## Title

Essays on Structural Breaks and Forecasting in Econometric Models

## Permalink

https://escholarship.org/uc/item/0gx089z4

## Author

Baek, Yaein
Publication Date
2019
Peer reviewed|Thesis/dissertation

## UNIVERSITY OF CALIFORNIA SAN DIEGO

## Essays on Structural Breaks and Forecasting in Econometric Models

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy
in

Economics
by

Yaein Baek

Committee in charge:

Professor Graham Elliott, Chair
Professor Brendan K. Beare
Professor James D. Hamilton
Professor Dimitris N. Politis
Professor Allan Timmermann

## Copyright

Yaein Baek, 2019
All rights reserved.

The dissertation of Yaein Baek is approved, and it is acceptable in quality and form for publication on microfilm and electronically:
$\qquad$
$\qquad$
$\qquad$
$\qquad$
$\qquad$ Chair

University of California San Diego

To my parents, Seung-Gwan Baek and Kyung-Hee Cho.

## TABLE OF CONTENTS

Signature Page ..... iii
Dedication ..... iv
Table of Contents ..... v
List of Figures ..... vii
List of Tables ..... viii
Acknowledgements ..... X
Vita ..... xi
Abstract of the Dissertation ..... xii
Chapter 1 Estimation of Structural Break Point in Linear Regression Models ..... 1
1.1 Introduction ..... 1
1.2 Structural Break Point Estimator ..... 5
1.3 Partial Break with Multiple Regressors ..... 13
1.4 In-fill Asymptotic Distribution ..... 24
1.4.1 Partial break in a stationary process ..... 25
1.4.2 Break in an autoregressive model ..... 33
1.5 Monte Carlo Simulation ..... 35
1.5.1 Univariate stationary process ..... 35
1.5.2 Autoregressive process ..... 40
1.6 Empirical Application ..... 45
1.6.1 U.S. real GDP growth rate ..... 45
1.6.2 Stock return prediction models ..... 49
1.6.3 Oil price shock and output growth ..... 55
1.7 Conclusion ..... 60
Chapter 2 Tests for Break in Coefficients in Linear Regression when the Direction of the Break is Known ..... 63
2.1 Introduction ..... 63
2.2 Directional Break Tests ..... 65
2.2.1 The model and test statistics ..... 65
2.2.2 Derivation and asymptotic distribution of test statistics ..... 70
2.3 Monte Carlo Simulation ..... 76
2.4 Empirical Application ..... 93
2.4.1 U.S. real GDP growth ..... 94
2.4.2 U.S. labor productivity ..... 97
2.5 Conclusion ..... 99
Chapter 3 Forecasting in Long Horizons using Smoothed Direct Forecast ..... 101
3.1 Introduction ..... 101
3.2 Smoothness Mechanism on Direct Forecasts ..... 104
3.2.1 Direct versus iterated forecasting ..... 104
3.2.2 Smoothed direct forecasts ..... 107
3.3 Empirical Applications ..... 111
3.3.1 Real oil prices ..... 111
3.3.2 Macroeconomic variables ..... 117
3.4 Concluding Remarks ..... 125
Appendix ..... 127
A. 1 Proofs for Chapter 1 ..... 127
A. 2 Appendix for Chapter 2 ..... 142
A.2.1 Comparing methods ..... 142
A.2.2 Proofs for Chapter 2 ..... 144
A. 3 Appendix for Chapter 3 ..... 148
Bibliography ..... 159

## LIST OF FIGURES

Figure 1.1: Finite sample distribution of $\hat{\rho}_{L S}$ ..... 6
Figure 1.2: Mean and one standard deviation of $\left|V_{T}(k)\right|$ and $V_{T}(k)$ ..... 8
Figure 1.3: Finite sample distribution of the break point estimator $\hat{\rho}$ ..... 9
Figure 1.4: Conditional prior distribution $\delta \mid \rho \sim N(0, \rho(1-\rho))$ ..... 13
Figure 1.5: $\quad\left(\rho_{0}=0.15\right)$ Finite sample distribution of $\hat{\rho}_{N E W}$ (left) and $\hat{\rho}_{L S}$ (right) ..... 37
Figure 1.6: $\quad\left(\rho_{0}=0.30\right)$ Finite sample distribution of $\hat{\rho}_{N E W}$ (left) and $\hat{\rho}_{L S}$ (right) ..... 38
Figure 1.7: $\quad\left(\rho_{0}=0.85\right)$ Finite sample distribution of $\hat{\rho}_{N E W}$ (left) and $\hat{\rho}_{L S}$ (right) ..... 39
Figure 1.8: (Stationary to stationary) Finite sample distributions of $\hat{\rho}_{N E W}$ (left) and $\hat{\rho}_{L S}$ (right) ..... 43
Figure 1.9: (Local-to-unity to stationary) Finite sample distributions of $\hat{\rho}_{N E W}$ (left) and $\hat{\rho}_{L S}$ (right) ..... 44
Figure 1.10: U.S. dividend yield (left) and term spread (right), 1952:7-2003:12 ..... 55
Figure 2.1: (Model 1, $\bar{\delta}$ correct) Finite sample power of the $5 \%$ level structural break tests when $\bar{\delta}=(1,-1)^{\prime}$ ..... 79
Figure 2.2: (Model 1) Finite sample rejection probability of the $5 \%$ level structural break tests when $\bar{\delta}=(1,1)^{\prime}$ ..... 80
Figure 2.3: (Model 1) Finite sample rejection probability of the $5 \%$ level structural break tests when $\bar{\delta}=(-1,-1)^{\prime}$ ..... 81
Figure 2.4: (Model 2) Finite sample power of the $5 \%$ level structural break tests when $\bar{\delta}=1$ ..... 83
Figure 2.5: (Model 2) Finite sample rejection probability of the 5\% level structural break tests when $\bar{\delta}=-1$ ..... 84
Figure 2.6: (Model 3) Finite sample power of the $5 \%$ level structural break tests when $\bar{\delta}=(1,1)^{\prime}$ ..... 87
Figure 2.7: (Model 3) Finite sample power of the $5 \%$ level structural break tests when $\bar{\delta}=(-1,1)^{\prime}$ ..... 88
Figure 2.8: (Model 4) Finite sample power of the $5 \%$ level structural break tests when $\bar{\delta}=-1$ ..... 90
Figure 2.9: (Model 5) Finite sample power of the $5 \%$ level structural break tests when $\bar{\delta}=1$ ..... 92
Figure 3.1: Government spending response to a news shock ..... 103

## LIST OF TABLES

Table 1.1: Finite sample RMSE, bias, and standard error of the new estimator and the LS estimator ..... 40
Table 1.2: In-fill asymptotic RMSE, bias, and standard error of the new estimator and the LS estimator ..... 41
Table 1.3: Finite sample RMSE, bias, and standard error of the new estimator and the LS estimator of the break point under the AR(1) model ..... 42
Table 1.4: In-fill asymptotic RMSE, bias, and standard error of the new estimator and the LS estimator of the break point under the AR(1) model ..... 45
Table 1.5: Structural break date estimates of postwar U.S. real GDP growth rate in a AR(1) model ..... 48
Table 1.6: Structural break date estimates of postwar U.S. real GDP growth rate in a AR(1) model using 54 sub-samples ..... 50
Table 1.7: Structural break date estimates of the U.S. stock return (S\&P 500) prediction model ..... 53
Table 1.8: $\quad$ Structural break date estimates of the UK (FTSE) stock return prediction model ..... 54
Table 1.9: Structural break point estimates of the relation between oil price change and the U.S. real GDP growth rates ..... 56
Table 1.10: Structural break date estimates of the relation between oil price change and the U.S. real GDP growth rates from 50 sub-samples ..... 59
Table 1.11: Structural break point estimates of the relation between oil price change and the U.S. real GDP growth rates from 50 sub-samples ..... 60
Table 2.1: Asymptotic critical values for directional test statistics $d b_{T}^{a}$ and $d b_{T}^{b}$ ..... 75
Table 2.2: (Model 1) Pitman efficiency (or ARE) ..... 77
Table 2.3: (Model 1) Rejection rate of structural break tests under no break ..... 78
Table 2.4: (Model 2) Rejection rate of structural break tests under no break ..... 85
Table 2.5: (Model 3) Rejection rate of structural break tests under no break ..... 86
Table 2.6: (Model 4) Rejection rate of structural break tests under no break ..... 89
Table 2.7: (Model 4) Pitman efficiency (or ARE) ..... 91
Table 2.8: (Model 5) Rejection rate of structural break tests under no break ..... 93
Table 2.9: (Model 5) Rejection rate of WLS directional break tests under no break . ..... 93
Table 2.10: Autoregressive Model of U.S. Real GDP Growth 1947Q1-2017Q2: Direc- tional break test ..... 95
Table 2.11: Autoregressive Model of U.S. Real GDP Growth 1947Q1-2017Q2: WLS directional break test ..... 96
Table 2.12: Autoregressive Model of U.S. Real GDP Growth 1947Q1-2017Q2; variance is estimated pre- and post-1984Q1 ..... 96
Table 2.13: Autoregressive Model of U.S. Real GDP Growth 1960Q1-2017Q2: WLS directional break test ..... 96
Table 2.14: Autoregressive Model of U.S. labor productivity Feb 1947-Apr 2001 ..... 98
Table 2.15: Autoregressive Model of U.S. labor productivity Jan 1964-Dec 1994 ..... 99
Table 3.1: Relative MSFEs of univariate direct forecast and iterated forecast (RAC) ..... 113
Table 3.2: Relative MSFEs of univariate direct forecast and iterated forecast (WTI). ..... 113
Table 3.3: Pointwise MSFE of univariate direct forecast and smoothed direct forecast (RAC). ..... 116
Table 3.4: Pointwise MSFE of univariate direct forecast and smoothed direct forecast (WTI). ..... 117
Table 3.5: Category (A) MSFE ratio of univariate smoothed direct forecast and direct forecast. ..... 120
Table 3.6: Category (B) MSFE ratio of univariate smoothed direct forecast and direct forecast. ..... 122
Table 3.7: Category (C) MSFE ratio of univariate smoothed direct forecast and direct forecast. ..... 123
Table 3.8: Category (D) MSFE ratio of univariate smoothed direct forecast and direct forecast. ..... 124
Table A.9: Category (A) Income, Output, Sales, Capacity Utilization ..... 153
Table A.10: Category (B) Employment and Unemployment ..... 155
Table A.11: Category (C) Construction, Inventories and Orders ..... 157
Table A.12: Category (D) Interest Rate and Asset Prices ..... 158

## ACKNOWLEDGEMENTS

I would like to thank my advisor Professor Graham Elliott, for his invaluable guidance and advice throughout my graduate studies. I would also like to express gratitude to other members of my committee, Professors James Hamilton, Brendan Beare, Allan Timmermann and Dimitris Politis for their insightful comments in developing the dissertation.

I am also grateful to my friends in UCSD for their encouragement. Last but not the least, I thank my parents and sister for their endless support along the way.

Chapter 1 , in full, is currently being prepared for submission for publication of the material. Baek, Yaein. The dissertation author was the sole author of this paper.

Chapter 2, in full, is currently being prepared for submission for publication of the material. Baek, Yaein.; Elliott, Graham. The dissertation author was a primary author of this paper.

Chapter 3, in full, is a reprint of the material that has been accepted for publication at Journal of Forecasting. Baek, Yaein. The dissertation author was the sole author of this paper.

## VITA

2011
B. A. in Applied Statistics and Economics, Yonsei University, South Korea M. A. in Economics, Yonsei University, South Korea

Ph. D. in Economics, University of California San Diego

## PUBLICATIONS

"Testing Linearity Using Power Transforms of Regressors" with Jin Seo Cho and Peter C. B. Phillips, Journal of Econometrics, 187(1), 376-384, 2015.
"Forecasting in Long Horizons Using Smoothed Direct Forecast", Journal of Forecasting, Forthcoming.

# ABSTRACT OF THE DISSERTATION 

# Essays on Structural Breaks and Forecasting in Econometric Models 

by<br>Yaein Baek<br>Doctor of Philosophy in Economics<br>University of California San Diego, 2019<br>Professor Graham Elliott, Chair

Instability of parametric models is a common problem in many fields of economics. In econometrics, these changes in the underlying data generating process are referred to as structural breaks. Although there is an extensive literature on estimation and statistical tests of structural breaks, existing methods fail to adequately capture a break. This dissertation consists of three papers on developing econometric methods for structural breaks and forecasting.

The first chapter develops a new method in estimating the location of a structural break in a linear model and provide theoretical results and empirical applications of the estimator. In finite sample the conventional least-squares estimates a break occurred at either ends of the sample with high probability, regardless of the true break point. I suggest an estimator of the break point that
resolves this pile up issue and thus, provide a more accurate estimate of the break. The second chapter constructs a statistical test to test existence of a structural break when the direction of the parameter shift is known. In practice it is likely that a researcher is interested in testing for a structural break in a particular direction because the direction is known, such as policy change or historical data. We incorporate this information in constructing three tests that have higher power when direction is correctly specified. The last chapter proposes a multi-period forecasting method that is robust to model misspecification. When we are interested in obtaining long horizon ahead forecasts, the direct forecast method is more favorable than the iterated forecast because it is more robust to misspecification. However, direct forecast estimates tend to have jagged shapes across horizons. I use a mechanism analogous to ridge regression on the direct forecast model to maintain robustness while smoothing out erratic estimates.

## Chapter 1

## Estimation of Structural Break Point in Linear Regression Models

### 1.1 Introduction

Parameter instability in models is widely addressed in many economic fields. In macroeconomics and finance it is a common empirical problem, such as decrease in output growth volatility in the 1980s known as "the Great Moderation", oil price shocks, labor productivity change, inflation uncertainty and stock return prediction models. It is often reasonable to assume that a change occurs over a long period of time or some historical event affects the dynamics of a structural model. Hence, interpretation of structural model dynamics or prediction models would rely heavily on estimation and testing of parameter instability. In econometrics literature these changes in the underlying data generating process (DGP) of time-series are referred to as structural breaks. The timing of the break as a fraction of the sample size is called the break point.

Estimation methods in the structural break literature have been used to analyze threshold effects and tipping points. Studies of policy change, income inequality dynamics and social interaction models have used estimation methods from the structural break literature. Card, Mas,
and Rothstein (2008) estimate a tipping point of segregation arising in neighborhoods with white preferences. The tipping point indicates the minority racial share of a neighborhood in which all whites in the neighborhood would leave if the share exceeds the tipping point. González-Val and Marcén (2012) explore the effect of child custody law reforms and Child Support Enforcement on U.S. divorce rates using the method of Bai and Perron (1998, 2003). Adapting structural break tests allows testing without imposing any a priori timing which confirms the effect of policy changes and also provides estimated break dates that can be matched with actual reform dates.

There is an extensive literature on structural break estimation methods, starting with maximum likelihood estimators (MLE) on break points. Hinkley (1970), Bhattacharya (1987) and Yao (1987) provide asymptotic theory of the MLE of the break point in a sequence of independent and identically distributed random variables. Asymptotic theory of least-squares (LS) estimation of a one-time break in a linear regression model has been developed by Bai (1994, 1997), with extension to multiple breaks in Bai and Perron (1998) and Bai, Lumsdaine, and Stock (1998). However there are few alternatives in literature to LS estimation of the break point, which is equivalent to MLE in linear regression models. The main issue of LS estimation of the break point is that its finite sample behavior depends on the size of the parameter shift. In many cases, break magnitudes that are empirically relevant are "small" in a statistical sense. For instance, quarterly U.S. real gross domestic product (GDP) growth rate from 1970Q1 to 2018Q2 has mean 0.68 and a standard deviation of 0.8 in percentage points. A break that decreases the quarterly mean growth rate by 0.25 percentage point is less than a half standard deviation change, but it is equivalent to a 1 percentage point decrease in annual growth which is a significant event for the economy.

In asymptotic analysis a small break can be represented by a magnitude $O\left(T^{-1 / 2}\right)$ that shrinks with sample size $T$ so that structural breaks tests have asymptotic power strictly less than one (Elliott and Müller, 2007). In the presence of such small breaks, the LS estimate of the break date is either the start or the end date of the sample period with high probability. In
other words, the LS estimator of the break point has a finite sample distribution that exhibits tri-modality with one mode at the true value and two modes at zero and one. Break points at zero or one do not give us any information about a structural break, nor is it likely to be true in practice. Therefore, inference in practical applications based on LS estimation of structural breaks would seem unreliable.

In this paper I provide a estimator of the structural break point that has a unique mode at the true break and flat tails in finite sample. This is achieved by the modification of the conventional LS objective function using a weight scheme. The LS break point estimates pick zero or one with high probability due to the functional form of the objective, which is the sum of two sub-samples partitioned by each potential break date $k=1, \ldots, T-1$. For $k$ near the 1 or $T-1$, the objective function has large estimation uncertainty due to small sub-sample size. I construct a weight function of the break point $\rho=k / T$ on the unit interval and impose it on the LS objective function to incorporate different estimation uncertainty across potential break dates. Small weights shrink the variance of the objective at ends of the sample toward zero. Construction of the weight function is explained intuitively and motivated by the Fisher information under a Gaussian assumption. The new break point estimator is asymptotically equivalent to the mode of the Bayesian posterior distribution ${ }^{1}$ when the prior depends on the Fisher information.

The new break point estimator is consistent with the same rate of convergence as the LS estimator (Bai, 1997) under regularity conditions on the weight functional form, in a linear regression model with a structural break on a subset (or all) coefficients. I provide a limit distribution of the break point estimator when the break magnitude is small, under a in-fill asymptotic ${ }^{2}$ framework. I follow the approach of Jiang, Wang, and Yu $(2017,2018)$ that shows the in-fill asymptotic distribution captures the asymmetric and tri-modal finite sample properties of the LS estimator in contrast to the conventional long-span asymptotic theory. The in-fill asymptotic

[^0]distribution of the new estimator is also asymmetric due to dependence on the true break point, but has flat tails with a unique mode. Thus, the new estimation method accurately estimates the structural break point compared to LS estimation. Moreover, Monte Carlo simulations show that the break point estimator has smaller mean squared error (MSE) than the LS estimator in finite sample for all break point values considered.

I provide three empirical applications of my method: structural breaks on post-war U.S. real GDP growth rate, the relation between oil price shock and U.S. output growth and one finance application on U.S. and UK stock return prediction models. For the quarterly U.S. real GDP growth rate under different sample periods, the new method estimates a break in early 1970s whereas the LS estimates varies from 1970s to 1952 or 2000, which are near ends of the sample. The break date estimate in early 1970s is matched with the "productivity growth slowdown" suggested in literature such as Perron (1989) and Hansen (2001). Thus, the new method gives reasonable break point estimates compared to the LS estimates, which is sensitive to trimming of the sample. For the estimation of structural break on stock return prediction models and oil price shocks, I follow the approach of Paye and Timmermann (2006) and Hamilton (2003); both use LS estimation on structural break points. Similar to the first application, the new break point estimates are not close to ends of the sample period and are robust to trimming.

The remainder of the paper proceeds as follows. Section 1.2 motivates and constructs the new break point estimator for a mean shift in a linear process. Section 1.3 provides a generalized linear regression model with multiple regressors, and proves consistency of the break point estimator. Section 1.4 contains in-fill asymptotic theory for stationary and local-to-unit root processes. Monte Carlo simulation results are in Section 1.5 and Section 1.6 provides three empirical applications of the new structural break estimation method. Concluding remarks are provided in Section 1.7. Additional theoretical results and proofs are in the Appendix.

### 1.2 Structural Break Point Estimator

We consider a linear regression model with multiple regressors under a one-time structural break at an unknown date, which allows for partial break in parameters. Theoretical results are provided in Section 1.3 under the general linear regression model. In this section we consider the simplest regression model with a constant term to provide intuitive explanation on the construction of the break point estimator. Suppose a single break occurs at time $k_{0}=\left[\rho_{0} T\right]$ where $\rho_{0} \in(0,1)$, $[\cdot]$ is the greatest smaller integer function, and $\mathbf{1}\left\{t>k_{0}\right\}$ is an indicator function that equals one if $t>k_{0}$ and zero otherwise.

$$
\begin{equation*}
y_{t}=\mu+\delta \mathbf{1}\left\{t>k_{0}\right\}+\varepsilon_{t}, \quad t=1, \ldots, T \tag{1.1}
\end{equation*}
$$

The disturbances $\left\{\varepsilon_{t}\right\}$ are independent and identically distributed (i.i.d.) with mean zero with $E \varepsilon_{t}^{2}=\sigma^{2}$. The pre-break mean is $\mu$ and the post-break mean is $\mu+\delta$. Assume we know a one-time break occurs but the break point $\rho_{0}$ and parameters $\left(\mu, \delta, \sigma^{2}\right)$ are unknown.

The LS estimator of the break date is obtained by finding a value $k$ that minimizes the objective function $S_{T}(k)^{2}$, which is the sum of squared residuals (SSR) under the assumption that $k$ is the break date, $S_{T}(k)^{2}=\sum_{t=1}^{k}\left(y_{t}-\bar{y}_{k}\right)^{2}+\sum_{t=k+1}^{T}\left(y_{t}-\bar{y}_{k}^{*}\right)^{2}$, where $\bar{y}_{k}=k^{-1} \sum_{j=1}^{k} y_{j}$ and $\bar{y}_{k}^{*}=(T-k)^{-1} \sum_{j=k+1}^{T} y_{j}$ are pre- and post-break LS estimates under break date $k$, respectively. Following the expression of Bai (1994), I use the identity $\sum_{t=1}^{T}\left(y_{t}-\bar{y}\right)^{2}=S_{T}(k)^{2}+T V_{T}(k)^{2}$ (Amemiya, 1985) where $V_{T}(k)^{2}=k / T(1-k / T)\left(\bar{y}_{k}^{*}-\bar{y}_{k}\right)^{2}$, to substitute the SSR. Then the LS estimator of the break date is equivalent to

$$
\begin{equation*}
\hat{k}_{L S}=\underset{k=1, \ldots, T-1}{\arg \max }\left|V_{T}(k)\right|, \quad \hat{\rho}_{L S}=\hat{k}_{L S} / T . \tag{1.2}
\end{equation*}
$$

Denote $\rho=k / T$ and $\rho_{0}=k_{0} / T$. Under a small break magnitude $|\delta|$, the LS estimator $\hat{\rho}_{L S}$ has a finite distribution that is tri-modal, which has two modes at ends of the unit interval and one
mode at the true point $\rho_{0}$. A break magnitude that is statistically small is not necessarily small in an economic sense. For example, quarterly U.S. real gross domestic product (GDP) growth rate from 1970Q1 to 2018Q2 has mean 0.68 and a standard deviation around 0.8 in percentage points. A break that decreases the mean quarterly growth rate by 0.3 percentage point ( 1.2 percentage point decrease in yearly growth) is a significant event for the economy. Suppose model (1.1) has parameter values similar to the U.S. real GDP growth rate: assume $\rho_{0}=0.3$, the pre-break mean is $\mu=0.88$ percentage points and the shift in the mean of growth rate is $\delta=-0.29$. The expectation of $y_{t}$ is $\mu+\left(1-\rho_{0}\right) \delta=0.68$, which matches the quarterly U.S. real GDP growth rate. The finite sample distribution of the LS estimator of $\rho$ under this model is provided in Figure 1.1 from a Monte Carlo simulation with 2,000 replications, assuming Gaussian disturbances $\varepsilon_{t} \stackrel{\text { i.i.d. }}{\sim} N\left(0,0.8^{2}\right)$ and $T=100$ observations. The finite distribution of $\hat{\rho}_{L S}$ shows tri-modality with modes at $\{0.01,0.30,0.99\}$. The LS estimator fails to accurately detect the break that occurs in the constant term of a univariate linear regression model. Thus we would expect that in practice, structural breaks that are economically important are not large enough for the LS estimator to detect in many cases. We can understand the finite sample property of LS estimator having two


Figure 1.1: Finite sample distribution of $\hat{\rho}_{L S}$ when $\left(\rho_{0}, \delta\right)=(0.3,-0.29), T=100$ and $\varepsilon_{t} \sim$ i.i.d $N\left(0,0.8^{2}\right)$ with 2,000 replications.
modes at ends $\rho \in\{0,1\}$ intuitively by examining the LS objective function in (1.2). For each
potential break date $k=1, \ldots, T-1$, the objective function $\left|V_{T}(k)\right|$ is constructed by partitioning the sample into two sub-samples, before and after $k$. Each sub-sample is used to estimate two different means, $\bar{y}_{k}$ and $\bar{y}_{k}^{*}$. If $k$ is near 1 , the pre-break sub-sample size $k$ is small and likewise, if $k$ is near $T-1$, the post-break sub-sample size $T-k$ is small. Hence, when the potential break date $k$ of $\left|V_{T}(k)\right|$ is near either ends of the sample, estimates of pre- or post-break mean is imprecise due to small sub-sample size. This implies large variance of $\left|V_{T}(k)\right|$ at boundaries so that $\hat{k}_{L S}$ is equal to 1 or $T-1$ with high probability.

The left plot of Figure 1.2 shows the mean and one standard deviation band of the LS objective function $\left|V_{T}(k)\right|$, and the function without the absolute term $V_{T}(k)$ under standard normal disturbances $\varepsilon_{t} \stackrel{i . i . d .}{\sim} N(0,1)$. Due to large estimation error the variance of $\left|V_{T}(k)\right|$ is large when $k$ is near ends of the sample. In contrast, $V_{T}(k)$ has a constant variance $T^{-1}$ across $k$. Although both functions have a unique maximum at true break, the value of $\left|V_{T}(k)\right|$ at boundaries is only slightly less than its value at $k_{0}$. Thus, it is likely that $\left|V_{T}(k)\right| \geq\left|V_{T}\left(k_{0}\right)\right|$ when $k$ is near ends such that the LS estimator is $\hat{k}_{L S}=1$ or $T-1$ with high probability.

Because the issue arises from large variance of the objective function at boundaries, we can think of shrinking the variance accordingly. Suppose we impose non-negative "weights" $\omega_{k}$, for each $k$ on the LS objective function $\left|V_{T}(k)\right|$, so that $k$ with large estimation error has smaller weights than $k$ with small estimation error. For ends of the sample period $k=1$ and $T-1$, weights near zero are imposed, which implies the variance of the weighted objective function $\omega_{k}\left|V_{T}(k)\right|$ would shrink toward zero. If we normalize the sample period into a unit interval so that $\rho=k / T \in\{1 / T, \ldots,(T-1) / T\}$, the weights are represented by a continuous function $\omega(\rho)$ on $\rho \in[0,1]$ that is zero at $\rho \in\{0,1\}$ and has positive values otherwise. Functions with such properties would look like an inverse U-shaped (or concave downward) function on the unit interval. The right plot of Figure 1.2 shows an example of a weight function, $\omega(\rho)=(\rho(1-\rho))^{1 / 2}$. The new break point estimator is the maximizing value of the objective function $\left|Q_{T}(k)\right|$,


Figure 1.2: Mean and one standard deviation of $\left|V_{T}(k)\right|$ and $V_{T}(k)$ as a function of $k,\left(\rho_{0}, \boldsymbol{\delta}\right)=$ $\left(0.3,4 T^{-1 / 2}\right)$ with $T=100$ and $\varepsilon_{t} \sim$ i.i.d $N(0,1)$. The blue (red) solid line is the mean of $\left|V_{T}(k)\right|$ $\left(V_{T}(k)\right)$ and the blue (red) dotted line corresponds to the mean plus one standard deviation. The right plot is weight function $\omega(\rho)=(\rho(1-\rho))^{1 / 2}$.
defined by multiplying weights $\omega_{k}$ to the LS objective $\left|V_{T}(k)\right|$.

$$
\begin{gather*}
\hat{k}=\underset{k=1, \ldots, T-1}{\arg \max }\left|Q_{T}(k)\right|, \quad \hat{\rho}=\hat{k} / T  \tag{1.3}\\
\left|Q_{T}(k)\right|:=\omega_{k}\left|V_{T}(k)\right|=\omega_{k}\left(\frac{k(T-k)}{T^{2}}\right)^{1 / 2}\left|\bar{y}_{k}^{*}-\bar{y}_{k}\right| .
\end{gather*}
$$

The distribution of $\left|Q_{T}(k)\right|$ has smaller variance at ends of the sample. Hence, the weight function eliminates small sample uncertainty of $\left|V_{T}(k)\right|$ when $k$ is near boundaries. Due to smaller variance of the objective function at ends of the sample, the maximizing value $\hat{k}$ is less likely to pick either ends.

Figure 1.3 shows the finite sample distribution of the break point estimator in (1.3), under the same DGP of Figure 1.1. As expected, the break point estimator has flat tails at ends of the unit interval with a mode at true break point $\rho_{0}=0.3$, whereas the LS estimator has modes at zero and one. Additional Monte Carlo simulations are provided in Section 1.5. It shows that the break point estimator has a finite sample distribution with a unique mode at true break and flat tails regardless of the actual break location and magnitude. In contrast, the finite sample
distribution of the LS estimator is tri-modal for any true break point values.


Figure 1.3: Finite sample distribution of the break point estimator $\hat{\rho}$ with weight function $\omega(\rho)=(\rho(1-\rho))^{1 / 2}$ when $\left(\rho_{0}, \delta\right)=(0.3,-0.29), T=100$ and $\varepsilon_{t} \sim$ i.i.d $N\left(0,0.8^{2}\right)$ with 2,000 replications.

It is intuitive to guess the inverse U-shaped functional form of weights, but where does the particular weight function $\omega(\rho)=(\rho(1-\rho))^{1 / 2}$ come from? This is equivalent to the square root of the Fisher information of $\delta$ conditional on $\rho$, assuming Gaussian disturbances. Suppose $\varepsilon_{t} \stackrel{\text { i.i.d. }}{\sim} N\left(0, \sigma^{2}\right)$ in model (1.1). Lets fix the break point $\rho$ and denote the conditional log-likelihood function as $l_{T}(\delta \mid \rho)$, and the information matrix as $I(\delta \mid \rho)$, which is

$$
\begin{equation*}
I(\delta \mid \rho):=E\left[-\frac{\partial^{2} l_{T}(\delta \mid \rho)}{\partial \delta \partial \delta^{\prime}}\right]=\sigma^{-2} T \rho(1-\rho) \tag{1.4}
\end{equation*}
$$

and thus, $\omega(\rho) \propto[I(\delta \mid \rho)]^{1 / 2}$. Note that the Fisher information $I(\delta \mid \rho)$ is interpreted as a way of measuring the amount of information about the unknown parameter $\delta$, given $\rho$. In this case, the Fisher information depends only on $\rho$ (omit $\sigma^{2}$ for simplicity). Given two different values $\rho_{1} \neq \rho_{2}$, the inequality $I\left(\delta \mid \rho_{1}\right)>I\left(\delta \mid \rho_{2}\right)$ reflects the fact that observations carry more information on the break magnitude if a break occurs at $\rho_{1}$, compared to $\rho_{2}$. In other words, we have more information on the structural break at $\rho_{1}$ than when it happens at $\rho_{2}$. If a break occurs with high probability, the magnitude of $\delta$ is far away from zero. If it is less likely, then $\delta$ is close to zero.

Therefore, we can make use of the information on structural breaks by constructing a Bayesian prior distribution of $\delta$ conditional on $\rho$, depend on $I(\delta \mid \rho)$.

$$
\begin{equation*}
\delta \mid \rho \sim N(0, I(\delta \mid \rho)), \quad I(\delta \mid \rho)=\rho(1-\rho) \tag{1.5}
\end{equation*}
$$

When we have more information of a break occurring at some fixed $\rho$, the variance of the prior distribution is large. Then $\delta$ is more spread out from zero and has large magnitude with high probability. If we have less information of a break at $\rho$, then $\delta$ is centered toward mean zero and the break magnitude is likely to be small. Similarly, a prior belief on $\rho$ can be expressed using the Fisher information ${ }^{3}$; a break is less likely to occur near ends of the unit interval.

$$
\begin{equation*}
f(\delta) \propto \operatorname{det}[I(\delta \mid \rho)]^{1 / 2}=(\rho(1-\rho))^{1 / 2} \tag{1.6}
\end{equation*}
$$

The break point estimator in (1.3) with weight function $\omega(\rho)=(\rho(1-\rho))^{1 / 2}$ can be motivated by a Bayesian framework because it is asymptotically equivalent to the mode of the Bayesian posterior distribution of $\rho$ with priors (1.5) and (1.6). Estimation of a structural break model is nonstandard and hence, it is likely that estimators such as the mode of the posterior, would have smaller variance than the maximum likelihood estimator (MLE). Asymptotic efficiency of the MLE of $\rho$ has not been established under an unknown break magnitude ${ }^{4} \delta$. A brief explanation of why the MLE no longer has optimal properties is as follows. Denote the likelihood function of model (1.1) as $f(\mathbf{y} \mid \rho, \delta)$ where $\mathbf{y}=\left(y_{1}, \ldots, y_{T}\right)^{\prime}$. The nuisance parameter $\mu$ is eliminated by using the maximal invariant, $y_{t}-\bar{y}$.

$$
\begin{equation*}
\ln f(\mathbf{y} \mid \rho, \delta)=T^{-1} \sum_{t=1}^{T}\left[\mathbf{1}\{t \leq[\rho T]\} \ln f\left(y_{t}\right)+\mathbf{1}\{t>[\rho T]\} \ln f\left(y_{t} ; \delta\right)\right] \tag{1.7}
\end{equation*}
$$

Because both $\delta$ and $\rho$ are unknown, this is a conditional likelihood function assuming $\rho$ is equal

[^1]to the true break point. That is, the log-likelihood function in (1.7) is equivalent to the average of the conditional density of $y_{t}$ on the regressor, $\ln f\left(y_{t} \mid v_{t} ; \rho, \delta\right)$, where $v_{t}=\mathbf{1}\{t \leq[\rho T]\}$ is a latent variable. The unconditional log-likelihood is the average of the joint density function $f\left(y_{t} \mid v_{t} ; \rho, \delta\right) f_{T}\left(v_{t} ; \rho\right)$, where the density of $v_{t}$ depends on $\rho$ and $T$. The two parameters $\delta$ and $\rho$ are related to each other ( $\rho$ is not identified if $\delta=0$ ), so the conditional ML estimators of the break magnitude $\hat{\delta}(\rho)$ and $\hat{\rho}$ that maximizes the concentrated likelihood $f(\mathbf{y} \mid \rho, \hat{\delta}(\rho))$ is not asymptotically efficient.

Under the prior distributions (1.5) and (1.6), the posterior distribution of $\rho$ is obtained when $\delta$ is integrated out. The LS estimator of $\delta$ under a fixed $\rho$ is $\left(\bar{y}_{k}^{*}-\bar{y}_{k}\right)$, which has the following normal distribution under Gaussian disturbances.

$$
\left(\bar{y}_{k}^{*}-\bar{y}_{k}\right) \mid \rho, \delta \sim N\left(\delta, \frac{\sigma^{2}}{T \rho(1-\rho)}\right) .
$$

From (1.2), let $V_{T}(k)=(\rho(1-\rho))^{1 / 2}\left(\bar{y}_{k}^{*}-\bar{y}_{k}\right)$ with $\rho=k / T$, and the prior of $\delta \mid \rho$ in (1.5) is normal so the joint distribution of $\left(V_{T}(k), \delta\right)$ conditional on $\rho$ is also a normal distribution.

$$
\left[\begin{array}{c}
V_{T}(k) \\
\delta
\end{array}\right] \left\lvert\, \rho \sim N\left(\left[\begin{array}{l}
0 \\
0
\end{array}\right],\left[\begin{array}{cc}
\frac{\sigma^{2}}{T}+\rho^{2}(1-\rho)^{2} & \rho^{3 / 2}(1-\rho)^{3 / 2} \\
\rho^{3 / 2}(1-\rho)^{3 / 2} & \rho(1-\rho)
\end{array}\right]\right)\right.
$$

Then the marginal distribution of data conditional on the break point $V_{T}(k) \mid \rho$ is normal with mean zero and variance $\left(\sigma^{2} / T+\rho^{2}(1-\rho)^{2}\right)$. The posterior distribution of $\rho$ is proportional to the conditional distribution $V_{T}(k) \mid \rho$. Denote the posterior distribution of $\rho$ conditional on data $\mathbf{y}$ as $f(\rho \mid \mathbf{y})$ and assume that $\rho$ is bounded away from $\{0,1\}$ so that $\sigma^{2} / T=O\left(T^{-1}\right)$ is substituted into $o(1)$. Assuming prior of the break point in (1.6), we have

$$
\ln f(\rho \mid \mathbf{y}) \propto \frac{-V_{T}(k)^{2}}{2 \rho^{2}(1-\rho)^{2}}
$$

Given the function of data $V_{T}(k)$, the argmax function of the monotone transformation of the log-likelihood is asymptotically equivalent to the $\operatorname{argmax}$ of $Q_{T}(k)^{2}$ defined in (1.3) with $\omega_{k}=$ $(k / T(1-k / T))^{1 / 2}$.

$$
\underset{\rho}{\arg \max } \ln f(\rho \mid \mathbf{y})=\underset{\rho}{\arg \max } Q_{T}(k)^{2}, \quad Q_{T}(k)^{2} \equiv \rho^{2}(1-\rho)^{2} V_{T}(k)^{2} .
$$

The log-likelihood function is equivalent to the objective function of the break point estimator in (1.3) up to order $O\left(T^{-1}\right)$. Hence, the mode of the $\log$ posterior density $\ln f(\rho \mid \mathbf{y})$ which is denoted as $\hat{\rho}_{\text {map }}$, is asymptotically equivalent to the break point estimator that maximizes $Q_{T}(k)^{2}$.

$$
\begin{gathered}
\hat{\rho}_{\text {map }}=\underset{\rho \in[\alpha, 1-\alpha]}{\arg \max } \ln f(\rho \mid \mathbf{y}) \\
f(\delta, \rho) \propto(\rho(1-\rho))^{1 / 2} \exp \left[-\frac{\delta^{2}}{2 \rho(1-\rho)}\right]=\omega(\rho) \exp \left[-\frac{1}{2}\left(\frac{\delta}{\omega(\rho)}\right)^{2}\right]
\end{gathered}
$$

If the weight is equivalent to the square root of the Fisher information $\omega(\rho)=(\rho(1-\rho))^{1 / 2}$, then the joint prior density $f(\delta, \rho)$ can be expressed as a function of the weight function ${ }^{5}$. Figure 1.4 shows the conditional prior distribution $\delta \mid \rho \sim N(0, I(\delta \mid \rho))$ on $\delta \in[-1.2,1.2]$ for $\rho \in\{0.1,0.3,0.5,0.7,0.9\}$, which visualizes our prior belief on the break based on the Fisher information. A structural break where $\rho$ is close to the median has $|\delta|$ spread out from zero, whereas when $\rho$ is near zero or one $|\delta|$ is close to zero. Therefore, we are uncertain of the presence of a break $(\delta=0)$ and if a break occurs, it does not happen at either end of the sample period.

In the next section I show that the break point estimator is consistent for any true break point $\rho_{0} \in[\alpha, 1-\alpha]$ where $0<\alpha<1 / 2$. Consistency holds under a general functional form of weights under regularity conditions, but if $\alpha$ is arbitrary close to zero it may restrict the functional form. In a univariate model with a break in the mean, a weight function that is sufficient for

[^2]consistency would be a concave downward (inverse U-shaped) function that is differentiable on the unit interval with restrictions on the slope magnitude. Thus, assuming that $\alpha$ is arbitrarily close to zero may restrict the choice of the weight function unless we choose $\omega(\rho)=(\rho(1-\rho))^{1 / 2}$. Details on the restriction of the weight function is provided in Section 1.3.


Figure 1.4: Conditional prior distribution $\delta \mid \rho \sim N(0, \rho(1-\rho))$. Parameter values are $\rho \in$ $\{0.1,0.3,0.5,0.7,0.9\}$ and $\delta \in[-1.2,1.2]$

### 1.3 Partial Break with Multiple Regressors

In this section, the assumptions and proof of consistency of the break point estimator are provided under a general linear regression model with multiple regressors. The model incorporates a partial break in coefficients and assumes a one-time break occurs at an unknown date $k_{0}=\left[\rho_{0} T\right]$ with $\rho_{0} \in(0,1)$. I follow the notations of Bai (1997); denote the vector of variables associated with a stable coefficient as $w_{t}$ and variables associated with coefficients under a break as $z_{t}$. Let $x_{t}=\left(w_{t}^{\prime}, z_{t}^{\prime}\right)^{\prime}$ be a $(p \times 1)$ vector where the variable $z_{t}$ is a $(q \times 1)$ vector and $q \leq p$,

$$
y_{t}= \begin{cases}x_{t}^{\prime} \beta+\varepsilon_{t} & \text { if } t=1, \ldots, k_{0}  \tag{1.8}\\ x_{t}^{\prime} \beta+z_{t}^{\prime} \delta_{T}+\varepsilon_{t} & \text { if } t=k_{0}+1, \ldots, T\end{cases}
$$

where $\varepsilon_{t}$ is a mean zero error term. In general, $z_{t}$ can be expressed as a linear function of $x_{t}$ so that $z_{t}=R^{\prime} x_{t}$ where $R$ is a $(p \times q)$ matrix with full column rank. Let $Y=\left(y_{1}, \ldots, y_{T}\right)^{\prime}$ and define $X_{k}:=\left(0, \ldots, 0, x_{k+1}, \ldots, x_{T}\right)^{\prime}$ and $X_{0}:=\left(0, \ldots, 0, x_{k_{0}+1}, \ldots, x_{T}\right)^{\prime}$. Define $Z_{k}$ and $Z_{0}$ analogously so that $Z_{k}=X_{k} R$ and $Z_{0}=X_{0} R$. Let $M:=I-X\left(X^{\prime} X\right)^{-1} X^{\prime}$ and use the maximal invariant to eliminate the nuisance parameter $\beta$.

The subscript on the break magnitude $\delta_{T}$ shows that it may depend on the sample size. For consistency we assume the break magnitude is outside the local $T^{-1 / 2}$ neighborhood of zero. This is because the break point is not consistently estimable if the break magnitude is in the local $T^{-1 / 2}$ neighborhood of zero such that $\delta_{T}=O\left(T^{-1 / 2}\right)$. In this case structural break tests have asymptotic power that is strictly less than one. Hence, we proceed assuming that $\delta_{T}$ is fixed, or it converges to zero at a rate slower than $T^{-1 / 2}$ so that power of structural break tests converge to one (Assumption 3).

Let $\bar{S}=Y^{\prime} M Y$, and denote $S_{T}(k)^{2}$ as the SSR regressing $M Y$ on $M Z_{k}$. The LS estimator of break date $\hat{k}_{L S}$ is the value that minimizes $S_{T}(k)^{2}$ and thus, maximizes $V_{T}(k)^{2}$ from the identity $\bar{S}=S_{T}(k)^{2}+V_{T}(k)^{2}$ (Amemiya, 1985),

$$
\begin{gathered}
\hat{k}_{L S}=\underset{k=1, \ldots, T-1}{\arg \max } V_{T}(k)^{2}, \quad \hat{\rho}_{L S}=\hat{k}_{L S} / T \\
V_{T}(k)^{2}:=\hat{\delta}_{k}^{\prime}\left(Z_{k}^{\prime} M Z_{k}\right) \hat{\delta}_{k},
\end{gathered}
$$

where $\hat{\delta}_{k}$ is the LS estimate of $\delta_{T}$ by regressing $M Y$ on $M Z_{k}$.
Note that $V_{T}(k)^{2}$ is non-negative from the inner product of the vector $\left(Z_{k}^{\prime} M Z_{k}\right)^{1 / 2} \hat{\delta}_{k}$. The LS objective function can be modified by multiplying a positive definite weight matrix $\Omega_{k}$ to the vector. Decompose the weight matrix so that $\Omega_{k}=\Omega_{k}^{1 / 2 \prime} \Omega_{k}^{1 / 2}$, then $\Omega_{k}^{1 / 2}$ is multiplied to the vector $\left(Z_{k}^{\prime} M Z_{k}\right)^{1 / 2} \hat{\delta}_{k}$. Take the inner product and obtain the objective function $Q_{T}(k)^{2}:=$
$\hat{\delta}_{k}^{\prime}\left(Z_{k}^{\prime} M Z_{k}\right)^{1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{1 / 2} \hat{\delta}_{k}$. Then the estimator of the break point is

$$
\begin{equation*}
\hat{k}=\underset{k=1, \ldots, T-1}{\arg \max } Q_{T}(k)^{2}, \quad \hat{\rho}=\hat{k} / T \tag{1.9}
\end{equation*}
$$

An example of the weight matrix is $\Omega_{k}=T^{-1} Z_{k}^{\prime} M Z_{k}$ which is proportional to the Fisher information matrix under a Gaussian assumption. This is analogous to the squared weight function $\omega(\rho)^{2}=\rho(1-\rho)$ in model (1.1). When $R=I$ and $X$ is a $(T \times 1)$ vector of ones, $\Omega_{k}=T^{-1} Z_{k}^{\prime} M Z_{k}$ is equal to $\omega_{k}^{2}=k / T(1-k / T)$. Thus, the weight matrix $\Omega_{k}$ is a generalization of $\omega_{k}$ for a linear regression model with multiple regressors. Similar to $\omega_{k}$, the matrix $T^{-1} Z_{k}^{\prime} M Z_{k}$ "decreases" as $k$ is approaches the either end of the sample from the following rearrangement of terms.

$$
\begin{align*}
T^{-1} Z_{k}^{\prime} M Z_{k} & =T^{-1}\left[Z_{k}^{\prime} Z_{k}-Z_{k}^{\prime} X\left(X^{\prime} X\right)^{-1} X^{\prime} Z_{k}\right] \\
& =T^{-1} R^{\prime}\left(X_{k}^{\prime} X_{k}\right)\left(X^{\prime} X\right)^{-1}\left(X^{\prime} X-X_{k}^{\prime} X_{k}\right) R \tag{1.10}
\end{align*}
$$

I prove consistency of the break point estimator $\hat{\rho}$ in (1.9) under regularity conditions on model (1.8) and weight matrix $\Omega_{k}$. The notation $\|\cdot\|$ denotes the Euclidean norm $\|x\|=\left(\sum_{i=1}^{p} x_{i}^{2}\right)^{1 / 2}$ for $x \in \mathbb{R}^{p}$. For a matrix $A,\|A\|$ represents the vector induced norm $\|A\|=\sup _{x}\|A x\| /\|x\|$ for $x \in \mathbb{R}^{p}$ and $A \in \mathbb{R}^{p \times p}$.

