}(\theta, \gamma) = \sum_{t=1}^T \left(y_t - f(x_{t-1}'\theta, \gamma) \right)^2 I(\|x_{t-1}\| \leq M_T) + \lambda (\|\theta\|^2 - 1), \quad (2.7)$$

where $I(\cdot)$ denotes the indicator function, λ is a Lagrange multiplier and $\|\cdot\|$ is the Euclidean norm. In equation (2.7), the truncation condition $I(\|x_{t-1}\| \leq M_T)$ discards

observations whose Euclidean norm is bigger than M_T . Similar condition has been considered by Li et al. (2016) and Ling (2007). M_T is a positive and increasing sequence satisfying $M_T \rightarrow \infty$ as $T \rightarrow \infty$. Following Li et al. (2016), we choose $M_T = C_\alpha n^{1-\beta}$ with $\beta = 1/2$ and $C_\alpha = 1$. Therefore, $M_T = \sqrt{T}$ in our study.

Then we can start with step 1 above to rearrange the equations and the constrained least squares (denoted CLS) estimators $\bar{\theta}$ and $\bar{\gamma}$ can be obtained by minimizing $Q_{T,M}(\theta, \gamma)$ over $\theta \in \Theta$ and $\gamma \in \Gamma$ such that the two restrictions hold; that is:

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

And the CLS estimator of β can be written as:

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

Monte Carlo Simulation

We investigate the finite sample properties of the NLS and the proposed CLS estimators for partially nonlinear model in multivariate co-integrated and non-cointegrated settings. In order to achieve better convergence, we design a unique procedure to find initial values for nonlinear least square estimation.

From the simulation results, we find that:

1. The Taylor initial values can help improve the finite sample performances of the estimators.
2. Both the NLS and CLS estimators converge when sample size increases. But the constrained least square method tend to give better convergence in most cases.
3. 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.
4. Among models with the 7 different functional forms, the one with polynomial functional form tend to provide the best results, but models with the two exponential functions do not perform as well as other ones.

5. Models using cointegrated x_t have a better finite sample performance than using the non-cointegrated x_t .

Empirical Study

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

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

Chapter 6: Conclusion

The final chapter of my thesis will contain all the concluding remarks and possible future extensions.

3 Timetable

3.1 Statement of Progress

Most of the materials for chapter 1 and 2 are already included in the separated chapters and papers. I will need to reorganize the materials I have and chapter 1 and 2 can be completed within a month. Chapter 3 and 4 are practically completed. For chapter 5, the simulations and out-of-sample prediction has been completed, we will need to add significance test results and the realised utility gains. This can also be finished in one month. The concluding chapter will be completed at the end.

3.2 Timetable

March 2022 - June 2022

- Test the significance of R_{OOS}^2 in the empirical study following [Clark and West \(2007\)](#).
- Calculate the realised utility gain following [Campbell and Thompson \(2008\)](#) and [Neely et al. \(2014\)](#).
- Completing Chapter 5 on the partially nonlinear model.

July 2022 - August 2022

- Completing the introduction, literature review, and conclusion chapters and submit the thesis.

3.3 Difficulties

Bibliography

- Campbell, J. Y. and Shiller, R. J. (1988), ‘The dividend-price ratio and expectations of future dividends and discount factors’, *The Review of Financial Studies* **1**(3), 195–228.
- Campbell, J. Y. and Thompson, S. B. (2008), ‘Predicting excess stock returns out of sample: Can anything beat the historical average?’, *The Review of Financial Studies* **21**(4), 1509–1531.
- Clark, T. E. and West, K. D. (2007), ‘Approximately normal tests for equal predictive accuracy in nested models’, *Journal of econometrics* **138**(1), 291–311.
- Crump, R., Goda, G. S. and Mumford, K. J. (2011), ‘Fertility and the personal exemption: comment’, *American Economic Review* **101**(4), 1616–28.
- Dong, C., Gao, J., Tjøstheim, D. et al. (2016), ‘Estimation for single-index and partially linear single-index integrated models’, *The Annals of Statistics* **44**(1), 425–453.
- Fama, E. F. (1990), ‘Stock returns, expected returns, and real activity’, *The journal of finance* **45**(4), 1089–1108.
- Lettau, M. and Ludvigson, S. (2001), ‘Consumption, aggregate wealth, and expected stock returns’, *the Journal of Finance* **56**(3), 815–849.
- Li, D., Tjøstheim, D., Gao, J. et al. (2016), ‘Estimation in nonlinear regression with harris recurrent markov chains’, *The Annals of Statistics* **44**(5), 1957–1987.
- Ling, S. (2007), ‘Self-weighted and local quasi-maximum likelihood estimators for arma-garch/igarch models’, *Journal of Econometrics* **140**(2), 849–873.
- Neely, C. J., Rapach, D. E., Tu, J. and Zhou, G. (2014), ‘Forecasting the equity risk premium: the role of technical indicators’, *Management science* **60**(7), 1772–1791.
- Park, J. Y. (2002), ‘Nonstationary nonlinear heteroskedasticity’, *Journal of econometrics* **110**(2), 383–415.
- Park, J. Y. and Phillips, P. C. (2001), ‘Nonlinear regressions with integrated time series’, *Econometrica* **69**(1), 117–161.

- Pesaran, M. H. and Timmermann, A. (1995), ‘Predictability of stock returns: Robustness and economic significance’, *The Journal of Finance* **50**(4), 1201–1228.
- Phillips, P. C. (2015), ‘Halbert white jr. memorial jfec lecture: Pitfalls and possibilities in predictive regression’, *Journal of Financial Econometrics* **13**(3), 521–555.
- Phillips, P. C., Li, D. and Gao, J. (2017), ‘Estimating smooth structural change in cointegration models’, *Journal of Econometrics* **196**(1), 180–195.
- Welch, I. and Goyal, A. (2008), ‘A comprehensive look at the empirical performance of equity premium prediction’, *The Review of Financial Studies* **21**(4), 1455–1508.
- Whittington, L. A., Alm, J. and Peters, H. E. (1990), ‘Fertility and the personal exemption: implicit pronatalist policy in the united states’, *The American Economic Review* **80**(3), 545–556.
- Wooldridge, J. M. (1994), ‘Estimation and inference for dependent processes’, *Handbook of econometrics* **4**, 2639–2738.