Assumption 1. (i) $k_{0}=\left[\rho_{0} T\right]$ where $\rho_{0} \in[\alpha, 1-\alpha], 0<\alpha<\frac{1}{2}$;
(ii) The data $\left\{y_{t T}, x_{t T}, z_{t T}: 1 \leq t \leq T, T \geq 1\right\}$ form a triangular array. The subscript $T$ is omitted for simplicity. In addition, $z_{t}=R^{\prime} x_{t}$, where $R$ is $p \times q, \operatorname{rank}(R)=q, z_{t} \in \mathbb{R}^{q}, x_{t} \in \mathbb{R}^{p}$ and $q \leq p$;
(iii) The matrices $\left(j^{-1} \sum_{t=1}^{j} x_{t} x_{t}^{\prime}\right),\left(j^{-1} \sum_{t=T-j+1}^{T} x_{t} x_{t}^{\prime}\right),\left(j^{-1} \sum_{t=k_{0}-j+1}^{k_{0}} x_{t} x_{t}^{\prime}\right)$ and $\left(j^{-1} \sum_{t=k_{0}+1}^{k_{0}+j} x_{t} x_{t}^{\prime}\right)$ have minimum eigenvalues bounded away from zero in probability for all large $j$. For simplicity we assume these matrices are invertible when $j \geq p$. In addition,
these four matrices have stochastically bounded norms uniformly in $j$. That is, for example, $\sup _{j \geq 1}\left\|j^{-1} \sum_{t=1}^{j} x_{t} x_{t}^{\prime}\right\|$ is stochastically bounded;
(iv) $T^{-1} \sum_{t=1}^{[s T]} x_{t} x_{t}^{\prime} \xrightarrow{p} s \Sigma_{x}$ uniformly in $s \in[0,1]$, where $\Sigma_{x}$ is a nonrandom positive definite matrix;
(v) For random regressors, $\sup _{t} E\left\|x_{t}\right\|^{4+\gamma} \leq K$ for some $\gamma>0$ and $K<\infty$;
(vi) The disturbance $\varepsilon_{t}$ is independent of the regressor $x_{s}$ for all $t$ and $s$. For an increasing sequence of $\sigma$-fields $\mathcal{F}_{t},\left\{\varepsilon_{t}, \mathcal{F}_{t}\right\}$ form a sequence of $L^{r}$-mixingale sequence with $r=4+\gamma$ for some $\gamma>0$ (McLeish (1975) and Andrews (1988)). That is, there exists nonnegative constants $\left\{c_{t}: t \geq 1\right\}$ and $\left\{\psi_{j}: j \geq 0\right\}$ such that $\psi_{j} \downarrow 0$ as $j \rightarrow \infty$ and for all $t \geq 1$ and $j \geq 0$, we have: (a) $E\left|E\left(\varepsilon_{t} \mid \mathcal{F}_{t-j}\right)\right|^{r} \leq c_{t}^{r} \Psi_{j}^{r}$, (b) $E\left|\varepsilon_{t}-E\left(\varepsilon_{t} \mid \mathcal{F}_{t+j}\right)\right|^{r} \leq c_{t}^{r} \psi_{j+1}^{r}$, (c) $\max _{j}\left|c_{j}\right|<K<\infty$, (d) $\sum_{j} j^{1+\kappa} \psi_{j}<\infty$ for some $\kappa>0$.

Assumption 2. $\Omega_{k}$ is a positive definite $(q \times q)$ matrix $\left(q=\operatorname{dim}\left(z_{t}\right)\right)$ that is a continuous function of data $\left\{y_{t}, x_{t}, z_{t}, 1 \leq t \leq T\right\}$ and have stochastically bounded norms uniformly in $k=1, \ldots, T-1$. In addition, for any nonzero vector $c \in \mathbb{R}^{q}$,

$$
\left\|\Omega_{k_{0}}^{1 / 2}\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} c\right\|>\left\|\Omega_{k}^{1 / 2}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2}\left(Z_{k}^{\prime} M Z_{0}\right) c\right\|
$$

holds for all $k$ and $k_{0}$, where $M=I-X\left(X^{\prime} X\right)^{-1} X^{\prime}$. When $k / T \rightarrow \rho$ as $T \rightarrow \infty$, then $\Omega_{k} \xrightarrow{p} \bar{\Omega}(\rho)$ where $\bar{\Omega}(\rho)$ is a differentiable function of $\rho$, element-wise.

Assumption 1 conditions are similar to assumptions A1 to A6 in Bai (1997), with additional restrictions on (iv) and (vi). Assumption $1(v i)$ allows for general serial correlation in disturbances and requires $x_{t}$ to be strict exogeneous. This is because $\Omega_{k}$ depends on the moments of regressors and we want to impose zero weights on the ends of the unit interval. For instance, if the second moments of $z_{t}$ changes at the true break point $\rho_{0}$, then $\Omega_{k}$ depends on the ratio of the pre- and post-break second moments and $\rho_{0}$. The ends of the unit interval may have positive
weights that depends on the distribution of $z_{t}$. These cases are avoided under strict exogeneity because $\Omega_{k}$ converges in probability to a nonrandom matrix that varies across $\rho$ only. Note that if $\Omega_{k}$ is a non-stochastic matrix that satisfies the norm inequality in Assumption 2, consistency holds under weakly exogeneous regressors (see Assumption 4).

Assumption 2 guarantees that the matrix

$$
\begin{align*}
A_{T}(k):= & \frac{1}{\left|k_{0}-k\right|}\left[\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2}\right. \\
& \left.-\left(Z_{0}^{\prime} M Z_{k}\right)\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2}\left(Z_{k}^{\prime} M Z_{0}\right)\right] \tag{1.11}
\end{align*}
$$

is positive definite and hence $\left\|A_{T}(k)\right\| \geq \lambda_{\min }\left(A_{T}(k)\right)>0$ where $\lambda_{\text {min }}$ denotes the minimum eigenvalue of $A_{T}(k)$. Under the univariate model (1.1), this condition is equivalent to $\left|\omega^{\prime}(\rho) / \omega(\rho)\right|<$ $(2 \rho(1-\rho))^{-1}$ for all $\rho$, where $\omega^{\prime}(\rho)=\partial \omega(x) /\left.\partial x\right|_{x=\rho}$. Thus, I assume $\alpha$ in $1(i)$ is strictly greater than zero because if it is arbitrarily close to zero, it may restrict the functional of $\omega(\cdot)$. If $\omega(\rho)=\left(\rho(1-\rho)^{1 / 2}\right.$, then Assumption 2 is satisfied regardless of the value of $\alpha$. However, for the model with multiple regressors, $\Omega_{k}$ may be close to a singular matrix in finite sample if $\alpha$ is extremely close to zero. Under Assumption 1(iv), the weight matrix converges in probability to a function of $\rho$ and $\Sigma_{x}$ as $T$ increases. Because $\bar{\Omega}(\rho)$ is a differentiable function of $\rho$ element-wise, we have $\left\|\Omega_{k}-\Omega_{k_{0}}\right\| \leq b\left|k-k_{0}\right| / T$ for some finite $b>0$ and all $k$.

Assumption 3. $\delta_{T} \rightarrow 0$ and $T^{1 / 2-\gamma} \delta_{T} \rightarrow \infty$ for some $\gamma \in\left(0, \frac{1}{2}\right)$.

As mentioned previously, we assume that the break magnitude is outside the local $T^{-1 / 2}$ neighborhood of zero in order to establish consistency of the break point estimator. The consistency of the break point estimator is proved by showing that if $\delta_{T} \neq 0$, then with high probability $Q_{T}(k)^{2}$ can only be maximized near the true break $k_{0}$. The objective function $Q_{T}(k)^{2}$ is defined in (1.9) and $\hat{\delta}_{k}$ is the LS estimator of the break magnitude assuming that $k$ is the break date: $\hat{\delta}_{k}=\left(Z_{k}^{\prime} M Z_{k}\right)^{-1}\left(Z_{k}^{\prime} M Z_{0}\right) \delta_{T}+\left(Z_{k}^{\prime} M Z_{k}\right)^{-1} Z_{k}^{\prime} M \varepsilon$. If $k=k_{0}$ then $\hat{\delta}_{k_{0}}=\delta_{T}+\left(Z_{0}^{\prime} M Z_{0}\right)^{-1} Z_{0}^{\prime} M \varepsilon$.

Theorem 1. Under Assumptions 1 and 2, suppose $\delta_{T}$ is fixed or shrinking toward zero such that Assumption 3 is satisfied. Then $\hat{k}=k_{0}+O_{p}\left(\left\|\delta_{T}\right\|^{-2}\right)$ and the break point estimator $\hat{\rho}$ in (1.9) is consistent.

$$
\left|\hat{\rho}-\rho_{0}\right|=O_{p}\left(T^{-1}\left\|\delta_{T}\right\|^{-2}\right)=o_{p}(1)
$$

Proof. By rearranging terms we have

$$
\begin{equation*}
Q_{T}(k)^{2}-Q_{T}\left(k_{0}\right)^{2}=-\left|k_{0}-k\right| G_{T}(k)+H_{T}(k), \tag{1.12}
\end{equation*}
$$

where $G_{T}(k)$ and $H_{T}(k)$ are defined as follows.

$$
\begin{align*}
G_{T}(k):= & \frac{1}{\left|k_{0}-k\right|} \delta_{T}^{\prime}\left[\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2}\right. \\
& \left.-\left(Z_{0}^{\prime} M Z_{k}\right)\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2}\left(Z_{k}^{\prime} M Z_{0}\right)\right] \delta_{T}  \tag{1.13}\\
H_{T}(k):= & \varepsilon^{\prime} M Z_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} Z_{k}^{\prime} M \varepsilon \\
& -\varepsilon^{\prime} M Z_{0}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} Z_{0}^{\prime} M \varepsilon \\
& +2 \delta_{T}^{\prime}\left(Z_{0}^{\prime} M Z_{k}\right)\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} Z_{k}^{\prime} M \varepsilon  \tag{1.14}\\
& -2 \delta_{T}^{\prime}\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} Z_{0}^{\prime} M \varepsilon
\end{align*}
$$

Proofs of lemma 1 and lemma 3 are in the Appendix. Lemma 2 is equivalent to lemma A. 3 from Bai (1997), which is the generalized Hájek-Rényi inequality for martingale differences to mixingales. For the proof see Bai and Perron (1998).

Lemma 1. Under Assumptions 1 and 2, for every $\varepsilon>0$, there exists $\lambda>0$ and $C<\infty$ such that $\inf _{\left|k-k_{0}\right|>C\left\|\delta_{T}\right\|^{-2}} G_{T}(k) \geq \lambda\left\|\delta_{T}\right\|^{2}$ with probability at least $1-\varepsilon$.

Lemma 2. Under Assumption 1, there exist a $M<\infty$ such that for every $c>0$ and $m>0$,

$$
P\left(\sup _{m \leq k \leq T} \frac{1}{k}\left\|\sum_{t=1}^{k} z_{t} \varepsilon_{t}\right\|>c\right) \leq \frac{M}{c^{2} m}
$$

Lemma 3. Under Assumptions 1 and 2, suppose $\delta_{T}$ is fixed or shrinking toward zero such that Assumption 3 is satisfied. Then the break point estimator $\hat{\rho}$ in (1.9) is consistent. That is, for every $\varepsilon>0$ and $\eta>0$, there exists $T_{0}>0$ such that when $T>T_{0}$

$$
P\left(\left|\hat{\rho}-\rho_{0}\right|>\eta\right)<\varepsilon
$$

Moreover, $\left|\hat{\rho}-\rho_{0}\right|=O_{p}\left(T^{-1 / 2}\left\|\delta_{T}\right\| \sqrt{\ln T}\right)$.
The rate of convergence of the break point estimator $\hat{\rho}$ in (1.9) can be improved from lemma 3. For a fixed $\varepsilon>0$ and $\eta>0$, the following inequality holds for any true break point $\rho_{0} \in[\alpha, 1-\alpha]$ when $T$ is large.

$$
\begin{equation*}
P\left(\sup _{\left|k-k_{0}\right|>T \eta} Q_{T}(k)^{2} \geq Q_{T}\left(k_{0}\right)^{2}\right)<\varepsilon \tag{1.15}
\end{equation*}
$$

This is equivalent to lemma 3 because given the estimator $\hat{k}$, then $Q_{T}(\hat{k})^{2}-Q_{T}\left(k_{0}\right)^{2} \geq 0$ by definition. This implies that to prove the improved rate of convergence $O_{p}\left(T^{-1}\left\|\delta_{T}\right\|^{-2}\right)$, it is sufficient to show that for all $\varepsilon>0$, there exists a finite $C>0$ so that for all $T>T_{\varepsilon}$,

$$
P\left(\sup _{k \in K_{T, \varepsilon}(C)} Q_{T}(k)^{2} \geq Q_{T}\left(k_{0}\right)^{2}\right)<\varepsilon
$$

Here, $K_{T}(C)=\left\{k:\left|k-k_{0}\right|>C\left\|\delta_{T}\right\|^{-2},\left|k-k_{0}\right| \leq T \eta\right\}$ for some small fraction $\eta$. From identity (1.12), $Q_{T}(k)^{2} \geq Q_{T}\left(k_{0}\right)^{2}$ is equivalent to $H_{T}(k) /\left|k-k_{0}\right| \geq G_{T}(k)$. From lemma 1, it is sufficient to prove that

$$
\begin{equation*}
P\left(\sup _{k \in K_{T}(C)}\left|\frac{H_{T}(k)}{k_{0}-k}\right|>\lambda\left\|\delta_{T}\right\|^{2}\right)<\varepsilon \tag{1.16}
\end{equation*}
$$

Use the expression $Z_{0}=Z_{k}-Z_{\Delta} \operatorname{sgn}\left(k_{0}-k\right)$ to rewrite the third and fourth terms of $H_{T}(k)$ given
in (1.14) as

$$
\begin{align*}
2 \delta_{T}^{\prime}[ & \left.\left(Z_{0}^{\prime} M Z_{k}\right)\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} Z_{k}^{\prime} M \varepsilon-\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} Z_{0}^{\prime} M \varepsilon\right] \\
= & 2 \delta_{T}^{\prime}\left[\left(Z_{k}^{\prime} M Z_{k}\right)^{1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} Z_{k}^{\prime} M \varepsilon-\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} Z_{k}^{\prime} M \varepsilon\right] \\
& +2 \delta_{T}^{\prime}\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2}\left(Z_{\Delta}^{\prime} M \varepsilon\right) \operatorname{sgn}\left(k_{0}-k\right)  \tag{1.17}\\
& -2 \delta_{T}^{\prime}\left(Z_{\Delta}^{\prime} M Z_{k}\right)\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2}\left(Z_{k}^{\prime} M \varepsilon\right) \operatorname{sgn}\left(k_{0}-k\right) .
\end{align*}
$$

Note that for nonsingular matrices $S$ and $A$ with bounded norms, $S A S^{-1}=A+o_{p}(1)$. Also, $\left(Z_{k}^{\prime} M Z_{k}\right)^{-1} Z_{k}^{\prime} M \varepsilon=O_{p}\left(T^{-1 / 2}\right)$ and $\left(Z_{0}^{\prime} M Z_{0}\right)^{-1}\left(Z_{k}^{\prime} M Z_{k}\right)=O_{p}(1)$ uniformly on $K_{T}(C)$. We use this to find the order of the first line of the right-side in (1.17).

$$
\begin{aligned}
& \left\|2 \delta_{T}^{\prime}\left\{\left(Z_{k}^{\prime} M Z_{k}\right)^{1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} Z_{k}^{\prime} M \varepsilon-\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} Z_{k}^{\prime} M \varepsilon\right\}\right\| \\
& \leq\left\|2 \delta_{T}^{\prime}\left\{\left(Z_{k}^{\prime} M Z_{k}\right)^{1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{1 / 2}-\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} O_{p}(1)\right\}\right\| \\
& \quad \times\left\|\left(Z_{k}^{\prime} M Z_{k}\right)^{-1} Z_{k}^{\prime} M \varepsilon\right\|+o_{p}(1) \\
& \leq\left\|2 \delta_{T}\right\|\left\|Z_{k}^{\prime} M Z_{k} \Omega_{k}-Z_{0}^{\prime} M Z_{0} \Omega_{k_{0}}\right\| O_{p}\left(T^{-1 / 2}\right)+o_{p}(1) \\
& =\left\|2 \delta_{T}\right\|\left\|\left(Z_{k}^{\prime} M Z_{k}-Z_{0}^{\prime} M Z_{0}\right) \Omega_{k}-Z_{0}^{\prime} M Z_{0}\left(\Omega_{k_{0}}-\Omega_{k}\right)\right\| O_{p}\left(T^{-1 / 2}\right)+o_{p}(1)
\end{aligned}
$$

Then the second norm can be rearranged by subtracting and adding $Z_{0}^{\prime} M Z_{k}$ to the term $\left(Z_{k}^{\prime} M Z_{k}-\right.$ $\left.Z_{0}^{\prime} M Z_{0}\right)$ and Assumption 2.

$$
\begin{align*}
& Z_{k}^{\prime} M Z_{k}-Z_{0}^{\prime} M Z_{0} \\
& \quad= \begin{cases}R^{\prime}\left[X_{\Delta}^{\prime} X_{\Delta}\left(X^{\prime} X\right)^{-1}\left(X^{\prime} X-X_{k}^{\prime} X_{k}\right)-X_{0}^{\prime} X_{0}\left(X^{\prime} X\right)^{-1} X_{\Delta}^{\prime} X_{\Delta}\right] R & \text { if } k \leq k_{0} \\
R^{\prime}\left[X_{\Delta}^{\prime} X_{\Delta}\left(X^{\prime} X\right)^{-1} X_{k}^{\prime} X_{k}-\left(X^{\prime} X-X_{0}^{\prime} X_{0}\right)\left(X^{\prime} X\right)^{-1} X_{\Delta}^{\prime} X_{\Delta}\right] R & \text { if } k>k_{0}\end{cases} \\
& \quad=\left|k_{0}-k\right| O_{p}(1)
\end{aligned} \begin{aligned}
& \Omega_{k_{0}}-\Omega_{k}=\left|k_{0}-k\right| T^{-1} O_{p}(1) \tag{1.18}
\end{align*}
$$

The norm $\left\|\left(k_{0}-k\right)^{-1} X_{\Delta}^{\prime} X_{\Delta}\right\|$ is bounded by assumption, hence the first line of (1.17) has order $\left|k_{0}-k\right|\left\|\delta_{T}\right\| O_{p}\left(T^{-1 / 2}\right)$. The second and third lines of (1.17) are

$$
\begin{aligned}
& 2 \delta_{T}^{\prime}\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2}\left(Z_{\Delta}^{\prime} M \varepsilon\right) \operatorname{sgn}\left(k_{0}-k\right) \\
&=2 \delta_{T}^{\prime} \Omega_{k_{0}}\left(Z_{\Delta}^{\prime} M \varepsilon\right) \operatorname{sgn}\left(k_{0}-k\right)+o_{p}(1) \\
&=2 \delta_{T}^{\prime} \Omega_{k_{0}}\left(Z_{\Delta}^{\prime} \varepsilon-Z_{\Delta}^{\prime} X\left(X^{\prime} X\right)^{-1} X^{\prime} \varepsilon\right) \operatorname{sgn}\left(k_{0}-k\right)+o_{p}(1) \\
&=2 \delta_{T}^{\prime} \Omega_{k_{0}} Z_{\Delta}^{\prime} \varepsilon \operatorname{sgn}\left(k_{0}-k\right)+\left|k_{0}-k\right| T^{-1 / 2}\left\|\delta_{T}\right\| O_{p}(1)+o_{p}(1), \\
&-2 \delta_{T}^{\prime}\left(Z_{\Delta}^{\prime} M Z_{k}\right)\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2}\left(Z_{k}^{\prime} M \varepsilon\right) \operatorname{sgn}\left(k_{0}-k\right) \\
&=-2 \delta_{T}^{\prime}\left(Z_{\Delta}^{\prime} M Z_{k}\right) \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1}\left(Z_{k}^{\prime} M \varepsilon\right) \operatorname{sgn}\left(k_{0}-k\right)+o_{p}(1) \\
&=\left|k_{0}-k\right| T^{-1 / 2}\left\|\delta_{T}\right\| O_{p}(1)
\end{aligned}
$$

The first and second terms of $H_{T}(k)$ in (1.14) are $O_{p}(1)$ uniformly in $K_{T}(C)$ under Assumptions 1 and 2. Therefore $H_{T}(k)$ divided by $\left|k_{0}-k\right|$ is

$$
\begin{equation*}
\frac{H_{T}(k)}{\left|k_{0}-k\right|}=2 \delta_{T}^{\prime} \Omega_{k_{0}} \frac{1}{\left|k_{0}-k\right|} Z_{\Delta}^{\prime} \varepsilon \operatorname{sgn}\left(k_{0}-k\right)+T^{-1 / 2}\left\|\delta_{T}\right\| O_{p}(1)+\frac{O_{p}(1)}{\left|k_{0}-k\right|} \tag{1.19}
\end{equation*}
$$

Now we can prove (1.16) using the above expression. Let $1 /\left\|\Omega_{k_{0}}\right\|=A$ where $A<\infty$ by

Assumption 2. Without loss of generality, consider the case $k<k_{0}$. The first term of (1.19) is bounded by lemma 2 .

$$
\begin{aligned}
P\left(\sup _{k \in K(C)}\right. & \| \\
& \left.2 \delta_{T}^{\prime} \Omega_{k_{0}} \frac{1}{k_{0}-k} \sum_{t=k+1}^{k_{0}} z_{t} \varepsilon_{t} \|>\frac{\lambda\left\|\delta_{T}\right\|^{2}}{3}\right) \\
& \leq P\left(\sup _{k_{0}-k \geq C\left\|\delta_{T}\right\|^{-2}}\left\|\frac{1}{k_{0}-k} \sum_{t=k+1}^{k_{0}} z_{t} \varepsilon_{t}\right\|>\frac{\lambda A\left\|\delta_{T}\right\|}{6}\right) \\
& \leq M\left(\frac{\lambda A\left\|\delta_{T}\right\|}{6}\right)^{-2} \frac{1}{C\left\|\delta_{T}\right\|^{-2}} \\
& =\frac{36 M}{\lambda^{2} A^{2} C}<\frac{\varepsilon}{3}
\end{aligned}
$$

The probability is negligible for large $T$ because we can choose a large $C$ value accordingly. For any $\varepsilon>0$ and $\eta>0$, we proved that the probability in (1.15) is negligible for large $T$. Thus we can choose $C$ such that $K_{T}(C)$ is non-empty and the inequality above is satisfied for all $\varepsilon>0$ and $T>T_{\varepsilon}$. The second term of (1.19) is bounded due to the assumption $\left(T^{1 / 2}\left\|\delta_{T}\right\|\right)^{-1} \rightarrow 0$.

$$
P\left(T^{-1 / 2}\left\|\delta_{T}\right\| O_{p}(1)>\frac{\lambda\left\|\delta_{T}\right\|^{2}}{3}\right)=P\left(\frac{O_{p}(1)}{T^{1 / 2}\left\|\delta_{T}\right\|}>\frac{\lambda}{3}\right)<\frac{\varepsilon}{3}
$$

The third term of (1.19) is bounded for $k_{0}-k \geq C\left\|\delta_{T}\right\|^{-2}$ since $O_{p}(1) /\left|k_{0}-k\right| \leq O_{p}(1)\left\|\delta_{T}\right\|^{2} / C$,

$$
P\left(\sup _{k_{0}-k \geq C\|\delta\|^{-2}} \frac{O_{p}(1)}{\left|k_{0}-k\right|}>\frac{\lambda\left\|\delta_{T}\right\|^{2}}{3}\right) \leq P\left(\frac{O_{p}(1)}{C}>\frac{\lambda}{3}\right)<\frac{\varepsilon}{3}
$$

where $O_{p}(1) / C$ is small for large $T$, by choosing a large constant $C$. Hence the bound (1.16) holds and the rate of convergence of the break point estimator $\hat{\rho}=\hat{k} / T$ in Theorem 1 is proved: $\left|\hat{\rho}-\rho_{0}\right|=O_{p}\left(T^{-1}\left\|\delta_{T}\right\|^{2}\right)$.

For weakly exogenous regressors $x_{t}$ in model (1.8), the break point estimator is consistent with the same rate of convergence in Theorem 1, under the following conditions that substitutes Assumption 1 and Assumption 2.

Assumption 4. Assume the following conditions in model (1.8) with Assumption 1(i)-(iii) and (v).
(i) $\left(X^{\prime} X\right) / T$ converges in probability to a nonrandom positive definite matrix, as $T \rightarrow \infty$;
(ii) $\left\{\varepsilon_{t}, \mathcal{F}_{t}\right\}$ form a sequence of martingale differences for $\mathcal{F}_{t}=\sigma$-field $\left\{\varepsilon_{s}, x_{s+1}: s \leq t\right\}$. Moreover, for all $t, E\left|\varepsilon_{t}\right|^{4+\gamma}<K$ for some $K<\infty$ and $\gamma>0$;
(iii) The weight matrix $\Omega_{k}$ is a nonrandom $(q \times q)$ positive definite matrix, and for any nonzero vector $c \in \mathbb{R}^{q}$,

$$
\left\|\Omega_{k_{0}}^{1 / 2}\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} c\right\|>\left\|\Omega_{k}^{1 / 2}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2}\left(Z_{k}^{\prime} M Z_{0}\right) c\right\|
$$

holds for all $k$ and $k_{0}$, where $M=I-X\left(X^{\prime} X\right)^{-1} X^{\prime} . \Omega_{k}$ converges to $\bar{\Omega}(\rho)$ as $k / T \rightarrow \infty$, which is a differentiable function of $\rho$ on the unit interval.

Theorem 2. Under Assumption 4, suppose $\delta_{T}$ is fixed or shrinking toward zero that satisfies $\delta_{T} \rightarrow 0$ and $T^{1 / 2-\gamma} \delta_{T} \rightarrow \infty$ for some $\gamma \in\left(0, \frac{1}{2}\right)$. Then $\hat{k}=k_{0}+O_{p}\left(\left\|\delta_{T}\right\|^{-2}\right)$ and the break point estimator $\hat{\rho}$ in (1.9) is consistent.

$$
\left|\hat{\rho}-\rho_{0}\right|=O_{p}\left(T^{-1}\left\|\delta_{T}\right\|^{-2}\right)=o_{p}(1) .
$$

The proof of Theorem 2 is similar to the proof of Theorem 1, hence omitted. Under Assumptions $1(v), 2(i)$ and $2(i i)$, the strong law of large numbers holds for $x_{t} \varepsilon_{t}$ because the conditions in Hansen (1991) are satisfied. The weight matrix $\Omega_{k}$ in Assumption 4(iii) depends on $k / T$ but not on the data $\left\{x_{t}, \varepsilon_{t}\right\}$. Thus, by setting $\rho=k / T$, it is a function of $\rho$ which is assumed to be differentiable with respect to $\rho$. Then the bound $\left\|\Omega_{k_{1}}-\Omega_{k_{2}}\right\| \leq c\left|k_{1}-k_{2}\right| / T$ holds for any $k_{1}$ and $k_{2}$, for some finite $c>0$. Using these properties, proving consistency of the estimator under Assumption 4 follows the same process as in the proof under Assumptions 1 and 2.

Given the consistency of the break point estimator from Theorem 1 or 2 , the estimator of the break magnitude corresponding to $\hat{k}$ is consistent and asymptotically normally distributed. Let $\hat{\delta}(\hat{\rho})=\hat{\delta}_{\hat{k}}$, then the following results hold. The proof is provided in Appendix A.1.

Corollary 1. Under Assumptions 1 and 2 , suppose $\delta_{T}$ is fixed or shrinking toward zero such that Assumption 3 is satisfied. Let $\hat{\delta}(\hat{\rho})$ be a consistent estimator of $\delta_{T}$ corresponding to $\hat{k}$, which is defined in (1.9). Then,

$$
\sqrt{T}\left(\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\rho}})-\boldsymbol{\delta}_{T}\right) \xrightarrow{d} N\left(0, \boldsymbol{V}^{-1} \boldsymbol{U} \boldsymbol{V}^{-1}\right)
$$

where

$$
\boldsymbol{V}:=\operatorname{plim}_{T \rightarrow \infty} T^{-1} Z_{0}^{\prime} M Z_{0}, \quad \boldsymbol{U}:=\lim _{T \rightarrow \infty} E\left[\left(T^{-1 / 2} Z_{0}^{\prime} M \varepsilon\right)^{2}\right]
$$

### 1.4 In-fill Asymptotic Distribution

Under the conventional long-span asymptotic framework, Bai (1997) shows that the limit distribution of the break point estimator is symmetric if the second moment of variables associated with coefficients under break ( $z_{t}$ in Section 1.3) do not change before and after break. However, the finite distribution of the break point estimator is asymmetric even though second moment values do not change across regimes.

In order to provide a better approximation of the finite distribution of the break point estimator, a continuous record asymptotic framework have been employed by Jiang, Wang, and Yu $(2017,2018)$ and Casini and Perron (2017). By assuming that a continuous record is available, a continuous time approximation to the discrete time model is constructed and a in-fill asymptotic distribution is developed. In contrast to the long-span asymptotic where the time span of the data increases, the in-fill asymptotic assumes a fixed time span with shrinking sampling intervals. For instance, if there are $T$ equally spaced observations of data available over a fixed time horizon $[0, N]$, then $N=T h$ denotes the time span of the data where $h$ is the sampling interval. Asymptotic inference is conducted by shrinking the time interval $h$ to zero while keeping $N$ fixed, which is
equivalent to $T$ increasing. Jiang et al. (2017) obtains a in-fill asymptotic distribution of the MLE of structural break in the drift function of a continuous time model that is analogous to discrete time models where a break occurs in the coefficient of an autoregressive (AR) model ${ }^{6}$, and Jiang et al. (2018) derives it for a break in the mean. The asymptotic distribution is asymmetric, tri-modal and dependent on the initial condition, which are also the properties of the finite sample distribution of the LS break point estimator in discrete time models.

The structural break point estimator in this paper is no longer tri-modal but is asymmetric, depending on the true break point. The long-span asymptotic theory does not capture this because the sample size before and after break increases proportionally as $T$ increases, eliminating the asymmetry of information. Therefore, I use the in-fill asymptotic framework, following Jiang et al. (2018) in deriving the limit distribution of the break point estimator.

### 1.4.1 Partial break in a stationary process

Consider the linear regression model (1.8) with continuous time process $\left\{W_{s}, Z_{s}, \mathcal{E}_{s}\right\}_{s \geq 0}$ defined on a filtered probability space $\left(\Omega, \mathcal{F},\left(\mathcal{F}_{s}\right)_{s \geq 0}, P\right)$, where $s$ can be interpreted as a continuous time index. Assume that we observe at discrete points of time so that $\left\{Y_{t h}, W_{t h}, Z_{t h}: t=\right.$ $0,1, \ldots, T=N / h\}$ where $N$ is the time span. For simplicity normalize the time span $N=1$. Denote the increment of processes as $\Delta_{h} Y_{t}:=Y_{t h}-Y_{(t-1) h}$. Let $X_{t h}=\left(W_{t h}^{\prime}, Z_{t h}^{\prime}\right)^{\prime}$ so that $Z_{t h}=R^{\prime} X_{t h}$. The model (1.8) can be expressed as

$$
\Delta_{h} Y_{t}= \begin{cases}\left(\Delta_{h} X_{t}\right)^{\prime} \beta_{h}+\Delta_{h} \mathcal{E}_{t} & \text { if } t=1, \ldots,\left[\rho_{0} T\right] \\ \left(\Delta_{h} X_{t}\right)^{\prime} \beta_{h}+\left(\Delta_{h} Z_{t}\right)^{\prime} \delta_{h}+\Delta_{h} \mathcal{E}_{t} & \text { if } t=\left[\rho_{0} T\right]+1, \ldots, T\end{cases}
$$

Divide both sides by $\sqrt{h}$ so that the error term variance is $O(1)$. Let $\varepsilon_{t}:=\Delta_{h} \mathcal{E}_{t} / \sqrt{h}, y_{t}:=$ $\Delta_{h} Y_{t} / \sqrt{h}, x_{t}:=\Delta_{h} X_{t} / \sqrt{h}, z_{t}:=\Delta_{h} Z_{t} / \sqrt{h}=R^{\prime} x_{t}$,

[^3]\[

y_{t}= $$
\begin{cases}x_{t}^{\prime} \beta_{h}+\varepsilon_{t} & \text { if } t=1, \ldots,\left[\rho_{0} T\right]  \tag{1.20}\\ x_{t}^{\prime} \beta_{h}+z_{t}^{\prime} \delta_{h}+\varepsilon_{t} & \text { if } t=\left[\rho_{0} T\right]+1, \ldots, T\end{cases}
$$
\]

Assumption 5. $\left\{z_{t}, \varepsilon_{t}\right\}$ is a covariance stationary process that satisfies the functional central limit theorem as $T=1 / h \rightarrow \infty$,

$$
T^{-1 / 2} \sum_{t=1}^{[s T]} z_{t} \varepsilon_{t} \Rightarrow B_{1}(s)
$$

where $B_{1}(s)$ is a multivariate Gaussian process on $[0,1]$ with mean zero and covariance $E\left[B_{1}(u) B_{1}(v)^{\prime}\right]=\min \{u, v\} \Xi$, and $\Xi:=\lim _{T \rightarrow \infty} E\left[\left(T^{-1 / 2} \sum_{t=1}^{T} z_{t} \varepsilon_{t}\right)^{2}\right]$.

Assumption 6. The break magnitude is $\delta_{h}=d_{0} \lambda_{h}$ where $d_{0} \in \mathbb{R}^{q}$ is a fixed vector and $\lambda_{h}$ is a scalar that depends on the sampling interval $h$. Assume one of the following cases on $\lambda_{h}$ as $h \rightarrow 0$,
(i) $\lambda_{h}=O\left(h^{1 / 2}\right)$ so that $\delta_{h}=d_{0} \sqrt{h}$;
(ii) $\lambda_{h}=O\left(h^{1 / 2-\gamma}\right)=O\left(T^{-1 / 2+\gamma}\right)$ where $0<\gamma<1 / 2$ so that $\delta_{h} / \sqrt{h} \rightarrow \infty$ simultaneously with $\delta_{h} \rightarrow 0$.

The same notations from Section 1.3 are used for model (1.20): $M Y=M Z_{0} \delta_{h}+M \varepsilon$, where $\varepsilon=\left(\varepsilon_{1}, \ldots, \varepsilon_{T}\right)^{\prime}$ and $M=I-X\left(X^{\prime} X\right)^{-1} X^{\prime}$. The estimator of break date $k_{0}=\left[\rho_{0} T\right]$ is written in (1.9). The objective functions of the estimator in (1.9) is restated below.

$$
\begin{equation*}
Q_{T}(k)^{2}=\sqrt{T} \hat{\delta}_{k}^{\prime}\left(T^{-1} Z_{k}^{\prime} M Z_{k}\right)^{1 / 2} \Omega_{k}\left(T^{-1} Z_{k}^{\prime} M Z_{k}\right)^{1 / 2} \sqrt{T} \hat{\delta}_{k} \tag{1.21}
\end{equation*}
$$

The in-fill asymptotic distribution is derived for the two different magnitudes of $\delta_{h}$ in Assumption 6 . Theorem 3 provides the limit distribution under $6(i)$, which represents small breaks.

Theorem 3. Consider the model (1.20) with unknown parameters $\left(\beta_{h}, \delta_{h}\right)$. Assumption 1, 2, 5, and $\sigma(i)$ holds. Then the break point estimator $\hat{\rho}=\hat{k} / T$ defined in (1.9) has the following in-fill
asymptotic distribution as $h \rightarrow 0$,

$$
T\left\|\delta_{h}\right\|^{2} \hat{\rho} \xrightarrow{d}\left\|d_{0}\right\|^{2} \underset{\rho \in(0,1)}{\arg \max } \widetilde{W}(\rho)^{\prime} \bar{\Omega}(\rho) \widetilde{W}(\rho)
$$

with

$$
\begin{array}{rlrl}
\widetilde{W}(\rho) & :=\Sigma_{z}^{-1 / 2} \frac{B_{1}(\rho)-\rho B_{1}(1)}{\sqrt{\rho(1-\rho)}}-\left(1-\rho_{0}\right)\left(\frac{\rho}{1-\rho}\right)^{1 / 2} \Sigma_{z}^{1 / 2} d_{0} & & \text { if } \rho \leq \rho_{0} \\
& :=\Sigma_{z}^{-1 / 2} \frac{B_{1}(\rho)-\rho B_{1}(1)}{\sqrt{\rho(1-\rho)}}-\rho_{0}\left(\frac{1-\rho}{\rho}\right)^{1 / 2} \Sigma_{z}^{1 / 2} d_{0} & \text { if } \rho>\rho_{0}
\end{array}
$$

where $B_{1}(\cdot)$ is a Brownian motion defined in Assumption 5.

Proof. When $\delta_{h}=d_{0} \sqrt{h}$, the break point estimator $\left\|\delta_{h}\right\|^{2}\left(\hat{k}-k_{0}\right)=\left\|d_{0}\right\|^{2}\left(\hat{\rho}-\rho_{0}\right)=O_{p}(1)$ has values in the interval $\left(-\rho_{0}\left\|d_{0}\right\|^{2},\left(1-\rho_{0}\right)\left\|d_{0}\right\|^{2}\right)$. Therefore, we only need to examine the behavior of the objective function $Q_{T}(k)^{2}$ for those $k$ in the neighborhood of $k_{0}$ such that $k=\left[k_{0}+s\left\|d_{0} \sqrt{h}\right\|^{-2}\right]$ with $s \in\left(-\rho_{0}\left\|d_{0}\right\|^{2},\left(1-\rho_{0}\right)\left\|d_{0}\right\|^{2}\right)$. Then for any fixed $s$, when $h \rightarrow 0$ it has $k \rightarrow \infty$ with $k / T \rightarrow \rho=\rho_{0}+u$, and $T-k \rightarrow \infty$ with $(T-k) / T \rightarrow 1-\rho=1-\rho_{0}-u$, where $u=s\left\|d_{0}\right\|^{-2} \in\left(-\rho_{0}, 1-\rho_{0}\right)$. From the objective function (1.21), we have

$$
\left(T^{-1} Z_{k}^{\prime} M Z_{k}\right)^{1 / 2} \sqrt{T} \hat{\delta}_{k}=\left(T^{-1} Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2}\left(T^{-1} Z_{k}^{\prime} M Z_{0}\right) d_{0}+\left(T^{-1} Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} T^{-1 / 2} Z_{k}^{\prime} M \varepsilon
$$

Consider each of the terms as $h \rightarrow 0$, which is equivalent to $T \rightarrow \infty$.

$$
\begin{aligned}
T^{-1} Z_{k}^{\prime} M Z_{0} & =T^{-1} \sum_{t=\max \left\{k, k_{0}\right\}+1}^{T} z_{t} z_{t}^{\prime}-\left(T^{-1} \sum_{t=k+1}^{T} R^{\prime} x_{t} x_{t}^{\prime}\right)\left(T^{-1} X^{\prime} X\right)^{-1}\left(T^{-1} \sum_{t=k_{0}+1}^{T} x_{t} x_{t}^{\prime} R\right) \\
& \longrightarrow\left(1-\max \left\{\rho, \rho_{0}\right\}\right) \Sigma_{z}-(1-\rho)\left(1-\rho_{0}\right) R^{\prime} \Sigma_{x} \Sigma_{x}^{-1} \Sigma_{x} R \\
& =\left(\min \left\{\rho, \rho_{0}\right\}-\rho \cdot \rho_{0}\right) \Sigma_{z}
\end{aligned}
$$

$$
\begin{aligned}
T^{-1} Z_{k}^{\prime} M Z_{k} & =T^{-1} \sum_{t=k+1}^{T} z_{t} z_{t}^{\prime}-\left(T^{-1} \sum_{t=k+1}^{T} R^{\prime} x_{t} x_{t}^{\prime}\right)\left(T^{-1} X^{\prime} X\right)^{-1}\left(T^{-1} \sum_{t=k+1}^{T} x_{t} x_{t}^{\prime} R\right) \\
& \longrightarrow \rho(1-\rho) \Sigma_{z} \\
T^{-1 / 2} Z_{k}^{\prime} M \varepsilon & =T^{-1 / 2} \sum_{t=k+1}^{T} z_{t} \varepsilon_{t}-\left(T^{-1} Z_{k}^{\prime} X\right)\left(T^{-1} X^{\prime} X\right)^{-1}\left(T^{-1 / 2} \sum_{t=1}^{T} x_{t} \varepsilon_{t}\right) \\
& \Rightarrow B_{1}(\rho)-\rho B_{1}(1) .
\end{aligned}
$$

By assumption, $\Omega_{k} \xrightarrow{p} \bar{\Omega}(\rho)$ as $k / T \rightarrow \rho$. This implies that for a fixed $d_{0}$, the objective function $Q_{T}(k)^{2}$ weakly converges as follows. For $\rho \leq \rho_{0}$,

$$
\begin{aligned}
Q_{T}(k)^{2} \Rightarrow \frac{1}{\rho(1-\rho)}\left[B_{1}(\rho)-\rho B_{1}(1)\right. & \left.-\rho\left(1-\rho_{0}\right) \Sigma_{z} d_{0}\right]^{\prime} \\
& \times \Sigma_{z}^{-1 / 2} \bar{\Omega}(\rho) \Sigma_{z}^{-1 / 2}\left[B_{1}(\rho)-\rho B_{1}(1)-\rho\left(1-\rho_{0}\right) \Sigma_{z} d_{0}\right]
\end{aligned}
$$

and for $\rho>\rho_{0}$,

$$
\begin{aligned}
Q_{T}(k)^{2} \Rightarrow \frac{1}{\rho(1-\rho)}\left[B_{1}(\rho)-\rho B_{1}(1)\right. & \left.-\rho_{0}(1-\rho) \Sigma_{z} d_{0}\right]^{\prime} \\
& \times \Sigma_{z}^{-1 / 2} \bar{\Omega}(\rho) \Sigma_{z}^{-1 / 2}\left[B_{1}(\rho)-\rho B_{1}(1)-\rho_{0}(1-\rho) \Sigma_{z} d_{0}\right]
\end{aligned}
$$

By continuous mapping theorem, the the in-fill asymptotic distribution of $T\left\|\delta_{h}\right\|^{2} \hat{\rho}$ is the $\operatorname{argmax}$ functional of the limit of $Q_{T}(k)^{2}$, stated in Theorem 3.

An equivalent representation of the in-fill asymptotic distribution is (let $\rho=\rho_{0}+u$ )

$$
T\left\|\delta_{h}\right\|^{2}\left(\hat{\rho}-\rho_{0}\right) \xrightarrow{d}\left\|d_{0}\right\|^{2} \underset{u \in\left(\rho_{0}, 1-\rho_{0}\right)}{\arg \max } \widetilde{W}\left(\rho_{0}+u\right)^{\prime} \bar{\Omega}\left(\rho_{0}+u\right) \widetilde{W}\left(\rho_{0}+u\right)
$$

where $\widetilde{W}(\cdot)$ is defined in Theorem 3.
In the special case where the weight matrix is equivalent to the sample information matrix
under Gaussian disturbances $\Omega_{k}=T^{-1} Z_{k}^{\prime} M Z_{k} \xrightarrow{p} \bar{\Omega}(\rho)=\rho(1-\rho) \Sigma_{Z}$ uniformly, as $k / T \rightarrow \rho$. Then the limit of $Q_{T}(k)^{2}$ is simplified as follows.

$$
\begin{aligned}
& Q_{T}(k)^{2} \Rightarrow\left[B_{1}(\rho)-\rho B_{1}(1)-\left(\min \left\{\rho, \rho_{0}\right\}-\rho \cdot \rho_{0}\right) \Sigma_{z} d_{0}\right]^{\prime} \\
& \times\left[B_{1}(\rho)-\rho B_{1}(1)-\left(\min \left\{\rho, \rho_{0}\right\}-\rho \cdot \rho_{0}\right) \Sigma_{z} d_{0}\right]
\end{aligned}
$$

In this case, the difference between the in-fill asymptotic distribution of our estimator and the LS estimator is evident because

$$
\begin{aligned}
V_{T}(k)^{2} \Rightarrow & {\left[\frac{B_{1}(\rho)-\rho B_{1}(1)}{\sqrt{\rho(1-\rho)}}-\frac{\left(\min \left\{\rho, \rho_{0}\right\}-\rho \cdot \rho_{0}\right)}{\sqrt{\rho(1-\rho)}} \Sigma_{z} d_{0}\right]^{\prime} } \\
& \quad \times \Sigma_{z}^{-1}\left[\frac{B_{1}(\rho)-\rho B_{1}(1)}{\sqrt{\rho(1-\rho)}}-\frac{\left(\min \left\{\rho, \rho_{0}\right\}-\rho \cdot \rho_{0}\right)}{\sqrt{\rho(1-\rho)}} \Sigma_{z} d_{0}\right],
\end{aligned}
$$

thus, the LS objective function $V_{T}(k)^{2}$ weakly converges to a squared function of a normalized Brownian bridge $(\rho(1-\rho))^{-1 / 2}\left(B_{1}(\rho)-\rho B_{1}(1)\right)$ that has a covariance matrix that does not depend on $\rho$.

Next, consider the case of Assumption $6(i i)$, where $\lambda_{h}=O\left(T^{-1 / 2+\gamma}\right)$ with $0<\gamma<1 / 2$ so that $h^{-1 / 2} \delta_{h}$ increases as sampling interval shrinks but at a slower rate than $\sqrt{h}$. In this case the break point estimator is asymptotically equivalent to the LS estimator and the in-fill asymptotic distribution is equivalent to the long-span asymptotic distribution of Bai (1997).

Theorem 4. Consider the model (1.20) with unknown parameters $\left(\beta_{h}, \delta_{h}\right)$. Assumption 1, 2, 5, and $\sigma$ (ii) holds. For simplicity denote $\bar{\Omega}_{0}$ for $\bar{\Omega}\left(\rho_{0}\right)$. Then the break point estimator $\hat{\rho}=\hat{k} / T$ defined in (1.9) has the following in-fill asymptotic distribution as $h \rightarrow 0$,

$$
\frac{\left(\delta_{h}^{\prime} \Sigma_{z} \bar{\Omega}_{0} \delta_{h}\right)^{2}}{\left(\delta_{h}^{\prime} \bar{\Omega}_{0} \Xi \bar{\Omega}_{0} \delta_{h}\right)} T\left(\hat{\rho}-\rho_{0}\right) \xrightarrow{d} \underset{u \in(-\infty, \infty)}{\arg \max }\left\{W(u)-\frac{|u|}{2}\right\}
$$

where $W(u)=W_{1}(-u)$ for $u \leq 0$ and $W(u)=W_{2}(u)$ for $u>0 . W_{1}(\cdot)$ and $W_{2}(\cdot)$ are two indepen-
dent Wiener processes on $[0, \infty)$.
Proof. We omit the proof of consistency of the break point estimator $\hat{\rho} \rightarrow \rho_{0}$, because it follows the same procedure as the proof of Theorem 1. Given the rate of convergence $\hat{\rho}-\rho_{0}=$ $O_{p}\left(T^{-1} \lambda_{h}^{-2}\right)$, we only need to examine the behavior of $Q_{T}(k)^{2}-Q_{T}\left(k_{0}\right)^{2}$ for those $k$ in the neighborhood of $k_{0}$ such that $k \in K(C)$, where $K(C)=\left\{k:\left|k-k_{0}\right| \leq C \lambda_{h}^{-2}\right\}$ for some $C>0$.

Lemma 4. Consider the model (1.20) and the weight matrix $\Omega_{k}$ that satisfies Assumption 2. For the break magntiude $\delta_{h}=d_{0} \lambda_{h}$ that satisfies Assumption $6(i i)$,

$$
Q_{T}(k)^{2}-Q_{T}\left(k_{0}\right)^{2}=-\lambda_{h}^{2} d_{0}^{\prime} Z_{\Delta}^{\prime} Z_{\Delta} \Omega_{k_{0}} d_{0}+2 \lambda_{h} d_{0}^{\prime} \Omega_{k_{0}} Z_{\Delta}^{\prime} \operatorname{sgg}\left(k_{0}-k\right)+o_{p}(1)
$$

where $Z_{\Delta}:=\operatorname{sgn}\left(k_{0}-k\right)\left(Z_{k}-Z_{0}\right)$ and $o_{p}(1)$ is uniform on $K(C)$.
For the proof of Lemma 4, see Appendix A.1. Because $\delta_{h}=d_{0} \lambda_{h}$, for any constant $C$ of $K(C)$, we consider the limiting process of $Q_{T}(k)^{2}-Q_{T}\left(k_{0}\right)^{2}$ for $k=\left[k_{0}+v \lambda_{h}^{-2}\right]$ and $v \in[-C, C]$. Consider $v \leq 0$ (i.e., $\rho \leq \rho_{0}$ ). From Lemma 4,

$$
Q_{T}(k)^{2}-Q_{T}\left(k_{0}\right)^{2}=-d_{0}^{\prime}\left(\lambda_{h}^{2} \sum_{t=k+1}^{k_{0}} z_{t} z_{t}^{\prime}\right) \Omega_{k_{0}} d_{0}+2 d_{0}^{\prime} \Omega_{k_{0}}\left(\lambda_{h} \sum_{t=k+1}^{k_{0}} z_{t} \varepsilon_{t}\right)+o_{p}(1)
$$

For $k_{0}-k=\left[-v \lambda_{h}^{-2}\right]$,

$$
\lambda_{h}^{2} \sum_{t=k+1}^{k_{0}} z_{t} z_{t}^{\prime} \xrightarrow{p}|v| \Sigma_{z} .
$$

In addition, the partial sum of $z_{t} \varepsilon_{t}$ weakly converges to a Brownian motion process $B_{1}(-v)$ on $[0, \infty)$ that has variance $|v| \Xi$.

$$
\lambda_{h} \sum_{t=k+1}^{k_{0}} z_{t} \varepsilon_{t} \Rightarrow B_{1}(-v)
$$

By assumption, $\Omega_{k_{0}} \xrightarrow{p} \bar{\Omega}_{0}$. Therefore,

$$
Q_{T}\left(\left[k_{0}+v \lambda_{h}^{-2}\right]\right)^{2}-Q_{T}\left(k_{0}\right)^{2} \Rightarrow-|v| d_{0}^{\prime} \Sigma_{z} \bar{\Omega}_{0} d_{0}+2 d_{0}^{\prime} \bar{\Omega}_{0} B_{1}(-v)
$$

Let $W_{1}(\cdot)$ and $W_{2}(\cdot)$ be an Wiener processes that are independent of each other on $[0, \infty)$. Define

$$
\widetilde{G}(v):= \begin{cases}-\frac{|v|}{2}\left(d_{0}^{\prime} \Sigma_{z} \bar{\Omega}_{0} d_{0}\right)+\left(d_{0}^{\prime} \bar{\Omega}_{0} \Xi \bar{\Omega}_{0} d_{0}\right)^{1 / 2} W_{1}(-v) & \text { if } v \leq 0 \\ -\frac{|v|}{2}\left(d_{0}^{\prime} \Sigma_{z} \bar{\Omega}_{0} d_{0}\right)+\left(d_{0}^{\prime} \bar{\Omega}_{0} \Xi \bar{\Omega}_{0} d_{0}\right)^{1 / 2} W_{2}(v) & \text { if } v>0 .\end{cases}
$$

From the continuous mapping theorem, the in-fill asymptotic distribution of the break point estimator is $\lambda_{h}^{2}\left(\hat{k}-k_{0}\right) \Rightarrow \arg \max _{v} \widetilde{G}(v)$. Let $v=c u$, where $c=\left(d_{0}^{\prime} \bar{\Omega}_{0} \Xi \bar{\Omega}_{0} d_{0}\right) /\left(d_{0}^{\prime} \Sigma_{z} \bar{\Omega}_{0} d_{0}\right)^{2}$ and $u \in(-\infty, \infty)$. For $v \leq 0$,

$$
\begin{aligned}
\underset{v \in(-\infty, 0]}{\arg \max } \widetilde{G}(v) & =\underset{c u \in(-\infty, 0]}{\arg \max }-\frac{|u|}{2} c\left(d_{0}^{\prime} \Sigma_{z} \bar{\Omega}_{0} d_{0}\right)+c^{1 / 2}\left(d_{0}^{\prime} \bar{\Omega}_{0} \Xi \bar{\Omega}_{0} d_{0}\right)^{1 / 2} W_{1}(-u) \\
& =\underset{c u \in(-\infty, 0]}{\arg \max }\left\{W_{1}(-u)-\frac{|u|}{2}\right\}=\underset{u \in(-\infty, 0]}{\arg \max }\left\{W_{1}(-u)-\frac{|u|}{2}\right\}
\end{aligned}
$$

where the second equality is from $c\left(d_{0}^{\prime} \Sigma_{z} \bar{\Omega}_{0} d_{0}\right)=c^{1 / 2}\left(d_{0}^{\prime} \bar{\Omega}_{0} \Xi \bar{\Omega}_{0} d_{0}\right)^{1 / 2}$. For $v>0$ we have $\arg \max _{v \in(0, \infty)} \widetilde{G}(v)=c \arg \max _{u \in(0, \infty)}\left\{W_{2}(u)-|u| / 2\right\}$. Thus,

$$
\begin{gathered}
c^{-1} \lambda_{h}^{2}\left(\hat{k}-k_{0}\right) \Rightarrow \underset{u \in(-\infty, \infty)}{\arg \max }\left\{W(u)-\frac{|u|}{2}\right\} \\
c^{-1} \lambda_{h}^{2}=\left(\delta_{h}^{\prime} \Sigma_{z} \bar{\Omega}_{0} \delta_{h}\right)^{2} /\left(\delta_{h}^{\prime} \bar{\Omega}_{0} \Xi \bar{\Omega}_{0} \delta_{h}\right)
\end{gathered}
$$

for $W(\cdot)$ defined in Theorem 4.

Under the assumptions of Theorem 4, we can construct confidence intervals of the break point from consistent estimators of $\Sigma_{z}, \Xi, \delta_{h}$ and $\bar{\Omega}_{0}$. Let $\hat{\Sigma}_{z}=T^{-1} \sum_{t=1}^{T} z_{t} z_{t}^{\prime}$. For serially correlated and heteroskedastic disturbances, use a heteroskedasty and autocorrelation consistent (HAC) estimator of $T^{-1 / 2} \sum_{t=1}^{T} z_{t} \varepsilon_{t}$ and denote it as $\hat{\Xi}$. Denote the break magnitude estimator corresponding to $\hat{\rho}$ as $\hat{\delta}$. From Corollary 1 we have $\hat{\delta}=\delta_{h}+o_{p}(1)$. Lastly, $\bar{\Omega}_{0}$ is consistently estimated by replacing $k$ to $\hat{k}$ from the continuous mapping theorem, which we denote as $\hat{\Omega}_{k}$ instead of $\Omega_{\hat{k}}$. Define $\hat{L}:=\left(\hat{\delta}^{\prime} \hat{\Sigma}_{z} \hat{\Omega}_{k} \hat{\delta}\right)^{2} /\left(\hat{\delta}^{\prime} \hat{\Omega}_{k} \hat{\Xi} \hat{\Xi} \hat{\Omega}_{k} \hat{\delta}\right)$, and $L$ as the parameter analogue. Then
we can show that $(\hat{L}-L)\left(\hat{k}-k_{0}\right) \xrightarrow{p} 0$, and the $100(1-\alpha) \%$ confidence interval is given by $\left[\hat{k}-\left[c_{\alpha} / \hat{L}\right]-1, \hat{k}+\left[c_{\alpha} / \hat{L}\right]+1\right]$ where $c_{\alpha}$ is the $(1-\alpha / 2)$ th quantile of the random variable $\arg \max _{u}\{W(u)-|u| / 2\}$, and $\left[c_{\alpha} / \hat{L}\right]$ is the integer part of $c_{\alpha} / \hat{L}$. The quantile $c_{\alpha}$ can be computed from the cumulative distribution function formula (B.4) in appendix B of Bai (1997).

Under small break magnitudes, the limit distribution of the break point estimator depends on nuisance parameters in a complicated way (see Theorem 3). In this case we can use bootstrap methods to approximate the distribution of the break point estimator. Under the assumption that the errors of model (1.8) are i.i.d., we can use a residual-based method to estimate the model. Let $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}})$ be OLS estimates of coefficients corresponding to break date estimate $\hat{k}=[\hat{\rho} T]$. The estimated residuals are $\hat{\varepsilon}_{t}=y_{t}-x_{t}^{\prime} \hat{\beta}-z_{t}^{\prime} \hat{\mathbf{\delta}}\{t>\hat{k}\}$ for $t=1, \ldots, T$. We draw a random sample from $\left\{\hat{\varepsilon}_{1}-\tilde{\varepsilon}, \ldots, \hat{\varepsilon}_{T}-\tilde{\varepsilon}\right\}$ with replacement and label it as $\hat{\mathcal{E}}_{(b)}^{*}=\left\{\hat{\varepsilon}_{1}^{*}, \ldots, \hat{\varepsilon}_{T}^{*}\right\}$ for $b=1, \ldots, B$, where $\tilde{\varepsilon}:=T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_{t}$ and $B$ is the number of bootstrap replications. We can construct a new process $\left\{y_{t}^{*}\right\}$ as

$$
y_{t}^{*}=x_{t}^{\prime} \hat{\beta}+z_{t}^{\prime} \hat{\delta} \mathbf{1}\{t>\hat{k}\}+\hat{\varepsilon}_{t}^{*}
$$

for $t=1, \ldots, T$ and obtain break date estimate $\hat{k}_{(b)}^{*}$ associated each bootstrap sample $\hat{\mathcal{E}}_{(b)}^{*}$ for $b=1, \ldots, B$. To account for heteroskedasticity of errors, we can use the wild bootstrap method of Liu (1988). To construct the confidence interval of break date, first sort the estimated break dates $\hat{k}_{(b)}^{*}$ in ascending order with the estimate $\hat{k}$ included. The $100(1-\alpha) \%$ confidence interval is obtained by finding the $\alpha / 2$ and $(1-\alpha / 2)$ quantiles.

The residual-based bootstrap method is also valid when regressors include lags of the dependent variable. We construct $y_{t}^{*}$ recursively by using $y_{0}^{*}=y_{0}$. To account serially correlated errors we can use the sieve bootstrap by Bühlmann (1997); for heteroskedasticity use the wild bootstrap method by Liu (1988).

### 1.4.2 Break in an autoregressive model

In this section I derive the in-fill asymptotic distribution of an autoregressive (AR) model with a structural break in its lag coefficient, using a deterministic weight function $\omega(\cdot)$. As mentioned in Section 1.3, Assumption 1 excludes lagged dependent variables, due to dependence of weight function on regressors. We relax this condition to allow weakly exogeneous regressors by assuming non-stochastic weights. I follow the approach of Jiang et al. (2017) of using a discrete model closely related to the Ornstein-Uhlenbeck process with a break in the drift function:

$$
d x(t)=-\left(\mu+\delta \mathbf{1}\left\{t>\rho_{0}\right\}\right) x(t) d t+\sigma d B(t)
$$

where $t \in[0,1]$ and $B(\cdot)$ denotes a standard Brownian motion. The discrete time model has the form of

$$
x_{t}=\left(\beta_{1} \mathbf{1}\left\{t \leq k_{0}\right\}+\beta_{2} \mathbf{1}\left\{t>k_{0}\right\}\right) x_{t-1}+\sqrt{h} \varepsilon_{t}, \varepsilon_{t} \stackrel{i . i . d .}{\sim}\left(0, \sigma^{2}\right), x_{0}=O_{p}(1)
$$

where $\beta_{1}=\exp \{-\mu / T\}$ and $\beta_{2}=\exp \{-(\mu+\delta) / T\}$ are the AR roots before and after the break. Denote $y_{t}=x_{t} / \sqrt{h}$ so that the order of errors is $O_{p}(1)$ as in model (1.8). Then, we have for $t=1, \ldots, T$,

$$
\begin{equation*}
y_{t}=\left(\beta_{1} \mathbf{1}\left\{t \leq k_{0}\right\}+\beta_{2} \mathbf{1}\left\{t>k_{0}\right\}\right) y_{t-1}+\varepsilon_{t}, \varepsilon_{t} \stackrel{i . i . d .}{\sim}\left(0, \sigma^{2}\right), y_{0}=x_{o} / \sqrt{h}=O_{p}\left(T^{1 / 2}\right) . \tag{1.22}
\end{equation*}
$$

The initial condition of $y_{t}$ in (1.22) diverges at rate $T^{1 / 2}$, thus the in-fill asymptotic distribution will depend explicitly on the initial value $x_{0}$. The break size is $\beta_{2}-\beta_{1}=O\left(T^{-1}\right)$, whereas in literature using long-span asymptotics it is assumed to be $O\left(T^{-\gamma}\right)$ with $0<\gamma<1$. Also, note that $\beta_{1}=\exp \{-\mu / T\} \rightarrow 1$ and $\boldsymbol{\beta}_{2}=\exp \{-(\mu+\delta) / T\} \rightarrow 1$ as $T \rightarrow \infty$ for any finite $(\mu, \boldsymbol{\delta})$. Hence, the $\mathrm{AR}(1)$ model (1.22) is a local-to-unit root process. In contrast, long-span asymptotic theory incorporates stationary $\operatorname{AR}(1)$ processes where $\left|\beta_{1}\right|<1$ and $\left|\beta_{2}\right|<1$ in model (1.22), where

Chong (2001) derives the long-span distribution under $\left|\beta_{2}-\beta_{1}\right|=O\left(T^{-1 / 2+\gamma}\right)$ with $0<\gamma<1 / 2$. Because this paper focuses on the break point under small break magnitudes, the in-fill asymptotic theory is adapted instead of long-span asymptotics. Moreover, Jiang et al. (2017) provides simulation results that the in-fill asymptotic theory works well even when $\beta_{1}$ and/or $\beta_{2}$ are distant from unity in finite sample.

The new break point estimator and the LS estimator in model (1.22) takes the form

$$
\begin{gather*}
S(k)^{2}=\sum_{t=1}^{k}\left(y_{t}-\hat{\beta}_{1}(k) y_{t-1}\right)^{2}+\sum_{t=k+1}^{T}\left(y_{t}-\hat{\beta}_{2}(k) y_{t-1}\right)^{2} \\
\hat{k}=\underset{k=1, \ldots, T-1}{\arg \min } \omega_{k}^{2} S(k)^{2}, \quad \hat{\rho}=\hat{k} / T  \tag{1.23}\\
\hat{k}_{L S}=\underset{k=1, \ldots, T-1}{\arg \min } S(k)^{2}, \quad \hat{\rho}_{L S}=\hat{k}_{L S} / T
\end{gather*}
$$

where $\hat{\beta}_{1}(k)=\sum_{t=1}^{k} y_{t} y_{t-1} / \sum_{t=1}^{k} y_{t-1}^{2}$ and $\hat{\beta}_{2}(k)=\sum_{t=k+1}^{T} y_{t} y_{t-1} / \sum_{t=k+1}^{T} y_{t-1}^{2}$ are LS estimates of $\beta_{1}$ and $\beta_{2}$ under break at $k$, respectively.

Theorem 5. Consider the model (1.22) with fixed parameters $(\mu, \delta)$ so that $\ln \beta_{1}=O\left(T^{-1}\right)$ and $\ln \beta_{2}=O\left(T^{-1}\right)$. Assume the weight function $\omega_{k}$ is nonrandom and bounded on the unit interval with $\omega_{k} \rightarrow \omega(\rho)$ as $k / T \rightarrow \rho$. Then the break point estimator $\hat{\rho}=\hat{k} / T$ in (1.23) has the in-fill asymptotic distribution as

$$
\hat{\rho} \Longrightarrow \underset{\rho \in(0,1)}{\arg \max } \omega(\rho)^{2}\left[\frac{\left(\widetilde{J}_{0}(\rho)^{2}-\widetilde{J}_{0}(0)^{2}-\rho\right)^{2}}{\int_{0}^{\rho} \widetilde{J}_{0}(r)^{2} d r}+\frac{\left(\widetilde{J}_{0}(1)^{2}-\widetilde{J}_{0}(\rho)^{2}-(1-\rho)\right)^{2}}{\int_{\rho}^{1} \widetilde{J}_{0}(r)^{2} d r}\right]
$$

where $\widetilde{J}_{0}(r)$, for $r \in[0,1]$ is a Gaussian process defined by

$$
\begin{equation*}
d \widetilde{J}_{0}(r)=-\left(\mu+\delta \boldsymbol{1}\left\{r>\rho_{0}\right\}\right) \widetilde{J}_{0}(r) d r+d B(r) \tag{1.24}
\end{equation*}
$$

with the initial condition $\widetilde{J}_{0}(0)=y_{0} / \sigma=x_{0} /(\sigma \sqrt{h})$, and $B(\cdot)$ is a standard Brownian motion.

The results of Theorem 5 are derived from applying the continuous mapping theorem to the limit distribution $S(k)^{2}$ in Theorem 4.1 of Jiang et al. (2017). See Appendix A. 1 for proof. The difference of the two estimators' asymptotic distributions is the weight function $\omega(\rho)$ multiplied to the stochastic process in the argmax function. Both estimators are asymmetrically distributed around the true point $\rho_{0} \neq 1 / 2$ and biased. In Section 1.5 we see that the variance of the in-fill distribution of the break point estimator is smaller than that of the LS estimator.

### 1.5 Monte Carlo Simulation

Finite sample distributions of the new estimator and the LS estimator are compared by Monte Carlo simulation. I consider structural breaks in two different models: a break in the mean of a univariate regression model and a break in the lag coefficient of a $\operatorname{AR}(1)$ process. The root mean squared error (RMSE), bias and standard errors of the two estimators are compared in finite sample and in-fill asymptotics.

### 1.5.1 Univariate stationary process

The first model is when a structural break occurs in model (1.8) where $x_{t}=z_{t}=1$ for all $t$. The break magnitude $\delta_{T}=d_{0} T^{-1 / 2}$ is in the local $T^{-1 / 2}$ neighborhood of zero to represent small break magnitudes.

$$
\begin{equation*}
y_{t}=\mu+\delta_{T} \mathbf{1}\left\{t>\left[\rho_{0} T\right]\right\}+\varepsilon_{t} \tag{1.25}
\end{equation*}
$$

where $\sigma=1$ and $\varepsilon_{t} \stackrel{i . i . d .}{\sim} N\left(0, \sigma^{2}\right)$. Parameter values are $\rho_{0} \in\{0.15,0.3,0.5,0.7,0.85\}, \mu=4, d_{0} \in$ $\{1,2,4\}$ and $T=100$ with 5,000 replications. The weight function is $\omega_{k}=(k / T(1-k / T))^{1 / 2}$, which is the representative weight function motivated in Section $1.2^{7}$. The break point estimator

[^4]$\hat{\rho}_{N E W}$ is defined in (1.3) and the LS estimator $\hat{\rho}_{L S}$ in (1.2).
Table 1.1 provides the RMSE, the bias and the standard error for the finite sample distribution. The values are extremely close to those computed using the in-fill asymptotic distribution in Table 1.2. We can see that for all $\rho_{0} \in\{0.15,0.3,0.5,0.7,0.85\}$ and $d_{0} \in\{1,2,4\}$, the RMSE of the estimator $\hat{\rho}_{N E W}$ is smaller than that of $\hat{\rho}_{L S}$ in finite sample. The RMSE in limit distribution shows the same results. The break point estimator outperforms the LS estimator uniformly in terms of the RMSE. The difference in asymptotic RMSE of the two estimators are minimized when $\left(\rho_{0}, d_{0}\right)=(0.15,4)$ and maximized when $\left(\rho_{0}, d_{0}\right)=(0.5,1)$. We can see a trade-off of slightly larger bias but a large decrease in standard error for $\hat{\rho}_{N E W}$ compared to $\hat{\rho}_{L S}$, that leads to a decrease in RMSE.

Figures 1.5, 1.6 and 1.7 shows the finite sample distribution of the two estimator under $\rho_{0}=0.15,0.30$ and 0.85 . In finite sample the LS estimator performs particularly worse when $\rho_{0}=0.15$ or 0.85 , and $d_{0}=1$ is small. Under these parameter values, the tri-modal LS estimator distribution becomes bi-modal with modes at zero and one. In contrast, the new break point estimator has an unique mode at the true break $\rho_{0}$, or at a point very close to true value. This finite sample property of the break point estimator holds uniformly across breaks $\rho_{0} \in[0.15,0.85]$ for all $d_{0}$ values considered; the distributions look similar to Figures 1.5, 1.6 and 1.7 and are omitted here. This implies that regardless of the true break point, the new estimator outperforms, or is not worse off than the LS estimator in finite sample under a small break magnitude. This result is quite strong because the performance of the new estimator is superior or at least as reliable as the LS estimator.


Figure 1.5: $\left(\rho_{0}=0.15\right)$ Finite sample distribution of $\hat{\rho}_{N E W}$ (left) and $\hat{\rho}_{L S}$ (right) under model (1.25) with parameter values $\left(\rho_{0}, \delta_{T}\right)=\left(0.15, T^{-1 / 2}\right),\left(0.15,2 T^{-1 / 2}\right)$ and $\left(0.15,4 T^{-1 / 2}\right)$ and $T=100$, respectively.


Figure 1.6: $\left(\rho_{0}=0.30\right)$ Finite sample distribution of $\hat{\rho}_{N E W}$ (left) and $\hat{\rho}_{L S}$ (right) under model (1.25) with parameter values $\left(\rho_{0}, \delta_{T}\right)=\left(0.3, T^{-1 / 2}\right),\left(0.3,2 T^{-1 / 2}\right)$ and $\left(0.3,4 T^{-1 / 2}\right)$ and $T=$ 100 , respectively.


Figure 1.7: $\left(\rho_{0}=0.85\right)$ Finite sample distribution of $\hat{\rho}_{N E W}$ (left) and $\hat{\rho}_{L S}$ (right) under model (1.25) with parameter values $\left(\rho_{0}, \delta_{T}\right)=\left(0.85, T^{-1 / 2}\right),\left(0.85,2 T^{-1 / 2}\right)$ and $\left(0.85,4 T^{-1 / 2}\right)$ and $T=100$, respectively.

Table 1.1: Finite sample RMSE, bias, and the standard error of the new estimator and the LS estimator of the break point under model (1.25) with parameter values ( $\rho_{0}, d_{0}$ ) and $T=100$. The number of replications is 5,000 .

|  |  | RMSE |  | Bias |  | Standard error |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\rho_{0}$ | $d_{0}$ | NEW | LS | NEW | LS | NEW | LS |
|  | 1 | 0.4086 | 0.4970 | 0.3442 | 0.3387 | 0.2206 | 0.3636 |
| 0.15 | 2 | 0.3949 | 0.4789 | 0.3262 | 0.3181 | 0.2226 | 0.3580 |
|  | 4 | 0.3468 | 0.4099 | 0.2678 | 0.2368 | 0.2204 | 0.3346 |
|  | 1 | 0.2862 | 0.4104 | 0.1876 | 0.1935 | 0.2161 | 0.3618 |
| 0.30 | 2 | 0.2672 | 0.3821 | 0.1667 | 0.1686 | 0.2087 | 0.3429 |
|  | 4 | 0.2041 | 0.2972 | 0.1137 | 0.1009 | 0.1695 | 0.2795 |
|  | 1 | 0.2104 | 0.3563 | -0.0041 | -0.0138 | 0.2103 | 0.3561 |
| 0.50 | 2 | 0.1908 | 0.3333 | 0.0043 | 0.0036 | 0.1907 | 0.3333 |
|  | 4 | 0.1375 | 0.2592 | -0.0007 | -0.0043 | 0.1375 | 0.2592 |
|  | 1 | 0.2866 | 0.4061 | -0.1886 | -0.1896 | 0.2158 | 0.3591 |
| 0.70 | 2 | 0.2693 | 0.3827 | -0.1707 | -0.1673 | 0.2083 | 0.3442 |
|  | 4 | 0.2073 | 0.3093 | -0.1165 | -0.1127 | 0.1715 | 0.2880 |
|  | 1 | 0.4096 | 0.5030 | -0.3467 | -0.3459 | 0.2181 | 0.3652 |
| 0.85 | 2 | 0.3959 | 0.4800 | -0.3255 | -0.3180 | 0.2253 | 0.3595 |
|  | 4 | 0.3496 | 0.4123 | -0.2693 | -0.2370 | 0.2230 | 0.3374 |

### 1.5.2 Autoregressive process

For the AR(1) process, I replicate two experiments from Jiang et al. (2017). The first experiment is a break in the lag coefficient so that the stationary process changes to another stationary AR(1) process. The second case is a change from a local-to-unit root to a stationary $\operatorname{AR}(1)$ process. Each experiment is generated from model (1.22) with $h=1 / 200(T=200), \sigma=1$, $\varepsilon_{t} \stackrel{i . i . d .}{\sim} N(0,1), \rho_{0} \in\{0.3,0.5,0.7\}$ and different combinations of $\mu$ and $\delta$ with $\beta_{1}=\exp (-\mu / T)$ and $\beta_{2}=\exp (-(\mu+\delta) / T)$.

1. Stationary to stationary: $(\mu, \delta)=(138,55)$ which implies $\left(\beta_{1}, \beta_{2}\right)=(0.5,0.38)$;
2. Local-to-unity to stationary: $(\mu, \delta)=(1,5)$ which implies $\left(\beta_{1}, \beta_{2}\right)=(0.995,0.97)$.

The stochastic integrals of in-fill asymptotic distributions are approximated over a grid size

Table 1.2: In-fill asymptotic RMSE, bias, and standard error of the new estimator and the LS estimator of the break point under model (1.25) with parameter values ( $\rho_{0}, d_{0}$ ) and $T=100$. The number of replications is 5,000 .

|  |  | RMSE |  | Bias |  | Standard error |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\rho_{0}$ | $d_{0}$ | NEW | LS | NEW | LS | NEW | LS |
| 0.15 | 1 | 0.4138 | 0.4951 | 0.3486 | 0.3400 | 0.2228 | 0.3598 |
|  | 2 | 0.3947 | 0.4780 | 0.3270 | 0.3140 | 0.2210 | 0.3604 |
|  | 4 | 0.3493 | 0.4126 | 0.2699 | 0.2385 | 0.2217 | 0.3367 |
|  | 1 | 0.2928 | 0.4090 | 0.1953 | 0.1989 | 0.2181 | 0.3574 |
|  | 2 | 0.2670 | 0.3849 | 0.1689 | 0.1694 | 0.2068 | 0.3456 |
|  | 4 | 0.2040 | 0.3025 | 0.1128 | 0.1045 | 0.1700 | 0.2839 |
|  | 1 | 0.2109 | 0.3562 | 0.0023 | 0.0011 | 0.2109 | 0.3562 |
|  | 2 | 0.1933 | 0.3355 | 0.0011 | -0.0068 | 0.1933 | 0.3354 |
|  | 4 | 0.1372 | 0.2612 | 0.0007 | -0.0002 | 0.1372 | 0.2612 |
|  | 1 | 0.2909 | 0.4091 | -0.1958 | -0.1979 | 0.2150 | 0.3581 |
| 0.70 | 2 | 0.2649 | 0.3797 | -0.1657 | -0.1638 | 0.2067 | 0.3426 |
|  | 4 | 0.2069 | 0.3065 | -0.1150 | -0.1072 | 0.1720 | 0.2871 |
|  | 1 | 0.4066 | 0.4948 | -0.3429 | -0.3347 | 0.2185 | 0.3644 |
| 0.85 | 2 | 0.3935 | 0.4768 | -0.3250 | -0.3108 | 0.2218 | 0.3616 |
|  | 4 | 0.3503 | 0.4151 | -0.2699 | -0.2406 | 0.2232 | 0.3383 |

$h=0.005$. The break point estimator $\hat{\rho}_{N E W}$ of the $\operatorname{AR}(1)$ model is defined in (1.23) and its asymptotic distribution is stated in Theorem 3. The in-fill asymptotic distribution of the LS estimator $\hat{\rho}_{L S}$ is stated Theorem 4.1 of Jiang et al. (2017).

Tables 1.3 and 1.4 provide the RMSE, bias and the standard error of $\hat{\rho}_{N E W}$ and $\hat{\rho}_{L S}$ for the finite sample and the asymptotic distribution, respectively. Similar to the structural break in the mean of a stationary model, the RMSE of $\hat{\rho}_{N E W}$ is smaller than that of $\hat{\rho}_{L S}$ for all parameter values $\left(\beta_{1}, \beta_{2}, \rho_{0}\right)$ considered. This is also holds in the limit. In addition, the asymptotic distribution approximates the finite sample better for the local-to-unit root process change. We can see that the decrease in the RMSE of $\hat{\rho}_{N E W}$ is from the trade-off a relatively large decrease in variance compared to the increase in the squared bias.

Figures 1.8 and 1.9 are finite sample distributions of the break point in the stationary and
the local-to-unity $\operatorname{AR}(1)$ processes, respectively. For the stationary to another stationary process change, the LS estimator $\hat{\rho}_{L S}$ mode at the true break point is almost negligible unless it is the median $\rho_{0}=0.5$. In contrast, the estimator $\hat{\rho}_{N E W}$ has a unique mode at the true break point for all $\rho_{0} \in\{0.3,0.5,0.7\}$. For the local-to-unit root to a stationary $\operatorname{AR}(1)$ change, both estimators have a higher probability at the true break point. However, the LS estimator continues to exhibit tri-modality with modes at ends whereas our estimator has a unique mode at $\rho_{0}$. Thus, simulation results of the $\mathrm{AR}(1)$ model gives the same conclusion as section 1.5.1; the performance of the break point estimator is superior or at least as reliable as the LS estimator.

Table 1.3: Finite sample RMSE, bias, and standard error of the new estimator and the LS estimator of the break point under the $\operatorname{AR}(1)$ model (1.22) with parameter values $\left(\beta_{1}, \beta_{1}, \rho_{0}\right)$ and $T=200$. The number of replications is 5,000 .

|  |  |  | RMSE |  | Bias |  | Standard error |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\beta_{1}$ | $\beta_{2}$ | $\rho_{0}$ | NEW | LS | NEW | LS | NEW | LS |
|  |  | 0.3 | 0.2627 | 0.3091 | 0.1821 | 0.1657 | 0.1893 | 0.2610 |
| 0.5 | 0.38 | 0.5 | 0.1763 | 0.2452 | 0.0204 | 0.0282 | 0.1751 | 0.2436 |
|  |  | 0.7 | 0.2285 | 0.2725 | -0.1379 | -0.1223 | 0.1822 | 0.2435 |
|  |  | 0.3 | 0.2369 | 0.2780 | 0.1319 | 0.1279 | 0.1967 | 0.2469 |
| 0.995 | 0.97 | 0.5 | 0.1754 | 0.2328 | -0.0042 | -0.0047 | 0.1754 | 0.2327 |
|  |  | 0.7 | 0.2375 | 0.2784 | -0.1358 | -0.1336 | 0.1948 | 0.2442 |



Figure 1.8: (Stationary to stationary) Finite sample distributions of $\hat{\rho}_{N E W}$ (left) and $\hat{\rho}_{L S}$ (right) when the lag coefficient pre- and post-break are $\left(\beta_{1}, \beta_{2}\right)=(0.5,0.38)$ at break points $\rho_{0}=$ $0.3,0.5$, and 0.7 , respectively.


Figure 1.9: (Local-to-unity to stationary) Finite sample distributions of the new estimator (left) and the LS estimator (right) when the lag coefficient pre- and post-break are $\left(\beta_{1}, \beta_{2}\right)=$ $(0.995,0.97)$ at break points $\rho_{0}=0.3,0.5$, and 0.7 , respectively.

Table 1.4: In-fill asymptotic RMSE, bias, and standard error of the new estimator and the LS estimator of the break point under the $\operatorname{AR}(1)$ model (1.22) with parameter values ( $\left.\beta_{1}, \beta_{1}, \rho_{0}\right)$ and $T=200$. The number of replications is 5,000 .

|  |  |  | RMSE |  | Bias |  | Standard error |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\beta_{1}$ | $\beta_{2}$ | $\rho_{0}$ | NEW | LS | NEW | LS | NEW | LS |
|  |  | 0.3 | 0.1911 | 0.2944 | 0.1888 | 0.0422 | 0.0298 | 0.2913 |
| 0.5 | 0.38 | 0.5 | 0.0270 | 0.2969 | -0.0053 | -0.0518 | 0.0264 | 0.2923 |
|  |  | 0.7 | 0.1995 | 0.3553 | -0.1970 | -0.1623 | 0.0316 | 0.3160 |
|  |  | 0.3 | 0.1325 | 0.2391 | 0.0472 | -0.1148 | 0.1238 | 0.2097 |
| 0.995 | 0.97 | 0.5 | 0.1118 | 0.3650 | -0.0205 | -0.2383 | 0.1099 | 0.2765 |
|  |  | 0.7 | 0.1975 | 0.5229 | -0.1022 | -0.3897 | 0.1690 | 0.3487 |

### 1.6 Empirical Application

In this section I use the new estimation method for structural breaks in three empirical applications. I analyze the performance of the break point estimator by comparing with the LS estimator and historical events documented in literature. Furthermore, I show that the estimator is robust to trimming the sample period whereas the LS estimator varies significantly depending on the trimmed sample. The first application is about the structural break in postwar U.S. real GDP growth rate, where an autoregressive model is used to estimate the break. The second application is estimating the break date on the U.S. and the UK stock returns using the return prediction model of Paye and Timmermann (2006). Lastly, I analyze the structural break of the relation between oil price shocks and the U.S. output growth rate studied in Hamilton (2003).

### 1.6.1 U.S. real GDP growth rate

In macroeconomics literature, shocks that affect mean growth rate are often modelled as a one-time structural break because of its rare occurrence. However, existing estimation methods fail to capture the graphical evidence of postwar European and U.S. growth slowing down sometime in the 1970s, known as the "productivity growth slowdown". For instance, Bai,

Lumsdaine, and Stock (1998) show that for the U.S., most test statistics reject the no-break hypothesis, but the estimated confidence interval does not contain the slowdown in the 1970s.

I estimate a structural break of an autoregressive model using postwar quarterly U.S. real GDP growth rate. Real GDP in chained dollars (base year 2012) data are obtained from the Bureau of Economic Analysis (BEA) website for the sample period 1947Q1-2018Q2, seasonally adjusted at annual rates. Annualized quarterly growth rates are calculated as 400 times the first differences of the natural logarithms of the levels data. I assume that log output has a stochastic trend with a drift and a finite-order representation. Following the approach of Eo and Morley (2015), I use the modified Bayesian information criterion (BIC) of Kurozumi and Tuvaandorj (2011) for lag selection in order to account for structural breaks. The highest lag order selected is 1 for output growth given an upper bound of four lags and four breaks. The $\operatorname{AR}(1)$ model is estimated under three cases. First case is a break in the drift term only (the constant term $\gamma=0$ ), second case is a break in the coefficient of lags only, the "propagation term" $\left(\delta_{1}=0\right)$, and lastly a break in both constant and coefficient.

$$
\begin{equation*}
\Delta y_{t}=\beta+\phi_{1} \Delta y_{t-1}+\mathbf{1}\left\{t>k_{0}\right\}\left(\gamma+\delta_{1} \Delta y_{t-1}\right)+\varepsilon_{t} . \tag{1.26}
\end{equation*}
$$

Assume the error term $\left\{\varepsilon_{t}\right\}$ are serially uncorrelated mean zero disturbances. If a structural break occurs in both constant and lag coefficient, the long-run growth rate of log output will change from $E\left[\Delta y_{t}\right]=\beta /\left(1-\phi_{1}\right)$ to $(\beta+\gamma) /\left(1-\phi_{1}-\delta_{1}\right)$ and the volatility of growth rate will change from $\operatorname{Var}\left[\Delta y_{t}\right]=\sigma^{2} /\left(1-\phi_{1}^{2}\right)$ to $\sigma^{2} /\left(1-\left(\phi_{1}+\delta_{1}\right)^{2}\right)$ at time $k_{0}$.

Using the notations of model (1.8), we have $x_{t}=\left(1, \Delta y_{t-1}\right)^{\prime}$, the dependent variable is $\Delta y_{t}$ and for each model $z_{t}=R^{\prime} x_{t}$ is as follows.

- M1: $R=(1,0)^{\prime}, z_{t}=1$
- M2: $R=(0,1)^{\prime}, z_{t}=\Delta y_{t-1}$
- M3: $R=(1,1)^{\prime}, z_{t}=\left(1, \Delta y_{t-1}\right)^{\prime}$

The break point estimator $\hat{k}_{N E W}$ defined in (1.9) is obtained by using the weight $\Omega_{k}=\omega_{k}^{2} I_{q}$ where $\omega_{k}=(k / T(1-k / T))^{1 / 2}$ and $q=\operatorname{dim}\left(z_{t}\right)^{8}$. The full sample 1947Q3-2018Q2 $(T=284)$ is used to estimate the structural break date, then a shorter sub-sample is used to see if break date estimates change. $k$ is searched over a trimmed sample with fraction $\alpha=0.1$ at both ends; the grid starts at 1954Q2 and ends at 2011Q1 for the full sample. The second and third columns of Table 1.5 shows break date estimates for the full sample. The two estimates are extremely different from each other for M1 and M3, $\hat{k}_{N E W}$ is 1973Q1 whereas $\hat{k}_{L S}$ is 2000Q2. The break point estimates on the unit interval are approximately 0.36 and 0.75 , respectively. Without any knowledge of historical events, one might think that the finite sample properties of $\hat{\rho}_{L S}$ do not appear here because it is not close to $\alpha=0.1$ or $1-\alpha=0.9$.

However, the LS estimate switches to boundary of search grid if we consider a sub-sample that is one decade shorter. Consider a sub-sample that ends at 2007Q1 with starting date 1947Q3, so that the search grid includes $\hat{k}_{L S}$ from all models. The LS estimate of M1 changes drastically to 1953 Q 1 which is the end of the search grid, $\hat{\rho}_{L S}=0.1$. In contrast, our estimator under M1 provides the same break date estimate $\hat{k}_{N E W}=1973$ Q1. For M3 both estimates change so that $\hat{k}_{N E W}=1966 \mathrm{Q} 1$ and $\hat{k}_{L S}=1958 \mathrm{Q} 1$. Compared to the full sample estimate, the change of the new estimate is 7 years whereas the LS estimate change is over 40 years.

The $95 \%$ confidence interval is obtained under each model and samples by residual-based bootstrap (see Section 1.4) with the number of bootstrap samples set to 999 . Under M1 the LS estimator switches to a break date estimate outside the confidence interval when the sample is trimmed (under M3 it switches to boundary value of the confidence interval). This implies that LS estimates vary significantly depending on trimming and it is likely to be at boundaries due to its finite sample behavior. In addition, we can check that the confidence interval of our estimation method has shorter length than LS for all cases.

[^5]Table 1.5: Structural break date estimates of postwar U.S. real GDP growth rate in a $\operatorname{AR}(1)$ model. For each model the first row is the break date estimate and the third row is the break point estimate (fraction within corresponding sample). The second and fourth rows are $95 \%$ confidence intervals obtained by bootstrap with 999 replications.

|  | 1947Q3-2018Q2 |  | 1947Q3-2007Q1 |  |
| :--- | :---: | :---: | :---: | :---: |
| Model | NEW | LS | NEW | LS |
|  | 1973Q1 | 2000 Q 2 | 1973 Q 1 | 1953 Q 1 |
|  | [59Q4, 02Q2] | [57Q4, 10Q3] | $[56 \mathrm{Q} 1,96 \mathrm{Q} 4]$ | $[53 \mathrm{Q} 1,00 \mathrm{Q} 2]$ |
|  | 0.36 | 0.75 | 0.43 | 0.10 |
|  | $[0.18,0.77]$ | $[0.15,0.89]$ | $[0.15,0.83]$ | $[0.10,0.89]$ |
|  | 1966 Q 1 | 1966 Q 1 | 1966 Q 1 | 1966 Q 1 |
|  | $[57 \mathrm{Q} 2,07 \mathrm{Q} 4]$ | $[54 \mathrm{Q} 3,10 \mathrm{Q} 4]$ | $[55 \mathrm{Q} 2,97 \mathrm{Q} 1]$ | $[53 \mathrm{Q} 2,00 \mathrm{Q} 3]$ |
| M2 | 0.26 | 0.26 | 0.32 | 0.32 |
|  | $[0.14,0.85]$ | $[0.10,0.89]$ | $[0.13,0.84]$ | $[0.10,0.90]$ |
|  | 1973 Q 1 | 2000 Q 2 | 1966 Q 1 | 1958 Q 1 |
|  | $[59 \mathrm{Q} 4,02 \mathrm{Q} 1]$ | $[57 \mathrm{Q} 4,10 \mathrm{Q} 3]$ | $[56 \mathrm{Q} 1,96 \mathrm{Q} 4]$ | $[53 \mathrm{Q} 1,00 \mathrm{Q} 3]$ |
| M3 | 0.36 | 0.75 | 0.32 | 0.18 |
|  | $[0.18,0.77]$ | $[0.15,0.89]$ | $[0.15,0.83]$ | $[0.10,0.90]$ |

The break date estimate $\hat{k}_{N E W}=1973$ Q1 under M1 corresponds to the productivity growth slowdown in early 1970s, which is widely hypothesized in macroeconomics literature. The U.S. labor productivity experienced a slowdown in growth after the oil shock in 1973 (see Perron (1989) and Hansen (2001)). None of the models estimate a break date in 1980s, which is known as "the Great Moderation". It refers to a empirical fact of a large reduction of volatility of U.S. real GDP growth in 1984Q1, established by Kim and Nelson (1999) and McConnell and Perez-Quiros (2000). In model (1.26), the change in volatility is not a linear function of the change in the lag coefficient. Hence, we focus on events that affect the mean rather than the volatility of growth rate.

Because the sample is over 70 years, it is likely that there exists multiple structural breaks in the output growth rate. Estimates that vary depending on the sub-sample could be evidence of more than one break in the sample period. The break date estimate $\hat{k}_{L S}=2000 \mathrm{Q} 2$ can be associated with the tech bubble, also known as the dot-com crash in 2000. In relation to business
cycles, $\hat{k}_{L S}=1953$ Q1 and 1958Q1 are both in recession, at 1953 is the Korean war ended. Under M2 both estimates from the full sample are 1966Q1, and the closest historical event that is likely to affect the output growth rate is the Vietnam war.

Estimating multiple structural breaks using the weighting scheme is beyond the scope of this paper. For LS estimation of multiple breaks see Bai and Perron (1998) and Bai, Lumsdaine, and Stock (1998). Instead, I compute the two estimators for sub-samples that end at different dates from 2005Q1 to 2018Q2, for total 54 sub-samples. Because the sample is trimmed one quarter at a time, switching to a different estimate that is far apart implies that the estimator is sensitive to trimming, rather than suggesting multiple breaks. Table 1.6 shows that the LS estimates of M1 and M3 vary from $\hat{\rho}_{L S}=0.1$ to 0.9 , which is equivalent to the trim fraction $\alpha=0.1$. However, the new estimates are either mid 1960s or early 1970s, which is always in the fraction interval $\hat{\rho} \in[0.2,0.5]$.

In short, estimating a structural break of postwar U.S. real GDP growth rate using our estimation method provides evidence of a break occurring in 1973Q1, which corresponds to the productivity growth slowdown period. However, the LS estimates a break occurs in 2000 or 1953, depending on the time interval. Break date estimates are obtained for sub-samples with end dates 2005Q1 to 2018Q2 for both methods; the LS estimates vary considerably, with $\hat{\rho}_{L S}$ near 0.1 and 0.9 for almost $40 \%$ of the sub-samples considered under M1. In contrast, our estimates are 1966Q1 or 1973Q1 for all sub-samples and models. This suggests that the difference in LS estimates depending on the sample period is due to its finite sample behavior (tri-modality) rather than evidence of multiple structural breaks.

### 1.6.2 Stock return prediction models

Paye and Timmermann (2006) studies the instability in models of ex-post predictable components in stock returns by examining structural breaks in the coefficients of state variables.

Table 1.6: Structural break date estimates of postwar U.S. real GDP growth rate in a $\operatorname{AR}(1)$ model using 54 sub-samples. The entries are the fraction of the number of sub-samples that has break date estimates corresponding to the first column. The second column is the interval of break point estimates that depend on the sub-sample size. The start date is 1947Q3 and the end dates of sub-samples change across 2005Q1 to 2018Q2.

|  |  | M1 |  | M2 |  | M3 |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Break date | $\hat{\rho}$ interval | NEW | LS | NEW | LS | NEW |  |
| LS |  |  |  |  |  |  |  |
| 1953Q1 | 0.10 |  | 0.17 |  |  |  |  |
| 1958Q1 | $[0.17,0.19]$ |  |  |  |  |  |  |
| 1966Q1 | $[0.26,0.33]$ |  | 0.06 | 1 | 1 | 0.30 |  |
| 1973Q1 | $[0.36,0.45]$ | 1 | 0.05 |  |  | 0.70 |  |
| 2000Q2 | $[0.74,0.86]$ |  | 0.52 |  |  |  |  |
| 2006Q1 | $[0.85,0.90]$ |  | 0.20 |  |  |  |  |

The regression model (1.27) is specified with four state variables: the lagged dividend yield, short interest rate, term spread and default premium. The model allows for all coefficients to change since there is no strong reason to believe that the coefficient on any of the regressors should be immune from shifts. The multivariate model with a one-time structural break at $k$ with $t=1, \ldots, T$ is

$$
\begin{align*}
\text { Ret }_{t}= & \beta_{0}+\beta_{1} \text { Div }_{t-1}+\beta_{2} \text { Tbill }_{t-1}+\beta_{3} \text { Spread }_{t-1}+\beta_{4} \text { Def }_{t-1}  \tag{1.27}\\
& +\mathbf{1}\{t>k\}\left(\delta_{0}+\delta_{1} \text { Div }_{t-1}+\delta_{2} \text { Tbill }_{t-1}+\delta_{3} \text { Spread }_{t-1}+\delta_{4} \text { Def }_{t-1}\right)+\varepsilon_{t},
\end{align*}
$$

where Ret $_{t}$ represents the excess return for the international index in question during month $t, D i v_{t-1}$ is the lagged dividend yield, Tbill $_{t-1}$ is the lagged local country short interest rate, Spread $_{t-1}$ is the lagged local country spread and $D e f_{t-1}$ is the lagged US default premium. From the notation of model (1.8), $y_{t}=\operatorname{Ret}_{t}$ and for the multivariate model, $x_{t}=z_{t}=\left(1\right.$, Div $_{t-1}$, Tbill $_{t-1}$, $\left.\operatorname{Spread}_{t-1}, \operatorname{Def}_{t-1}\right)$. For the univariate model with dividend yield $x_{t}=z_{t}=\left(1, \operatorname{Div}_{t-1}\right)$, which is defined analogously for other univariate models. The weight matrix is $\Omega_{k}=T^{-1} Z_{k}^{\prime} M Z_{k}$ where $Z_{k}=\left(0, \ldots, 0, z_{k+1}, \ldots, z_{T}\right)^{\prime}$ and $M=I-X\left(X^{\prime} X\right)^{-1} X^{\prime}$. Following the approach of Paye and

Timmermann (2006), I examine univariate models to facilitate interpretation of coefficients, in addition to the multivariate model (1.27).

Under the univariate model with the lagged dividend yield as a single forecasting regressor, the LS estimate of the break point for the S\&P 500 is close to the end date of the trimmed sample. Paye and Timmermann notes that the NYSE or S\&P 500 indices have the same estimated break date when the trimming window is shortened, and thus the discrepancy is not the sole explanation of the timing of the break. However, it is likely that estimated break dates of the return prediction model are near the end of the sample due to the finite sample behavior of the LS estimator. I check whether the new estimator provides a different break date estimate of the model (1.27) using data similar to the first dataset of Paye and Timmermann, which is monthly data on the U.S. and the UK stock returns from 1952:7 to 2003:12. The trimming window is also identical to fraction 15\%, thus potential break date starts from February, 1960 and ends at March, 1996. For comparison I also estimate the break using a shorter period 1970:1-2003:12, which is equivalent to the sample period of their second dataset. For each model and sample the $95 \%$ confidence interval is obtained by wild bootstrap (Liu, 1988) with the number of bootstrap samples set to 499.

Data are collected from Global Financial Data and Federal Reserve Economic Data (FRED). The indices to which the total return and dividend yield series are the S\&P 500 for the U.S. and the FTSE All-share for the UK. The dividend yields is expressed as an annual rate and is constructed as the sum of dividends over the preceding 12 months, divided by the current price. For both countries, a 3-month Treasury bill (T-bill) rate is used as a measure of the short interest rate and the 20-year government bond yield is the measure of the long interest rate. Excess returns are computed as the total return on stocks in the local currency less the total return on T-bills. The term spread is constructed as the difference between the long and the short local country interest rate. U.S. default premium is defined as the differences in yields between Moody's Baa and Aaa rated bonds. For each sample period the break date is obtained from a grid that is trimmed by
fraction $\alpha=0.15$. For the full sample the grid is 1960:2-1996:3, and for the sub-sample it is 1975:1-1998:10.

Table 1.7 provides the estimates of the two samples using the $\mathrm{S} \& \mathrm{P} 500$ index. One notable feature is that the LS estimates a break occurred in December 1994, with break point $\hat{\rho}_{L S}=0.85$ whereas the new method estimates a break in the mid 1980s and $\hat{\rho}_{N E W}=0.62$. Although the LS estimate is close to the end of trimmed sample, it gives the same break estimate in the subsample starting at 1970. This suggests that a break may have occurred multiple times. Paye and Timmermann uses the method of Bai and Perron (1998) and find that two structural breaks occur in the return model (1.27) using S\&P 500, where each break occurs at 1987:7 and 1995:3. They note that the break at 1987 appears to be an isolated break not appearing in other international markets. These two break date estimates are similar to estimates in Table 1.7 that assume a one-time structural break.

Another explanation of the break in the early 1980s is that our estimation method captures a change in the individual state variable itself rather than the coefficient of the prediction model (1.27), because it is extremely difficult to detect a break due to the noisy nature of stock market returns. For instance, the estimate $\hat{\beta}_{1}$ could be capturing noise caused by the movement in $\operatorname{Div} v_{t-1}$. Figure 1.10 plots the two state variables, U.S. dividend yield and term spread. Both series have a change in its trend in the early 1980s. If we compare the two break date estimates from the univariate model in Table 1.7, our estimate is closer to the date of the change in trend occurred.

For UK stock returns, both estimation methods obtain break date estimate that is (or close to) 1975:1 under all models and sample periods. In addition, the LS estimator has a slightly shorter length of the $95 \%$ confidence interval compared to our estimator for all models except the term spread univariate model. This is different from the result using S\&P 500 index series because the excess return for the FTSE All-share index increases near 10 standard deviations from 1975:1 to 1975:2 ${ }^{9}$. Hence, the change in excess returns is large enough for the LS to detect

[^6]the break point appropriately. Paye and Timmermann relates the break in mid 1970s to the large macroeconomic shocks reflecting large oil price increases; the breaks in the underlying economic fundamentals process can explain breaks in financial return models. If this is the case, then the break magnitude is large enough so that both methods accurately estimates the break date 1975:1.

Table 1.7: Structural break date estimates of the U.S. stock return (S\&P 500) prediction model for samples 1952:7-2003:12, and 1970:1-2003:12. For each model the first row is the break date estimate and the third row is the break point estimate (fraction within corresponding sample). The second and fourth rows are $95 \%$ bootstrap confidence intervals obtained by 499 replications.

|  | $1952: 7-2003: 12$ |  | 1970:1-2003:12 |  |
| :--- | :---: | :---: | :---: | :---: |
| Model | NEW | LS | NEW | LS |
|  | $1984: 8$ | $1994: 12$ | $1982: 8$ | $1994: 12$ |
| Multi. | $[77: 4,88: 9]$ | $[83: 5,95: 1]$ | $[79: 5,88: 3]$ | $[75: 5,95: 1]$ |
|  | 0.62 | 0.83 | 0.37 | 0.74 |
|  | $[0.48,0.70]$ | $[0.60,0.83]$ | $[0.28,0.54]$ | $[0.16,0.74]$ |
|  | $1982: 8$ | $1995: 1$ | $1982: 8$ | $1996: 9$ |
| Div. yield | $[68: 9,84: 8]$ | $[74: 11,95: 2]$ | $[76: 12,84: 3]$ | $[75: 1,98: 9]$ |
|  | 0.59 | 0.83 | 0.37 | 0.79 |
|  | $[0.32,0.62]$ | $[0.44,0.83]$ | $[0.21,0.42]$ | $[0.15,0.85]$ |
|  | $1974: 10$ | $1974: 10$ | $1982: 8$ | $1975: 1$ |
|  | $[74: 7,81: 11]$ | $[70: 7,81: 1]$ | $[77: 6,88: 6]$ | $[75: 1,90: 7]$ |
| T-bill | 0.43 | 0.43 | 0.37 | 0.15 |
|  | $[0.43,0.57]$ | $[0.35,0.56]$ | $[0.22,0.55]$ | $[0.15,0.61]$ |
|  | $1983: 5$ | $1976: 2$ | $1987: 9$ | $1976: 2$ |
|  | $[80: 9,92: 4]$ | $[72: 6,83: 7]$ | $[82: 7,92: 9]$ | $[75: 7,86: 10]$ |
| Spread | 0.60 | 0.46 | 0.52 | 0.18 |
|  | $[0.55,0.77]$ | $[0.39,0.60]$ | $[0.37,0.67]$ | $[0.16,0.50]$ |
|  | $1968: 12$ | $1965: 11$ | $1982: 8$ | $1975: 7$ |
|  | $[63: 1,92: 2]$ | $[62: 4,95: 12]$ | $[78: 1,91: 12]$ | $[75: 3,86: 9]$ |
| Def.prem. | 0.32 | 0.26 | 0.37 | 0.16 |
|  | $[0.21,0.77]$ | $[0.19,0.84]$ | $[0.24,0.65]$ | $[0.15,0.49]$ |

change is approximately 9.6 standard deviations.

Table 1.8: Structural break date estimates of the UK (FTSE) stock return prediction model for samples 1952:7-2003:12 and 1970:1-2003:12. For each model the first row is the break date estimate and the third row is the break point estimate (fraction within corresponding sample). The second and fourth rows are $95 \%$ bootstrap confidence intervals with 499 replications.

|  | 1952:7-2003:12 |  | 1970:1-2003:12 |  |
| :--- | :---: | :---: | :---: | :---: |
| Model | NEW | LS | NEW | LS |
|  | $1975: 1$ | $1975: 1$ | $1975: 1$ | $1975: 1$ |
| Multi. | $[77: 4,88: 9]$ | $[83: 5,95: 1]$ | $[75: 1,88: 10]$ | $[75: 1,76: 11]$ |
|  | 0.44 | 0.44 | 0.15 | 0.15 |
|  | $[0.44,0.74]$ | $[0.44,0.45]$ | $[0.15,0.55]$ | $[0.15,0.20]$ |
|  | $1975: 1$ | $1975: 1$ | $1975: 1$ | $1975: 1$ |
| Div. yield | $[67: 3,76: 11]$ | $[75: 1,77: 12]$ | $[75: 1,77: 1]$ | $[75: 1,75: 9]$ |
|  | 0.44 | 0.44 | 0.15 | 0.15 |
|  | $[0.29,0.47]$ | $[0.44,0.50]$ | $[0.15,0.21]$ | $[0.15,0.17]$ |
|  | $1975: 1$ | $1974: 12$ | $1975: 1$ | $1975: 1$ |
|  | $[74: 10,81: 11]$ | $[73: 11,77: 1]$ | $[73: 1,81: 12]$ | $[73: 11,76: 10]$ |
| T-bill | 0.44 | 0.29 | 0.15 | 0.15 |
|  | $[0.43,0.57]$ | $[0.42,0.48]$ | $[0.15,0.35]$ | $[0.15,0.20]$ |
|  | $1975: 1$ | $1975: 6$ | $1975: 1$ | $1975: 3$ |
|  | $[73: 1,78: 2]$ | $[68: 2,89: 12]$ | $[75: 1,77: 12]$ | $[75: 1,78: 2]$ |
| Spread | 0.44 | 0.45 | 0.15 | 0.15 |
|  | $[0.40,0.50]$ | $[0.31,0.73]$ | $[0.15,0.24]$ | $[0.15,0.24]$ |
|  | $1979: 5$ | $1975: 6$ | $1975: 1$ | $1975: 3$ |
|  | $[69: 2,90: 4]$ | $[70: 11,80: 8]$ | $[75: 1,91: 4]$ | $[75: 1,76: 6]$ |
| Def.prem. | 0.52 | 0.45 | 0.52 | 0.15 |
|  | $[0.32,0.73]$ | $[0.36,0.55]$ | $[0.15,0.63]$ | $[0.15,0.19]$ |



Figure 1.10: U.S. dividend yield (left) and term spread (right), 1952:7-2003:12. Red and blue dotted lines are the new and LS break date estimates from the univariate model, respectively.

### 1.6.3 Oil price shock and output growth

Hamilton (2003) tests the existence of a structural break on the relation between real U.S. GDP growth rates and nonlinear transforms of oil price measures. I use the same model and data ${ }^{10}$, but assume that the nonlinear relation is unstable. That is, I assume there is a structural break in the nonlinear oil price transforms and GDP growth rate and estimate the structural break date using the new estimation method. Let $y_{t}$ denote the real output, which is the quarterly growth rate of chain-weighted real U.S. GDP. The oil price series is denoted as $o_{t}$, which is 100 times the quarterly logarithmic growth rate of nominal crude oil producer price index, seasonally unadjusted. The sample used for estimation starts from 1949 Q 2 to 2001 Q 3 , for a total of $T=210$ observations, not including the lagged initial values for conditioning.

Four different oil price measures are used, the percentage change rate in nominal crude oil prices and three measures that are nonlinear transforms of the percentage change rates of the oil price. The formulation of Mork (1989) uses the positive values $o_{t}^{+}$, where $o_{t}^{+}=o_{t}$ if $o_{t}>0$ and 0 if $o_{t} \leq 0$. The annual net oil price is the amount by which the oil price in quarter $t$ exceeds its peak value over the previous 12 months, and the 3-year net oil price is defined analogously for horizon 3 years instead of one year. For details on the oil price measures see Hamilton

[^7](2003). Assume model (1.8) with lags of the oil price measure as $z_{t}=\left(o_{t-1}, o_{t-2}, o_{t-3}, o_{t-4}\right)^{\prime}$ and $x_{t}=\left(1, y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, z_{t}\right)^{\prime}$. The weight matrix is $\Omega_{k}=\omega_{k}^{2} I_{4}$ where $\omega_{k}=(k / T(1-k / T))^{1 / 2}$ and $I_{4}$ is a ( $4 \times 4$ ) identity matrix. The search grid of $k$ is trimmed by fraction $\alpha=0.15$ on both ends.

Table 1.9: Structural break point estimates of the relation between oil price change and the U.S. real GDP growth rates, 1949Q2-2001Q3. The oil price measures are percentage change rate of nominal crude oil price, the positive percentage change rate of nominal crude oil price (Mork), the annual net oil price increase, and the 3 -year net oil price increase, respectively. For each model the first row is the break date estimate and the third row is the break point estimate $\hat{\rho}$. The second and fourth rows are $95 \%$ bootstrap confidence intervals with 999 replications.

| Oil price measure | NEW | LS |
| :--- | :---: | :---: |
|  | 1982 Q 1 | 1985 Q 3 |
| \% rate | $[74 \mathrm{Q} 2,87 \mathrm{Q} 1]$ | $[74 \mathrm{Q} 1,89 \mathrm{Q} 2]$ |
|  | 0.64 | 0.70 |
|  | $[0.49,0.73]$ | $[0.49,0.77]$ |
|  | 1982 Q 1 | 1991 Q 1 |
| Mork | $[71 \mathrm{Q} 1,89 \mathrm{Q} 3]$ | $[70 \mathrm{Q} 2,93 \mathrm{Q} 2]$ |
|  | 0.64 | 0.80 |
|  | $[0.43,0.78]$ | $[0.42,0.85]$ |
|  | 1970 Q 4 | 1990 Q 4 |
| Net 1-yr | $[67 \mathrm{Q} 3,82 \mathrm{Q} 1]$ | $[70 \mathrm{Q} 2,93 \mathrm{Q} 2]$ |
|  | 0.43 | 0.80 |
|  | $[0.36,0.64]$ | $[0.42,0.85]$ |
|  | 1970 Q 4 | 1970 Q 2 |
| Net 3-yr | $[67 \mathrm{Q} 3,82 \mathrm{Q} 1]$ | $[55 \mathrm{Q} 4,91 \mathrm{Q} 3]$ |
|  | 0.43 | 0.42 |
|  | $[0.36,0.64]$ | $[0.15,0.81]$ |

Table 1.9 shows the break location estimates for the full sample 1949Q2-2001Q3 (excluding lagged initial values for conditioning). Our method estimates are either 1982Q1 or 1970Q4, whereas the LS estimates vary across 1970, 1985 and early 1990s. The $95 \%$ confidence interval is obtained from residual based bootstrap with number of bootstrap samples set to 999 . For all models our method has a shorter interval length compared to LS. Because the LS estimates vary
substantially depending on the start date of the sample, we obtain break date estimates for various sub-samples. The start date of the sub-sample changes across 1949Q3 to 1957Q2 and 1961Q3 to 1965Q4 ${ }^{11}$. All sub-samples end at 2001Q3 and the estimates from the full sample are included in the search grid trimmed by fraction $\alpha=0.15$. Total 50 sub-samples are considered. Table 1.10 provides the fraction of sub-samples (the number of sub-samples divided by 50) that gives the same break date estimates.

From 1949Q2 to 2001Q3, there are five military conflicts in the Middle East that have significantly disrupted world petroleum supplies (Hamilton, 2003). The Suez crisis in November 1956, the Arab-Israel war in November 1973, the Iraninan revolution in November 1978, the IranIraq war in October 1980 and the Persian Gulf war in August 1990, where the month indicates the largest observed drop in oil production. The events are in the first column of Table 1.10 next to the corresponding break date in the second column. If an oil supply shock occurred in the quarter, it is categorized with the next four quarters due to the number of lags in the regression. For the Arab-Israel war there were no break date estimates equivalent to 1973Q4, so it is with 1974Q1-1974Q4. There are no break date estimates close to the earliest event at 1956Q4, the Suez crisis, for all oil price measures and estimation methods. For a large oil supply shock at quarter $t$, the annual net and the 3-year net oil price measures would be affected up to quarter $t+4$ and $t+12$, respectively. Thus, for a event occurring at $t$, the break date estimates are categorized for $t+5$ to $t+16$ due to the construction of the net oil price measures and four lags in the regression.

For all oil price measures, the LS estimator has a higher fraction of sub-samples with estimates that are near the start date of the trim, which corresponds to the first row. However both estimation methods did not have sub-sample estimates near the end date of the trim; the latest estimate date is 1991Q2. Our estimator has a higher fraction of sub-samples with estimates within the following 16 quarters from the Iraninan revolution and the Iran-Iraq war, for all oil price measure. For the Arab-Israel war, our estimator using oil price measures except the annual

[^8]net oil price also has a higher fraction of sub-samples with estimates in the following 16 quarters. In this period the difference in the number of sub-samples is only one between the two estimation methods using the 3-year net oil price.

Table 1.11 summarizes the fraction of sub-samples with break point estimates closer or further away from the end of the unit interval. For all measures there are sub-samples with $\hat{\rho}_{L S} \in[0.15,0.25)$, which corresponds to dates earlier than the Arab-Israel war, but almost ten years past the Suez crisis in 1956. In contrast, the break point estimator $\hat{\rho}_{N E W}$ does not have any sub-samples in $[0.15,0.25)$ except one sub-sample using the 3-year net oil price measure. This also holds for the interval close to 0.85 ; for all oil price measures none of the sub-samples have $\hat{\rho}_{N E W}$ in $[0.75,0.85]$ whereas the LS estimator has one or three sub-samples with $\hat{\rho}_{L S}$ included in the interval. Although this interval include dates following the Persian Gulf war, the fraction 0.02 and 0.06 of sub-samples seems quite small to think that the LS estimator actually estimates the break date of this event. Rather, for these sub-samples the finite sample behavior of the LS estimator leads to estimates near the ends.

Kilian (2008) argues that nonlinear transforms of oil price data do not identify the exogenous component of oil price changes. Hence, I also find the structural break point using the measure of exogenous oil supply shocks constructed in Kilian (2008). A quarterly measure of the OPEC oil production shock series ${ }^{12}$ from 1971Q1 to 2004Q3 is employed as $o_{t}$. The regression model and weight matrix is the same as before; four lags of oil shock measures are associated with coefficients under structural break ${ }^{13}$. Both methods estimates a break occurred at 1980Q3, corresponding to break point of 0.27 .

Overall, the difference between our estimator and the LS estimator of the break point is less prominent than the applications in section 1.6.1 and 1.6.2 but similar to previous results: our

[^9]estimator is robust to trimming of the sample in estimating the structural break of the relation between oil price shock and output growth. Furthermore, in obtaining estimates across different trimmed samples, our estimator has a higher number of sub-samples that estimates the three exogenous events that caused a drop in total world crude oil production: the Arab-Israel war in 1973, the Iranian revolution in 1978, and the Iran-Iraq war in 1980. Although a more rigorous hypothesis test is required to conclude that the new estimation method accurately estimates a break caused by these military conflicts, the results suggest that the relation of nonlinear oil price measures and output growth in not stable. This is different from the results of Hamilton (2003) in that a stable nonlinear relation between oil price and output can be represented using positive changes of oil price. That is, even when nonlinear transforms of measures are used, it is likely that a structural break occurs from a oil price shock.

Table 1.10: Structural break date estimates of the relation between oil price change and the U.S. real GDP growth rates from 50 sub-samples. The start date changes across 1948Q3 to 1956Q2; 1960Q3 to 1964Q4 and the end date is 2001Q3 for all sub-samples. The entries are the fraction of the number of sub-samples with break date estimates in the first column.

| Event | Break date | Oil price measure |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | \% rate |  | Mork |  | Net 1-yr |  | Net 3-yr |  |
|  |  | NEW | LS | NEW | LS | NEW | LS | NEW | LS |
| Arab-Israel | 65Q1-66Q4 | 0.02 | 0.08 | 0.02 | 0.06 |  | 0.02 | 0.02 | 0.02 |
|  | 67Q1-73Q3 | 0.10 | 0.14 | 0.08 | 0.08 | 0.16 | 0.16 | 0.14 | 0.18 |
|  | 74Q1-74Q4 | 0.06 | 0.08 | 0.12 | 0.08 | 0.12 | 0.12 | 0.12 | 0.08 |
|  | 75Q2-77Q4 | 0.24 | 0.20 | 0.24 | 0.24 | 0.22 | 0.24 | 0.18 | 0.20 |
|  | 78Q1-78Q3 | 0.08 | 0.06 | 0.04 | 0.04 | 0.12 | 0.12 | 0.06 | 0.06 |
| Iranian | 78Q4-79Q4 | 0.12 | 0.10 | 0.04 | 0.04 | 0.08 | 0.04 | 0.14 | 0.06 |
|  | 80Q1-80Q3 | 0.02 |  | 0.06 | 0.06 | 0.12 | 0.08 | 0.10 | 0.08 |
| Iran-Iraq | 80Q4-81Q4 | 0.16 | 0.12 | 0.16 | 0.14 | 0.10 | 0.08 | 0.12 | 0.10 |
|  | 82Q1-84Q4 | 0.10 | 0.10 | 0.10 | 0.10 |  |  | 0.06 | 0.04 |
|  | 85Q1-87Q3 | 0.08 | 0.10 | 0.10 | 0.10 | 0.02 | 0.06 |  |  |
|  | 88Q1-90Q2 | 0.02 |  | 0.04 | 0.04 |  |  | 0.06 | 0.10 |
| Persian Gulf | 90Q3-91Q2 |  | 0.02 |  | 0.02 | 0.02 | 0.04 |  | 0.08 |

Table 1.11: Structural break point estimates of the relation between oil price change and the U.S. real GDP growth rates from 50 sub-samples. The start date changes across 1948Q3 to 1956Q2; 1960Q3 to 1964Q4 and the end date is 2001Q3 for all sub-samples. The entries are the fraction of the number of sub-samples with break point estimates included in the first column interval.

|  | Oil price measure |  |  |  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | \% rate |  | Mork |  | Net 1-yr | Net 3-yr |  |  |
| $\hat{\rho}$ interval | NEW | LS | NEW | LS | NEW | LS | NEW | LS |
| $[0.15,0.25)$ | 0 | 0.12 | 0 | 0.04 | 0 | 0.02 | 0.02 | 0.02 |
| $[0.25,0.35)$ | 0.18 | 0.20 | 0.16 | 0.16 | 0.18 | 0.24 | 0.06 | 0.14 |
| $[0.35,0.45)$ | 0.24 | 0.18 | 0.24 | 0.20 | 0.44 | 0.36 | 0.42 | 0.32 |
| $[0.45,0.55)$ | 0.22 | 0.18 | 0.22 | 0.22 | 0.12 | 0.12 | 0.24 | 0.20 |
| $[0.55,0.65)$ | 0.28 | 0.18 | 0.28 | 0.26 | 0.18 | 0.12 | 0.20 | 0.14 |
| $[0.65,0.75)$ | 0.08 | 0.12 | 0.10 | 0.10 | 0.08 | 0.08 | 0.06 | 0.12 |
| $[0.75,0.85]$ | 0 | 0.02 | 0 | 0.02 | 0 | 0.06 | 0 | 0.06 |

### 1.7 Conclusion

This paper provides a estimation method of the structural break point in multivariate linear regression models, when a one-time break occurs in a subset of (or all) coefficients. In particular, this paper focuses on break magnitudes that are empirically relevant. That is, in practice it is likely that the shift in parameters are small in a statistical sense. The least-squares estimation widely used in literature fails to accurately estimate the break point under small break magnitudes, which motivates the construction of the estimation method in this paper.

I show that the functional form of the objective function leads to tri-modality of the finite sample distribution of the LS estimator. A weight function is constructed on the sample period normalized to the unit interval, which assigns small weights on the LS objective for potential break points with large estimation uncertainty. The break point estimator is the argmax functional of the objective that is equivalent to the LS objective function multiplied by weights. The weight function is proportional to the Fisher information under a Gaussian assumption on the model. The Fisher information reflects a belief that a structural break is less likely to occur near ends of
the sample period. I show that the break point estimator is asymptotically equivalent to the mode of the Bayesian posterior distribution if we use a prior that depends on the Fisher information.

The break point estimator is consistent under regularity conditions on a general weight function, with the same rate of convergence as the LS estimator from Bai (1997). The limit distribution under a small break magnitude is derived under a in-fill asymptotic framework, following the approach of Jiang, Wang, and Yu (2017, 2018). For a structural break in a stationary linear process with a break magnitude that is inside the local $T^{-1 / 2}$ neighborhood of zero, the asymptotic distribution of the new estimator explicitly depends on the weight function. However, if the break magnitude is outside the local $T^{-1 / 2}$ neighborhood of zero, the limit distribution of the estimator is equivalent to that of the LS estimator. The in-fill asymptotic distribution is also derived for a break in a local-to-unit root process, assuming the break magnitude is $O\left(T^{-1}\right)$. This is smaller than the break sizes assumed in conventional long-span asymptotic theory. Monte Carlo simulation results show that under a small break the break point estimator reduces the RMSE compared to the LS estimator for all parameter values considered.

The paper provides three empirical applications: structural breaks on the U.S. real GDP growth, the U.S. and the UK stock return prediction models, the relation between oil price shocks and the U.S. output growth. The break point estimates are robust to trimming of the sample, in contrast to LS estimates. In particular, our method estimates the break date 1973Q1 in U.S. real GDP growth rates, which LS estimation has failed to confirm. In macroeconomics literature the "productivity growth slowdown" in early 1970s is a widely known empirical fact.

In short, this paper provides a estimation method that accurately estimates the timing of the structural break in linear regression models, under empirically relevant break sizes. The new estimator resolves the tri-modality issue of the least-squares estimation in finite sample. To my knowledge it is the first in the literature for a break point estimator to have a uni-modal finite sample distribution under statistically small break magnitudes. The paper provides theoretical results of consistency of the estimator and a asymptotic distribution that represents finite sample
behavior. If the break magnitude is small, the new estimator outperforms the least-squares estimation and if the magnitude is large, it becomes similar to the LS estimator. Thus, the estimator of this paper provides reliable inference of the change point in models, and does not perform worse than LS estimator uniformly. The estimation method can be generalized to estimate multiple structural breaks, which is for future research.

Chapter 1 , in full, is currently being prepared for submission for publication of the material. Baek, Yaein. The dissertation author was the sole author of this paper.

## Chapter 2

## Tests for Break in Coefficients in Linear

## Regression when the Direction of the Break

## is Known

### 2.1 Introduction

One of the main concerns in modelling parametric models that represent econometric relationships is that parameters are likely to be unstable. In time series regression we refer the change in parameters over time as structural breaks in the model. In statistics and econometrics literature there is an extensive amount of work on testing for structural breaks. Earlier works such as Chow (1960) tests for parameter stability under a known break date using a $F$-statistic, and Quandt $(1958,1960)$ suggests using a maximum $F$-statistic over all values of the potential break date when the break date is unknown. Andrews (1993) studies the properties of such tests and derives the asymptotic distribution of the test statistic. Brown, Durbin, and Evans (1978) provides a test on the stability of parameters by considering partial sums of the standardized forecast errors of rolling regressions, referred as the CUSUM test. Ploberger, Krämer, and Kontrus (1989) and

Ploberger and Krämer (1992) provides test that are functions of the partial sums of OLS residuals from a regression ignoring the breaks. Papers that discuss optimality of tests are Nyblom (1989), which provides small sample locally optimal tests that are valid for a single structural break model. Andrews and Ploberger (1994) derive asymptotically optimal tests by employing an weighted average power criterion function.

Previous literature on structural break tests, including all tests stated above, studies a twosided test where the null hypothesis sets the parameter change equal to zero and the alternative hypothesis states it is nonzero. That is, existing methods tests for breaks of either direction of a parameter change. However in practice there are instances where the researcher is interested in testing for breaks of a particular direction. For example, in Lucas critique type problems the direction of the policy change may be known and hence the direction of the subsequent effect of the reduced form regressions are also known. Or the direction may be known based on historical data of relevant series. Through not directing power towards uninteresting alternatives, the power of a test aimed in the correct direction should improve power. To the best of our knowledge, tests for a known direction of break has not been discussed in literature.

In this paper we provide three tests where the direction of the break, i.e., the sign of the parameter change in a linear regression model is known. By incorporating the information of known directions, the test statistics are directed toward the alternative which leads to increase in power. The first two tests have optimality properties under restrictive assumptions. The third test, although lacking optimality properties, has an asymptotic null distribution equivalent to a standard normal distribution and thus simple to calculate. We refer these test with known direction of breaks as "directional break tests". The directional break tests also accommodate cases where the variance of the error term changes simultaneously with coefficients.

We compare the three directional break tests with other structural break tests widely used in literature, the Nyblom (1989) test, the average exponential Lagrange multiplier (LM) test from Andrews and Ploberger (1994), and the $q L L$ test from Elliott and Müller (2006). Monte Carlo
simulations show that directional break tests with correct directions have higher power than these tests. Furthermore, our empirical application on testing for a break in postwar U.S. real GDP growth rate show that our tests suggest significant evidence of a break whereas other tests show insignificance or weak evidence.

The plan of the paper is as follows. Section 2.2 provides our basic model and derives the three directional break tests with their null asymptotic distribution. Section 2.3 discusses finite sample power of tests from Monte Carlo simulations. Section 2.4 applies directional break tests on testing for a break in U.S. real GDP growth rate series, U.S. labor productivity and compare results with other structural break tests. Concluding remarks are in Section 2.5.

### 2.2 Directional Break Tests

### 2.2.1 The model and test statistics

Consider a linear regression model where a structural break occurs at time $\tau_{0}$.

$$
\begin{equation*}
y_{t}=X_{t}^{\prime} \beta+X_{t}^{\prime} \boldsymbol{\delta}\left\{t \geq \tau_{0}\right\}+Z_{t}^{\prime} \gamma+u_{t}, \quad t=1, \ldots, T \tag{2.1}
\end{equation*}
$$

where $y_{t}$ is a scalar, $X_{t}, \beta, \delta$ are $(k \times 1)$ vectors, $Z_{t}$ and $\gamma$ are $(m \times 1)$ vectors, $u_{t}$ is a mean zero disturbance with long-run variance $\sigma^{2}$ that is possibly heterogeneous and autocorrelated. $\mathbf{1}\{t \geq$ $\left.\tau_{0}\right\}$ is an indicator function that equals one if $t \geq \tau_{0}$ and zero otherwise. The break magnitude $\delta$ and break date $\tau_{0}$ are unknown to the researcher but the direction of the break $\bar{\delta}:=\operatorname{sgn}\left(\delta_{a}\right)$ is known; $\bar{\delta}$ corresponds to the signs of the coefficient under the alternative hypothesis $\delta_{a}$ where each element is 1 or -1 . Denote $\delta_{j}$ and $\bar{\delta}_{j}$ as the $j$ th element of $\delta$ and $\bar{\delta}$, respectively.

$$
\begin{align*}
& \mathcal{H}_{0}: \bar{\delta}_{j} \delta_{j} \leq 0, \quad \forall j=1, \ldots, k  \tag{2.2}\\
& \mathcal{H}_{1}: \bar{\delta}_{j} \delta_{j}>0, \quad \exists j
\end{align*}
$$

Under the null hypothesis all elements of vector $\delta$ are either equal to zero (no structural break) or its direction is incorrectly specified. The alternative hypothesis indicates that at least one (nonzero) element of $\delta$ is correctly specified. For asymptotic results, we impose the following regularity condition on model (2.1). Let $Q_{t}=\left(X_{t}^{\prime}, Z_{t}^{\prime}\right)^{\prime}$ and $[\cdot]$ is the greatest smaller integer function.

Condition 1. (i) $\tau_{0}=\left[r_{0} T\right]$ for some $0<r_{0}<1$.
(ii) $T^{-1 / 2} \sum_{t=1}^{[s T]} X_{t} u_{t} \Rightarrow \Omega_{1}^{1 / 2} W(s)$ for $0 \leq s \leq r_{0}$ and $T^{-1 / 2} \sum_{t=\tau_{0}+1}^{[s T]} X_{t} u_{t} \Rightarrow \Omega_{2}^{1 / 2}\left(W(s)-W\left(r_{0}\right)\right.$ for $r_{0} \leq s \leq 1$ with $\Omega_{1}$ and $\Omega_{2}$ some symmetric positive definite $k \times k$ matrices and $W(\cdot)$ a $k \times 1$ standard Wiener process.
(iii) $\sup _{0 \leq s \leq 1}\left\|T^{-1 / 2} \sum_{t=1}^{[s T]} Z_{t} u_{t}\right\|=O_{p}(1)$.
(iv) $T^{-1} \sum_{t=1}^{[s T]} Q_{t} Q_{t}^{\prime} \rightarrow s \Sigma_{Q}=s\left(\begin{array}{cc}\Sigma_{X} & \Sigma_{X Z} \\ \Sigma_{Z X} & \Sigma_{Z}\end{array}\right)$ uniformly in $0 \leq s \leq 1$ where $\Sigma_{Q}$ is full rank.

These assumptions are standard in the literature on tests for structural breaks. Condition $1(i)$ assumes the break point $r_{0}$ is bounded away from end points. Conditions (ii)-(iv) are standard high-level time series conditions, that allow for heterogeneous and serially correlated $\left\{u_{t}\right\}$ and regressors $\left\{Q_{t}\right\}$. As in Bai (1997), we allow for the long-run variance of $\left\{X_{t} u_{t}\right\}$ to change at the break date $\tau_{0}$.

We denote the tests of break under known directions as "directional break tests". First, consider the case of no break in the variance of $\left\{X_{t} u_{t}\right\}$, such that $\Omega=\Omega_{1}=\Omega_{2}$. Denote $\hat{v}_{t}:=X_{t} \hat{u}_{t}$ where $\hat{u}_{t}$ are the OLS residuals from regressing $\left\{y_{t}\right\}$ on $\left\{Q_{t}\right\}$. If we relax the conditional homoskedasticity assumption such that the error term has a variance-covariance matrix $E\left[U U^{\prime}\right]=$ $\Omega_{u}$ that is not necessary diagonal, the directional test statistics are as follows. Note that $\bar{d}$ is a $(k \times 1)$ vector that needs to be specified for the first two test statistics. It represents the "normalized" local break magnitude of each element of $\delta$, which we further discuss in section
2.2.2.

$$
\begin{align*}
& d b_{T}^{a}=T^{-1} \sum_{t=1}^{T} \exp \left[\left(\frac{\bar{d}^{\prime} \bar{d}}{\bar{\delta}^{\prime} \hat{\Sigma}_{X}^{-1 / 2} \hat{\Omega} \hat{\Sigma}_{X}^{-1 / 2} \bar{\delta}}\right)^{1 / 2} \bar{\delta}^{\prime} \hat{\Sigma}_{X}^{-1 / 2}\left(T^{-1 / 2} \sum_{s=t+1}^{T} \hat{v}_{s}\right)-\frac{1}{2} \bar{d}^{\prime} \bar{d}\left(\frac{t}{T}\right)\left(1-\frac{t}{T}\right)\right]  \tag{2.3}\\
& d b_{T}^{b}=T^{-1} \sum_{t=1}^{T} \exp \left[\left(\frac{\bar{d}^{\prime} \bar{d}}{\bar{\delta}^{\prime} \hat{\Sigma}_{X}^{-1 / 2} \hat{\Omega} \hat{\Sigma}_{X}^{-1 / 2} \bar{\delta}}\right)^{1 / 2} \bar{\delta}^{\prime} \hat{\Sigma}_{X}^{-1 / 2}\left(T^{-1 / 2} \sum_{s=t+1}^{T} \hat{v}_{s}\right)\right]  \tag{2.4}\\
& d b_{T}^{c}=\left(\frac{12}{\bar{\delta}^{\prime} \hat{\Sigma}_{X}^{-1 / 2} \hat{\Omega} \hat{\Sigma}_{X}^{-1 / 2} \bar{\delta}}\right)^{1 / 2} T^{-1} \sum_{t=1}^{T} \bar{\delta}^{\prime} \hat{\Sigma}_{X}^{-1 / 2}\left(T^{-1 / 2} \sum_{s=t+1}^{T} \hat{v}_{s}\right) \tag{2.5}
\end{align*}
$$

where $\hat{\Omega}$ is the long-run variance estimator of $\left\{X_{t} u_{t}\right\}$ and $\hat{\Sigma}_{X}$ is a consistent estimator of the second moment of $X_{t}$. Note that under conditional homoskedasticity, $\hat{\Omega}=\hat{\sigma}^{2} \hat{\Sigma}_{X}$ and the denominator of the first term simplifies to $\left(\hat{\sigma}^{2} \bar{\delta}^{\prime} \bar{\delta}\right)^{1 / 2}=\hat{\sigma} \sqrt{k}$. Each test statistic depends on the known direction vector $\bar{\delta}$. The first two test statistics depend on the local break magnitude $\bar{d}$. In section 2.2.2 we show that these two test statistics are derived from a likelihood ratio test statistic and thus have optimality properties under i.i.d Gaussian disturbances and strictly exogenous regressors. In contrast, the third test statistic $d b_{T}^{c}$ do not depend on $\bar{d}$ and follow a standard normal distribution in the limit under no break. The construction of the test statistic is not based on a likelihood test statistic and hence do not share optimal properties. However, the limit is invariant of $\bar{d}$ which makes it easier to use the test statistic.

The tests that assume $\Omega_{1}=\Omega_{2}$ and allow for heteroskedastic and serially correlated errors $\left\{u_{t}\right\}$ are constructed as follows. Choose $\bar{\delta}$, where each element $\overline{\boldsymbol{\delta}}_{j}$ corresponds to the sign of the alternative we are testing for.

1. Compute the least squares regression of $\left\{y_{t}\right\}_{t=1}^{T}$ on $\left\{X_{t}, Z_{t}\right\}_{t=1}^{T}$ and compute the OLS residuals $\hat{u}_{t}$.
2. Construct $\left\{\hat{v}_{t}\right\}_{t=1}^{T}=\left\{X_{t} \hat{u}_{t}\right\}_{t=1}^{T}$.
3. Compute the long-run variance estimator $\hat{\Omega}$ of $\{\hat{\nu}\}_{t=1}^{T}$. An example would be to use the estimator of Newey and West $(1987,1994)$.
4. Compute test statistics as in (2.3), (2.4) or (2.5).
5. All tests reject for large values regardless of the choice of $\bar{\delta}$. Critical values for (2.3) and (2.4) are in Table 2.1 for $\bar{d} \in\{1.5,2,2.5\}$.

Now suppose there is a break in the variance of $X_{t} u_{t}$. We assume conditional homoskedasticity and use a weighted least square (WLS) approach to construct test statistics. Let $\Omega_{1}=\sigma_{1}^{2} \Sigma_{X}$ and $\Omega_{2}=\sigma_{2}^{2} \Sigma_{X}$ in Condition $1(i i)$, where $\sigma_{1} \neq \sigma_{2}$. In matrix form the model (2.1) is $Y=Q \theta+\left(X-X\left(\tau_{0}\right)\right) \delta+U$, where $Q=(X, Z), X\left(\tau_{0}\right)=\left(X_{1}, \ldots, X_{\tau_{0}}, 0 \ldots, 0\right)^{\prime}$ and $\theta=\left(\beta^{\prime}, \gamma^{\prime}\right)^{\prime}$. Let $\Omega_{U, t}:=\operatorname{diag}\left\{\sigma_{1}^{2} I_{t}, \sigma_{2}^{2} I_{T-t}\right\}$ and transform (2.1) by pre-multiplying $\Omega_{U, t}^{-1 / 2}$.

$$
\begin{aligned}
\Omega_{U, t}^{-1 / 2} Y= & \Omega_{U, t}^{-1 / 2} Q \theta+\Omega_{U, t}^{-1 / 2}\left(X-X\left(\tau_{0}\right)\right) \delta+\Omega_{U, t}^{-1 / 2} U \\
& \Leftrightarrow \tilde{Y}=\tilde{Q} \theta+\left(\tilde{X}-\tilde{X}\left(\tau_{0}\right)\right) \delta+\tilde{U}
\end{aligned}
$$

For a fixed $t$, Condition $1(i i)$ and (iv) holds for the transformed model.

$$
\begin{gathered}
T^{-1 / 2} \sum_{j=1}^{[s T]} \tilde{X}_{j} \tilde{u}_{j} \Rightarrow \Sigma_{X, l}^{1 / 2} W(s) \text { for } 0 \leq s \leq 1, \\
T^{-1} \sum_{j=1}^{[s T]} \tilde{Q}_{j} \tilde{Q}_{j}^{\prime} \rightarrow s \Sigma_{Q, l}=s\left(\begin{array}{cc}
\Sigma_{X, l} & \Sigma_{X Z, l} \\
\Sigma_{Z X, l} & \Sigma_{Z, l}
\end{array}\right)
\end{gathered}
$$

Note that subscript $l=t / T$ of $\Sigma_{Q, l}$ denotes its dependence on $\Omega_{U, t}$. Let $\tilde{v}_{j}$ be the OLS residuals from regressing $\tilde{Y}$ on $\tilde{Q}$. We have $T^{-1} \tilde{X}^{\prime} \tilde{X} \xrightarrow{p} \Sigma_{X, l}$, which is well-defined for all $l \in[\varepsilon, 1-\varepsilon]$ for some fraction $\varepsilon>0$. Denote $\hat{\Sigma}_{X, t}$ as a consistent estimator of $\Sigma_{X, l}$. Then $\hat{\Sigma}_{X, t}=T^{-1} X^{\prime} \hat{\Omega}_{U, t} X$, which is obtained from $\hat{\Omega}_{U, t}=\operatorname{diag}\left\{\hat{\sigma}_{1}^{2} I_{t}, \hat{\sigma}_{2}^{2} I_{T-t}\right\}$ where $\hat{\sigma}_{1}^{2}=\sum_{j=1}^{t} \hat{u}_{t}^{2}$ and $\hat{\sigma}_{2}^{2}=\sum_{j=t+1}^{T} \hat{u}_{t}^{2}$. The
three directional break test statistics using WLS estimators are as follows.

$$
\begin{align*}
& d b_{T}^{a}=T^{-1} \sum_{t=1}^{T} \exp \left[\left(\frac{\bar{d}^{\prime} \bar{d}}{\bar{\delta}^{\prime} \bar{\delta}}\right)^{1 / 2} \bar{\delta}^{\prime} \hat{\Sigma}_{X, t}^{-1 / 2}\left(T^{-1 / 2} \sum_{s=t+1}^{T} \tilde{v}_{s}\right)-\frac{1}{2} \bar{d}^{\prime} \bar{d}\left(\frac{t}{T}\right)\left(1-\frac{t}{T}\right)\right]  \tag{2.6}\\
& d b_{T}^{b}=T^{-1} \sum_{t=1}^{T} \exp \left[\left(\frac{\bar{d}^{\prime} \bar{d}}{\bar{\delta}^{\prime} \bar{\delta}}\right)^{1 / 2} \bar{\delta}^{\prime} \hat{\Sigma}_{X, t}^{-1 / 2}\left(T^{-1 / 2} \sum_{s=t+1}^{T} \tilde{v}_{s}\right)\right]  \tag{2.7}\\
& d b_{T}^{c}=T^{-1} \sum_{t=1}^{T}\left(\frac{12}{\bar{\delta}^{\prime} \bar{\delta}}\right)^{1 / 2} \bar{\delta}^{\prime} \hat{\Sigma}_{X, t}^{-1 / 2}\left(T^{-1 / 2} \sum_{s=t+1}^{T} \tilde{v}_{s}\right) . \tag{2.8}
\end{align*}
$$

The tests that allow for a break in the long-run variance of $\left\{X_{t} u_{t}\right\}$ is constructed as follows.

1. Compute the least squares regression of $\left\{y_{j}\right\}_{j=1}^{T}$ on $\left\{X_{j}, Z_{j}\right\}_{j=1}^{T}$ and compute the OLS residuals $\hat{u}_{j}$.
2. For a fixed break date $t$, compute the pre-break variance estimate $\hat{\sigma}_{1}^{2}=\sum_{j=1}^{t} \hat{u}_{j}^{2}$ and the post-break variance estimate $\hat{\sigma}_{2}^{2}=\sum_{j=t+1}^{T} \hat{u}_{j}^{2}$.
3. Construct transformed series $\left\{\tilde{y}_{j}\right\}_{j=1}^{t}=\left\{\hat{\sigma}_{1}^{-1} y_{j}\right\}_{j=1}^{t},\left\{\tilde{X}_{j}, \tilde{Z}_{j}\right\}_{j=1}^{t}=\left\{\hat{\sigma}_{1}^{-1} X_{j}, \hat{\sigma}_{1}^{-1} Z_{j}\right\}_{j=1}^{t}$, $\left\{\tilde{y}_{j}\right\}_{j=t+1}^{T}=\left\{\hat{\sigma}_{2}^{-1} y_{j}\right\}_{j=t+1}^{T}$, and $\left\{\tilde{X}_{j}, \tilde{Z}_{j}\right\}_{j=t+1}^{T}=\left\{\hat{\sigma}_{2}^{-1} X_{j}, \hat{\sigma}_{2}^{-1} Z_{j}\right\}_{j=t+1}^{T}$
4. Compute $\hat{\Sigma}_{X, t}=T^{-1} \sum_{j=1}^{T} \tilde{X}_{j} \tilde{X}_{j}^{\prime}$ and the OLS residuals $\left\{\tilde{u}_{j}\right\}_{j=1}^{T}$ from the transformed regression of $\left\{\tilde{y}_{j}\right\}_{j=1}^{T}$ on $\left\{\tilde{X}_{j}, \tilde{Z}_{j}\right\}_{j=1}^{T}$.
5. Construct $\left\{\tilde{v}_{j}\right\}_{j=1}^{T}=\left\{\tilde{X}_{j} \tilde{u}_{j}\right\}_{j=1}^{T}$ and compute the partial sum $T^{-1 / 2} \sum_{j=t+1}^{T} \tilde{v}_{j}$.
6. Repeat steps 2 to 5 for $t=[\varepsilon T], \ldots,[(1-\varepsilon) T]$ for some $\varepsilon>0$ and compute test statistics (2.6), (2.7) and (2.8).

Likewise, the three directional break tests reject for large values and critical values are in Table 2.1. We restrict the break point such that $l=t / T \in[\varepsilon, 1-\varepsilon]$ in order to have enough observations for identification in finite sample.

### 2.2.2 Derivation and asymptotic distribution of test statistics

We derive the first two test statistics $d b_{T}^{a}$ and $d b_{T}^{b}$ under i.i.d Gaussian disturbances $\left\{u_{t}\right\}$ and strictly exogenous regressors. Under restricted assumptions the test statistics have optimality properties because they are derived from a likelihood ratio test statistic. Assume $u_{t} \stackrel{i . i . d}{\sim} N\left(0, \sigma^{2}\right)$, conditionally homoskedastic so that $\Omega=\sigma^{2} \Sigma_{X}$, and $Q$ is strictly exogeneous. Let $M:=I_{T}-Q\left(Q^{\prime} Q\right)^{-1} Q^{\prime}$ and pre-multiply to the model so that $M Y=M(X-X(\tau)) \delta+M U$ with $M U \sim N\left(\mathbf{0}, \sigma^{2} M\right)$. Under these assumptions the likelihood ratio test statistic of hypotheses, $\mathcal{H}_{0}: \delta=0$ against $\mathcal{H}_{1}: \delta=\delta_{a}$ with break at time $t$, have optimal properties. The alternative hypothesis is a weighted average with $w_{t}$ as the weight on the alternative that a break occurs at time $t$.

$$
\begin{align*}
L R_{T}= & \sum_{t=1}^{T} w_{t} \exp \left[-\frac{1}{2} \sigma^{-2}\left(M Y-M(X-X(t)) \delta_{a}\right)^{\prime}\left(M Y-M(X-X(t)) \delta_{a}\right)\right. \\
& \left.\quad+\frac{1}{2} \sigma^{-2} Y^{\prime} M Y\right] \\
= & \sum_{t=1}^{T} w_{t} \exp \left[\sigma^{-2} \delta_{a}^{\prime}(X-X(t))^{\prime} M Y-\frac{1}{2} \sigma^{-2} \delta_{a}^{\prime}(X-X(t))^{\prime} M(X-X(t)) \delta_{a}\right] \tag{2.9}
\end{align*}
$$

The directional break test statistics tests the null of no break against the alternative hypothesis that is a weighted average of a break occurring at time $t$. For any sample size, and rejecting for large values, the test statistic (2.9) along with the appropriate critical values results in a test that is optimal for testing null hypothesis of no break against the alternative hypothesis a break occurs at time $t$ with weights $w_{t}$ (under restricted assumptions stated above).

We consider the asymptotic properties of the test statistic (2.9) by setting the magnitude of the break $\delta_{a}=T^{-1 / 2} \sigma \Sigma_{X}^{-1 / 2}(\bar{d} \circ \bar{\delta})$, where $\circ$ indicates the Hadamard product. The vector $\bar{d}>0$ represents a "normalized" local break magnitude, so that $\bar{d} \circ \bar{\delta}$ represents the $(k \times 1)$ vector of local parameter change (positive or negative) under the alternative hypothesis. We set $\bar{d}=\bar{d}_{1} l_{k}$ where $\boldsymbol{l}_{k}$ is a $(k \times 1)$ vector of ones so that each element is equal to a constant $\bar{d}_{1}$. Specifically,
in sections 2.3 and 2.4 we set $\bar{d}_{1}=2$ from setting $\delta_{a}=2 \sigma_{\hat{\beta}}$, where $\sigma_{\hat{\beta}}$ is the standard deviation of the OLS coefficient estimator under the null hypothesis. If the break magnitude is at least twice the standard error of $\hat{\beta}$, the break is significant to reject the null of no break. Hence, we can rewrite the local alternative as $\delta_{a}=T^{-1 / 2} \sigma \bar{d}_{1} \Sigma_{X}^{-1 / 2} \bar{\delta}$. Note that under the null hypothesis we have $M Y=M U$. We use the following lemma to derive the asymptotic distribution. For proof see Appendix A.2.2.

## Lemma 5. Under Condition 1,

$$
T^{-1 / 2}(X-X(t))^{\prime} M U \Rightarrow \Omega^{1 / 2}(W(l)-l W(1))
$$

where $W(\cdot)$ is a $(k \times 1)$ standard Brownian Motion on the unit interval and $2 \pi \Omega$ is the spectral density of $\left\{X_{t} u_{t}\right\}$

Under the null hypothesis, the first term of the exponential function in (2.9) weakly converges as follows, by Lemma 5.

$$
\begin{aligned}
T^{-1 / 2} \sigma^{-1}(\bar{d} \circ \bar{\delta})^{\prime} \Sigma_{X}^{-1 / 2}(X-X(t))^{\prime} M Y & \Rightarrow \sigma^{-1}(\bar{d} \circ \bar{\delta})^{\prime} \Sigma_{X}^{-1 / 2} \Omega^{1 / 2}(W(l)-l W(1)) \\
& =\bar{d}^{\prime}(W(l)-l W(1)),
\end{aligned}
$$

where the last equation is from $\Omega^{1 / 2}=\sigma \Sigma_{X}^{1 / 2}$ and $W(l)-l W(1)$ is symmetric. The second term of the exponential function in (2.9) converges in probability to $\frac{1}{2} \bar{d}^{\prime} \bar{d} l(1-l)$, which we show in the proof of Theorem 6. Under the assumption that weights $w_{t}$ satisfy $T w_{t} \rightarrow w(l)$ uniformly on $l \in(0,1)$ and by continuous mapping theorem, the likelihood ratio test statistic has the null asymptotic distribution,

$$
\begin{equation*}
L R_{T} \Rightarrow \int_{0}^{1} w(l) \exp \left[\bar{d}^{\prime}(W(l)-l W(1))-\frac{1}{2} \bar{d}^{\prime} \bar{d} l(1-l)\right] d l . \tag{2.10}
\end{equation*}
$$

The first two directional break tests $d b_{T}^{a}$ and $d b_{T}^{b}$ are motivated from these results and are given
by specifying the weights $w_{t}$. For $d b_{T}^{a}$, the weights are uniformly distributed over possible break dates, i.e., $w_{t}=T^{-1}$. The second term inside the exponential function is by a term that is asymptotically equivalent, $\frac{1}{2} \bar{d}^{\prime} \bar{d}\left(\frac{t}{T}\right)\left(1-\frac{t}{T}\right)$.

$$
\begin{equation*}
d b_{T}^{a}=T^{-1} \sum_{t=1}^{T} \exp \left[\left(\frac{\bar{d}^{\prime} \bar{d}}{\hat{\sigma}^{2} \bar{\delta}^{\prime} \bar{\delta}}\right)^{1 / 2} \hat{\delta}^{\prime} \hat{\Sigma}_{X}^{-1 / 2}\left(T^{-1 / 2} \sum_{s=t+1}^{T} \hat{v}_{S}\right)-\frac{1}{2} \bar{d}^{\prime} \bar{d}\left(\frac{t}{T}\right)\left(1-\frac{t}{T}\right)\right] \tag{2.11}
\end{equation*}
$$

Note that $\left(\bar{d}^{\prime} \bar{d} /\left(\hat{\sigma}^{2} \bar{\delta}^{\prime} \bar{\delta}\right)\right)^{1 / 2}=\hat{\sigma}^{-1} \bar{d}_{1}$. The test statistic (2.3) simplifies to (2.11) under conditional homoskedasticity. For the second test $d b_{T}^{b}$, the weights are selected to offset the second term inside the exponential term of (2.10), which is independent of the data asymptotically.

$$
w_{t}=T^{-1} \exp \left[\frac{1}{2} \bar{d}^{\prime} \bar{d}\left(\frac{t}{T}\right)\left(1-\frac{t}{T}\right)\right]
$$

The weights here depend on the potential break date, and is maximized at the center of the sample. Thus the weights place more emphasis on finding breaks in the center of the sample rather than the edges of the data ${ }^{1}$.

Before providing the null limit distribution of directional test statistics, we provide Proposition 1 which states that $\delta=0$ under the null hypothesis is the least favorable distribution.

Proposition 1. Denote the null set of parameters $\Delta_{0}=\left\{\delta \mid \bar{\delta}_{j} \delta_{j} \leq 0, \forall j=1, \ldots, k\right\}$, the boundary set as $\Delta_{b}=\left\{\delta \mid \bar{\delta}_{j} \delta_{j}=0, \forall j=1, \ldots, k\right\}$, and $\Delta_{i}=\Delta_{0} \backslash \Delta_{b}$. For some fixed constant $c>0$,

$$
\lim _{T \rightarrow \infty} \operatorname{Pr}\left[d b_{T}^{a}>c \mid \delta \in \Delta_{i}\right] \leq \lim _{T \rightarrow \infty} \operatorname{Pr}\left[d b_{T}^{a}>c \mid \delta \in \Delta_{b}\right]
$$

The inequality also holds for $d b_{T}^{b}$ and $d b_{T}^{c}$.
This result is explained in detail after we present Theorem 7 on the local alternative distribution of test statistics. Under the least favorable distribution of no break, we have the

[^10]null limit distribution of the three directional break test statistics in Theorem 6. For proof see Appendix A.2.2.

Theorem 6. Under Condition 1 and the least favorable distribution $\delta=0$ under the null hypothesis, the directional break tests (2.3), (2.4) and (2.5) have the following asymptotic distributions.

$$
\begin{align*}
d b_{T}^{a} & \Rightarrow \int_{0}^{1} \exp \left[\left(\bar{d}^{\prime} \bar{d}\right)^{1 / 2} B B(l)-\frac{1}{2} \bar{d}^{\prime} \bar{d} l(1-l)\right] d l  \tag{2.12}\\
d b_{T}^{b} & \Rightarrow \int_{0}^{1} \exp \left[\left(\bar{d}^{\prime} \bar{d}\right)^{1 / 2} B B(l)\right] d l  \tag{2.13}\\
d b_{T}^{c} & \Rightarrow \int_{0}^{1} \sqrt{12} B B(l) d l \tag{2.14}
\end{align*}
$$

$B B(l)$ is an univariate Brownian Bridge, $W(l)-l W(1)$ on the unit interval $l \in[0,1]$. The asymptotic distribution in (2.14) is a standard normal distribution.

If $Q$ is weakly exogenous, the three directional test statistics have equivalent asymptotic distributions but loses the optimal property of the likelihood ratio test statistic. Previously when we assumed serially uncorrelated and homoskedastic disturbances, we chose $\bar{d}$ in test statistics $d b_{T}^{a}$ and $d b_{T}^{b}$ by setting $\delta_{a}=2 \hat{\sigma}_{\beta}$, which gives us $\bar{d}_{1}$ equal to 2 . Although this is not true under relaxed assumptions $\left(\Omega \neq \sigma^{2} \Sigma_{X}\right)$ we will maintain $\bar{d}_{1}=2$ to avoid finite sample estimation error.

Now we derive the local alternative limit distribution of test statistics assuming $\delta=$ $T^{-1 / 2} \sigma \Sigma_{X}^{-1 / 2} d$, such that the break magnitude is inside the local $T^{-1 / 2}$ neighborhood of zero (note that $d$ can be either positive or negative). For proof see Appendix A.2.2.

Theorem 7. Under Condition 1, suppose the break magnitude is $\delta=T^{-1 / 2} \sigma \Sigma_{X}^{-1 / 2} d$ with fixed $d$.

Let $\bar{d}=\bar{d}_{1} 1_{k}$. Then the local alternative limit distribution of directional break test statistics are

$$
\begin{aligned}
& d b_{T}^{a} \Rightarrow \int_{0}^{1} \exp \left[\frac{\sigma\left(\bar{d}^{\prime} \bar{d}\right)^{1 / 2}\left(\min \left\{l, r_{0}\right\}-l r_{0}\right)}{\left(\bar{\delta}^{\prime} \Sigma_{X}^{-1 / 2} \Omega \Sigma_{X}^{-1 / 2} \bar{\delta}\right)^{1 / 2}} \bar{\delta}^{\prime} d+\left(\bar{d}^{\prime} \bar{d}\right)^{1 / 2} B B(l)-\frac{1}{2} \bar{d}^{\prime} \bar{d} l(1-l)\right] d l, \\
& d b_{T}^{b} \Rightarrow \int_{0}^{1} \exp \left[\frac{\sigma\left(\bar{d}^{\prime} \bar{d}\right)^{1 / 2}\left(\min \left\{l, r_{0}\right\}-l r_{0}\right)}{\left(\bar{\delta}^{\prime} \Sigma_{X}^{-1 / 2} \Omega \Sigma_{X}^{-1 / 2} \bar{\delta}\right)^{1 / 2}} \bar{\delta}^{\prime} d+\left(\bar{d}^{\prime} \bar{d}\right)^{1 / 2} B B(l)\right] d l \\
& d b_{T}^{c} \Rightarrow \int_{0}^{1}\left(\frac{\sigma \sqrt{12}\left(\min \left\{l, r_{0}\right\}-l r_{0}\right)}{\left(\bar{\delta}^{\prime} \Sigma_{X}^{-1 / 2} \Omega \Sigma_{X}^{-1 / 2} \bar{\delta}\right)^{1 / 2}} \bar{\delta}^{\prime} d+\sqrt{12} B B(l)\right) d l .
\end{aligned}
$$

The limit distributions in Theorem 7 shifts depending on the sign of $\bar{\delta}^{\prime} d$. Under the case of no break $\delta=0$ we have $\bar{\delta}^{\prime} d=0$. In other cases of the null hypothesis where $\bar{\delta}_{j} \delta_{j} \leq 0$ holds for all $j$ with strict inequality for at least one $j$, we have $\bar{\delta}^{\prime} d<0$. Because the test statistic rejects for large values, given some critical value $c$ and $\bar{\delta}^{\prime} d<0, \operatorname{Pr}\left[d b_{T}^{a}>c\right]$ is smaller than that when $\bar{\delta}^{\prime} d=0$ in the limit. Hence we have the results of Proposition 1, which is analogous to Lemma 1 of Wolak (1989). The paper examines tests when imposing linear inequality restrictions on parameters of a linear model. In our case $\bar{\delta}^{\prime} \delta=0$ is the unique least favorable value of $\bar{\delta}^{\prime} \delta$ that specifies the the null hypothesis to obtain critical values for any size test. When $\bar{\delta}^{\prime} \delta<0$, the test statistic shifts away from the alternative and the rejection probability is less than $\alpha$ in the limit. Therefore we obtain critical values of any size test from the limit distribution under $\delta=0$.

Note that in some cases of the alternative hypothesis, we have $\bar{\delta}^{\prime} d<0$ and thus the tests do not have power. That is, if the directions are incorrectly specified for coefficient that has magnitude dominating the correctly specified magnitude, we lose power. In the opposite case, where correctly specified coefficients' magnitude is larger than those that are incorrectly specified, we have $\bar{\delta}^{\prime} d>0$ and the tests will still have power.

The asymptotic power function and power envelope of the directional test statistics can be computed under the local alternative distributions in Theorem 7. For simplicity assume conditional homoskedastiticy, then the asymptotic power function of the three directional test
statistics are as follows.

$$
\begin{align*}
& \pi^{a}(d, \bar{d}):=\operatorname{Pr}\left[\int_{0}^{1} \exp \left(\bar{d}_{1}\left(\min \left\{l, r_{0}\right\}-l r_{0}\right) \bar{\delta}^{\prime} d+\left(\bar{d}^{\prime} \bar{d}\right)^{1 / 2} B B(l)-\frac{1}{2} \bar{d}^{\prime} \overline{d l}(1-l)\right) d l>c v^{a}(\bar{d})\right]  \tag{2.15}\\
& \pi^{b}(d, \bar{d}):=\operatorname{Pr}\left[\int_{0}^{1} \exp \left(\bar{d}_{1}\left(\min \left\{l, r_{0}\right\}-l r_{0}\right) \bar{\delta}^{\prime} d+\left(\bar{d}^{\prime} \bar{d}\right)^{1 / 2} B B(l)\right) d l>c v^{b}(\bar{d})\right]  \tag{2.16}\\
& \pi^{c}(d, \bar{d}):=\operatorname{Pr}\left[\int_{0}^{1}\left(\sqrt{12}\left(\min \left\{l, r_{0}\right\}-l r_{0}\right) \bar{\delta}^{\prime} d+\sqrt{12} B B(l)\right) d l>1.645\right] . \tag{2.17}
\end{align*}
$$

The critical values $c v^{a}(\bar{d})$ and $c v^{b}(\bar{d})$ are determined by $\pi^{a}(0, \bar{d})=0.05$ and $\pi^{b}(0, \bar{d})=0.05$, respectively. Table 2.1 presents asymptotic critical values of $d b_{T}^{a}$ and $d b_{T}^{b}$ at $1 \%, 5 \%$ and $10 \%$ depending on the value and dimension of $\bar{d}$.

Table 2.1: Asymptotic critical values for directional test statistics $d b_{T}^{a}$ and $d b_{T}^{b}$

| $\operatorname{dim}(\bar{d})$ |  | $\bar{d}=1.5$ |  |  | $\bar{d}=2$ |  |  | $\bar{d}=2.5$ |  |  |
| :--- | :---: | :---: | :---: | :---: | ---: | :---: | :---: | :---: | :---: | :---: |
|  | $c v^{a}$ | 2.542 | 1.883 | 1.589 | 3.328 | 2.213 | 1.756 | 4.264 | 2.534 | 1.889 |
|  | $c v^{b}$ | 2.843 | 2.094 | 1.760 | 4.146 | 2.699 | 2.122 | 6.059 | 3.495 | 2.552 |
| 2 | $c v^{a}$ | 3.546 | 2.294 | 1.793 | 4.935 | 2.740 | 1.951 | 6.497 | 3.090 | 2.016 |
|  | $c v^{b}$ | 4.542 | 2.876 | 2.222 | 7.876 | 4.155 | 2.879 | 14.010 | 5.962 | 3.704 |
| 3 | $c v^{a}$ | 4.454 | 2.602 | 1.907 | 6.347 | 3.061 | 2.011 | 8.190 | 3.235 | 1.933 |
|  | $c v^{b}$ | 6.552 | 3.677 | 2.653 | 13.208 | 5.763 | 3.616 | 23.360 | 8.913 | 4.843 |

In sections 2.3 and 2.4 we compare the three directional break test statistics with the $L$ statistic from Nyblom (1989) (denote as Nyblom), the (average) exponential $L M_{T}$ test statistic from Andrews and Ploberger (1994) (denote as AP) and $q L L$ test statistic from Elliott and Müller (2006). The three test statistics and asymptotic distributions under the null and local alternatives are described in Appendix A.2.1.

### 2.3 Monte Carlo Simulation

This section examines finite sample rejection probabilities of the three directional break tests, Nyblom, AP, and $q L L$ test statistics. The rejection probabilities are computed from 5,000 replications, except Model 2 that is computed from 10,000 replications. We consider five different cases of model (2.1) with $T=100$. For simplicity assume $Z_{t}=0$ for all five cases.

1. Break in the mean with i.i.d. Gaussian disturbances
2. Break in the mean with serially correlated disturbances
3. Misspecification of number of parameters under break
4. Break in the drift of a $\operatorname{AR}(1)$ model
5. Break in coefficient and variance

## Model 1: i.i.d. Gaussian disturbances

The data generating process is a bivariate model (2.1) where a break occurs in both coefficients of regressor $X_{t}$ simultaneously; $X_{t}=\left(1, g_{t}\right)^{\prime}, g_{t} \stackrel{i . i . d}{\sim} N(2,1), \beta=(0,0)^{\prime}, \delta=T^{-1 / 2} \Sigma_{X}^{-1 / 2} d$, $d=(2 c,-c)^{\prime}, c \in[0,10], u_{t} \stackrel{i . i . d}{\sim} N(0,1)$ and $T=100$. For directional break test statistics $d b_{T}^{a}$ and $d b_{T}^{b}$ we use $\bar{d}=(2,2)^{\prime}$. Table 2.3 shows the size of each test statistic and Figure 2.1 shows the power of test statistics. Figures 2.2 and 2.3 are rejection probabilities of directional break tests when one direction is correct but the other is incorrect.

To assess the power differences in Figure 2.1, we compute the Pitman efficiency of the directional test and comparative tests. The Pitman efficiency (or asymptotic relative efficiency, ARE) is the ratio of sample sizes giving, asymptotically, the same power for that sequence (Stock, 1994). In our case, suppose that $d b_{T}^{a}$ and $\mathrm{AP}(\infty)$ tests achieve 50 percent power against the local alternative $c_{1}$ and $c_{2}$, respectively. Then the $\operatorname{ARE}$ of the $\operatorname{AP}(\infty)$ test relative to the $d b_{T}^{a}$ test is
$c_{2} / c_{1}=1.0645$ under $r_{0}=0.50$. To achieve 50 percent power against a local alternative using the $\mathrm{AP}(\infty)$ test statistic asymptotically requires 6.45 percent more observations than are needed using the directional test statistic $d b_{T}^{a}$. For all comparative methods relative to directional test statistics, the ARE range between 1.03 to 1.08 under the bivariate model (2.1) with i.i.d Gaussian disturbances. Because the values are similar across different $r_{0}$ values, we show the case for $r_{0} \in\{0.20,0.50\}$.

Table 2.2: (Model 1) Pitman efficiency (or ARE) of directional break tests, Nyblom, AP and $q L L$ tests.

|  |  | ARE |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $r_{0}$ |  | Test | Nyblom | $\operatorname{AP}(\infty)$ | $\operatorname{AP}(0)$ |
| 0.2 | $d b_{T}^{a}$ | 1.0833 | 1.0536 | 1.0714 | 1.0536 |
|  | $d b_{T}^{b}$ | 1.0769 | 1.0473 | 1.0651 | 1.0473 |
|  | $d b_{T}^{c}$ | 1.0643 | 1.0351 | 1.0526 | 1.0351 |
| 0.5 | $d b_{T}^{a}$ | 1.0452 | 1.0645 | 1.0581 | 1.0581 |
|  | $d b_{T}^{b}$ | 1.0519 | 1.0714 | 1.0649 | 1.0649 |
|  | $d b_{T}^{c}$ | 1.0452 | 1.0645 | 1.0581 | 1.0581 |

Table 2.3: (Model 1) Rejection rate of structural break tests under no break $\delta=0, T=100$.

|  |  | Test Statistic |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :---: |
| $r_{0}$ | $\bar{\delta}$ | $d b_{T}^{a}$ | $d b_{T}^{b}$ | $d b_{T}^{c}$ | Nyblom | $\mathrm{AP}(0)$ | $\mathrm{AP}(\infty)$ | $q L L$ |  |
|  | $(1,-1)$ | 0.0530 | 0.0534 | 0.0482 |  |  |  |  |  |
| 0.2 | $(1,1)$ | 0.0490 | 0.0500 | 0.0516 | 0.0520 | 0.0468 | 0.0462 | 0.0626 |  |
|  | $(-1,-1)$ | 0.0520 | 0.0518 | 0.0546 |  |  |  |  |  |
|  | $(-1,1)$ | 0.0518 | 0.0524 | 0.0526 |  |  |  |  |  |
|  | $(1,-1)$ | 0.0444 | 0.0450 | 0.0462 |  |  |  |  |  |
| 0.3 | $(1,1)$ | 0.0476 | 0.0458 | 0.0446 | 0.0492 | 0.0440 | 0.0430 | 0.0586 |  |
|  | $(-1,-1)$ | 0.0458 | 0.0466 | 0.0456 |  |  |  |  |  |
|  | $(-1,1)$ | 0.0516 | 0.0522 | 0.0556 |  |  |  |  |  |
|  | $(1,-1)$ | 0.0424 | 0.0438 | 0.0446 |  |  |  |  |  |
| 0.4 | $(1,1)$ | 0.0488 | 0.0500 | 0.0512 | 0.0442 | 0.0384 | 0.0376 | 0.0514 |  |
|  | $(-1,-1)$ | 0.0498 | 0.0514 | 0.0540 |  |  |  |  |  |
|  | $(-1,1)$ | 0.0478 | 0.0476 | 0.0472 |  |  |  |  |  |
|  | $(1,-1)$ | 0.0420 | 0.0444 | 0.0438 |  |  |  |  |  |
| 0.5 | $(1,1)$ | 0.0500 | 0.0496 | 0.0534 | 0.0478 | 0.0430 | 0.0414 | 0.0594 |  |
|  | $(-1,-1)$ | 0.0504 | 0.0514 | 0.0526 | 0.047 |  |  |  |  |
|  | $(-1,1)$ | 0.0488 | 0.0502 | 0.0476 |  |  |  |  |  |
|  | $(1,-1)$ | 0.0460 | 0.0470 | 0.0496 |  |  |  |  |  |
| 0.6 | $(1,1)$ | 0.0494 | 0.0492 | 0.0492 | 0.0450 | 0.0422 | 0.0420 | 0.0560 |  |
|  | $(-1,-1)$ | 0.0482 | 0.0516 | 0.0482 |  |  |  |  |  |
|  | $(-1,1)$ | 0.0448 | 0.0452 | 0.0442 |  |  |  |  |  |
|  | $(1,-1)$ | 0.0430 | 0.0448 | 0.0452 |  |  |  |  |  |
| 0.7 | $(1,1)$ | 0.0462 | 0.0484 | 0.0508 | 0.0452 | 0.0410 | 0.0374 | 0.0556 |  |
|  | $(-1,-1)$ | 0.0468 | 0.0478 | 0.0508 |  |  |  |  |  |
|  | $(-1,1)$ | 0.0464 | 0.0466 | 0.0458 |  |  |  |  |  |
|  | $(1,-1)$ | 0.0514 | 0.0494 | 0.0510 |  |  |  |  |  |
|  | $(1,1)$ | 0.0456 | 0.0462 | 0.0492 | 0.0446 | 0.0376 | 0.0422 | 0.0562 |  |
|  | $(-1,-1)$ | 0.0454 | 0.0470 | 0.0472 | 0.053 |  |  |  |  |
|  | $(-1,1)$ | 0.0540 | 0.0538 | 0.0538 |  |  |  |  |  |



Figure 2.1: (Model 1, $\bar{\delta}$ correct) Finite sample power of the $5 \%$ level structural break tests when $\bar{\delta}=(1,-1)^{\prime}$ across local parameter $d=(2 c,-c)^{\prime}, c \in[0,10]$. The plots correspond to break location $r_{0}=0.20,0.30,0.40,0.50,0.60$ and 0.70 , respectively.


Figure 2.2: (Model 1) Finite sample rejection probability of the 5\% level structural break tests when $\bar{\delta}=(1,1)^{\prime}$ across local parameter $d=(2 c,-c)^{\prime}, c \in[0,10]$. The plots correspond to break location $r_{0}=0.20,0.30,0.40,0.50,0.60$ and 0.70 , respectively.


Figure 2.3: (Model 1) Finite sample rejection probability of the $5 \%$ level structural break tests when $\bar{\delta}=(-1,-1)^{\prime}$ across local parameter $d=(2 c,-c)^{\prime}, c \in[0,10]$. The plots correspond to break location $r_{0}=0.20,0.30,0.40,0.50,0.60$ and 0.70 , respectively.

## Model 2: Serially correlated disturbances

Consider an univariate model (2.1) where the error term is an $\operatorname{AR}(1)$ process: $u_{t}=$ $\phi u_{t-1}+\varepsilon_{t}, \phi=0.3$ and $\varepsilon_{t} \stackrel{i . i . d}{\sim} N\left(0, \sigma_{\varepsilon}^{2}\right), \sigma_{\varepsilon}=0.9$. Hence the long-run variance of the error term is $\sigma^{2}=\sigma_{\varepsilon}^{2} /\left(1-\phi^{2}\right) \approx 1.29$. The regressor is a constant term $X_{t}=1, \beta=0$ and $\delta=T^{-1 / 2} \sigma d$, $d \in[0,20]$. Under correctly specified $\bar{\delta}=1$ we compare the power of the structural break test statistics. Estimation of $\Omega=\sigma^{2}=\lim _{T \rightarrow \infty} \operatorname{Var}\left(T^{-1 / 2} \sum_{t=1}^{T} v_{t}\right)$ is conducted by the NeweyWest estimator with a bandwidth $q(T)=4(T / 100)^{2 / 9}$ (Newey and West, 1987, 1994) and for directional break test statistics $d b_{T}^{a}$ and $d b_{T}^{b}$ we choose $\bar{d}=2$.

Figures 2.4 and 2.5 show finite sample rejection probabilities of three directional tests under directions $\bar{\delta}=1$ (correct) and $\bar{\delta}=-1$ (incorrect), respectively. In Figure 2.4 we see there is power improvement compared to the Nyblom and the $\operatorname{AP}(\infty)$ tests for all break locations considered, although the power increase is not as large as the model with i.i.d disturbances. Figure 2.5 show that tests have rejection rate less than size under the wrong direction.


Figure 2.4: (Model 2) Finite sample power of the $5 \%$ level structural break tests when $\bar{\delta}=1$ across local parameter $d \in[0,20]$. The plots correspond to break location $r_{0}=$ $0.20,0.30,0.40,0.50,0.60$ and 0.70 , respectively.


Figure 2.5: (Model 2) Finite sample rejection probability of the 5\% level structural break tests when $\bar{\delta}=-1$ across local parameter $d \in[0,20]$. The plots correspond to break location $r_{0}=0.20,0.30,0.40,0.50,0.60$ and 0.70 , respectively.

Table 2.4: (Model 2) Rejection rate of structural break tests under no break $\delta=0, T=100$.

|  |  | Test Statistic |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :---: |
| $r_{0}$ | $\bar{\delta}$ | $d b_{T}^{a}$ | $d b_{T}^{b}$ | $d b_{T}^{c}$ | Nyblom | $\mathrm{AP}(0)$ | $\mathrm{AP}(\infty)$ | $q L L$ |  |
| 0.2 | 1 | 0.0725 | 0.0738 | 0.0698 | 0.0778 | 0.0637 | 0.0633 | 0.0656 |  |
|  | -1 | 0.0751 | 0.0754 | 0.0729 |  |  |  |  |  |
| 0.3 | 1 | 0.0745 | 0.0738 | 0.0716 | 0.0748 | 0.0609 | 0.0610 | 0.0637 |  |
|  | -1 | 0.0713 | 0.0712 | 0.0701 |  |  |  |  |  |
| 0.4 | 1 | 0.0716 | 0.0726 | 0.0711 | 0.0723 | 0.0603 | 0.0594 | 0.0633 |  |
|  | -1 | 0.0693 | 0.0699 | 0.0712 |  |  |  |  |  |
| 0.5 | 1 | 0.0719 | 0.0726 | 0.0715 | 0.0745 | 0.0623 | 0.0608 | 0.0657 |  |
|  | -1 | 0.0741 | 0.0751 | 0.0698 |  |  |  |  |  |
| 0.6 | 1 | 0.0702 | 0.0714 | 0.0693 | 0.0757 | 0.0614 | 0.0630 | 0.0637 |  |
|  | -1 | 0.0738 | 0.0746 | 0.0722 |  |  |  |  |  |
| 0.7 | 1 | 0.0713 | 0.0719 | 0.0695 | 0.0744 | 0.0597 | 0.0584 | 0.0613 |  |
|  | -1 | 0.0700 | 0.0709 | 0.0675 |  |  |  |  |  |
| 0.8 | 1 | 0.0726 | 0.0728 | 0.0706 | 0.0783 | 0.0653 | 0.0653 | 0.0657 |  |

## Model 3: Misspecification of number of parameters under break

As mentioned in Section 2.2.2, we check whether the directional break tests have power when there exists a structural break in only one (unknown) coefficient of bivariate variable $X_{t}$. Consider model (2.1) where $X_{t}=\left(1, g_{t}\right)^{\prime}, g_{t} \stackrel{i . i . d}{\sim} N(1,0.25)$ and assume conditionally homoskedastic and serially uncorrelated errors $u_{t} \stackrel{i . i . d}{\sim} N\left(0, \sigma^{2}\right), \sigma=1$. There are two cases of a structural break in one coefficient, $\delta_{1}=0$ or $\delta_{2}=0$ where $\delta=\left(\delta_{1}, \delta_{2}\right)^{\prime}$. We consider the case of a positive break in the second variable $g_{t}$ such that $\delta=T^{-1 / 2} d$ and $d=\left(0, d_{g}\right)^{\prime} . \beta$ is a vector of zeros and $d_{g} \in[0,30]$. All structural break test statistics are computed under the possibility that there is a break in any coefficient parameter. If we specify both coefficients under break, there are two cases in which the direction $\bar{\delta}$ is correct: $\bar{\delta}=(1,1)^{\prime}$ and $\bar{\delta}=(-1,1)^{\prime}$. We plot the finite sample power under each case and compare with Nyblom and $\mathrm{AP}(\infty)$ tests.

Under $\bar{\delta}=(1,1)^{\prime}$, Figure 2.6 shows that the three directional break tests have higher
power than Nyblom or $\operatorname{AP}(\infty)$ tests. In contrast, Figure 2.7 under $\bar{\delta}=(-1,1)^{\prime}$ shows the power of directional break tests are significantly smaller than the power of Nyblom or AP tests. The asymptotic power function of directional break test statistics depend on the term $\bar{\delta}^{\prime} \Sigma_{X}^{1 / 2} d$. Because it depends on a positive definite matrix $\Sigma_{X}^{1 / 2}$, the term $\bar{\delta}^{\prime} \Sigma_{X}^{1 / 2} d$ is larger when $\bar{\delta}=(1,1)^{\prime}$. This is also true for the case when the constant term is under break, $d=\left(d_{1}, 0\right)^{\prime}$. Hence, under a bivariate model there is power improvement if $\bar{\delta}$ has both elements equal to the direction of coefficient under break.

Table 2.5: (Model 3) Rejection rate of structural break tests under no break $\delta=0, T=100$.

|  |  | Test Statistic |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $r_{0}$ | $\bar{\delta}$ | $d b_{T}^{a}$ | $d b_{T}^{b}$ | $d b_{T}^{c}$ | Nyblom | $\mathrm{AP}(0)$ | $\mathrm{AP}(\infty)$ | $q L L$ |
|  | $(1,-1)$ | 0.0470 | 0.0468 | 0.0460 |  |  |  |  |
| 0.2 | $(1,1)$ | 0.0462 | 0.0472 | 0.0490 | 0.0444 | 0.0438 | 0.0410 | 0.0586 |
|  | $(-1,-1)$ | 0.0452 | 0.0474 | 0.0472 |  |  |  |  |
|  | $(-1,1)$ | 0.0462 | 0.0478 | 0.0472 |  |  |  |  |
|  | $(1,-1)$ | 0.0490 | 0.0512 | 0.0480 |  |  |  |  |
| 0.3 | $(1,1)$ | 0.0524 | 0.0526 | 0.0528 | 0.0408 | 0.0364 | 0.0358 | 0.0512 |
|  | $(-1,-1)$ | 0.0478 | 0.0494 | 0.0510 |  |  |  |  |
|  | $(-1,1)$ | 0.0430 | 0.0444 | 0.0398 |  |  |  |  |
|  | $(1,-1)$ | 0.0432 | 0.0440 | 0.0432 |  |  |  |  |
| 0.4 | $(1,1)$ | 0.0468 | 0.0466 | 0.0476 | 0.0442 | 0.0350 | 0.0354 | 0.0524 |
|  | $(-1,-1)$ | 0.0476 | 0.0478 | 0.0472 | 0.04 |  |  |  |
|  | $(-1,1)$ | 0.0516 | 0.0522 | 0.0516 |  |  |  |  |
|  | $(1,-1)$ | 0.0448 | 0.0450 | 0.0430 |  |  |  |  |
| 0.5 | $(1,1)$ | 0.0510 | 0.0514 | 0.0512 | 0.0444 | 0.0398 | 0.0398 | 0.0564 |
|  | $(-1,-1)$ | 0.0444 | 0.0452 | 0.0466 |  |  |  |  |
|  | $(-1,1)$ | 0.0444 | 0.0444 | 0.0436 |  |  |  |  |



Figure 2.6: (Model 3) Finite sample power of the 5\% level structural break tests when $\bar{\delta}=(1,1)^{\prime}$ across local parameter $d_{g} \in[0,30]$. The plots correspond to break location $r_{0}=0.20,0.30,0.40$ and 0.50 , respectively.


Figure 2.7: (Model 3) Finite sample power of the $5 \%$ level structural break tests when $\bar{\delta}=$ $(-1,1)^{\prime}$ across local parameter $d_{g} \in[0,30]$. The plots correspond to break location $r_{0}=$ $0.20,0.30,0.40$ and 0.50 , respectively.

## Model 4: AR(1) with a drift

The DGP is an autoregressive model with one lag where the constant term and lag coefficient is under structural break simultaneously. In model (2.1) the regressor is $X_{t}=1$ and $Z_{t}=y_{t-1}$.

$$
\begin{equation*}
y_{t}=\beta+\delta \mathbf{1}\left\{t \geq\left[r_{0} T\right]\right\}+\phi y_{t-1}+u_{t}, \quad t=1, \ldots, T \tag{2.18}
\end{equation*}
$$

where $\beta=0, \phi=0.3, u_{t} \stackrel{i . i . d .}{\sim} N\left(0, \sigma^{2}\right), \sigma=0.7, \delta=T^{-1 / 2} \sigma d, d \in[-15,0]$, and $T=100$. For test statistics $d b_{T}^{a}$ and $d b_{T}^{b}$ we choose $\bar{d}=2$.

Table 2.6 shows that $\mathrm{AP}(\infty)$ and $q L L$ tests have rejection probability less than 0.03 under
the null hypothesis, given $5 \%$ significance level. The loss of power in $\operatorname{AP}(\infty)$ test is shown in Figure 2.8, where we can see that the power improvement of directional break tests compared to comparative methods is quite large. Pitman efficiency in Table 2.7 ranges between 1.10 to 1.18, which are larger than the values in Table 2.2 under i.i.d disturbances. This is because other break tests are conservative under the DGP (2.18) while directional break tests maintain their performance.

Table 2.6: (Model 4) Rejection rate of structural break tests under no break $\delta=0, T=100$.

|  |  | Test Statistic |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :---: |
| $r_{0}$ | $\bar{\delta}$ | $d b_{T}^{a}$ | $d b_{T}^{b}$ | $d b_{T}^{c}$ | Nyblom | $\mathrm{AP}(0)$ | $\mathrm{AP}(\infty)$ | $q L L$ |  |
| 0.2 | 1 | 0.0446 | 0.0456 | 0.0456 | 0.0408 | 0.0338 | 0.0260 | 0.0238 |  |
|  | -1 | 0.0454 | 0.0458 | 0.0480 |  |  |  |  |  |
| 0.3 | 1 | 0.0406 | 0.0418 | 0.0428 | 0.0372 | 0.0276 | 0.0246 | 0.0252 |  |
|  | -1 | 0.0404 | 0.0424 | 0.0418 |  |  |  |  |  |
| 0.4 | 1 | 0.0460 | 0.0460 | 0.0498 | 0.0386 | 0.0326 | 0.0296 | 0.0256 |  |
|  | -1 | 0.0460 | 0.0474 | 0.0468 |  |  |  |  |  |
| 0.5 | 1 | 0.0520 | 0.0536 | 0.0512 | 0.0402 | 0.0302 | 0.0288 | 0.0248 |  |
|  | -1 | 0.0400 | 0.0402 | 0.0442 |  |  |  |  |  |
| 0.6 | 1 | 0.0410 | 0.0418 | 0.0436 | 0.0408 | 0.0312 | 0.0268 | 0.0296 |  |
|  | -1 | 0.0516 | 0.0534 | 0.0540 |  |  |  |  |  |
| 0.7 | 1 | 0.0406 | 0.0424 | 0.0426 | 0.0386 | 0.0276 | 0.0280 | 0.0276 |  |
|  | -1 | 0.0460 | 0.0472 | 0.0472 |  |  |  |  |  |
| 0.8 | 1 | 0.0426 | 0.0424 | 0.0432 | 0.0446 | 0.0312 | 0.0278 | 0.0276 |  |
|  | -1 | 0.0516 | 0.0528 | 0.0546 |  |  |  |  |  |



Figure 2.8: (Model 4) Finite sample power of the $5 \%$ level structural break tests when $\bar{\delta}=-1$ across local parameter $d \in[-15,0]$. The plots correspond to break location $r_{0}=$ $0.20,0.30,0.40,0.50,0.60$ and 0.70 , respectively.

Table 2.7: (Model 4) Pitman efficiency (or ARE) of directional break tests, Nyblom, AP and $q L L$ tests.

|  |  |  | ARE |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :---: |
| $r_{0}$ | Test | Nyblom | $\operatorname{AP}(\infty)$ | $\operatorname{AP}(0)$ | $q L L$ |  |
| 0.2 | $d b_{T}^{a}$ | 1.1308 | 1.1017 | 1.1477 | 1.1477 |  |
|  | $d b_{T}^{b}$ | 1.1363 | 1.1071 | 1.1533 | 1.1533 |  |
|  | $d b_{T}^{c}$ | 1.1172 | 1.0885 | 1.1340 | 1.1340 |  |
| 0.5 | $d b_{T}^{a}$ | 1.1667 | 1.1574 | 1.1358 | 1.1821 |  |
|  | $d b_{T}^{b}$ | 1.1560 | 1.1468 | 1.1254 | 1.1713 |  |
|  | $d b_{T}^{c}$ | 1.1631 | 1.1538 | 1.1323 | 1.1785 |  |

## Model 5: Break in variance and coefficient

We consider the case where there is an one-time break in the variance of the disturbance and coefficient simultaneously. We compute our "usual" directional break test statistics using OLS estimators and test statistics using WLS estimators, which takes into account the break in variance of the error term $\left\{u_{t}\right\}$. The DGP is model (2.1) where $X_{t}=1, Z_{t}=0, u_{t} \stackrel{i . i . d}{\sim} N\left(0, \sigma_{1}^{2}\right)$ when $t \leq\left[r_{0} T\right]$ and $u_{t} \stackrel{i . i . d}{\sim} N\left(0, \sigma_{2}^{2}\right)$ when $t>\left[r_{0} T\right]$. Let $\sigma_{1}=2$ and $\sigma_{2}=1$ so that $\Omega_{1} \neq \Omega_{2}$; $\beta=0, \delta=T^{-1 / 2} \sigma_{2} d, d \in[0,30]$ and $T=100$.

None of the structural break tests in Table 2.8 accommodate a break in the variance of $u_{t}$, which results in rejection rate larger than size under the null hypothesis. WLS directional break tests in Table 2.9 shows size control. Thus, we cannot directly compare the finite sample power of WLS directional break tests and other tests. Nonetheless we plot the power of tests in Figure 2.9. The left plot shows directional break tests using OLS estimators have larger power in finite sample than Nyblom and $\mathrm{AP}(\infty)$ tests for all break location considered. The right plot shows that WLS directional break test statistics have smaller power than comparing methods, as expected, except the case when a break occurs at the median of the sample $r_{0}=0.5$.


Figure 2.9: (Model 5) Finite sample power of the $5 \%$ level structural break tests when $\bar{\delta}=1$ across local parameter $d \in[0,30]$ and $\sigma_{1} / \sigma_{2}=2$. The left (right) panel shows directional break test statistics using OLS (WLS). The plots correspond to break location $r_{0}=0.20,0.30$, and 0.50 respectively.

Table 2.8: (Model 5) Rejection rate of structural break tests under no break $\delta=0, T=100$.

|  |  | Test Statistic |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :---: |
| $r_{0}$ | $\bar{\delta}$ | $d b_{T}^{a}$ | $d b_{T}^{b}$ | $d b_{T}^{c}$ | Nyblom | $\mathrm{AP}(0)$ | $\mathrm{AP}(\infty)$ | $q L L$ |  |
| 0.2 | 1 | 0.0764 | 0.0749 | 0.0788 | 0.0934 | 0.0937 | 0.1750 | 0.0946 |  |
|  | -1 | 0.0827 | 0.0807 | 0.0822 |  |  |  |  |  |
| 0.3 | 1 | 0.0704 | 0.0697 | 0.0719 | 0.0805 | 0.0752 | 0.1280 | 0.0832 |  |
|  | -1 | 0.0730 | 0.0722 | 0.0736 |  |  |  |  |  |
| 0.4 | 1 | 0.0658 | 0.0658 | 0.0653 | 0.0700 | 0.0618 | 0.1069 | 0.0726 |  |
|  | -1 | 0.0643 | 0.0637 | 0.0647 |  |  |  |  |  |
| 0.5 | 1 | 0.0493 | 0.0491 | 0.0487 | 0.0479 | 0.0409 | 0.0775 | 0.0651 |  |

Table 2.9: (Model 5) Rejection rate of WLS directional break tests under no break $\delta=0$, $T=100$.

|  |  | Test Statistic |  |  |
| :--- | :--- | :--- | :--- | :--- |
| $r_{0}$ | $\bar{\delta}$ | $d b_{T}^{a}(\mathrm{WLS})$ | $d b_{T}^{b}(\mathrm{WLS})$ | $d b_{T}^{c}(\mathrm{WLS})$ |
| 0.2 | 1 | 0.0454 | 0.0451 | 0.0496 |
|  | -1 | 0.0468 | 0.0473 | 0.0513 |
| 0.3 | 1 | 0.0420 | 0.0412 | 0.0451 |
|  | -1 | 0.0447 | 0.0446 | 0.0487 |
| 0.4 | 1 | 0.0449 | 0.0436 | 0.0499 |
|  | -1 | 0.0443 | 0.0423 | 0.0471 |
| 0.5 | 1 | 0.0415 | 0.0406 | 0.0429 |
|  | -1 | 0.0435 | 0.0427 | 0.0456 |

### 2.4 Empirical Application

Previous literature has attempted to test the existence of breaks in mean macroeconomic growth rates, particularly because shocks that affect mean growth rates occur rarely and hence useful to model them as a one-time event. Bai, Lumsdaine, and Stock (1998) tests existence of a break in the output growth in postwar European and U.S., in which graphical evidence points to growth slowing down sometime in the 1970s. The authors finds that for the U.S. most of the
test statistics reject for the no-break hypothesis but the estimated confidence interval does not contain the growth slowdown in the 1970s. Eo and Morley (2015) applies an inverted likelihood ratio test and the testing method of Qu and Perron (2007) on postwar quarterly U.S. real GDP and consumption on nondurable goods and services. Under the assumption of no-break in the unconditional mean of the co-integrating relationships of output and consumption, they test for the break in the long-run growth rate. The results suggest weak evidence in the timing of the "productivity growth slowdown" in the early 1970s and the Great Moderation in the mid-1980s. It is likely that the directional break tests will be able to find stronger evidence of a structural break by specifying its direction. Section 2.4.1 investigates a structural break in U.S. real GDP growth rate and Section 2.4.2 considers U.S. labor productivity based on the work of Hansen (2001).

### 2.4.1 U.S. real GDP growth

We consider a simple univariate model to test for a structural break in postwar quarterly U.S. real GDP growth rate. Data are obtained from the Bureau of Economic Analysis (BEA) website for the sample period 1947Q1-2017Q2. Quantity indexes are seasonally adjusted and the base year is $2009(2009=100)$. We assume that $\log$ output has a stochastic trend with a drift and a finite-order representation. We test for a structural break in a $\operatorname{AR}(1)$ model for three cases. The lag order is selected using Kurozumi and Tuvaandorj's (2011) modified Bayesian information criterion (BIC), following Eo and Morley (2015) approach. The model selection method takes into account structural breaks, given an upper bound of four lags and four breaks in output growth. The first case is a break in the drift term only (the constant term $\gamma=0$ ), second case is a break in the lag coefficient only, i.e., "propagation term" $(\delta=0)$, and lastly a break in both constant and coefficient terms $\theta=(\gamma, \delta)^{\prime}$.

$$
\begin{equation*}
\Delta y_{t}=\beta+\phi \Delta y_{t-1}+\mathbf{1}\left\{t>\tau_{0}\right\}\left(\gamma+\delta \Delta y_{t-1}\right)+u_{t} . \tag{2.19}
\end{equation*}
$$

We assume the error term $\left\{u_{t}\right\}$ are serially uncorrelated mean zero disturbances with a break in variance: $E\left[u_{t}^{2}\right]=\sigma_{1}^{2}$ at $t \leq \tau_{0}$ and $E\left[u_{t}^{2}\right]=\sigma_{2}^{2}$ at $t>\tau_{0}$. At break date $\tau_{0}$, the long-run growth rate of $\log$ output change from $E\left[\Delta y_{t}\right]=\beta /(1-\phi)$ to $(\beta+\gamma) /(1-\phi-\delta)$ and the volatility of growth rate change from $\operatorname{Var}\left[\Delta y_{t}\right]=\sigma_{1}^{2} /\left(1-\phi^{2}\right)$ to $\sigma_{2}^{2} /\left(1-(\phi+\delta)^{2}\right)$. Previous literature have investigated the decrease in volatility of U.S. real GDP growth rate occurring in mid-1980s (Stock and Watson, 2002). We incorporate the change in variance by using WLS in addition to OLS estimation for directional break tests.

We would like to see if directional break tests reject the null hypothesis of no break for three cases. In the first two cases when a break occurs in one coefficient of an $\operatorname{AR}(1)$ model, we would expect a negative direction in $\operatorname{break}^{2} \bar{\delta}=-1$. For the third case when both coefficients are under a break, we set $\bar{\theta}=(-1,-1)^{\prime}$. For comparison we compute the Nyblom, AP and $q L L$ test statistics which do not allow break in variance. Table 2.10 computes all test statistics ignoring the break in volatility, using OLS residuals for directional test statistics (2.3), (2.4) and (2.5). Table 2.11 computes the WLS directional test statistics (2.6), (2.7) and (2.8).

Table 2.10: Autoregressive Model of U.S. Real GDP Growth 1947Q1-2017Q2: Directional break test statistics, Nyblom, AP and $q L L$ test statistics; **: Rejects the null hypothesis of no break at 5\% significance level; *: Rejects the null hypothesis of no break at $10 \%$ significance level.

|  | Test Statistic |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :---: |
| Null | $d b_{T}^{a}$ | $d b_{T}^{b}$ | $d b_{T}^{c}$ | Nyblom | $\operatorname{AP}(0)$ | $\operatorname{AP}(\infty)$ | $q L L$ |  |
| $\gamma=0$ | $2.5915^{* *}$ | $3.1663^{* *}$ | $2.0624^{* *}$ |  | $2.3120^{*}$ | 1.3797 | -5.8025 |  |
| $\delta=0$ | 1.5708 | 1.8752 | 1.2087 |  | 1.0295 | 0.6856 | -3.7586 |  |
| $\theta=0$ | $3.2003^{* *}$ | $4.8278^{* *}$ | $2.0725^{* *}$ | 0.5204 | 2.8558 | 1.7168 | -9.5463 |  |

For three different null hypotheses on the model (2.19), all three directional tests reject the null hypothesis of no break in the constant term, $\gamma=0$ and the null of no break in both coefficients,

[^11]Table 2.11: Autoregressive Model of U.S. Real GDP Growth 1947Q1-2017Q2: WLS directional break test statistics, Nyblom, AP and $q L L$ test statistics; *: Rejects the null hypothesis of no break at $10 \%$ significance level.

|  | Test Statistic |  |  |  |  |  |  |
| :---: | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Null | $d b_{T}^{a}(\mathrm{WLS})$ | $d b_{T}^{b}(\mathrm{WLS})$ | $d b_{T}^{c}(\mathrm{WLS})$ | Nyblom | $\operatorname{AP}(0)$ | $\operatorname{AP}(\infty)$ | $q L L$ |
| $\gamma=0$ | 1.5801 | 1.8595 | 1.2462 |  | $2.3120^{*}$ | 1.3797 | -5.8025 |
| $\delta=0$ | $1.9484^{*}$ | $2.2735^{*}$ | $1.5426^{*}$ |  | 1.0295 | 0.6856 | -3.7586 |
| $\theta=0$ | $2.1572^{*}$ | $2.9710^{*}$ | $1.5324^{*}$ | 0.5204 | 2.8558 | 1.7168 | -9.5463 |

Table 2.12: Autoregressive Model of U.S. Real GDP Growth 1947Q1-2017Q2; variance is estimated pre- and post-1984Q1: Directional break test statistics, Nyblom, AP and $q L L$ test statistics; **: Rejects the null hypothesis of no break at 5\% significance level; ***: Rejects the null hypothesis of no break at $1 \%$ significance level.

|  | Test Statistic |  |  |  |  |  |  |
| :---: | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Null | $d b_{T}^{a}$ | $d b_{T}^{b}$ | $d b_{T}^{c}$ | Nyblom | $\operatorname{AP}(0)$ | $\mathrm{AP}(\infty)$ | $q L L$ |
| $\gamma=0$ | 0.6799 | 0.7469 | -0.6634 |  | $2.9976^{* *}$ | $4.0594^{* * *}$ | $-9.4593^{* *}$ |
| $\delta=0$ | 0.6438 | 0.7017 | 0.6592 |  | 1.0860 | 0.6341 | -3.5761 |
| $\theta=0$ | 0.5896 | -0.6786 | -0.7701 | 0.5773 | 3.3682 | $4.2376^{* *}$ | -12.1004 |

Table 2.13: Autoregressive Model of U.S. Real GDP Growth 1960Q1-2017Q2: WLS directional break test statistics, Nyblom, AP, qLL test statistics and $\delta:=\left(\gamma, \delta_{1}\right)$. **: Rejects the null hypothesis of no break at 5\% significance level; *: Rejects the null hypothesis of no break at $10 \%$ significance level.

|  | Test Statistic |  |  |  |  |  |  |  |
| :---: | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :---: |
| Null | $d b_{T}^{a}(\mathrm{WLS})$ | $d b_{T}^{b}(\mathrm{WLS})$ | $d b_{T}^{c}(\mathrm{WLS})$ | Nyblom | $\mathrm{AP}(0)$ | $\mathrm{AP}(\infty)$ | $q L L$ |  |
| $\gamma=0$ | 1.6547 | 1.9511 | $1.3315^{*}$ |  | $2.5826^{*}$ | $1.6511^{*}$ | -7.1101 |  |
| $\delta=0$ | 1.5107 | 1.7033 | 0.9249 |  | 0.4966 | 0.4127 | -3.3896 |  |
| $\theta=0$ | $2.1731^{*}$ | $2.9902^{*}$ | $1.5238^{*}$ | 0.5662 | 3.0547 | 1.9861 | -10.3368 |  |

$\theta=0$ at $5 \%$ significance level. WLS directional tests reject the null hypothesis of no break in the lag coefficient, $\delta=0$ and the null $\theta=0$ at $10 \%$ significance level. Other structural break tests fail to reject the null of no break except for $\mathrm{AP}(0)$ which rejects the null $\gamma=0$ at $10 \%$ significance level. Therefore, all three directional break tests provide significant evidence of a break in U.S.
real GDP growth rate whereas Nyblom, AP and $q L L$ tests suggests weak or no evidence.
Allowing for $\sigma_{1}^{2} \neq \sigma_{2}^{2}$ using WLS estimation provides a weak evidence a break in the lag coefficient, in contrast to a significant break in the constant term using OLS. Nonetheless, it suggests a structural break while taking into account the shift in growth rate volatility, such as the Great Moderation, that might be caused from the change in second moment of disturbances.

If we assume a break in variance at a specific date we can allow volatility break in all structural break test statistics, including directional break test statistics using OLS estimators. Based on previous literature on the Great Moderation, we estimate the variance of the $\operatorname{AR}(1)$ disturbance term pre- and post-1984Q1. Hence $\hat{\sigma}_{1}^{2}=\tau^{-1} \sum_{t=1}^{\tau} \hat{u}_{t}^{2}$ where $\tau=1984 \mathrm{Q} 1$ and $\hat{\sigma}_{2}^{2}=$ $(T-\tau)^{-1} \sum_{t=\tau+1}^{T} \hat{u}_{t}^{2}$. The results in Table 2.12 are in striking contrast to Table 2.10. $\mathrm{AP}(0)$ and $q L L$ tests reject the null $\gamma=0$ at $5 \%$ significance level. $\operatorname{AP}(\infty)$ test rejects $\gamma=0$ at $1 \%$ significance level and $\theta=0$ at 5\% significance level. However, all directional break tests fail to find evidence of a break for all three hypotheses. We also compute tests for sub-sample 1960Q1-2017Q2. Table 2.13 shows that $\mathrm{AP}(0)$ and $\mathrm{AP}(\infty)$ test statistics reject the null hypothesis $\gamma=0$ at $10 \%$ significance level. WLS directional break tests reject the null $\theta=0$ at $10 \%$ significance level but $d b_{T}^{a}$ and $d b_{T}^{b}$ tests to reject $\gamma=0$.

Overall, allowing for a break in variance and assuming a negative direction in a $\operatorname{AR}(1)$ model of output growth rate leads to rejection of no break in the U.S. real GDP growth rate. Other structural break methods fail to reject the null of no break (or has weaker evidence of a break), unless we assume there is a break in the variance at time of Great Moderation. This suggests that using directional break tests on U.S. real GDP growth rate provides stronger evidence of the output growth rate slowdown.

### 2.4.2 U.S. labor productivity

We also consider the structural break in U.S. labor productivity, using the productivity data from Hansen (2001). Under a AR(1) model Hansen (2001) measures labor productivity
in the manufacturing/durables sector as the growth rate of the Industrial Production Index for manufacturing/durables to average weekly labor hours, a monthly time series from February 1947 to April 2001 (total 651 observations). The author finds that the test of Andrews (1993) rejects the null hypothesis of no structural break. Subsample estimates of mean growth rates (in annualized units) are $3.4 \%$ for 1947-1964, 2.5\% for 1964-1982, 4.2\% for 1982-1995 and 7.7\% for 1995-2001. Clearly there is an increase in the mean growth rate compared to previous history.

Under the model (2.19) and considering three cases of the change in coefficient parameters that causes a increase in the mean $E\left[\Delta y_{t}\right]$. We would expect a positive direction $\bar{\delta}=1$ for the change in the constant term or the lag coefficient. For the change in both terms, consider $\bar{\delta}=\left(\begin{array}{ll}1 & 1\end{array}\right)^{\prime}$. Table 2.14 shows that all three directional break tests and comparative break tests provides significant evidence of a break in the constant term $\gamma$ and both coefficients $\theta=(\gamma, \delta)^{\prime}$.

Table 2.14: Autoregressive Model of U.S. labor productivity Feb 1947-Apr 2001: Directional break test statistics, Nyblom, AP and $q L L$ test statistics; ***: Rejects the null hypothesis of no break at $1 \%$ significance level; **: Rejects the null hypothesis of no break at $5 \%$ significance level; *: Rejects the null hypothesis of no break at $10 \%$ significance level.

| Null | Test Statistic |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $d b_{T}^{a}$ | $d b_{T}^{b}$ | $d b_{T}^{c}$ | Nyblom | AP(0) | $\mathrm{AP}(\infty)$ | $q L L$ |
| $\gamma=0$ | $3.8714^{* * *}$ | 4.6545*** | 2.1862** |  | 4.0219*** | 4.2237*** | $-11.5064^{* * *}$ |
| $\delta=0$ | 0.7202 | 0.8299 | -0.0085 |  | 1.6134 | 1.4806 | -8.3101* |
| $\theta=0$ | 2.8498** | 3.6454* | 0.8081 | 1.0277** | 5.7627** | 4.4869** | $-19.6957 * * *$ |

We follow the sub-sample analysis of Hansen (2001) and partition the sample based on the break date estimate from the full sample (January 1982). All three directional break tests and four comparative tests fail to reject the null of no break in period [1947, 1982] but all tests find evidence of break during $[1982,2001]$ with results similar to Table 2.14. The break date estimate in $[1982,2001]$ is December 1994. For period [1947, 1994], only the Nyblom test finds evidence of a break at $10 \%$ significance level and others fail to reject the null of no break. The break date estimate in period $[1947,1994]$ is December 1963. For subsample period $[1964,1994]$ in Table
2.15, none of the tests finds evidence of a break at $10 \%$ significance level. Finally, taking the subsample [1964, 2001], all tests finds evidence of a structural break, similar to results in Table 2.14. In summary, directional break tests results are similar to other break tests in providing significant evidence of a break in U.S. labor productivity, for all sub-samples considered.

Table 2.15: Autoregressive Model of U.S. labor productivity Jan 1964-Dec 1994: Directional break test statistics, Nyblom, AP and $q L L$ test statistics; *: Rejects the null hypothesis of no break at $10 \%$ significance level.

|  | Test Statistic |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Null | $d b_{T}^{a}$ | $d b_{T}^{b}$ | $d b_{T}^{c}$ | Nyblom | $\operatorname{AP}(0)$ | $\operatorname{AP}(\infty)$ | $q L L$ |  |
| $\gamma=0$ | 1.6332 | 2.0089 | 1.1616 |  | 1.1319 | 0.7270 | -4.6487 |  |
| $\delta=0$ | 0.7011 | 0.7805 | -0.3196 |  | 0.3657 | 0.1685 | -3.0092 |  |
| $\theta=0$ | $1.9871^{*}$ | $3.1668^{*}$ | 1.2241 | 0.3006 | 1.5537 | 1.0235 | -7.9184 |  |

### 2.5 Conclusion

This paper provides three structural break tests when the direction of the coefficient change in a linear regression model is known. We construct test statistics that are directed toward the alternative by incorporating the known direction of a break, which leads to increase in power. We denote these break tests under known direction of a break as "directional break tests". The first two directional break test statistics are motivated from a likelihood ratio test statistic that tests the null of no break against the alternative hypothesis that is a weighted average of a break occurring at each potential break date. Thus, the first two directional break tests have optimality properties under i.i.d Gaussian disturbances and strictly exogeneous regressors. The third directional break test does not have optimality properties, but it has a null limit distribution equivalent to a standard normal distribution and thus simple to calculate.

The null asymptotic distribution of directional break test statistics are derived under
assumptions that allow for heterogeneous and autocorrelated error terms and regressors. We also allow the long-run variance of the error term to change simultaneously with coefficients, and provide directional break test statistics using weighted least squares.

In Monte Carlo simulations and empirical applications we compare the three directional break tests with other structural break tests widely used in literature. We consider five different DGPs under a structural break for Monte Carlo simulations. Results show that under correct directions, finite sample power of directional break tests are higher than that of conventional structural break tests. In particular, the case when a break occurs in the drift of a AR(1) model (Model 4) shows the most improvement in power. Two empirical applications are provided: postwar U.S. real GDP growth rate and U.S. labor productivity. Directional break tests suggest significant evidence of a break in U.S. real GDP growth rate whereas other break tests fail to find evidence of a break.

Chapter 2, in full, is currently being prepared for submission for publication of the material. Baek, Yaein.; Elliott, Graham. The dissertation author was a primary author of this paper.

## Chapter 3

## Forecasting in Long Horizons using Smoothed Direct Forecast

### 3.1 Introduction

In forecasting multiperiod time series we confront two different methods. The "iterated" forecast specifies a one-period-ahead model such as autoregression, then iterates forward to obtain the multiperiod horizon forecast. In contrast the "direct" forecast has each horizon specified in a model where the dependent variable on the left hand side of the regression is multiple periods ahead. The idea of direct forecasting goes back to Cox (1961) and Weiss (1991), where asymptotic properties of the direct forecasts under general conditions are established. Direct forecast methods are also used in estimating the impulse response of a dynamic system, referenced as local projections by Jordà (2005).

In theory the iterative method would provide more efficient estimates compared to the direct method if the model is correctly specified. For instance, suppose we want to forecast a time series by specifying an autoregression model with four lags for estimation. If the data-generating process (DGP) of this series is indeed a stationary autoregressive model with four lags or less,
then the iterated forecast will have smaller mean square forecast error (MSFE) than the direct method. Under a Gaussian process, the iterated method provides estimates that are asymptotically equivalent to the maximum likelihood estimator. Analytic expressions are provided by Bhansali (1996) and Ing (2003) under a non-Gaussian assumption.

Instead of the true lags of four, suppose we use an autoregressive model with two lags. Then the iterated forecasts are biased and the compounding misspecification error from recursive iteration would lead to larger bias in long horizons. In contrast, direct forecast is more robust to misspecification of an unknown DGP. The two forecasting methods involve a bias and variance trade-off; thus, a forecaster who is particularly interested in predicting long horizons would prefer the direct-forecast method over the iterated approach. Furthermore, the direct method is flexible because control variables can differ across horizons and it is relatively easy to estimate for non-linear dynamic systems.

Although obtaining long-horizon forecasts through the direct forecast method seems attractive due to its robustness, direct method estimates tend to be erratic across horizons. Because direct forecasting imposes less structure than the iterated method, the obtained forecasts are not "smooth" across horizons, contradicting what we would expect in theory. For example, Figure 3.1 is a replication of Owyang, Ramey, and Zubairy (2013) Figure 5, which shows the response of government spending to a news shock equal to 1 percent of GDP, based on quarterly data from 1920:I to 2011:IV for Canada. The direct method is used to estimate impulse responses instead of standard vector autoregressions (VAR) due to construction of government multipliers. The estimates show jagged shapes across time whereas in theory, impulse responses are smooth. Hence a forecaster using a direct method is likely to report unreliable multiperiod forecasts.

This view motivates us to provide a smoothness mechanism on direct forecasts to resolve the erratic behavior of estimates. We expect improvement in long-horizon forecast performance by imposing smoothness across multiperiod forecasts while maintaining flexibility of the direct method.


Figure 3.1: Government spending response to a news shock. The solid line is the impulse response estimated by the direct-forecast method. The dotted line is the corresponding two standard error band.

The main goal of this paper is to develop a method that can be implemented in direct forecasts to obtain long-horizon forecasts with improved performance. A smoothness prior is imposed across horizons of direct forecasts such that multiperiod forecasts show less jagged shapes as the horizon length increases. The new method is more robust to misspecification compared to the iterative method, conducted through a restricted regression in which we impose a smoothing parameter on the first differences of estimators, which is analogous to ridge regression. The smoothing parameter can be implemented as a prior distribution from a Bayesian perspective. Shiller (1973) introduced the concept of imposing a smoothness to the lag curve. We apply our method to time series where long-horizon forecasts are of interest: real oil prices and the U.S. macroeconomic time series from Marcellino, Stock, and Watson (2006). Both results show that our method shows improvement over the direct forecast approach in long horizons such as 3 to 5 years. For most series, forecasts based on our method have uniformly less MSFE than direct forecasts across horizons.

The rest of this paper is organized as follows. Section 3.2 briefly introduces the direct and iterated forecast methods and describes the estimation of our forecast model. Section 3.3 applies the method to forecasting real oil prices and the macroeconomic series and evaluates performance.

Section 3.4 includes an outline of further research and concluding remarks.

### 3.2 Smoothness Mechanism on Direct Forecasts

Section 3.2.1 introduces the difference in forecasts obtained from the direct and iterated methods in addition to the literature that compares the performance of the two forecasts. Section 3.2.2 describes the construction and intuition of our forecast method, which is based on the directforecast model. We explain how to choose the smoothing parameter used for our estimation.

### 3.2.1 Direct versus iterated forecasting

Several researchers have evaluated the performance of direct forecasts compared to iterated forecasts. Marcellino, Stock, and Watson (2006) compared the performance of iterated and direct forecasts using 170 U.S. monthly macroeconomic time series, spanning 1959 to 2002. A parametric bootstrap is conducted by assuming an autoregression as the DGP. This method allows examination of the spread of the distribution of MSFEs to see whether the direct method improves on the iterated method, on average, over the population of macroeconomic variables. Results show that iterated forecasts outperform the direct forecasts under correct specification. In contrast, Bhansali (1996) provided simulation results in which the direct method has a smaller MSFE than the iterated method if an underparameterized autoregressive model is fit on a generated autoregressive moving average process. Ing (2003) obtained asymptotic expressions of MSFE of the two methods in an $\operatorname{AR}(p)$ process with $1 \leq p \leq \infty$ and compared their performance. If the fitted order $k$ is such that $k<p$, the multistep MSFE of the direct forecast is less than that of the iterated forecast for almost all points in the parameter space. In addition, the paper shows that under certain nonstationary processes, the relative performance of direct and iterated forecasts are ambiguous.

We introduce properties of the direct and iterated forecast approach in the following
model. Assume observations $\left\{y_{t}\right\}$ are generated from a stationary autoregressive process and $\left\{y_{1}, \ldots, y_{t}\right\}$ are observable at time $t$. Suppose a forecaster is interested in obtaining $h=1, \ldots, H$ period ahead forecasts for this series, where $H$ is the maximum forecast horizon of interest. An iterated forecast would be obtained from an $\operatorname{AR}(p)$ model below.

$$
\begin{equation*}
y_{t+1}=\sum_{j=1}^{p} \phi_{j} y_{t+1-j}+\varepsilon_{t+1} . \tag{3.1}
\end{equation*}
$$

A one-period ahead regression model is specified and for multiperiod forecasts we iterate over equation (3.1). Suppose the companion form is expressed as $w_{t+1}=\psi w_{t}+v_{t+1}$. At time $T$, the iterated forecast is $\hat{w}_{T+h}=\hat{\psi}^{h} w_{T}$ for horizon $h=1, \ldots, H$.

In contrast, the direct forecasting method specifies different regression for each horizon in obtaining $h$ period-ahead forecasts. On the left side is the $h$ period-ahead variable of interest, and for the regressors we specify variables that are available at time $t$. In addition to the lags of $y_{t}$, we can include a multivariate exogenous variable $\mathbf{x}_{t}$ available at time $t$, unlike the iterated method.

$$
\begin{equation*}
y_{t+h}=\sum_{j=0}^{p-1} \beta_{h, j} y_{t-j}+\mathbf{x}_{t}^{\prime} \gamma_{h}+u_{t+h}, \quad h=1, \ldots, H \tag{3.2}
\end{equation*}
$$

The estimators $\left(\hat{\beta}_{h, 0}, \ldots, \hat{\beta}_{h, p-1}, \hat{\gamma}_{h}\right)$ depend on $h$ due to different regression models for each horizon. At time $T$ the direct forecast is $\hat{y}_{T+h}=\sum_{j=0}^{p-1} \hat{\beta}_{h, j} y_{T-j}+\mathbf{x}_{T}^{\prime} \hat{\gamma}_{h}$ for horizon $h$.

Suppose our data has $T$ periods of series $\left\{y_{t}, x_{t}\right\}$. Divide the sample into two subsamples: the first $R$ observations are included in the pseudo in-sample set and the remaining $T-R$ observations are in the pseudo out-of-sample set. Denote $\hat{y}_{t, t+h}$ as the $h$ horizon forecast using information available at time $t$ (this notation is necessary for our method due to joint estimation at each period). Forecasts are evaluated by comparing $h$ horizon MSFE for all horizons.

$$
\begin{equation*}
\operatorname{MSFE}(h)=\frac{1}{T-h-R+1} \sum_{t=R}^{T-h} \hat{u}_{t, t+h}^{2}, \quad \hat{u}_{t, t+h}=y_{t+h}-\hat{y}_{t, t+h} \tag{3.3}
\end{equation*}
$$

If the iterated forecast model (3.1) is correctly specified, the iterated MSFE is smaller than the direct MSFE, and under a Gaussian assumption, it is efficient. However if the model is misspecified, such as an incorrect lag order $p$, the misspecification error compounds across horizons in predicting multiperiod forecasts.

In practice we do not know the "true" lag order; thus, $p$ is likely to be misspecified. The advantage of the direct method is that it is robust to such misspecification. The drawback is when $h>1$, overlap in data affects the covariance of forecast errors, and serial correlation of the errors lead to erratic estimations across horizons. That is, the forecast error $\hat{u}_{t, t+h}$ in regression (3.2) is an MA $(h-1)$ process that needs to be incorporated in estimation.

To illustrate this, we use the VAR example from Jordà (2005). Suppose the true DGP is a $\operatorname{VAR}(p)$. Then the horizon $h$ forecast can be obtained by recursively substituting the companion form $\mathbf{W}_{t+h}=\Psi^{h} \mathbf{W}_{t}+\Psi^{h-1} \mathbf{v}_{t+1}+\cdots+\Psi \mathbf{v}_{t+h-1}+\mathbf{v}_{t+h}$, and therefore $\mathbf{y}_{t+h}$ is

$$
\begin{align*}
\mathbf{y}_{t+h}= & \Phi_{1}^{h} \mathbf{y}_{t}+\Phi_{2}^{h} \mathbf{y}_{t-1}+\cdots+\Phi_{p}^{h} \mathbf{y}_{t-p+1}  \tag{3.4}\\
& +\underbrace{\left(\mathbf{v}_{t+h}+\Phi_{1}^{1} \mathbf{v}_{t+h-1}+\cdots+\Phi_{1}^{h-1} \mathbf{v}_{t+1}\right)}_{\varepsilon_{t+h}}
\end{align*}
$$

where $\Phi_{i}^{h}$ is the $i$ th upper $(k \times k)$ block of the matrix $\Phi^{h}$. We can see that the error term at horizon $h$ is $\varepsilon_{t+h}$, a moving average of errors from time $t+1$ to $t+h$. Under known datagenerating processes, we can use this moving-average structure to construct the covariance matrix of horizon $h>1$ forecast errors and obtain an efficient estimate by using generalized least squares (GLS). In practice, however, the true DGP is unknown; hence, we cannot use the standard GLS formula. Because we do know that the forecast error $\varepsilon_{t+h}$ follows an $\mathrm{MA}(h-1)$ process in which heteroskedasticty and autocorrelation robust standard errors can be used.

Chang and Sakata (2007) provided asymptotic normality and consistency of direct forecast regression. The authors proposed the direct method to estimate impulse response functions using a two-stage procedure. First, innovations are estimated in a prior stage by a "long autoregression"
that fits an $\operatorname{AR}(p)$ model to the data. Then the data $y_{t}$ are regressed on the estimated innovation at lag $h$ to estimate the impulse response at horizon $h$. The local projection estimator by Jordà (2005) is a special case of this procedure; it uses an $h$ step-ahead linear prediction model which is equivalent to this two-stage method with the last $h$ observations discarded in the first stage.

### 3.2.2 Smoothed direct forecasts

To resolve the erratic behavior of direct forecasts across horizons, we impose a smoothness prior on estimators that is analogous to the method developed by Shiller (1973) imposing a Bayesian prior regarding the "smoothness" of the lag curve. The first degree smoothness priors will cause jagged shapes to be unlikely to occur; hence, we expect less erratic shapes by imposing a prior on first differences. A restricted regression that has a penalty term (smoothing parameter) on the first differences of coefficients are estimated, similar to the ridge regression. The resulting estimator is numerically equivalent to the Bayesian posterior mean if we assume a Gaussian prior distribution on parameters, where the smoothing parameter is embedded in the covariance matrix.

Consider the direct forecast regression of horizon $h$ that is univariate: $y_{t+h}=\mathbf{x}_{t}^{\prime} \beta_{h}+u_{t+h}$. The dependent variable $y_{t+h}$ is an element of multivariate variable $\mathbf{y}_{t+h}$, the regressor $\mathbf{x}_{t}$ consists of lags of $\mathbf{y}_{t}$ and a constant: $\mathbf{x}_{t}=\left(1, \mathbf{y}_{t}^{\prime}, \ldots, \mathbf{y}_{t-p+1}^{\prime}\right)^{\prime}$. Let $n=k p+1$ be the dimension of $\mathbf{x}_{t}$ and $\beta_{h}$ for each $h=1, \ldots, H$. We impose a smoothness prior through a smoothing parameter $\lambda(h, T)$ in the following restricted regression. The smoothing parameter $\lambda_{h}=\lambda(h, T)$ is a function of horizon $h$ and sample size $T$, which is described in detail later. Different units of variables are incorporated by standardizing the error term and first difference of coefficients. Thus we minimize the sum of squared terms that has a standard deviation of one. Denote $\sigma_{h}^{2}:=\lim _{T \rightarrow \infty} \operatorname{Var}\left(T^{-1 / 2} \sum_{t=p}^{T-h} u_{t+h}\right)$ and $\Omega_{h}:=\operatorname{diag}\left(\operatorname{Var}\left(\sqrt{T}\left(\hat{\beta}_{h}-\hat{\beta}_{h-1}\right)\right)\right)$ is a diagonal matrix that is equal to the diagonal elements of the covariance matrix of first differences.

$$
\begin{equation*}
\min _{\beta_{1}, \ldots, \beta_{H}}\left(\sum_{h=1}^{H} T^{-1} \sum_{t=p}^{T-p-H+1} \hat{\sigma}_{h}^{-2}\left(y_{t+h}-\mathbf{x}_{t}^{\prime} \beta_{h}\right)^{2}+\sum_{h=2}^{H} \lambda_{h}\left(\beta_{h}-\beta_{h-1}\right)^{\prime} \hat{\Omega}_{h}^{-1}\left(\beta_{h}-\beta_{h-1}\right)\right) . \tag{3.5}
\end{equation*}
$$

Since $\sigma_{h}$ and $\Omega_{h}$ are unknown, we replace them with the estimators $\hat{\sigma}_{h}$ and $\hat{\Omega}_{h}$ from the direct forecast regression. The FOC with respect to $\beta_{1}, \ldots, \beta_{H}$ results in the following estimator,

$$
\begin{gather*}
\left(\hat{\mathbf{V}}^{-1} \otimes T^{-1} \sum_{t} \mathbf{x}_{t} \mathbf{x}_{t}^{\prime}+\left[\begin{array}{ccccc}
\lambda_{2} \hat{\Omega}_{2}^{-1} & -\lambda_{2} \hat{\Omega}_{2}^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\
-\lambda_{2} \hat{\Omega}_{2}^{-1} & \left(\lambda_{2} \hat{\Omega}_{2}^{-1}+\lambda_{3} \hat{\Omega}_{3}^{-1}\right) & -\lambda_{3} \hat{\Omega}_{3}^{-1} & \cdots & \mathbf{0} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \lambda_{H} \hat{\Omega}_{H}^{-1}
\end{array}\right]\right)\left[\begin{array}{c}
\hat{\beta}_{1} \\
\hat{\beta}_{2} \\
\vdots \\
\hat{\beta}_{H}
\end{array}\right] \\
 \tag{3.6}\\
=\left[\begin{array}{c}
\hat{\sigma}_{1}^{-2} T^{-1} \sum_{t} \mathbf{x}_{t} y_{t+1} \\
\hat{\sigma}_{2}^{-2} T^{-1} \sum_{t} \mathbf{x}_{t} y_{t+2} \\
\vdots \\
\hat{\sigma}_{H}^{-2} T^{-1} \sum_{t} \mathbf{x}_{t} y_{t+H}
\end{array}\right],
\end{gather*}
$$

where $\mathbf{0}$ is a $(n \times n)$ matrix of zeros and $\hat{\mathbf{V}}:=\operatorname{diag}\left\{\hat{\sigma}_{1}^{2}, \hat{\sigma}_{2}^{2}, \ldots, \hat{\sigma}_{H}^{2}\right\}$. Denote the matrix with $\hat{\Omega}_{h}^{-1}$ terms as $\mathbf{M}^{-1}$. Then the above equation is equivalent to $\left(\hat{\mathbf{V}}^{-1} \otimes T^{-1} \sum_{t} \mathbf{x}_{t} \mathbf{x}_{t}^{\prime}+\mathbf{M}^{-1}\right) \hat{\boldsymbol{\beta}}=$ $\left(T^{-1} \sum_{t} \hat{\mathbf{V}}^{-1} \mathbf{Y}_{t} \otimes \mathbf{x}_{t}\right)$ where $\hat{\beta}=\left(\hat{\beta}_{1}^{\prime}, \ldots, \hat{\beta}_{H}^{\prime}\right)^{\prime}$ and $\mathbf{Y}_{t}:=\left(y_{t+1}, \ldots, y_{t+H}\right)^{\prime}$. From the FOC equation (3.6), note that all $H$ horizon coefficients are jointly estimated due to the first differences in the objective function (3.5). Thus a total Hn number of parameters are jointly estimated, which is likely to be large. However, even though there are a large number of coefficients to estimate, it is not difficult to compute. This is because estimated matrices with kronecker products $\hat{\mathbf{V}}^{-1}$ and $\hat{\Omega}_{h}^{-1}$ are diagonal for all $h$ so that $\mathbf{M}^{-1}$ consists of diagonal matrices. Therefore the FOC (3.6) looks complicated at first sight but easy to compute. Furthermore, the estimators can be interpreted as a Bayesian posterior mean where a smoothness prior is imposed on the direct
forecast model. Suppose we assume a Gaussian prior distribution on the coefficients $\beta$ with covariance matrix $\mathbf{M}$; that is, inverse of the matrix $\mathbf{M}^{-1}$ above. The prior mean is a $(\mathrm{nH} \times 1)$ vector $\mathbf{0}$ where all elements are zero. The posterior distribution of $\beta$ is

$$
\begin{equation*}
\beta \mid \mathbf{Y} \sim N\left(\mathbf{M}^{*}\left[T^{-1} \sum_{t} \hat{\mathbf{V}}^{-1} \mathbf{Y}_{t} \otimes \mathbf{x}_{t}\right], \mathbf{M}^{*}\right), \mathbf{M}^{*}=\left(\hat{\mathbf{V}}^{-1} \otimes T^{-1} \sum_{t} \mathbf{x}_{t} \mathbf{x}_{t}^{\prime}+\mathbf{M}^{-1}\right)^{-1} \tag{3.7}
\end{equation*}
$$

derived from conjugate Gaussian prior. The posterior mean is equivalent to the estimator from the restricted regression with a smoothing parameter on first differences. Thus we have used a Bayesian interpretation of the restricted least square with smoothing parameter $\lambda(h, T)$ that differs across standardized first difference of coefficients $\Omega_{h}^{-1 / 2} \Delta \beta_{h}$ with rate of convergence $O(1)$. If $\lambda_{h}$ is large, we are imposing a strong smoothness prior on horizon $h$ forecast toward horizon $h-1$ forecast.

The intuition of the smoothing parameter $\lambda(h, T)$ is to think of this as weights on the $h$ horizon loss function (the standardized square of first differences). A larger weight means more penalty on a large loss. That is, the smoothing parameter can be interpreted as the ratio of the weight on the sum of squared error and the weight on the square of standardized first differences. Suppose we impose weights $\left(1-\omega_{h}\right)$ and $\omega_{h}$ on each term of the objective function (3.5).

$$
\sum_{h=1}^{H}\left(1-\omega_{h}\right) T^{-1} \sum_{t=p}^{T-p-H+1} \hat{\sigma}_{h}^{-2}\left(y_{t+h}-\mathbf{x}_{t}^{\prime} \beta_{h}\right)^{2}+\sum_{h=2}^{H} \omega_{h}\left(\beta_{h}-\beta_{h-1}\right)^{\prime} \hat{\Omega}_{h}^{-1}\left(\beta_{h}-\beta_{h-1}\right)
$$

Then, for each horizon, we minimize a weighted average of loss functions: the objective function of the direct-method estimator and the first differences of coefficients across horizons. This objective function is identical to (3.5) with $\lambda_{h}=\omega_{h} /\left(1-\omega_{h}\right)$. If we assign equal weights $\omega_{h}=0.5$ on both objective functions then $\lambda_{h}=1$. In the two extreme cases, $\omega_{h} \rightarrow 0$ and $\omega_{h} \rightarrow 1$ implies $\lambda_{h} \rightarrow 0$ and $\lambda_{h} \rightarrow \infty$, respectively. If $\lambda_{h}$ approaches zero, we obtain estimates that are similar to the direct estimates whereas a large value of $\lambda_{h}$ results in constant coefficients across horizons.

To incorporate different penalties on different horizons and sample size, a linear and an exponential function of $h$ and $T$ is adapted. The exponential function is

$$
\lambda(h, T)=\exp (h / \sqrt{T})
$$

where weight on the sum of squared error and standardized first differences of coefficients are as below, respectively.

$$
1-\omega_{h}=\frac{1}{1+\exp (h / \sqrt{T})}, \quad \omega_{h}=\frac{\exp (h / \sqrt{T})}{1+\exp (h / \sqrt{T})} .
$$

The exponential smoothing parameter increases with horizon length and shrinks as sample size increases. The parameter increases with $h$ because we want to "pull" toward the previous horizon estimator that is more reliable and shorter horizons are estimated better, relative to longer horizon estimates. Furthermore, if $h$ is large relative to sample size $T$, we would like to rely on shorter horizon estimators due to lack of information.

In general we can modify the exponential smoothing parameter through multiplication of constants inside and outside the exponential term, such as $a \exp (b \cdot h / \sqrt{T})$. However this generalization is rather unnecessary; as the smoothing parameter increases exponentially in longer horizons, the weight $\omega_{h}$ on $\Delta \beta_{h}$ is close to one and different values of $(a, b)$ will not make a large difference in estimates. In addition, empirical results in section 3.3 show us that forecasts are robust to the choice of the smoothing parameter.

One may think of imposing smoothness after obtaining direct forecast estimates instead of incorporating smoothness in the estimation stage. This is equivalent to a linear function of direct forecast estimates $\hat{\beta}=\left(\hat{\beta}_{1}^{\prime}, \ldots, \hat{\beta}_{H}^{\prime}\right)^{\prime}$. Our method is a special case in which we take a particular linear function (see Appendix A. 3 for details). We choose to incorporate smoothing in the estimation stage because it is easier to interpret this linear function. For instance, if we want to smooth out the direct estimates in a simple weighted average of horizon $h$ and $h-1$ estimates,
it is unclear how to choose the weights intuitively.
From now on, we denote estimators obtained by imposing smoothness prior on direct forecast regressions through ridge regression (3.5) as "smoothed direct forecasts". In addition to the smoothing parameter on first differences, we can impose a penalty term on coefficients of stationary variables themselves. Suppose prior information shows that the impulse response to a shock dies out quickly. Then a prior that is analogous to a shrinkage estimator is imposed on long-horizon impulse responses. We add in an additional parameter $\gamma_{h}$ in regression (3.5).

$$
\begin{align*}
\min _{\beta_{1}, \ldots, \beta_{H}}\left(\sum_{h=1}^{H} T^{-1} \sum_{t=p}^{T-p-H+1}\right. & \hat{\sigma}_{h}^{-2}\left(y_{t+h}-\mathbf{x}_{t}^{\prime} \beta_{h}\right)^{2} \\
& \left.+\sum_{h=2}^{H} \lambda_{h}\left(\beta_{h}-\beta_{h-1}\right)^{\prime} \hat{\Omega}_{h}^{-1}\left(\beta_{h}-\beta_{h-1}\right)+\sum_{h=1}^{H} \gamma_{h} \beta_{h}^{\prime} \hat{\mathbf{W}}_{h}^{-1} \beta_{h}\right) \tag{3.8}
\end{align*}
$$

The parameter $\gamma_{h}$ is set to be zero for shorter horizons $h=1, \ldots, l-1$ and a function of $h$ and $T$ for longer horizons $h=l, \ldots, H$. The FOC with respect to $\beta_{1}, \ldots, \beta_{h}$ can be written similar to (3.6) so that the inverse prior covariance matrix $\mathbf{M}^{-1}$ is modified by adding a matrix that has $\left\{\gamma_{1} \hat{\mathbf{W}}_{1}^{-1}, \gamma_{2} \hat{\mathbf{W}}_{2}^{-1}, \ldots, \gamma_{H} \hat{\mathbf{W}}_{H}^{-1}\right\}$ as the diagonal $(n \times n)$ block elements and zero otherwise.

### 3.3 Empirical Applications

### 3.3.1 Real oil prices

Forecasts of the price of crude oil have been considered a key variable in predicting macroeconomic risks and generating forecasts of macroeconomic outcomes. In this section, direct forecasts and smoothed direct forecasts of the price of real oil at horizons up to 5 years are obtained and then the MSFE are compared.

First we check whether the direct forecasts of real oil prices obtain smaller MSFE relative to the iterated forecast in long horizons. The measures for real oil prices are obtained from
monthly nominal oil prices and the U.S. CPI as the deflator. We use two nominal oil price measures obtained from the U.S. Energy Information Administration (EIA). The West Texas Intermediate (WTI) price is the spot price for crude oil in dollars per barrel, and the U.S. refiners' acquisition cost (RAC) of imported crude oil in dollars per barrel ${ }^{1}$. The WTI price sample period is 1986.1-2016:7 with size $T=367$, and the RAC imported price sample period is 1974.1-2016:7 with size $T=511 .{ }^{2}$ The in-sample data set (estimation sample) is 1986:1-2001:12 for the WTI price with size $R=192$, and 1974:1-2001:12 for the RAC price with size $R=336$. The sub-sample used for to evaluate out-of-sample forecasts (validation sample) is 2002:1-2011:8 for both measures ${ }^{3}$, with size $P=175$. The out-of-sample forecasts are obtained from recursive estimation. For the WTI price, the iterated forecasts are obtained from an autoregressive model with lags selected by SIC $(p=2)$ and AIC $(p=8)$. Likewise, direct forecasts are computed under lags $p=2$ and $p=8$. Both criteria select the same lags $p=3$ for the RAC imported price. A bootstrap $p$-value of the mean relative MSFE ratios is constructed under the null hypothesis that the iterated model is efficient. ${ }^{4}$

If the MSFE ratio is less than one, then the direct forecast performs better than the iterated forecast. Tables 3.1 and 3.2 show that direct forecasts performs worse than iterated forecasts for all horizons and slightly less so in longer horizons for both measures. Bootstrap p-values shows that when rejecting the null hypothesis, the iterated forecast performs better and fails to reject all horizons; the $p$-value is almost one at short horizons but decreases to 0.6 at 5 years. Hence the MSFE difference of the two methods are statistically insignificant. If the smoothed direct-forecast

[^12]method is implemented in predicting in longer horizons, we expect to obtain smaller MSFE values compared to direct forecasts and also iterated forecasts in long horizons.

Table 3.1: Relative MSFEs of univariate direct forecast and iterated forecast (RAC). The entry in parentheses is the $p$-value of the hypothesis test that the iterated model is efficient against the alternative that the direct model is more efficient, using the parametric bootstrap algorithm.

|  | Forecast Horizon |  |  |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :---: | :---: | :---: |
| Lag | 3 | 6 | 12 | 24 | 36 | 48 | 60 |  |  |  |
| AR(3) | 1.4885 | 1.0250 | 1.0584 | 1.1014 | 1.0935 | 1.1587 | 1.1931 |  |  |  |
|  | $(1.000)$ | $(0.888)$ | $(0.758)$ | $(0.668)$ | $(0.558)$ | $(0.628)$ | $(0.656)$ |  |  |  |

Table 3.2: Relative MSFEs of univariate direct forecast and iterated forecast (WTI). The entry in parentheses is the $p$-value of the hypothesis test that the iterated model is efficient against the alternative that the direct model is more efficient, using the parametric bootstrap algorithm.

|  | Forecast Horizon |  |  |  |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :---: | :---: | :---: | :---: |
| Lag | 3 | 6 | 12 | 24 | 36 | 48 | 60 |  |  |  |  |
| AR(2) | 1.5141 | 1.3889 | 1.8894 | 2.1791 | 1.1410 | 1.1003 | 1.3120 |  |  |  |  |
|  | $(0.988)$ | $(0.924)$ | $(0.950)$ | $(0.930)$ | $(0.466)$ | $(0.400)$ | $(0.594)$ |  |  |  |  |
| AR(8) | 2.4617 | 2.2329 | 2.7092 | 2.5984 | 1.2709 | 1.3172 | 1.6789 |  |  |  |  |
|  | $(0.994)$ | $(0.972)$ | $(0.966)$ | $(0.934)$ | $(0.528)$ | $(0.582)$ | $(0.770)$ |  |  |  |  |

We compute the smoothed direct forecasts with different smoothing parameters and compare MSFE values with the direct forecasts. The direct and smoothed direct forecast models have the same lag length $p$. Let $y_{t}$ be the series of interest, in our case the real oil price. Denote $\mathbf{x}_{t}:=\left(1, y_{t}, \ldots, y_{t-p+1}\right)^{\prime}$. The direct forecast of horizon $h$ is obtained from estimating the following regression

$$
\begin{equation*}
y_{t+h}=\gamma_{0 h}+\gamma_{1 h} y_{t}+\cdots+\gamma_{p h} y_{t-p+1}+u_{t+h} \equiv \mathbf{x}_{t}^{\prime} \gamma_{h}+u_{t+h}, \quad t=1, \cdots, R . \tag{3.9}
\end{equation*}
$$

where the estimate of $\gamma_{h}$ is $\hat{\gamma}_{h, t}=\left(\sum_{j=1}^{t} \mathbf{x}_{j} \mathbf{x}_{j}^{\prime}\right)^{-1}\left(\sum_{j=1}^{t} \mathbf{x}_{j} y_{j+h}\right)$. The $h$ horizon direct forecast
at time $t$ is $\hat{y}_{t, t+h}=\mathbf{x}_{t}^{\prime} \hat{\gamma}_{h, t}$. The residual is $\hat{u}_{t, t+h}=y_{t+h}-\mathbf{x}_{t}^{\prime} \hat{\gamma}_{h, t}$ for $t=R+1, \ldots, T-h$ in which we compute the MSFE of horizons $h, P^{-1} \sum_{t=R+1}^{T-h} \hat{u}_{t, t+h}^{2}$ where $P=T-h-R$ and $h \in$ $\{3,6,12,24,36,48,60\}$. Denote the dimension of $\hat{\gamma}_{h, t}$ as $k:=p+1$.

The smoothed direct forecast is obtained from the restricted regression (3.5). The smoothed direct forecast also has the same model specification (3.9) but different coefficient estimates for $h>1$. Denote $\hat{\beta}_{h, t}$ as the vector of estimated coefficients so that $\hat{y}_{t, t+h}=\mathbf{x}_{t}^{\prime} \hat{\beta}_{h, t}$ is the $h$ horizon smoothed direct forecast at time $t$. The estimator $\hat{\beta}_{h, t}$ is the $(h-1) k+1$ to $h k$ elements of the estimator $\hat{\mathbf{B}}_{t}$.

$$
\hat{\mathbf{B}}_{t}=\left(\hat{\mathbf{V}}_{t}^{-1} \otimes t^{-1} \sum_{j=1}^{t} \mathbf{x}_{j} \mathbf{x}_{j}^{\prime}+\mathbf{M}^{-1}\right)^{-1}\left(t^{-1} \sum_{j=1}^{t} \hat{\mathbf{V}}_{t}^{-1} \mathbf{y}_{j} \otimes \mathbf{x}_{j}\right) .
$$

where $\mathbf{y}_{j}=\left(y_{j+1}, \cdots, y_{j+60}\right)^{\prime}$ and $\hat{\mathbf{V}}_{t}^{1 / 2}$ is a $(60 \times 60)$ diagonal matrix that consists of standard estimates $\hat{\sigma}_{h t}=\left(t^{-1} \sum_{j=1}^{t} \hat{u}_{j, j+h}^{2}\right)^{1 / 2}$ for $h=1, \ldots, 60$ from (3.9). $\mathbf{M}$ is the prior covariance matrix of $\mathbf{B}_{t}$ and the coefficients are jointly estimated for horizons $h=1, \ldots, 60$. The smoothing parameter is an exponential function $\lambda_{h}=a \exp (b h / \sqrt{T})$ and uses different constant combinations $(a, b)$ for $a \in\{1,50,500\}$ and $b \in\{1,2\}$.

Both forecasts are obtained from recursive estimation so that the in-sample size changes for each out-of-sample date, i.e. $\hat{y}_{t, t+h} \neq \hat{y}_{t+1, t+h}$, the forecasts for period $t+h$ are different when we forecast at time $t$ for horizon $h$ and at time $t+1$ for horizon $h-1$. We compare the forecast accuracy of direct and smoothed direct forecasts using West's (1996) test procedure, which provides a tool to test predictive accuracy when the predictions rely on regression estimates. Details of the test statistic and null hypothesis are in Appendix A.3.

Tables 3.3 and 3.4 exhibit the MSFE values of direct and smoothed direct forecast of RAC and WTI price, respectively, for different $(a, b)$ values of the smoothing parameter. Compared to direct forecasts, the MSFE of smoothed direct forecasts are smaller except for the short horizon $(h=3)$ where it increases for the WTI price. For the RAC real oil price, the MSFE of smoothed
direct forecasts are less than that of direct forecasts in all horizons. The parentheses show the $p$ values of the hypothesis test that the MSFE of direct and smoothed direct forecasts are equivalent. For both measures and all horizons, the $p$-values are larger than .9 ; thus the tables omit $p$-values other than $\lambda_{h}=\exp (h / \sqrt{T})$.

For both WTI and RAC real oil prices, the MSFE of smoothed direct forecasts are less than direct forecasts and for WTI prices, particularly at long horizons. However, West's test show that the MSFE differences of smoothed direct and direct forecasts are statistically insignificant at $10 \%$ significance level for all $(a, b)$ values considered and both measures.

Between the smoothed direct forecasts with different smoothing parameters, various constants $(a, b)$ in the exponential function have small differences and are sometimes almost identical in short horizons. Consider the MSFE values of the direct forecast $\left(\lambda_{h}=0\right)$ and smoothed direct forecast with parameter $\lambda_{h}=1$ and $\lambda_{h}=500 \exp (2 h / \sqrt{T})$ at horizon 60 months under $\operatorname{AR}(8)$ in table 3.4. The MSFE decreases approximately 0.91 to 0.64 when $\lambda_{h}=0$ changes to 1 , but when it changes from $\lambda_{h}=1$ to an exponential function with $a=500$, it decreases to 0.50 . The $p$-value of West's test statistic is approximately .9999 for both $\lambda_{h}=1$ and $\lambda_{h}=500 e^{2 h T^{-1 / 2}}$. Therefore the difference of smoothed direct forecast on direct forecast is robust to the smoothing parameter, which implies that we do not need to worry about determining $\lambda_{h}$ in estimation.

So far we have seen that the smooth direct forecast results in improvement at long horizon such as 4 or 5 years. Although West's test results imply that the difference in MSFE of the direct and the smoothed direct forecasts are statistically insignificant, it does not imply that the improvement is economically insignificant. For instance, consider the root mean squared error (RMSFE) in evaluating forecast accuracy at horizon 5 years. Using the WTI real oil price with 8 lags, the direct forecast RMSFE is 0.9554 and the smoothed direct forecast with $\lambda_{h}=500 e^{2 h T^{-1 / 2}}$ has RMSFE is 0.7049 , in $\log$ scale. In units of real oil price, the forecast accuracy of the direct and the smoothed direct method are approximately 2.60 and 2.02 (deflated) U.S. dollars per barrel, respectively. Then the difference between the two forecasts is 0.58 real U.S. dollars per barrel,
which is a significant difference considering that $2016: 7$ oil price is 0.056 real U.S dollars per barrel.

Furthermore the smoothed direct forecasts are robust to the choice of the smoothing parameter, whether it is a constant or an exponential function. Because we did not construct an optimal smoothing parameter that depends on data, we do not want forecasts to drastically deviate from direct forecasts due to our choice of parameter. In conclusion, our results show that the smoothed direct method is preferable in forecasting 4 to 5 years forward for real oil prices.

Table 3.3: Pointwise MSFE of univariate direct forecast and smoothed direct forecast (RAC). The entry in parentheses is the $p$-value of the hypothesis test that the MSFE of direct and smoothed direct forecasts at horizon $h$ are equivalent.

| Lag | $\lambda_{h}$ | Forecast Horizon |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 3 | 6 | 12 | 24 | 36 | 48 | 60 |
| AR(3) | Direct | 0.0856 | 0.1037 | 0.1496 | 0.2326 | 0.3441 | 0.4578 | 0.5211 |
|  | 1 | 0.0838 | 0.1118 | 0.1416 | 0.2178 | 0.3280 | 0.4296 | 0.5014 |
|  | $h T^{-1 / 2}$ | 0.0839 | 0.1120 | 0.1435 | 0.2220 | 0.3273 | 0.4268 | 0.5039 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 0.0839 \\ & (0.9988) \end{aligned}$ | $\begin{aligned} & 0.1119 \\ & (0.9968) \end{aligned}$ | $\begin{aligned} & 0.1415 \\ & (0.9981) \end{aligned}$ | $\begin{aligned} & 0.2126 \\ & (0.9978) \end{aligned}$ | $\begin{aligned} & 0.3114 \\ & (0.9978) \end{aligned}$ | $\begin{aligned} & 0.4178 \\ & (0.9978) \end{aligned}$ | $\begin{aligned} & 0.5101 \\ & (0.9994) \end{aligned}$ |
|  | $e^{2 h T^{-1 / 2}}$ | 0.0839 | 0.1119 | 0.1415 | 0.2107 | 0.3054 | 0.4115 | 0.5055 |
|  | $50 e^{2 h T^{-1 / 2}}$ | 0.0840 | 0.1116 | 0.1398 | 0.1976 | 0.2797 | 0.3792 | 0.4688 |
|  | $500 e^{2 h T^{-1 / 2}}$ | 0.0841 | 0.1114 | 0.1379 | 0.1901 | 0.2675 | 0.3630 | 0.4489 |

Table 3.4: Pointwise MSFE of univariate direct forecast and smoothed direct forecast (WTI). The entry in parentheses is the $p$-value of the hypothesis test that the MSFE of direct and smoothed direct forecasts at horizon $h$ are equivalent. The $p$-values under $\operatorname{AR}(8)$ are omitted; $p$-values are approximately 1 for all horizons and smoothing parameters.

| Lag |  | Forecast Horizon |  |  |  |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :---: | :---: | :---: | :---: |
|  | $\lambda_{h}$ | 3 | 6 | 12 | 24 | 36 | 48 | 60 |  |  |  |  |
|  | Direct | 0.0844 | 0.1649 | 0.3947 | 0.7555 | 0.5611 | 0.6333 | 0.7923 |  |  |  |  |
|  | 1 | 0.1072 | 0.1390 | 0.1857 | 0.2778 | 0.4179 | 0.5078 | 0.5927 |  |  |  |  |
|  | $h T^{-1 / 2}$ | 0.1055 | 0.1474 | 0.1994 | 0.2853 | 0.3945 | 0.4758 | 0.5575 |  |  |  |  |
|  | $e^{h T^{-1 / 2}}$ | 0.1124 | 0.1487 | 0.1920 | 0.2608 | 0.3455 | 0.4296 | 0.5108 |  |  |  |  |
|  |  | $(0.9999)$ | $(1.0000)$ | $(0.9997)$ | $(0.9996)$ | $(0.9999)$ | $(0.9999)$ | $(0.9999)$ |  |  |  |  |
|  | $e^{2 h T^{-1 / 2}}$ | 0.1153 | 0.1507 | 0.1902 | 0.2496 | 0.3353 | 0.4216 | 0.4791 |  |  |  |  |
|  | $50 e^{2 h T^{-1 / 2}}$ | 0.1156 | 0.1507 | 0.1901 | 0.2497 | 0.3378 | 0.4191 | 0.4839 |  |  |  |  |
|  | $500 e^{2 h T^{-1 / 2}}$ | 0.1151 | 0.1151 | 0.1918 | 0.2513 | 0.3392 | 0.4190 | 0.4810 |  |  |  |  |
|  | Direct | 0.1258 | 0.2336 | 0.4441 | 0.6560 | 0.4846 | 0.6344 | 0.9127 |  |  |  |  |
|  | 1 | 0.1465 | 0.1668 | 0.1881 | 0.2916 | 0.4150 | 0.5719 | 0.6414 |  |  |  |  |
|  | $h T^{-1 / 2}$ | 0.1509 | 0.1767 | 0.2037 | 0.3054 | 0.4135 | 0.4956 | 0.5767 |  |  |  |  |
| AR(8) | $e^{h T^{-1 / 2}}$ | 0.1565 | 0.1764 | 0.1956 | 0.2810 | 0.3690 | 0.4547 | 0.5365 |  |  |  |  |
|  | $e^{2 h T^{-1 / 2}}$ | 0.1590 | 0.1780 | 0.1941 | 0.2699 | 0.3564 | 0.4519 | 0.5242 |  |  |  |  |
|  | $50 e^{2 h T^{-1 / 2}}$ | 0.1599 | 0.1765 | 0.1896 | 0.2676 | 0.3573 | 0.4440 | 0.5061 |  |  |  |  |
|  | $500 e^{2 h T^{-1 / 2}}$ | 0.1592 | 0.1770 | 0.1913 | 0.2687 | 0.3565 | 0.4399 | 0.4969 |  |  |  |  |

### 3.3.2 Macroeconomic variables

In this section we replicate the simulation in Marcellino, Stock, and Watson (2006) using a partition of their data and conduct a test of forecast accuracy of direct and smoothed direct forecasts on each series. The data consists of 80 U.S. monthly macroeconomic series from the Basic Economics database (IHS Global Insight). Marcellino et al. (2006) categorized 170 series into five categories of which we use 80 series from four categories described below. Detailed description of the data is available in Appendix A.3.
(A) Income, output, sales and capacity use ( 26 series)
(B) Employment and unemployment (24 series)
(C) Construction, inventories and orders (17 series)
(D) Interest rate and asset prices (13 series).

For each series we obtain the MSFE of the smoothed direct forecast, direct forecast, and iterated forecast. Pseudo out-of-sample forecasts are obtained starting from 1990:1 and the final forecast date is the last available observation (2016:6) minus the forecast horizon $h$. A univariate model with 4 lags and a constant is specified for each series and the subsample used for estimation starts at 1974:4 for all 80 series. Marcellino et al. (2006) specified $\operatorname{AR}(p)$ using four different lag selections: $p=4, p=12$ and lags selected by AIC and BIC. Instead, we use a fixed lag $p=4$ for all series.

Of 80 series, we only display variables that have an MSFE ratio of the smoothed direct forecast relative to the direct forecast less than one, and MSFE ratio relative to the iterated forecast less than one for any horizon $h \in\{3,6,12,24,36,48,60\}$ to check for improvement on the iterated approach as well. A total 26 of 80 series showed improvement on direct and iterated forecasts, indicating our method improves on the direct forecast, but fails to outperform the iterated forecast for most macroeconomic series.

Tables 3.5, 3.6, 3.7 and 3.8 show the MSFE ratio of smoothed direct forecasts relative to direct forecasts for each category A, B, C and D, respectively. For most series in category B (employment and unemployment) and category C (construction, inventories and orders), direct forecasts have extremely large MSFE compared to iterative forecasts. Due to the poor performance of the direct method, the relative ratio of smoothed direct forecasts are quite small, although the MSFE ratio relative to the iterated forecast would be larger than one or close to one. These variables are also omitted.

A huge change emerged in the MSFE ratio for series in category $B$ in which the smoothed direct forecast performance improves as horizon length increases. The difference in MSFE of direct and smoothed direct forecasts remains statistically insignificant. Category A series such as Industrial Production Index business equipment or Personal Consumption Index services show statistically significant differences in MSFE, where the direct forecast outperforms in the
short horizon (3 months) and the smooth direct forecast outperforms in longer horizons. In outperforming direct and iterated forecasts, category D has 7 of 13 variables in which the MSFE of the smoothed direct forecast is less than iterated and direct forecasts, although the improvement is statistically insignificant for all horizons and series. The $p$-values are omitted except for the exponential smoothing parameter when $(a, b)=(1,1)$. Similar to the results in section 3.1, the MSFE ratios change only slightly as constants $(a, b)$ vary. Large values of the smoothing parameter, such as $a=500$, are omitted due to difficulty in computation. ${ }^{5}$

Although statistically insignificant, it is plausible that there is economically significant improvement in forecasts. For example, consider personal consumption expenditure of durable goods in Table 3.5. Under the smoothing parameter $\lambda_{h}=\exp (h / \sqrt{T})$, the 60 -month horizon MSFE ratio of smoothed direct and direct forecast is 0.2349 so the RMSFE is approximately 0.485. The difference in MSFE of the two methods is statistically insignificant with a $p$-value of .94 from West's test. However, it turns out that this difference is economically significant. Suppose personal consumption expenditure of durable goods (denote this variable as PCE for simplicity) has a $4 \%$ growth rate on average, and the direct forecast of PCE growth rate 5 years ahead is $10 \%$. Reverting the transform of log first differences, the smoothed direct forecast of PCE growth rate 5 years ahead is $6 \%$. Hence the forecast error of PCE growth rate is $2 \%$ for the smoothed direct forecast whereas it is $6 \%$ for the direct forecast, which is a large difference economically.

[^13]Table 3.5: Category (A) MSFE ratio of univariate smoothed direct forecast and direct forecast. The entry in parentheses is the $p$-value of hypothesis test that horizon $h$ MSFE of direct and smoothed direct forecasts are equivalent.

| Variable | $\lambda_{h}$ | Forecast Horizon |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 3 | 6 | 12 | 24 | 36 | 48 | 60 |
| IPI, Business equipment | 1 | 10.4844 | 8.8916 | 3.2269 | 0.9705 | 0.5272 | 0.4041 | 0.3242 |
|  | $h T^{-1 / 2}$ | 10.4394 | 8.8666 | 3.2278 | 0.9720 | 0.5281 | 0.4046 | 0.3244 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 10.6011 \\ & (0.009) \end{aligned}$ | $\begin{aligned} & 9.0112 \\ & (0.764) \end{aligned}$ | $\begin{aligned} & 3.2285 \\ & (0.450) \end{aligned}$ | $\begin{aligned} & 0.9726 \\ & (0.763) \end{aligned}$ | $\begin{aligned} & 0.5274 \\ & (0.933) \end{aligned}$ | $\begin{aligned} & 0.4034 \\ & (0.969) \end{aligned}$ | $\begin{aligned} & 0.3228 \\ & (0.986) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 10.6898 | 9.0917 | 3.2278 | 0.9728 | 0.5270 | 0.4028 | 0.3223 |
| IPI, Non-durable goods materials | 1 | 26.2042 | 16.6506 | 19.6070 | 3.2278 | 1.4184 | 0.7619 | 0.7229 |
|  | $h T^{-1 / 2}$ | 27.0694 | 16.7260 | 19.4242 | 3.2052 | 1.4096 | 0.7667 | 0.7587 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 24.8913 \\ & (0.912) \end{aligned}$ | $\begin{aligned} & 15.9290 \\ & (0.904) \end{aligned}$ | $\begin{aligned} & 19.3219 \\ & (0.975) \end{aligned}$ | $\begin{aligned} & 3.2057 \\ & (0.977) \end{aligned}$ | $\begin{aligned} & 1.4159 \\ & (0.983) \end{aligned}$ | $\begin{aligned} & 0.7799 \\ & (0.986) \end{aligned}$ | $\begin{aligned} & 0.7853 \\ & (0.992) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 23.2206 | 15.1815 | 19.1181 | 3.2154 | 1.4308 | 0.7914 | 0.7971 |
| IPI, <br> Manufacturing | 1 | 39.4572 | 8.9542 | 3.3762 | 0.9569 | 0.5387 | 0.3790 | 0.2163 |
|  | $h T^{-1 / 2}$ | 36.3910 | 9.1037 | 3.4054 | 0.9563 | 0.5299 | 0.3780 | 0.2194 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 43.0451 \\ & (0.839) \end{aligned}$ | $\begin{aligned} & 9.6027 \\ & (0.694) \end{aligned}$ | $\begin{aligned} & 3.4378 \\ & (0.724) \end{aligned}$ | $\begin{aligned} & 0.9413 \\ & (0.851) \end{aligned}$ | $\begin{aligned} & 0.5217 \\ & (0.930) \end{aligned}$ | $\begin{aligned} & 0.3745 \\ & (0.960) \end{aligned}$ | $\begin{aligned} & 0.2193 \\ & (0.906) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 53.5834 | 11.1153 | 3.4981 | 0.9253 | 0.5092 | 0.3673 | 0.2159 |
| IPI, <br> Nondurable <br> goods <br> manufacturing | 1 | 68.1065 | 14.0836 | 20.1583 | 3.2446 | 1.4569 | 0.7821 | 0.7895 |
|  | $h T^{-1 / 2}$ | 64.5383 | 41.6229 | 20.3190 | 3.2365 | 1.4397 | 0.7847 | 0.8061 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 70.0496 \\ & (0.870) \end{aligned}$ | $\begin{aligned} & 41.4488 \\ & (0.919) \end{aligned}$ | $\begin{aligned} & 20.1801 \\ & (0.955) \end{aligned}$ | $\begin{aligned} & 3.2016 \\ & (0.953) \end{aligned}$ | $\begin{aligned} & 1.4250 \\ & (0.966) \end{aligned}$ | $\begin{aligned} & 0.7857 \\ & (0.972) \end{aligned}$ | $\begin{aligned} & 0.8141 \\ & (0.984) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 75.5174 | 42.3381 | 20.2062 | 3.1782 | 1.4141 | 0.7811 | 0.8108 |
| IPI, Electricity and gas utilities | 1 | 4.7970 | 4.3853 | 1.9966 | 1.3671 | 0.2780 | 0.3252 | 0.2276 |
|  | $h T^{-1 / 2}$ | 4.8666 | 4.3738 | 1.9794 | 1.3666 | 0.2789 | 0.3263 | 0.2260 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 4.6399 \\ & (0.436) \end{aligned}$ | $\begin{aligned} & 4.2677 \\ & (0.638) \end{aligned}$ | $\begin{aligned} & 1.9563 \\ & (0.680) \end{aligned}$ | $\begin{aligned} & 1.3763 \\ & (0.896) \end{aligned}$ | $\begin{aligned} & 0.2818 \\ & (0.605) \end{aligned}$ | $\begin{aligned} & 0.3289 \\ & (0.893) \end{aligned}$ | $\begin{aligned} & 0.2269 \\ & (0.885) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 4.5103 | 4.1889 | 1.9381 | 1.3821 | 0.2837 | 0.3307 | 0.2280 |

Table 3.5: Category (A) MSFE ratio of univariate smoothed direct forecast and direct forecast. The entry in parentheses is the $p$-value of hypothesis test that horizon $h$ MSFE of direct and smoothed direct forecasts are equivalent; continued from previous page.

| Variable | $\lambda_{h}$ | Forecast Horizon |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 3 | 6 | 12 | 24 | 36 | 48 | 60 |
| Nominal <br> Personal <br> Income | 1 | 12.5942 | 3.1900 | 0.8517 | 0.1507 | 0.0637 | 0.0244 | 0.0115 |
|  | $h T^{-1 / 2}$ | 12.4946 | 3.1791 | 0.8538 | 0.1505 | 0.0645 | 0.0244 | 0.0114 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 12.7645 \\ & (0.104) \end{aligned}$ | $\begin{aligned} & 3.2389 \\ & (0.187) \end{aligned}$ | $\begin{aligned} & 0.8566 \\ & (0.371) \end{aligned}$ | $\begin{aligned} & 0.1493 \\ & (0.338) \end{aligned}$ | $\begin{aligned} & 0.0646 \\ & 0.476 \end{aligned}$ | $\begin{aligned} & 0.0242 \\ & 0 \end{aligned}$ | $\begin{aligned} & 0.0113 \\ & (0.304) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 13.0102 | 3.2860 | 0.8602 | 0.1486 | 0.0645 | 0.0241 | 0.0112 |
| PCE excluding <br> Food and Energy | 1 | 27.1339 | 8.4365 | 1.8213 | 0.3793 | 0.1576 | 0.0756 | 0.0452 |
|  | $h T^{-1 / 2}$ | 27.2527 | 8.4308 | 1.8087 | 0.3780 | 0.1565 | 0.0760 | 0.0461 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 26.5410 \\ & (0.432) \end{aligned}$ | $\begin{aligned} & 8.2703 \\ & (0.556) \end{aligned}$ | $\begin{aligned} & 1.7945 \\ & (0.649) \end{aligned}$ | $\begin{aligned} & 0.3793 \\ & (0.654) \end{aligned}$ | $\begin{aligned} & 0.1571 \\ & (0.645) \end{aligned}$ | $\begin{aligned} & 0.0768 \\ & (0.732) \end{aligned}$ | $\begin{aligned} & 0.0469 \\ & (0.762) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 26.0687 | 8.1547 | 1.7831 | 0.3805 | 0.1580 | 0.0773 | 0.0472 |
| PCE, Durable goods | 1 | 62.2185 | 43.3796 | 8.3233 | 1.3843 | 0.4970 | 0.2754 | 0.2426 |
|  | $h T^{-1 / 2}$ | 63.3413 | 46.6166 | 8.3697 | 1.3636 | 0.4894 | 0.2694 | 0.2397 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 63.0750 \\ & (0.881) \end{aligned}$ | $\begin{aligned} & 46.9324 \\ & (0.937) \end{aligned}$ | $\begin{aligned} & 8.5165 \\ & (0.944) \end{aligned}$ | $\begin{aligned} & 1.3857 \\ & (0.855) \end{aligned}$ | $\begin{aligned} & 0.4848 \\ & (0.819) \end{aligned}$ | $\begin{aligned} & 0.2639 \\ & (0.920) \end{aligned}$ | $\begin{aligned} & 0.2349 \\ & (0.941) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 63.0485 | 47.0669 | 8.5881 | 1.3924 | 0.4820 | 0.2615 | 0.2327 |
| PCE, <br> Nondurable goods | 1 | 27.4503 | 9.8219 | 2.2934 | 0.5576 | 0.1701 | 0.0577 | 0.0191 |
|  | $h T^{-1 / 2}$ | 27.5531 | 9.9546 | 2.3158 | 0.5625 | 0.1642 | 0.0567 | 0.0188 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 30.7323 \\ & (0.595) \end{aligned}$ | $\begin{aligned} & 10.9595 \\ & (0.671) \end{aligned}$ | $\begin{aligned} & 2.5638 \\ & (0.707) \end{aligned}$ | $\begin{aligned} & 0.5582 \\ & (0.689) \end{aligned}$ | $\begin{aligned} & 0.1556 \\ & (0.678) \end{aligned}$ | $\begin{aligned} & 0.0532 \\ & (0.680) \end{aligned}$ | $\begin{aligned} & 0.0173 \\ & (0.397) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 34.1423 | 12.0156 | 2.7311 | 0.5512 | 0.1488 | 0.0501 | 0.0160 |
| PCE, Services | 1 | 4.1523 | 0.9974 | 0.2436 | 0.0509 | 0.0209 | 0.0095 | 0.0048 |
|  | $h T^{-1 / 2}$ | 4.1386 | 0.9887 | 0.2403 | 0.0505 | 0.0209 | 0.0096 | 0.0050 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 4.0151 \\ & (<0.001) \end{aligned}$ | $\begin{aligned} & 0.9643 \\ & (0.001) \end{aligned}$ | $\begin{aligned} & 0.2365 \\ & (0.010) \end{aligned}$ | $\begin{aligned} & 0.0505 \\ & (0.061) \end{aligned}$ | $\begin{aligned} & 0.0210 \\ & (0.094) \end{aligned}$ | $\begin{aligned} & 0.0097 \\ & (0.114) \end{aligned}$ | $\begin{aligned} & 0.0051 \\ & (0.105) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 3.9469 | 0.9500 | 0.2344 | 0.0506 | 0.0211 | 0.0098 | 0.0051 |

Table 3.6: Category (B) MSFE ratio of univariate smoothed direct forecast and direct forecast. The entry in parentheses is the $p$-value of hypothesis test that horizon $h$ MSFE of direct and smoothed direct forecasts are equivalent.

| Variable | $\lambda_{h}$ | Forecast Horizon |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 3 | 6 | 12 | 24 | 36 | 48 | 60 |
| Employment of all employees, <br> Government | 1 | 17.6892 | 5.1690 | 2.6165 | 0.9015 | 0.6631 | 0.2554 | 0.2358 |
|  | $h T^{-1 / 2}$ | 17.4714 | 5.2605 | 2.7243 | 0.9326 | 0.6568 | 0.2469 | 0.2324 |
|  | $e^{h T^{-1 / 2}}$ | 20.2539 | 5.9734 | 2.8871 | 0.9511 | 0.6403 | 0.2351 | 0.2233 |
|  |  | (0.394) | (0.474) | (0.763) | (0.828) | (0.891) | (0.859) | (0.914) |
|  | $50 e^{h T^{-1 / 2}}$ | 23.1397 | 6.7610 | 3.0367 | 0.9529 | 0.6247 | 0.2270 | 0.2166 |
| Avg.weekly overtime hrs. of prod. \& nonsup. employees: Total private | 1 | 63.9248 | 17.6614 | 9.1001 | 5.8294 | 2.5567 | 0.7475 | 0.1623 |
|  | $h T^{-1 / 2}$ | 64.6093 | 17.5542 | 8.9325 | 5.7954 | 2.5339 | 0.7514 | 0.7132 |
|  | $e^{h T^{-1 / 2}}$ | 60.6430 | 16.5763 | 8.7989 | 5.8263 | 2.5563 | 0.7695 | 0.1827 |
|  |  | (0.963) | (0.927) | (0.982) | (0.973) | (0.982) | (0.974) | (0.985) |
|  | $50 e^{h T^{-1 / 2}}$ | 57.4704 | 15.7023 | 8.6404 | 5.8540 | 2.5884 | 0.7854 | 0.1882 |
| Civilian labor force: 16yrs and over | 1 | 56.1628 | 17.6579 | 4.2179 | 0.8942 | 0.4436 | 0.1787 | 0.1001 |
|  | $h T^{-1 / 2}$ | 54.9540 | 17.8853 | 4.2784 | 0.9150 | 0.4382 | 0.1705 | 0.0967 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 53.9402 \\ & (0.680) \end{aligned}$ | $\begin{aligned} & 18.4024 \\ & (0.857) \end{aligned}$ | $\begin{aligned} & 4.4965 \\ & (0.866) \end{aligned}$ | $\begin{aligned} & 0.9391 \\ & (0.877) \end{aligned}$ | $\begin{aligned} & 0.4322 \\ & (0.857) \end{aligned}$ | $\begin{aligned} & 0.1635 \\ & (0.748) \end{aligned}$ | $\begin{aligned} & 0.0926 \\ & (0.871) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 53.8839 | 18.7104 | 4.6202 | 0.9458 | 0.4276 | 0.1605 | 0.0907 |

Table 3.7: Category (C) MSFE ratio of univariate smoothed direct forecast and direct forecast. The entry in parentheses is the $p$-value of hypothesis test that horizon $h$ MSFE of direct and smoothed direct forecasts are equivalent.

|  |  | Forecast Horizon |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Variable | $\lambda_{h}$ | 3 | 6 | 12 | 24 | 36 | 48 | 60 |  |  |  |  |  |  |  |  |
|  | 1 | 0.2012 | 0.2579 | 0.3932 | 0.9365 | 0.8608 | 1.1646 | 1.1784 |  |  |  |  |  |  |  |  |
| Housing starts: | $h T^{-1 / 2}$ | 0.1715 | 0.2234 | 0.3451 | 0.5973 | 0.6993 | 0.8116 | 0.8174 |  |  |  |  |  |  |  |  |
| South | $e^{h T^{-1 / 2}}$ | 0.1861 | 0.2399 | 0.3594 | 0.5780 | 0.6826 | 0.7604 | 0.7869 |  |  |  |  |  |  |  |  |
|  |  | $(0.969)$ | $(0.982)$ | $(0.992)$ | $(0.997)$ | $(0.997)$ | $(0.997)$ | $(0.995)$ |  |  |  |  |  |  |  |  |
|  | $50 e^{h T^{-1 / 2}}$ | 0.2222 | 0.2795 | 0.4003 | 0.5556 | 0.6655 | 0.7389 | 0.7797 |  |  |  |  |  |  |  |  |
|  | 1 | 0.3342 | 0.3771 | 0.4496 | 0.5756 | 1.0481 | 1.3684 | 1.5429 |  |  |  |  |  |  |  |  |
| Housing starts: | $h T^{-1 / 2}$ | 0.2021 | 0.2253 | 0.2682 | 0.2883 | 0.2930 | 0.2803 | 0.3300 |  |  |  |  |  |  |  |  |
| West | $e^{h T^{-1 / 2}}$ | 0.2059 | 0.2273 | 0.2674 | 0.2758 | 0.2581 | 0.2679 | 0.3138 |  |  |  |  |  |  |  |  |
|  |  | $(0.998)$ | $(0.998)$ | $(0.999)$ | $(0.999)$ | $(0.999)$ | $(0.999)$ | $(0.999)$ |  |  |  |  |  |  |  |  |
|  | $50 e^{h T^{-1 / 2}}$ | 0.2095 | 0.2303 | 0.2681 | 0.2729 | 0.2562 | 0.2666 | 0.3124 |  |  |  |  |  |  |  |  |
|  | 1 | 25.2450 | 3.6875 | 0.9135 | 2.5459 | 1.3650 | 1.3829 | 0.3652 |  |  |  |  |  |  |  |  |
| Ratio of mfg. | $h T^{-1 / 2}$ | 24.3819 | 3.6896 | 0.9147 | 2.5280 | 1.3425 | 1.3614 | 0.3994 |  |  |  |  |  |  |  |  |
| and trade | $e^{h T^{-1 / 2}}$ | 24.7990 | 3.5447 | 0.8824 | 2.4703 | 1.3362 | 1.3841 | 0.4286 |  |  |  |  |  |  |  |  |
|  |  | $(0.994)$ | $(0.990)$ | $(0.972)$ | $(0.983)$ | $(0.987)$ | $(0.978)$ | $(0.997)$ |  |  |  |  |  |  |  |  |
|  | $50 e^{h T^{-1 / 2}}$ | 24.6399 | 3.4255 | 0.8646 | 2.4464 | 1.3510 | 1.4087 | 0.4400 |  |  |  |  |  |  |  |  |

Table 3.8: Category (D) MSFE ratio of univariate smoothed direct forecast and direct forecast. The entry in parentheses is the $p$-value of hypothesis test that horizon $h$ MSFE of direct and smoothed direct forecasts are equivalent.

| Variable | $\lambda_{h}$ | Forecast Horizon |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 3 | 6 | 12 | 24 | 36 | 48 | 60 |
| IR, U.S. treasury const maturities, $10-\mathrm{yr}$ | 1 | 1.9083 | 1.8183 | 0.2612 | 0.7633 | 4.1354 | 1.0791 | 1.3473 |
|  | $h T^{-1 / 2}$ | 1.8717 | 1.7988 | 0.2664 | 0.8214 | 4.4321 | 1.0751 | 1.0719 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 1.9807 \\ & (0.990) \end{aligned}$ | $\begin{aligned} & 1.8855 \\ & (0.993) \end{aligned}$ | $\begin{aligned} & 0.2565 \\ & (0.966) \end{aligned}$ | $\begin{aligned} & 0.8077 \\ & (0.998) \end{aligned}$ | $\begin{aligned} & 4.9284 \\ & (0.994) \end{aligned}$ | $\begin{aligned} & 1.0729 \\ & (0.999) \end{aligned}$ | $\begin{aligned} & 0.9017 \\ & (0.999) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 2.1515 | 2.0102 | 0.2528 | 0.7930 | 4.9587 | 1.0462 | 0.8612 |
| IR, federal funds (effective) | 1 | 0.6536 | 1.8021 | 1.1709 | 2.4904 | 1.2894 | 1.5329 | 0. 2540 |
|  | $h T^{-1 / 2}$ | 0.7507 | 1.6851 | 1.1036 | 2.2032 | 1.4234 | 1.5782 | 0.2393 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 0.6456 \\ & (0.986) \end{aligned}$ | $\begin{aligned} & 1.8964 \\ & (0.992) \end{aligned}$ | $\begin{aligned} & 1.2277 \\ & (0.998) \end{aligned}$ | $\begin{aligned} & 2.4122 \\ & (0.994) \end{aligned}$ | $\begin{aligned} & 1.3558 \\ & (0.994) \end{aligned}$ | $\begin{aligned} & 1.5669 \\ & (0.994) \end{aligned}$ | $\begin{aligned} & 0.2671 \\ & (0.977) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 0.4173 | 1.4742 | 1.3212 | 2.6719 | 1.4206 | 1.5365 | 0.2507 |
| IR, U.S. treasury const maturities, $1-\mathrm{yr}$ | 1 | 1.4115 | 4.0744 | 1.9826 | 27.0994 | 164.0314 | 90.3570 | 7.8115 |
|  | $h T^{-1 / 2}$ | 0.5214 | 1.3963 | 0.5995 | 2.2837 | 1.1878 | 1.5714 | 0.1408 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 0.5563 \\ & (0.984) \end{aligned}$ | $\begin{aligned} & 1.7176 \\ & (0.994) \end{aligned}$ | $\begin{aligned} & 0.7797 \\ & (0.992) \end{aligned}$ | $\begin{aligned} & 4.2135 \\ & (0.992) \end{aligned}$ | $\begin{aligned} & 0.6232 \\ & (0.998) \end{aligned}$ | $\begin{aligned} & 2.1706 \\ & (0.997) \end{aligned}$ | $\begin{aligned} & 0.2476 \\ & (0.998) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 0.4327 | 1.4335 | 0.6410 | 2.5178 | 0.9653 | 1.6965 | 0.1476 |
| EIR, conventional home mtge loans | 1 | 20.0875 | 7.1791 | 22.1647 | 4.4942 | 0.0989 | 0.3692 | 0.2959 |
|  | $h T^{-1 / 2}$ | 3.5337 | 1.0061 | 2.3216 | 5.8254 | 1.1788 | 1.7701 | 1.2316 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 5.5798 \\ & (0.924) \end{aligned}$ | $\begin{aligned} & 1.5891 \\ & (0.947) \end{aligned}$ | $\begin{aligned} & 3.0834 \\ & (0.976) \end{aligned}$ | $\begin{aligned} & 6.6599 \\ & (0.965) \end{aligned}$ | $\begin{aligned} & 0.8488 \\ & (0.982) \end{aligned}$ | $\begin{aligned} & 1.3041 \\ & (0.948) \end{aligned}$ | $\begin{aligned} & 0.8418 \\ & (0.985) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 8.6436 | 2.4601 | 4.2549 | 5.9025 | 0.7557 | 1.1659 | 0.7212 |
| FER, <br> Switzerland |  | 0.2467 | 0.2585 | 0.3576 | 0.2858 | 0.4190 | 0.8238 | 0.5038 |
|  | $h T^{-1 / 2}$ | 0.2466 | 0.2647 | 0.3709 | 0.2931 | 0.3657 | 6.5491 | 10.4281 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 0.2483 \\ & (0.898) \end{aligned}$ | $\begin{aligned} & 0.2582 \\ & (0.922) \end{aligned}$ | $\begin{aligned} & 0.3460 \\ & (0.946) \end{aligned}$ | $\begin{aligned} & 0.2403 \\ & (0.973) \end{aligned}$ | $\begin{aligned} & 0.2987 \\ & (0.978) \end{aligned}$ | $\begin{aligned} & 0.7498 \\ & (0.987) \end{aligned}$ | $\begin{aligned} & 0.5581 \\ & (0.990) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 0.2632 | 0.2703 | 0.3516 | 0.2385 | 0.2968 | 0.7473 | 0.5463 |
| FER: Japan | 1 | 0.1803 | 0.1913 | 0.2171 | 0.2600 | 0.2754 | 0.3562 | 0.3829 |
|  | $h T^{-1 / 2}$ | 0.1776 | 0.1925 | 0.2209 | 0.2665 | 0.2738 | 0.3480 | 0.3753 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 0.1888 \\ & (0.898) \end{aligned}$ | $\begin{aligned} & 0.2000 \\ & (0.922) \end{aligned}$ | $\begin{aligned} & 0.2245 \\ & (0.946) \end{aligned}$ | $\begin{aligned} & 0.2710 \\ & (0.973) \end{aligned}$ | $\begin{aligned} & 0.2698 \\ & (0.978) \end{aligned}$ | $\begin{aligned} & 0.3374 \\ & (0.987) \end{aligned}$ | $\begin{aligned} & 0.3707 \\ & (0.990) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 0.2004 | 0.2100 | 0.2306 | 0.2721 | 0.2662 | 0.3313 | 0.3664 |
| FER, Canada | 1 | 0.3022 | 0.9337 | 0.1911 | 0.5862 | 0.1799 | 0.7057 | 0.3536 |
|  | $h T^{-1 / 2}$ | 0.3042 | 0.7361 | 0.1945 | 0.5887 | 0.1793 | 0.7033 | 0.3485 |
|  | $e^{h T^{-1 / 2}}$ | $\begin{aligned} & 0.3174 \\ & (0.928) \end{aligned}$ | $\begin{aligned} & 0.7723 \\ & (0.992) \end{aligned}$ | $\begin{aligned} & 0.1995 \\ & (0.887) \end{aligned}$ | $\begin{aligned} & 0.5844 \\ & (0.986) \end{aligned}$ | $\begin{aligned} & 0.1768 \\ & (0.917) \end{aligned}$ | $\begin{aligned} & 0.6962 \\ & (0.995) \end{aligned}$ | $\begin{aligned} & 0.3464 \\ & (0.985) \end{aligned}$ |
|  | $50 e^{h T^{-1 / 2}}$ | 0.3283 | 0.7985 | 0.2032 | 0.5802 | 0.1749 | 0.6919 | 0.3463 |

### 3.4 Concluding Remarks

In this paper we construct a method to obtain long-horizon forecasts that outperform the conventional iterated or direct-forecast methods. The direct method estimation is modified by imposing a smoothing parameter on the first differences of parameters across horizons. This obtains estimators that are more robust than the iterated method, but less erratic across horizons compared to direct forecasts. This is an extension of the ridge regression where we impose a smoothness prior by restricted regression. Our application of the ridge regression by "smoothing out" estimates across horizons is similar to the smoothness method developed in Shiller (1973). The smoothing parameter (or penalty term) is either a constant or a exponential/linear function that depends on the horizon length and data size so that stronger restriction is imposed on longer horizon forecasts. Our forecasting method is denoted as the "smoothed direct forecast".

We apply the smoothed-direct-forecast method to two empirical applications where long horizon forecasts are of interest. Our first application on forecasting real oil prices shows improvement on direct forecasts for MSFE, from 3 to 5 years horizons. The difference in MSFE of direct and smoothed direct forecasts are statistically insignificant for all horizons using West's (1996) forecast accuracy test. Regardless of the test results, we argued that the improvement of real oil price forecasts of 5 years ahead are economically significant. In addition, due to our simple functional form of the smoothing parameter, it is preferable to have small deviations from direct forecasts. Smoothed direct forecasts show only slight differences for various exponential functions, which implies robustness of the choice of the smoothing parameter.

Our second empirical application of data is with the U.S. macroeconomic time series from Marcellino, Stock, and Watson (2006). We apply the smoothed-direct-forecasting method on 80 macroeconomic series and compare with the direct and iterated forecasts. Overall, the smoothed direct forecast for most series shows improvement on the direct forecast in long horizons, partly due to poor performance of the direct method. Similar to the real oil price application, the
differences in MSFE remain statistically insignificant but that does not mean the differences are economically insignificant.

In conclusion, the two empirical application results suggest that our forecast method provides long-horizon forecasts that improve on direct forecasts and might improve on iterated forecasts as well. In future research an optimal choice of the functional form of the smoothing parameter can provide further improvement in performance of forecasts.

Chapter 3, in full, is a reprint of the material that has been accepted for publication at Journal of Forecasting. Baek, Yaein. The dissertation author was the sole author of this paper.

## Appendix

## A. 1 Proofs for Chapter 1

## Proof of Lemma 1

Proof. Assumption 2 implies that $A_{T}(k)$ from (1.11) is positive definite and thus $G_{T}(k)=$ $\delta_{T}^{\prime} A_{T}(k) \delta_{T} \geq \lambda_{T}(k)\left\|\delta_{T}\right\|^{2}$ where $\lambda_{T}(k)$ is the minimum eigenvalue of $A_{T}(k)$. It is sufficient to argue that $\lambda_{T}(k)$ is bounded away from zero with probability tending to 1 as $\left|k_{0}-k\right|$ increases. The matrices $Z_{0}^{\prime} M Z_{0}$ and $Z_{0}^{\prime} M Z_{k}$ are rearranged as follows, similar to $Z_{k}^{\prime} M Z_{k}$ in (1.10).

$$
\begin{align*}
& Z_{0}^{\prime} M Z_{0}=R^{\prime}\left(X_{0}^{\prime} X_{0}\right)\left(X^{\prime} X\right)^{-1}\left(X^{\prime} X-X_{0}^{\prime} X_{0}\right) R \\
& Z_{0}^{\prime} M Z_{k}= \begin{cases}R^{\prime}\left(X_{0}^{\prime} X_{0}\right)\left(X^{\prime} X\right)^{-1}\left(X^{\prime} X-X_{k}^{\prime} X_{k}\right) R & \text { if } k \leq k_{0} \\
R^{\prime}\left(X^{\prime} X-X_{0}^{\prime} X_{0}\right)\left(X^{\prime} X\right)^{-1}\left(X_{k}^{\prime} X_{k}\right) R & \text { if } k>k_{0}\end{cases} \tag{A.1}
\end{align*}
$$

Without loss of generality assume $k \leq k_{0}$. The second term of $\left|k_{0}-k\right| A_{T}(k)$ from (1.11) is

$$
\begin{align*}
&\left(Z_{0}^{\prime} M Z_{k}\right)\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2}\left(Z_{k}^{\prime} M Z_{0}\right) \\
&= {\left[R^{\prime}\left(X_{0}^{\prime} X_{0}\right)\left(X^{\prime} X\right)^{-1}\left(X^{\prime} X-X_{k}^{\prime} X_{k}\right) R\right]\left[\left(Z_{k}^{\prime} M Z_{k}\right)^{1 / 2} \Omega_{k}^{-1}\left(Z_{k}^{\prime} M Z_{k}\right)^{1 / 2}\right]^{-1} }  \tag{A.2}\\
& \times\left[R^{\prime}\left(X^{\prime} X-X_{k}^{\prime} X_{k}\right)\left(X^{\prime} X\right)^{-1}\left(X_{0}^{\prime} X_{0}\right) R\right]
\end{align*}
$$

Define the following matrices.

$$
\begin{align*}
F_{k} & :=\left(X_{k}^{\prime} X_{k}\right)^{-1}-\left(X^{\prime} X\right)^{-1}=\left(X^{\prime} X\right)^{-1}\left(X^{\prime} X-X_{k}^{\prime} X_{k}\right)\left(X_{k}^{\prime} X_{k}\right)^{-1} \\
F_{0} & :=\left(X_{0}^{\prime} X_{0}\right)^{-1}-\left(X^{\prime} X\right)^{-1}=\left(X^{\prime} X\right)^{-1}\left(X^{\prime} X-X_{0}^{\prime} X_{0}\right)\left(X_{0}^{\prime} X_{0}\right)^{-1}  \tag{A.3}\\
\Omega_{X, k} & :=\left[\begin{array}{cc}
I_{(p-q)} & \mathbf{0}_{(p-q) \times q} \\
\mathbf{0}_{q \times(p-q)} & \Omega_{k}
\end{array}\right], \quad B:=\Omega_{X, k}^{-1 / 2} F_{k}^{1 / 2} X_{k}^{\prime} X_{k} .
\end{align*}
$$

Both $F_{k}$ and $F_{0}$ are positive definite matrices under Assumption 1. Hence, each matrix can be decomposed into $F_{k}=\left(F_{k}^{1 / 2}\right)^{2}$ and $F_{0}=\left(F_{0}^{1 / 2}\right)^{2}$ where $F_{k}^{1 / 2}$ and $F_{0}^{1 / 2}$ are nonsingular $(p \times p)$ matrices with $p=\operatorname{dim}\left(x_{t}\right) . \Omega_{X, k}$ is a $(p \times p)$ matrix where $I_{(p-q)}$ is an identity matrix with rank $(p-q)$, and zeros in non-diagonal block matrices such that $R^{\prime} \Omega_{X, k} R=\Omega_{k}$. The projection matrix of $B R$ is $I_{p}-B R\left(R^{\prime} B^{\prime} B R\right)^{-1} R^{\prime} B^{\prime}$, which is positive semi-definite. If we multiply $R^{\prime}\left(X_{0}^{\prime} X_{0}\right) F_{k}^{1 / 2} \Omega_{X, k}^{1 / 2}$ to the left and its transpose to the right of the projection matrix, the following inequality is obtained.

$$
\begin{aligned}
& R^{\prime}\left(X_{0}^{\prime} X_{0}\right) F_{k}^{1 / 2} \Omega_{X, k} F_{k}^{1 / 2}\left(X_{0}^{\prime} X_{0}\right) R \\
& \geq \\
& \geq R^{\prime}\left(X_{0}^{\prime} X_{0}\right) F_{k}\left(X_{k}^{\prime} X_{k}\right) R\left(R^{\prime} B^{\prime} B R\right)^{-1} R^{\prime}\left(X_{k}^{\prime} X_{k}\right) F_{k}\left(X_{0}^{\prime} X_{0}\right) R \\
& = \\
& \quad R^{\prime}\left(X_{0}^{\prime} X_{0}\right)\left(X^{\prime} X\right)^{-1}\left(X^{\prime} X^{\prime}-X_{k}^{\prime} X_{k}\right) R\left(R^{\prime} B^{\prime} B R\right)^{-1} \\
& \quad \times R^{\prime}\left(X^{\prime} X-X_{k}^{\prime} X_{k}\right)\left(X^{\prime} X\right)^{-1}\left(X_{0}^{\prime} X_{0}\right) R
\end{aligned}
$$

From (A.3), $R^{\prime} B^{\prime} B R=R^{\prime}\left(X_{k}^{\prime} X_{k}\right) F_{k}^{1 / 2} \Omega_{X, k}^{-1} F_{k}^{1 / 2}\left(X_{k}^{\prime} X_{k}\right) R=\left(Z_{k}^{\prime} M Z_{k}\right)^{1 / 2} \Omega_{k}^{-1}\left(Z_{k}^{\prime} M Z_{k}\right)^{1 / 2}$ and the right side of the inequality is equivalent to (A.2). Therefore it is sufficient to show that the right-side of inequality (A.4) is bounded away from zero for large $k_{0}-k$.

$$
\begin{equation*}
A_{T}(k) \geq \frac{1}{\left|k_{0}-k\right|}\left[\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2}-R^{\prime}\left(X_{0}^{\prime} X_{0}\right) F_{k}^{1 / 2} \Omega_{X, k} F_{k}^{1 / 2}\left(X_{0}^{\prime} X_{0}\right) R\right] \tag{A.4}
\end{equation*}
$$

Also from (A.3), $\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2}=R^{\prime}\left(X_{0}^{\prime} X_{0}\right) F_{0}^{1 / 2} \Omega_{X, k_{0}} F_{0}^{1 / 2}\left(X_{0}^{\prime} X_{0}\right) R$. Thus,

$$
\begin{aligned}
\left|k_{0}-k\right| A_{T}(k) & \geq R^{\prime}\left(X_{0}^{\prime} X_{0}\right)\left[F_{0}^{1 / 2} \Omega_{X, k_{0}} F_{0}^{1 / 2}-F_{k}^{1 / 2} \Omega_{X, k} F_{k}^{1 / 2}\right]\left(X_{0}^{\prime} X_{0}\right) R \\
& \equiv\left|k_{0}-k\right| R^{\prime} \tilde{A}_{T}(k) R
\end{aligned}
$$

Define the $(q \times q)$ matrix on the right-side as $\left|k_{0}-k\right| \tilde{A}_{T}(k)$. Then,

$$
\begin{aligned}
\left\|\tilde{A}_{T}(k)^{-1}\right\| & =\left|k_{0}-k\right|\left\|\left(X_{0}^{\prime} X_{0}\right)^{-1}\left[F_{0}^{1 / 2} \Omega_{X, k_{0}} F_{0}^{1 / 2}-F_{k}^{1 / 2} \Omega_{X, k} F_{k}^{1 / 2}\right]^{-1}\left(X_{0}^{\prime} X_{0}\right)^{-1}\right\| \\
& \leq\left|k_{0}-k\right| \frac{1}{\left\|X_{0}^{\prime} X_{0}\right\|} \frac{1}{\left\|F_{0}^{1 / 2} \Omega_{X, k_{0}} F_{0}^{1 / 2}-F_{k}^{1 / 2} \Omega_{X, k} F_{k}^{1 / 2}\right\|} \frac{1}{\left\|X_{0}^{\prime} X_{0}\right\|}
\end{aligned}
$$

Note that for a nonsingular, bounded $(p \times p)$ matrix $S$, the norm does not change by multiplying $S$ on the left and $S^{-1}$ on the right of a matrix: $\left\|\Omega_{X, k}\right\|=\left\|S \Omega_{X, k} S^{-1}\right\|$. By assumption, $\Omega_{X, k} \geq$ $\lambda_{\min }\left(\Omega_{X, k}\right)>0$ where $\lambda_{\min }$ denotes the minimum eigenvalue. Therefore, $\left(F_{0}^{1 / 2}\right)^{-1} \Omega_{X, k_{0}} F_{0}^{1 / 2}=$ $\Omega_{X, k_{0}}+o_{p}(1) \geq \lambda_{\min }\left(\Omega_{X, k_{0}}\right)+o_{p}(1)$. By subtracting and adding the matrix $F_{k}\left(F_{0}^{1 / 2}\right)^{-1} \Omega_{X, k_{0}} F_{0}^{1 / 2}$ to the denominator of the second term, the following inequality holds.

$$
\begin{aligned}
& \left\|F_{0}^{1 / 2} \Omega_{X, k_{0}} F_{0}^{1 / 2}-F_{k}^{1 / 2} \Omega_{X, k} F_{k}^{1 / 2}\right\| \\
= & \left\|\left(F_{0}-F_{k}\right)\left(F_{0}^{1 / 2}\right)^{-1} \Omega_{X, k_{0}} F_{0}^{1 / 2}-F_{k}\left\{\left(F_{k}^{1 / 2}\right)^{-1} \Omega_{X, k} F_{k}^{1 / 2}-\left(F_{0}^{1 / 2}\right)^{-1} \Omega_{X, k_{0}} F_{0}^{1 / 2}\right\}\right\| \\
\geq & \mid\left\|\left(F_{0}-F_{k}\right)\left(F_{0}^{1 / 2}\right)^{-1} \Omega_{X, k_{0}} F_{0}^{1 / 2}\right\|-\left\|F_{k}\left\{\left(F_{k}^{1 / 2}\right)^{-1} \Omega_{X, k} F_{k}^{1 / 2}-\left(F_{0}^{1 / 2}\right)^{-1} \Omega_{X, k_{0}} F_{0}^{1 / 2}\right\}\right\| \| \\
= & \mid\left\|\left(F_{0}-F_{k}\right) \Omega_{X, k_{0}}\right\|-\left\|F_{k}\left(\Omega_{X, k}-\Omega_{X, k_{0}}\right)\right\| \|+o_{p}(1)
\end{aligned}
$$

where the inequality is from the inverse triangular inequality. Let $\tilde{\lambda}$ be the minimum value of $\lambda_{\text {min }}\left(\Omega_{X, k}\right)$ and $\lambda_{\text {min }}\left(\Omega_{X, k_{0}}\right)$. From Assumption 2 we have $\left|k_{0}-k\right|^{-1}\left\|\Omega_{k}-\Omega_{k_{0}}\right\| \leq b / T$. Hence,

$$
\begin{aligned}
& \left|k_{0}-k\right|^{-1}\left\|F_{0}^{1 / 2} \Omega_{X, k_{0}} F_{0}^{1 / 2}-F_{k}^{1 / 2} \Omega_{X, k} F_{k}^{1 / 2}\right\| \\
\geq & \left|\left(k_{0}-k\right)^{-1}\left\|\left(F_{0}-F_{k}\right) \tilde{\lambda}\right\|-b / T\left\|F_{k}\right\|\right|+\left|k_{0}-k\right|^{-1} o_{p}(1) \\
\geq & \tilde{\lambda}\left\|\left(k_{0}-k\right)^{-1}\left(F_{0}-F_{k}\right)\right\|+o_{p}(1)
\end{aligned}
$$

Let $X_{\Delta}:=\operatorname{sgn}\left(k_{0}-k\right)\left(X_{k}-X_{0}\right)$, then by rearranging terms similar to (A.3), $F_{0}-F_{k}=\left(X_{0}^{\prime} X_{0}\right)^{-1}$ $\times\left(X_{\Delta}^{\prime} X_{\Delta}\right)\left(X_{k}^{\prime} X_{k}\right)^{-1}$ so that

$$
\begin{aligned}
\left\|\tilde{A}_{T}(k)^{-1}\right\| & \leq \frac{1}{\left\|X_{0}^{\prime} X_{0}\right\|^{2}} \tilde{\tilde{\lambda}\left|k_{0}-k\right|^{-1}\left\|F_{0}-F_{k}\right\|} \\
& \leq \frac{1}{\tilde{\lambda}\left\|X_{0}^{\prime} X_{0}\right\|^{2}\left\|\left(X_{0}^{\prime} X_{0}\right)^{-1}\left(k_{0}-k\right)^{-1} X_{\Delta}^{\prime} X_{\Delta}\left(X_{k}^{\prime} X_{k}\right)^{-1}\right\|}
\end{aligned}
$$

From Assumptions 1 and 2, the right-side of the inequality is bounded:
$\tilde{\lambda}\left\|T^{-1} X_{0}^{\prime} X_{0}\right\|^{2}\left\|T^{2}\left(X_{0}^{\prime} X_{0}\right)^{-1}\left(X_{k}^{\prime} X_{k}\right)^{-1}\right\|<M$ for some $M<\infty$. In addition, the minimum eigenvalue of $\left(k_{0}-k\right)^{-1}\left(X_{\Delta}^{\prime} X_{\Delta}\right)$ is bounded away from zero with large probability so that $1 /\left\|\left(k_{0}-k\right)^{-1} X_{\Delta}^{\prime} X_{\Delta}\right\|$ is bounded with large probability for all large $k_{0}-k$. Thus $\left\|\tilde{A}_{T}(k)^{-1}\right\|$ is bounded with large probability for all large $k_{0}-k$. This implies that the minimum eigenvalue of $\tilde{A}_{T}(k)$ is bounded away from zero for all large $k_{0}-k$ and this is also true for $A_{T}(k)=R^{\prime} \tilde{A}_{T}(k) R$ because $R$ has full column rank.

For the proof of Lemma 3 we use results from Proposition 2 and Lemma A.1. Hájek and Rényi (1955) proved the inequality assuming i.i.d. random variables, and was later generalized to martingales by Birnbaum and W. (1961). We use the generalized Hájek-Rényi for martingale difference sequences to prove Lemma A.1, where $\left\{\varepsilon_{t}, \mathcal{F}_{t}\right\}$ are mixingale sequences under Assumption 1.

Proposition 2. Let $\varepsilon_{1}, \varepsilon_{2}, \ldots$, be a sequence of martingale differences with $E\left[\varepsilon_{t}^{2}\right]=\sigma^{2}$ and $\left\{c_{k}\right\}$
be a decreasing positive sequence of constants. The Hájek-Rényi inequality takes the following form.

$$
P\left(\max _{m \leq k \leq T} c_{k}\left|\sum_{t=1}^{k} \varepsilon_{t}\right|>\alpha\right) \leq \frac{\sigma^{2}}{\alpha^{2}}\left(m c_{m}^{2}+\sum_{t=m+1}^{T} c_{t}^{2}\right) .
$$

Lemma A.1. Under Assumption 1, for every $\alpha>0$ and $m>0$ there exists $C<\infty$ such that

$$
P\left(\sup _{m \leq k \leq T} \frac{1}{\sqrt{k}}\left|\sum_{t=1}^{k} z_{t} \varepsilon_{t}\right|>\alpha\right) \leq \frac{C \ln T}{\alpha^{2}}
$$

and thus, $\sup _{k} T^{-1 / 2}\left\|Z_{k}^{\prime} \varepsilon\right\|=O_{p}(\sqrt{\ln T})$
Proof. Denote $\xi_{t}=z_{t} \varepsilon_{t}$ and proceed. Let $\left\{\xi_{t}, \mathcal{F}_{t}\right\}$ be $(q \times 1) L^{r}$-mixingales, $r=4+\gamma$ for some $\gamma>0$ satisfying Assumption $1(v i)$. Define $\xi_{j t}:=E\left[\xi_{t} \mid \mathcal{F}_{t-j}\right]-E\left[\xi_{t} \mid \mathcal{F}_{t-j-1}\right]$. Then $\xi_{t}=\sum_{j=-\infty}^{\infty} \xi_{j t}$, and hence $\sum_{t=1}^{k} \xi_{t}=\sum_{j=-\infty}^{\infty} \sum_{t=1}^{k} \xi_{j t}$. Denote $\|\cdot\|_{s}$ for the $L^{s}$-norm. For each $T>0$,

$$
\begin{equation*}
P\left(\sup _{m \leq k \leq T} \frac{1}{\sqrt{k}}\left\|\sum_{t=1}^{k} \xi_{t}\right\|_{2}>\alpha\right) \leq P\left(\sum_{j=-\infty m \leq k \leq T}^{\infty} \sup _{m \leq T} \frac{1}{\sqrt{k}}\left\|\sum_{t=1}^{k} \xi_{j t}\right\|_{2}>\alpha\right) . \tag{A.5}
\end{equation*}
$$

For each $j$, $\left\{\xi_{j t}, \mathcal{F}_{t-j}\right\}$ forms a sequence of martingale difference and the generalized HájekRényi inequality (Proposition 2) holds for this sequence. Let $b_{j}>0$ for all $j$ and $\sum_{j=-\infty}^{\infty} b_{j}=1$. The right-side of (A.5) is bounded by

$$
\begin{aligned}
& \sum_{j=-\infty}^{\infty} P\left(\sup _{m \leq k \leq T} \frac{1}{\sqrt{k}}\left\|\sum_{t=1}^{k} \xi_{j t}\right\|_{2}>b_{j} \alpha\right) \\
& \quad \leq \frac{1}{\alpha^{2}} \sum_{j=-\infty}^{\infty} \frac{1}{b_{j}^{2}}\left(m^{-1} \sum_{i=1}^{m} E\left\|\xi_{j i}\right\|_{2}^{2}+\sum_{i=m+1}^{T} i^{-1} E\left\|\xi_{j i}\right\|_{2}^{2}\right) .
\end{aligned}
$$

Note that $\left\|\xi_{j t}\right\|_{2} \leq\left\|\xi_{j t}\right\|_{r}$ by Liapounov's inequality. By definition, for $j \geq 0$, we have $\left\|\xi_{j t}\right\|_{r} \leq$ $\left\|E\left[\xi_{t} \mid \mathcal{F}_{t-j}\right]\right\|_{r}+\left\|E\left[\xi_{t} \mid \mathcal{F}_{t-j-1}\right]\right\|_{r}$ and for $j<0,\left\|\xi_{j t}\right\|_{r} \leq\left\|\xi_{t}-E\left[\xi_{t} \mid \mathcal{F}_{t-j-1}\right]\right\|_{r}+\left\|\xi_{t}-E\left[\xi_{t} \mid \mathcal{F}_{t-j}\right]\right\|_{r}$. Hence, from the definition of a mixingale, $E\left\|\xi_{j t}\right\|_{r}^{2} \leq 4 c_{t}^{2} \psi_{|j|}^{2}$ and with Assumption 1(vi)(c), this
implies $E\left\|\xi_{j i}\right\|_{2}^{2} \leq 4 c_{i}^{2} \psi_{|j|}^{2} \leq 4 K^{2} \psi_{|j|}^{2}$. Then the right-side of (A.5) is bounded by

$$
\begin{equation*}
\frac{1}{\alpha^{2}} \sum_{j=-\infty}^{\infty} 4 b_{j}^{-2} K^{2} \psi_{|j|}^{2}\left(1+\sum_{i=m+1}^{T} i^{-1}\right) \leq \frac{1}{\alpha^{2}} \sum_{j=-\infty}^{\infty} 4 b_{j}^{-2} K^{2} \psi_{|j|}^{2}(1+\ln T) \tag{A.6}
\end{equation*}
$$

We can choose appropriate $\left\{b_{j}\right\}$ so that $\sum_{j} b_{j}^{-2} \psi_{|j|}^{2}$ are bounded. From lemma A. 6 of Bai and Perron (1998), let $\mathrm{v}_{0}=1$ and $\mathrm{v}_{j}=j^{-1-\kappa}$ for $j \geq 1$, where $\kappa>0$ is given in Assumption $1(v i)(d)$. Let $b_{j}=\mathrm{v}_{j} /\left(1+2 \sum_{i=1}^{\infty} \mathrm{v}_{i}\right)$ and $b_{-j}=b_{j}$ for all $j \geq 0$. Then $\sum_{j=-\infty}^{\infty} b_{j}=1$. By Assumption $1(v i)(d)$, we have

$$
\sum_{j=-\infty}^{\infty} b_{j}^{-2} \psi_{|j|}^{2}=\left(\psi_{0}^{2}+2 \sum_{j=1}^{\infty} j^{2+2 \kappa} \psi_{j}^{2}\right)\left(1+2 \sum_{j=1}^{\infty} j^{-2-2 \kappa}\right)<\infty
$$

Hence, the right-side inequality of (A.6) is bounded by $\frac{C \ln T}{\alpha^{2}}$ for some $C>0$ and we obtain the result of Lemma A.1.

## Proof of Lemma 3

Proof. We use the expression (1.12); the estimator $\hat{k}$ must satisfy $Q_{T}(\hat{k})^{2} \geq Q_{T}\left(k_{0}\right)^{2}$ which is equivalent to $H_{T}(\hat{k}) \geq\left|k_{0}-\hat{k}\right| G_{T}(\hat{k})$. Therefore we have

$$
\begin{align*}
P\left(\left|\hat{\rho}-\rho_{0}\right|>\eta\right) & =P\left(\left|\hat{k}-k_{0}\right|>T \eta\right) \\
& \leq P\left(\sup _{\left|k-k_{0}\right|>T \eta}\left|H_{T}(k)\right| \geq \inf _{\left|k-k_{0}\right|>T \eta}\left|k_{0}-k\right| G_{T}(k)\right) \\
& \leq P\left(\sup _{\left|k-k_{0}\right|>T \eta}\left|H_{T}(k)\right| \geq T \eta \inf _{\left|k-k_{0}\right|>T \eta} G_{T}(k)\right) \\
& \leq P\left(\tilde{G}^{-1} \sup _{p \leq k \leq T-p} T^{-1}\left|H_{T}(k)\right| \geq \eta\right) \tag{A.7}
\end{align*}
$$

where $\tilde{G}:=\inf _{\left|k-k_{0}\right|>T \eta} G_{T}(k)$ which is positive and bounded away from zero by Lemma 1 and
the restriction $p \leq k \leq T-p$ is imposed to guarantee existence of $H_{T}(k)$. Thus consistency follows by showing that $T^{-1} \sup _{p \leq k \leq T-p}\left|H_{T}(k)\right|=o_{p}(1)$, where

$$
\begin{align*}
T^{-1}\left|H_{T}(k)\right| \leq & \left|T^{-1} \varepsilon^{\prime} M Z_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} Z_{k}^{\prime} M \varepsilon\right| \\
& +\left|T^{-1} \varepsilon^{\prime} M Z_{0}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} Z_{0}^{\prime} M \varepsilon\right| \\
& +2\left|T^{-1} \delta_{T}^{\prime}\left(Z_{0}^{\prime} M Z_{k}\right)\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} Z_{k}^{\prime} M \varepsilon\right|  \tag{A.8}\\
& +2\left|T^{-1} \delta_{T}^{\prime}\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} Z_{0}^{\prime} M \varepsilon\right|
\end{align*}
$$

Lemma A. 1 implies that $\sup _{k}\left\|T^{-1 / 2} Z_{k}^{\prime} M \varepsilon\right\|=O_{p}(\sqrt{\ln T})$. We use the following to verify uniform convergence for all $k$.

$$
\begin{equation*}
\sup _{p \leq k \leq T-p}\left\|\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} Z_{k}^{\prime} M \varepsilon\right\|=O_{p}(\sqrt{\ln T}) \tag{A.9}
\end{equation*}
$$

We show the third and fourth terms of (A.8) are $O_{p}\left(T^{-1 / 2}\left\|\delta_{T}\right\| \ln T\right)$ and $O_{p}\left(T^{-1 / 2}\left\|\delta_{T}\right\|\right)$, respectively. Denote $D_{T}:=T^{-1 / 2}\left(Z_{0}^{\prime} M Z_{k}\right)\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2}$. From (1.10) and (A.1), $Z_{0}^{\prime} M Z_{k} \leq$ $Z_{k}^{\prime} M Z_{k}$ for all $k$, and thus

$$
\sup _{p \leq k \leq T-p} D_{T}^{\prime} D_{T} \leq \sup _{p \leq k \leq T-p} T^{-1} Z_{k}^{\prime} M Z_{k}=O_{p}(1)
$$

then the third term of (A.8) is bounded by

$$
2\left\|T^{-1 / 2} \delta_{T}^{\prime}\left(T^{-1} Z_{k}^{\prime} M Z_{k}\right)^{1 / 2} \Omega_{k}\right\|\left\|\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} Z_{k}^{\prime} M \varepsilon\right\|=\left\|\delta_{T}\right\| O_{p}\left(T^{-1 / 2} \sqrt{\ln T}\right)
$$

For the true break date $k_{0},\left|\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} Z_{0}^{\prime} M \varepsilon\right|=O_{p}(1)$ under our regularity conditions. Hence, the fourth term of (A.8) has order

$$
\left|T^{-1} \delta_{T}^{\prime}\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} Z_{0}^{\prime} M \varepsilon\right|=\left\|\delta_{T}\right\| O_{p}\left(T^{-1 / 2}\right)
$$

From (A.9) and boundedness of $\Omega_{k}$, we have $\sup _{k}\left\|\Omega_{k}^{1 / 2}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} Z_{k}^{\prime} M \varepsilon\right\|=O_{p}(\sqrt{\ln T})$. Then the first and second terms of (A.8) are bounded as below, respectively.

$$
\begin{gathered}
\sup _{k} T^{-1}\left\|\Omega_{k}^{1 / 2}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} Z_{k}^{\prime} M \varepsilon\right\|^{2}=O_{p}\left(T^{-1} \ln T\right) \\
T^{-1}\left\|\Omega_{k_{0}}^{1 / 2}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} Z_{0}^{\prime} M \varepsilon\right\|^{2}=O_{p}\left(T^{-1}\right)
\end{gathered}
$$

By combining all four terms,

$$
T^{-1} \sup _{p \leq k \leq T-p}\left|H_{T}(k)\right|=O_{p}\left(T^{-1 / 2}\left\|\delta_{T}\right\| \sqrt{\ln T}\right)=o_{p}(1)
$$

hence, the probability (A.7) is negligible for large $T$.

## Proof of Corollary 1

Proof. Let $\hat{Z}_{0}$ denote $Z_{k}$ when $k$ is replaced by $\hat{k}$. Then the LS estimator of $\hat{\delta}(\hat{\rho})$ is obtained by regressing $M Y$ on $M \hat{Z}_{0}$. The true model (1.8) multiplied by $M$ can be rewritten as $M Y=$ $M \hat{Z}_{0} \delta_{T}+M \varepsilon^{*}$, where $\varepsilon^{*}=\varepsilon+\left(Z_{0}-\hat{Z}_{0}\right) \delta_{T}$. Then,

$$
\begin{aligned}
\sqrt{T}\left(\hat{\delta}(\hat{\rho})-\delta_{T}\right) & =\left(T^{-1} \hat{Z}_{0}^{\prime} M \hat{Z}_{0}\right)^{-1} T^{-1 / 2} \hat{Z}_{0}^{\prime} M \varepsilon^{*} \\
& =\left(T^{-1} \hat{Z}_{0}^{\prime} M \hat{Z}_{0}\right)^{-1}\left(T^{-1 / 2} \hat{Z}_{0}^{\prime} M \varepsilon+T^{-1 / 2} \hat{Z}_{0}^{\prime} M\left(Z_{0}-\hat{Z}_{0}\right) \delta_{T}\right)
\end{aligned}
$$

We show that the right side converges in probability to the same limit as when $\hat{Z}_{0}$ is replaced by $Z_{0}$. First, we show that $\operatorname{plim} T^{-1 / 2} \hat{Z}_{0}^{\prime} M\left(Z_{0}-\hat{Z}_{0}\right) \delta_{T}=0$. Without loss of generality, consider $k \leq k_{0}$.

$$
\begin{aligned}
\left\|T^{-1 / 2} \hat{Z}_{0}^{\prime} M\left(Z_{0}-\hat{Z}_{0}\right) \delta_{T}\right\| \leq & T^{-1 / 2}\left\|\hat{Z}_{0}^{\prime}\left(Z_{0}-\hat{Z}_{0}\right)-\hat{Z}_{0}^{\prime} X\left(X^{\prime} X\right)^{-1} X^{\prime}\left(Z_{0}-\hat{Z}_{0}\right)\right\|\left\|\delta_{T}\right\| \\
\leq & \frac{1}{\sqrt{T}\left\|\delta_{T}\right\|}\left\|\sum_{t=\hat{k}+1}^{k_{0}} z_{t} z_{t}^{\prime}\right\|\left\|\delta_{T}\right\|^{2} \\
& +\left\|\left(\sum_{t=\hat{k}+1}^{T} z_{t} x_{t}^{\prime}\right)\left(\sum_{t=1}^{T} x_{t} x_{t}^{\prime}\right)^{-1}\right\| \frac{1}{\sqrt{T}\left\|\delta_{T}\right\|}\left\|\sum_{t=\hat{k}+1}^{k_{0}} x_{t} z_{t}^{\prime}\right\|\left\|\delta_{T}\right\|^{2} \\
= & \frac{1}{\sqrt{T}\left\|\delta_{T}\right\|} O_{p}(1)=o_{p}(1)
\end{aligned}
$$

Note that the sum has $k_{0}-\hat{k}=O_{p}\left(\left\|\delta_{T}\right\|^{-2}\right)$ terms, so $\left\|\Sigma_{t=\hat{k}+1}^{k_{0}} z_{t} z_{t}^{\prime}\right\|\left\|\delta_{T}\right\|^{2}=O_{p}(1)$. Also,

$$
\begin{aligned}
T^{-1}\left\|\hat{Z}_{0}^{\prime} M \hat{Z}_{0}-Z_{0}^{\prime} M Z_{0}\right\| \leq & T^{-1}\left\|\hat{Z}_{0}^{\prime} \hat{Z}_{0}-Z_{0}^{\prime} Z_{0}\right\|+T^{-1}\left\|\left(\hat{Z}_{0}-Z_{0}\right)^{\prime} X\left(X^{\prime} X\right)^{-1} X^{\prime} \hat{Z}_{0}\right\| \\
& +T^{-1}\left\|Z_{0}^{\prime} X\left(X^{\prime} X\right)^{-1} X^{\prime}\left(\hat{Z}_{0}-Z_{0}\right)\right\| \\
\leq & \frac{1}{T\left\|\delta_{T}\right\|^{2}}\left\|\sum_{\hat{k}+1}^{k_{0}} z_{t} z_{t}^{\prime}\right\|\left\|\delta_{T}\right\|^{2}+\frac{1}{T\left\|\delta_{T}\right\|^{2}}\left\|\sum_{\hat{k}+1}^{k_{0}} x_{t} z_{t}^{\prime}\right\|\left\|\delta_{T}\right\|^{2} O_{p}(1) \\
= & \frac{1}{T\left\|\delta_{T}\right\|^{2}} O_{p}(1)=o_{p}(1)
\end{aligned}
$$

Thus, $\sqrt{T}\left(\hat{\delta}(\hat{\rho})-\delta_{T}\right)=\left(T^{-1} Z_{0}^{\prime} M Z_{0}\right)^{-1} T^{-1 / 2} Z_{0}^{\prime} M \varepsilon+o_{p}(1)$ and the normality follows from the central limit theorem.

## Proof of Lemma 4

Proof. Use equation (1.12) to express terms in $F_{k}, F_{0}$ and $\Omega_{X, k}$, defined in (A.3).

$$
\begin{aligned}
\left|k_{0}-k\right| G_{T}(k)= & \delta_{h}^{\prime}\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{1 / 2} \delta_{h} \\
& -\delta_{h}\left(Z_{0}^{\prime} M Z_{k}\right)\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2}\left(Z_{k}^{\prime} M Z_{0}\right) \delta_{h} \\
= & \lambda_{h}^{2} d_{0}^{\prime} R^{\prime}\left(X_{0}^{\prime} X_{0}\right)\left[F_{0}^{1 / 2} \Omega_{X, k_{0}} F_{0}^{1 / 2}-F_{k}^{1 / 2} \Omega_{X, k} F_{k}^{1 / 2}\right]\left(X_{0}^{\prime} X_{0}\right) R d_{0} .
\end{aligned}
$$

The second equality is from Lemma A. 2 and $\left(X_{0}^{\prime} M X_{k}\right)\left(X_{k}^{\prime} M X_{k}\right)^{-1}=\left(X_{0}^{\prime} X_{0}\right)\left(X_{k}^{\prime} X_{k}\right)^{-1}$.
Lemma A.2. Under Assumption 1(i)-(iii),
$\left(Z_{0}^{\prime} M Z_{k}\right)\left(Z_{k}^{\prime} M Z_{k}\right)^{-1}=R^{\prime}\left(X_{0}^{\prime} M X_{k}\right)\left(X_{k}^{\prime} M X_{k}\right)^{-1} R$.
Similar to the proof of Lemma 1, the norm of the middle matrix is bounded by rearranging terms and from Assumption 2 where we have $\left\|\Omega_{k}-\Omega_{k_{0}}\right\| \leq b\left|k_{0}-k\right| / T$ for some finite $b>0$ and all $k$.

$$
\begin{aligned}
& \left\|F_{0}^{1 / 2} \Omega_{X, k_{0}} F_{0}^{1 / 2}-F_{k}^{1 / 2} \Omega_{X, k} F_{k}^{1 / 2}\right\| \\
= & \left\|\left(F_{0}-F_{k}\right)\left(F_{0}^{1 / 2}\right)^{-1} \Omega_{X, k_{0}} F_{0}^{1 / 2}-F_{k}\left\{\left(F_{k}^{1 / 2}\right)^{-1} \Omega_{X, k} F_{k}^{1 / 2}-\left(F_{0}^{1 / 2}\right)^{-1} \Omega_{X, k_{0}} F_{0}^{1 / 2}\right\}\right\| \\
\leq & \left\|\left(F_{0}-F_{k}\right) \Omega_{X, k_{0}}\right\|+b\left\|F_{k}\right\|\left|k_{0}-k\right| / T+o_{p}(1)
\end{aligned}
$$

Because $\left\|F_{k}\right\|=O_{p}\left(T^{-1}\right)$ and $\left(F_{0}-F_{k}\right)=\left(X_{0}^{\prime} X_{0}\right)^{-1}\left(X_{\Delta}^{\prime} X_{\Delta}\right)\left(X_{k}^{\prime} X_{k}\right)^{-1}$, we have

$$
F_{0}^{1 / 2} \Omega_{X, k_{0}} F_{0}^{1 / 2}-F_{k}^{1 / 2} \Omega_{X, k} F_{k}^{1 / 2}=\left(X_{0}^{\prime} X_{0}\right)^{-1}\left(X_{\Delta}^{\prime} X_{\Delta}\right)\left(X_{k}^{\prime} X_{k}\right)^{-1} \Omega_{X, k_{0}}+o_{p}(1)
$$

Then,

$$
\begin{align*}
\left|k_{0}-k\right| G_{T}(k) & =\lambda_{h}^{2} d_{0}^{\prime}\left[R^{\prime}\left(X_{\Delta}^{\prime} X_{\Delta}\right)\left(X_{k}^{\prime} X_{k}\right)^{-1} \Omega_{X, k_{0}}\left(X_{0}^{\prime} X_{0}\right) R\right] d_{0}+o_{p}(1) \\
& =\lambda_{h}^{2} d_{0}^{\prime}\left(Z_{\Delta}^{\prime} Z_{\Delta}\right) \Omega_{k_{0}} d_{0}+o_{p}(1) \tag{A.10}
\end{align*}
$$

where the second line is from $\left(X_{k}^{\prime} X_{k}\right)^{-1} \Omega_{X, k_{0}}\left(X_{0}^{\prime} X_{0}\right)=\Omega_{X, k_{0}} O_{p}(1)+o_{p}(1)$ by assumption. Next, consider $H_{T}(k)$ in equation (1.19).

$$
H_{T}(k)=2 \lambda_{h} d_{0}^{\prime} \Omega_{k_{0}} Z_{\Delta}^{\prime} \varepsilon \operatorname{sgn}\left(k_{0}-k\right)+T^{-1 / 2}\left\|\delta_{h}\right\|\left|k_{0}-k\right| O_{p}(1)+O_{p}(1)
$$

Because $\left|k_{0}-k\right| \leq C\left\|\delta_{h}\right\|^{-2}$ on $K(C)$, the second term in the equation above is bounded by $C T^{-1 / 2}\left\|\delta_{h}\right\|^{-1} O_{p}(1)=o_{p}(1)$. The last term $O_{p}(1)$ is $o_{p}(1)$ uniformly on $K(C)$, which can be
verified by rearranging terms using $Z_{0}=Z_{k}-Z_{\Delta} \operatorname{sgn}\left(k_{0}-k\right)$.

$$
\begin{aligned}
& \varepsilon^{\prime} M Z_{k}\left(Z_{k} M Z_{k}\right)^{-1 / 2} \Omega_{k}\left(Z_{k} M Z_{k}\right)^{-1 / 2} Z_{k}^{\prime} M \varepsilon-\varepsilon^{\prime} M Z_{0}\left(Z_{0} M Z_{0}\right)^{-1 / 2} \Omega_{k_{0}}\left(Z_{0} M Z_{0}\right)^{-1 / 2} Z_{0}^{\prime} M \varepsilon \\
&= \varepsilon^{\prime} M Z_{k}\left[\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2} \Omega_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1 / 2}-\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2}\right] Z_{k}^{\prime} M \varepsilon \\
&+\varepsilon^{\prime} M Z_{k}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} Z_{\Delta}^{\prime} M \varepsilon \operatorname{sgn}\left(k_{0}-k\right) \\
&+\varepsilon^{\prime} M Z_{0}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} \Omega_{k_{0}}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1 / 2} Z_{\Delta}^{\prime} M \varepsilon \operatorname{sgn}\left(k_{0}-k\right) \\
&= O_{p}\left(T^{-1}\left\|\delta_{h}\right\|^{-2}\right)+O_{p}\left(T^{-1 / 2}\left\|\delta_{h}\right\|^{-1}\right)+o_{p}(1)
\end{aligned}
$$

The first line is $O_{p}\left(T^{-1}\left\|\delta_{h}\right\|^{-2}\right)$ is uniformly on $K(C)$ :

$$
\begin{aligned}
\varepsilon^{\prime} M Z_{k} & {\left[\left(Z_{k}^{\prime} M Z_{k}\right)^{-1} \Omega_{k}-\left(Z_{0}^{\prime} M Z_{0}\right)^{-1} \Omega_{k_{0}}\right] Z_{k}^{\prime} M \varepsilon+o_{p}(1) } \\
= & \varepsilon^{\prime} M Z_{k}\left(Z_{k}^{\prime} M Z_{k}\right)^{-1}\left(Z_{0}^{\prime} M Z_{0}-Z_{k}^{\prime} M Z_{k}\right)\left(Z_{0}^{\prime} M Z_{0}\right)^{-1} \Omega_{k} Z_{k}^{\prime} M \varepsilon \\
& \quad+\varepsilon^{\prime} M Z_{k}\left(Z_{0}^{\prime} M Z_{0}\right)^{-1}\left(\Omega_{k}-\Omega_{k_{0}}\right) Z_{k}^{\prime} M \varepsilon+o_{p}(1) \\
& =T^{-1 / 2}\left\|\delta_{h}\right\|^{-2} O_{p}\left(T^{-1 / 2}\right)+o_{p}(1)
\end{aligned}
$$

from (1.18) and $\left(Z_{k}^{\prime} M Z_{k}\right)^{-1} Z_{k}^{\prime} M \varepsilon=O_{p}\left(T^{-1 / 2}\right)$ uniformly on $K(C)$. The second and third lines are $O_{p}\left(T^{-1 / 2}\left\|\delta_{h}\right\|^{-1}\right)$ from $Z_{\Delta}^{\prime} M \varepsilon \operatorname{sgn}\left(k_{0}-k\right)=\left|k_{0}-k\right|^{1 / 2} O_{p}(1)=O_{p}\left(\left\|\delta_{h}\right\|^{-1}\right)$. Hence,

$$
H_{T}(k)=2 \lambda_{h} d_{0}^{\prime} \Omega_{k_{0}} Z_{\Delta}^{\prime} \varepsilon \operatorname{sgn}\left(k_{0}-k\right)+o_{p}(1)
$$

Combine this with (A.10), we obtain the expression in Lemma 4.

## Proof of Lemma A. 2

Proof. Use the block matrix inverse formula (below) on $\left(X_{k}^{\prime} M X_{k}\right)^{-1}$.

$$
\left[\begin{array}{ll}
A & B \\
C & D
\end{array}\right]^{-1}=\left[\begin{array}{cc}
\left(A-B D^{-1} C\right)^{-1} & -\left(A-B D^{-1} C\right)^{-1} B D^{-1} \\
-D^{-1} C\left(A-B D^{-1} C\right)^{-1} & D^{-1}+D^{-1} C\left(A-B D^{-1} C\right)^{-1} B D^{-1}
\end{array}\right]
$$

Note that $X_{k}=\left[W_{k} \vdots Z_{k}\right]$ for all $k$. Then,

$$
\begin{align*}
R^{\prime}\left(X_{0}^{\prime} M X_{k}\right)\left(X_{k}^{\prime} M X_{k}\right)^{-1} R=- & \left(Z_{0}^{\prime} M W_{k}\right)\left(W_{k}^{\prime} M W_{k}\right)^{-1}\left(W_{0}^{\prime} M Z_{k}\right) \\
& \times\left[Z_{k}^{\prime} M Z_{k}-\left(Z_{k}^{\prime} M W_{k}\right)\left(W_{k}^{\prime} M W_{k}\right)^{-1}\left(W_{k}^{\prime} M Z_{k}\right)\right]^{-1} \\
& +Z_{0}^{\prime} M Z_{k}\left[Z_{k}^{\prime} M Z_{k}-\left(Z_{k}^{\prime} M W_{k}\right)\left(W_{k}^{\prime} M W_{k}\right)^{-1}\left(W_{k}^{\prime} M Z_{k}\right)\right]^{-1} \\
=( & \left.Z_{0}^{\prime} M \bar{M} M Z_{k}\right)\left(Z_{k}^{\prime} M \bar{M} M Z_{k}\right)^{-1} \tag{A.11}
\end{align*}
$$

where $\bar{M}:=I-M W_{k}\left(W_{k}^{\prime} M W_{k}\right)^{-1} W_{k}^{\prime} M$. Note that this is the OLS estimate of coefficients of multiple equation regression when we regress $\bar{M} M Z_{0}$ on $\bar{M} M Z_{k}$. This is equivalent to the coefficient of $M Z_{k}$ when we regress $M Z_{0}$ on $\left[M W_{k} \vdots M Z_{k}\right]=M X_{k}$. Thus, (A.11) is equivalent to the corresponding $(q \times q)$ block matrix of OLS estimate of regression $M X_{0}$ on $M X_{k}$, which is $R^{\prime}\left(X_{0}^{\prime} M X_{k}\right)\left(X_{k}^{\prime} M X_{k}\right)^{-1} R$.

## Proof of Theorem 5

Proof. Lemma B. 1 from Jiang et al. (2017) is restated below, and used without proof.
Lemma A.3. For the process $y_{t}$ defined in (1.22) the following equations hold when $T=1 / h \rightarrow \infty$ with a fixed $\rho_{0}=k_{0} / T$, for any $\rho \in[0,1]$,
(a) $T^{-1} \sum_{t=1}^{[\rho T]} y_{t-1} \varepsilon_{t} \Rightarrow \sigma^{2} \int_{0}^{\rho} \widetilde{J}_{0}(r) d B(r)$;
(b) $T^{-2} \sum_{t=1}^{[\rho T]} y_{t-1}^{2} \Rightarrow \sigma^{2} \int_{0}^{\rho}\left[\widetilde{J}_{0}(r)\right]^{2} d r$;
(c) $\left[\widetilde{J}_{0}(\rho)\right]^{2}-\left[\widetilde{J}_{0}(0)\right]^{2}=2 \int_{0}^{\rho} \widetilde{J}_{0}(r) d B(r)-2 \int_{0}^{\rho}\left(\mu+\delta \mathbf{1}\left\{r>\rho_{0}\right\}\right)\left[\widetilde{J}_{0}(r)\right]^{2} d r+\rho$;
(d) $\left[\widetilde{J}_{0}(1)\right]^{2}-\left[\widetilde{J}_{0}(\rho)\right]^{2}=2 \int_{\rho}^{1} \widetilde{J}_{0}(r) d B(r)-2 \int_{\rho}^{1}\left(\mu+\delta \mathbf{1}\left\{r>\rho_{0}\right\}\right)\left[\widetilde{J}_{0}(r)\right]^{2} d r+(1-\rho)$;
where $\widetilde{J}_{0}(r)$ for $r \in[0,1]$ is a Gaussian process defined in (1.24) and B(•) is a standard Brownian motion.

Define the $(T \times 2)$ matrix $Y(k)=\left[Y_{1}(k) \vdots Y_{2}(k)\right]$ with $Y_{1}(k)=\left(y_{0}, \ldots, y_{k-1}, 0 \ldots, 0\right)^{\prime}, Y_{2}(k)=$ $\left(0 \ldots, 0, y_{k}, \ldots, y_{T-1}\right)^{\prime}$ and $Y=\left(y_{1}, \ldots, y_{T}\right)^{\prime}$. Then the LS objective function can be expressed as $S(k)^{2}=Y^{\prime} M Y$ where

$$
M=I_{T}-Y_{1}(k)\left[Y_{1}(k)^{\prime} Y_{1}(k)\right]^{-1} Y_{1}(k)^{\prime}-Y_{2}(k)\left[Y_{2}(k)^{\prime} Y_{2}(k)\right]^{-1} Y_{2}(k)^{\prime}
$$

where $I_{T}$ is a $(T \times T)$ identity matrix. The model (1.22) can be written as

$$
y_{t}=\beta_{1} y_{t-1}+\left(\beta_{2}-\beta_{1}\right) \mathbf{1}\left\{t>k_{0}\right\} y_{t-1}+\varepsilon_{t}=\beta_{1} y_{t-1}+\eta_{t}
$$

where $\eta_{t}:=\left(\beta_{2}-\beta_{1}\right) \mathbf{1}\left\{t>k_{0}\right\} y_{t-1}+\varepsilon_{t}$. Let $Y_{-}=\left(y_{0}, \ldots, y_{T-1}\right)^{\prime}$ and $\eta=\left(\eta_{1}, \ldots, \eta_{T}\right)^{\prime}$. Then we have $Y=Y_{-} \beta_{1}+\eta$, and the LS objective function is

$$
\begin{aligned}
S(k)^{2} & =\left(Y_{-} \beta_{1}+\eta\right)^{\prime} M^{\prime} M\left(Y_{-} \beta_{1}+\eta\right) \\
& =\eta^{\prime} \eta-\eta^{\prime} Y_{1}(k)\left[Y_{1}(k)^{\prime} Y_{1}(k)\right]^{-1} Y_{1}(k)^{\prime} \eta-\eta^{\prime} Y_{2}(k)\left[Y_{2}(k)^{\prime} Y_{2}(k)\right]^{-1} Y_{2}(k)^{\prime} \eta
\end{aligned}
$$

because $M$ is a idempotent matrix and $M Y_{-}=\mathbf{0}$. Note that

$$
\eta^{\prime} \eta=\sum_{t=1}^{k_{0}} \eta_{t}^{2}+\sum_{t=k_{0}+1}^{T} \eta_{t}^{2}=\sum_{t=1}^{k_{0}} \varepsilon_{t}^{2}+\sum_{t=k_{0}+1}^{T}\left(\left(\beta_{2}-\beta_{1}\right) y_{t-1}+\varepsilon_{t}\right)^{2}
$$

which holds regardless of the choice of $k$, and

$$
\begin{aligned}
& \eta^{\prime} Y_{1}(k)\left[Y_{1}(k)^{\prime} Y_{1}(k)\right]^{-1} Y_{1}(k)^{\prime} \eta=\frac{\left(\sum_{t=1}^{k} y_{t-1} \eta_{t}\right)^{2}}{\sum_{t=1}^{k} y_{t-1}^{2}} \\
& \eta^{\prime} Y_{2}(k)\left[Y_{2}(k)^{\prime} Y_{2}(k)\right]^{-1} Y_{2}(k)^{\prime} \eta=\frac{\left(\sum_{t=k+1}^{T} y_{t-1} \eta_{t}\right)^{2}}{\sum_{t=k+1}^{T} y_{t-1}^{2}}
\end{aligned}
$$

Therefore, the break point estimator is

$$
\begin{gather*}
\hat{\rho}=\underset{\rho \in(0,1)}{\arg \max } \omega(\rho)^{2} \mathcal{V}(\rho),  \tag{A.12}\\
\mathcal{V}(\rho):=\left[\frac{\left(\sum_{t=1}^{[\rho T]} y_{t-1} \eta_{t}\right)^{2}}{\sum_{t=1}^{[\rho T]} y_{t-1}^{2}}+\frac{\left(\sum_{t=[\rho T]+1}^{T} y_{t-1} \eta_{t}\right)^{2}}{\sum_{t=[\rho T]+1}^{T} y_{t-1}^{2}}\right] .
\end{gather*}
$$

When $\rho \leq \rho_{0}$, the terms in the numerator and denominator of $\mathcal{V}(\rho)$ weakly converges as follows.

$$
T^{-1} \sum_{t=1}^{[\rho T]} y_{t-1} \eta_{t}=T^{-1} \sum_{t=1}^{[\rho T]} y_{t-1} \varepsilon_{t} \Rightarrow \sigma^{2} \int_{0}^{\rho} \widetilde{J}_{0}(r) d B(r) .
$$

From Lemma A.3,

$$
\begin{aligned}
& T_{t=[\rho T]+1}^{T} \sum_{t-1}^{T} \eta_{t}=T^{-1}\left[\sum_{t=[\rho T]+1}^{\left[\rho_{0} T\right]} y_{t-1} \eta_{t}+\sum_{t=\left[\rho_{0} T\right]+1}^{T} y_{t-1} \eta_{t}\right] \\
& \\
& =T^{-1} \sum_{t=[\rho T]+1}^{T} y_{t-1} \varepsilon_{t}+T\left(\beta_{2}-\beta_{1}\right) T^{-2} \sum_{t=\left[\rho_{0} T\right]+1}^{T} y_{t-1}^{2} \\
& \\
& \Rightarrow \sigma^{2} \int_{\rho}^{1} \widetilde{J}_{0}(r) d B(r)-\delta \sigma^{2} \int_{\rho_{0}}^{1}\left[\widetilde{J}_{0}(r)\right]^{2} d r, \\
& T^{-2} \sum_{t=1}^{[\rho T]} y_{t-1}^{2} \Rightarrow \sigma^{2} \int_{0}^{\rho}\left[\widetilde{J}_{0}(r)\right]^{2} d r, \text { and } T^{-2} \sum_{t=[\rho T]+1}^{T} y_{t-1}^{2} \Rightarrow \sigma^{2} \int_{\rho}^{1}\left[\widetilde{J}_{0}(r)\right]^{2} d r .
\end{aligned}
$$

Then the LS objective function $\mathcal{V}(\rho)$ in (A.12) weakly converges to

$$
\mathcal{V}(\rho) \Rightarrow \sigma^{2}\left[\frac{\left(\int_{0}^{\rho} \widetilde{J}_{0}(r) d B(r)\right)^{2}}{\int_{0}^{\rho}\left[\widetilde{J}_{0}(r)\right]^{2} d r}+\frac{\left(\int_{\rho}^{1} \widetilde{J}_{0}(r) d B(r)-\delta \int_{\rho_{0}}^{1}\left[\widetilde{J}_{0}(r)\right]^{2} d r\right)^{2}}{\int_{\rho}^{1}\left[\widetilde{J}_{0}(r)\right]^{2} d r}\right]
$$

Lemma A. 3 (c) and (d) implies that each term is rearranged as follows.

$$
\begin{aligned}
\begin{aligned}
& \frac{\left(\int_{0}^{\rho} \widetilde{J}_{0}(r) d B(r)\right)^{2}}{\int_{0}^{\rho}\left[\widetilde{J}_{0}(r)\right]^{2} d r}= \frac{\left(\left[\widetilde{J}_{0}(\rho)\right]^{2}-\left[\widetilde{J}_{0}(0)\right]^{2}-\rho+2 \mu \int_{0}^{\rho}\left[\widetilde{J}_{0}(r)\right]^{2} d r\right)^{2}}{4 \int_{0}^{\rho}\left[\widetilde{J}_{0}(r)\right]^{2} d r} \\
&= \frac{\left(\left[\widetilde{J}_{0}(\rho)\right]^{2}-\left[\widetilde{J}_{0}(0)\right]^{2}-\rho\right)^{2}}{4 \int_{0}^{\rho}\left[\widetilde{J}_{0}(r)\right]^{2} d r}+\mu^{2} \int_{0}^{\rho}\left[\widetilde{J}_{0}(r)\right]^{2} d r \\
&+\mu\left(\left[\widetilde{J}_{0}(\rho)\right]^{2}-\left[\widetilde{J}_{0}(0)\right]^{2}-\rho\right) \\
& \frac{\int_{\rho}^{1}\left[\widetilde{J}_{0}(r)\right]^{2} d r}{\left(\int_{\rho}^{1} \widetilde{J}_{0}(r) d B(r)-\delta \int_{\rho_{0}}^{1}\left[\widetilde{J}_{0}(r)\right]^{2} d r\right)^{2}}=\frac{\left(\left[\widetilde{J}_{0}(1)\right]^{2}-\left[\widetilde{J}_{0}(\rho)\right]^{2}-(1-\rho)\right)^{2}}{4 \int_{\rho}^{1}\left[\widetilde{J}_{0}(r)\right]^{2} d r} \\
&+\mu^{2} \int_{\rho}^{1}\left[\widetilde{J}_{0}(r)\right]^{2} d r+\mu\left(\left[\widetilde{J}_{0}(1)\right]^{2}-\left[\widetilde{J}_{0}(\rho)\right]^{2}-(1-\rho)\right)
\end{aligned}
\end{aligned}
$$

As a result, the objective function of the break point estimator in (A.12) weakly converges to

$$
\begin{aligned}
\frac{\omega(\rho)^{2} \mathcal{V}(\rho)}{\sigma^{2}} \Rightarrow & \frac{\omega(\rho)^{2}\left(\left[\widetilde{J}_{0}(\rho)\right]^{2}-\left[\widetilde{J}_{0}(0)\right]^{2}-\rho\right)^{2}}{4 \int_{0}^{\rho}\left[\widetilde{J}_{0}(r)\right]^{2} d r}+\frac{\omega(\rho)^{2}\left(\left[\widetilde{J}_{0}(1)\right]^{2}-\left[\widetilde{J}_{0}(\rho)\right]^{2}-(1-\rho)\right)^{2}}{4 \int_{\rho}^{1}\left[\widetilde{J}_{0}(r)\right]^{2} d r} \\
& +\mu^{2} \omega(\rho)^{2} \int_{0}^{1}\left[\widetilde{J}_{0}(r)\right]^{2} d r+\mu \omega(\rho)^{2}\left(\left[\widetilde{J}_{0}(1)\right]^{2}-\left[\widetilde{J}_{0}(0)\right]^{2}-1\right)
\end{aligned}
$$

Following the same procedure above, $\omega(\rho)^{2} \mathcal{V}(\rho) / \sigma^{2}$ has the same limit when $\rho>\rho_{0}$. Therefore,
by deleting the terms which are independent of the choice of $\rho$, the in-fill asymptotic distribution of $\hat{\rho}$ in (A.12) is

$$
\begin{aligned}
\hat{\rho} & =\underset{\rho \in(0,1)}{\arg \max } \omega(\rho)^{2} \mathcal{V}(\rho) \\
& \Rightarrow \underset{\rho \in(0,1)}{\arg \max } \frac{\omega(\rho)^{2}\left(\left[\widetilde{J}_{0}(\rho)\right]^{2}-\left[\widetilde{J}_{0}(0)\right]^{2}-\rho\right)^{2}}{\int_{0}^{\rho}\left[\widetilde{J}_{0}(r)\right]^{2} d r}+\frac{\omega(\rho)^{2}\left(\left[\widetilde{J}_{0}(1)\right]^{2}-\left[\widetilde{J}_{0}(\rho)\right]^{2}-(1-\rho)\right)^{2}}{\int_{\rho}^{1}\left[\widetilde{J}_{0}(r)\right]^{2} d r},
\end{aligned}
$$

which is identical to the distribution in Theorem 5.

## A. 2 Appendix for Chapter 2

## A.2.1 Comparing methods

Based on model (2.1) the Nyblom test statistic is

$$
\begin{align*}
\hat{L} / T^{2} & =T^{-2} \operatorname{tr}\left[\hat{\Omega}_{Q}^{-1} \sum_{t=1}^{T}\left(\sum_{s=t}^{T} Q_{s} \hat{u}_{s}\right)\left(\sum_{s=t}^{T} Q_{s} \hat{u}_{s}\right)^{\prime}\right]  \tag{A.13}\\
& \Rightarrow \int(W(l)-l W(1))^{\prime}(W(l)-l W(1)) d l
\end{align*}
$$

where $\Omega_{Q}=\lim _{T \rightarrow \infty} \operatorname{Var}\left(T^{-1 / 2} \sum_{t=1}^{T} Q_{t} u_{t}\right)$ and the Wiener process $W(1)$ has dimension equal to $\operatorname{dim}(Q)=k+m$. The Nyblom test statistic null limit distribution is the square of independent scalar Brownian bridges. The asymptotic critical value of $\hat{L} / T^{2}$ depends only on the dimension of $Q_{t}=\left(X_{t}^{\prime}, Z_{t}^{\prime}\right)^{\prime}$. The Nyblom test does not assume that the direction of test statistic is known, hence the null hypothesis is $\delta=0$ against the alternative $\delta \neq 0$. The local alternative asymptotic distribution under $\delta=T^{-1 / 2} d$ is as follows. Denote the inner product as $A^{\otimes 2}=A^{\prime} A$ where $A$ is a $(k \times 1)$ vector.

$$
\left.\int\left[W(l)-l W(1)+\left(\min \left\{l, r_{0}\right\}\right)-l r_{0}\right) \Omega_{Q}^{-1 / 2} \Sigma_{Q} d\right]^{\otimes 2} d l .
$$

The (average) exponential LM test from Andrews and Ploberger (1994) has the same asymptotic properties to the exponential Wald and LR test under the null and local alternatives. The exponential LM test statistic (denote as Exp- $L M_{T}$ ) does not specify the direction of the structural break like the Nyblom test statistic, hence the null and alternative hypotheses are

$$
\begin{aligned}
& \mathcal{H}_{0}: \delta=0 \\
& \mathcal{H}_{1}: \delta \neq 0 \text { and the likelihood function depends on the parameter } l .
\end{aligned}
$$

$\operatorname{Exp}-L M_{T}$ test statistic is derived under local alternatives to the null $\Theta_{0}$ of the form $f_{T}\left(\theta_{0}+B_{T}^{-1} h, l\right)$ for some $l \in \Pi \subset(0,1)$, some $h \in \mathbb{R}^{p}$, and $B_{T}=\sqrt{T} I_{p}$ in our model without trending variables. The test has the greatest weighted average power asymptotically in the class of all test of asymptotic significance level $\alpha$ with weighting function over $h, Q_{l}(h)$ and over $l, J(l)$. Given a constant $c>0$ that depends on the weight functions $Q_{l}(\cdot)$, the asymptotically optimal test statistic $\operatorname{Exp}-L M_{T c}$ is

$$
\begin{equation*}
\operatorname{Exp}-L M_{T c}=(1+c)^{-p / 2} \int \exp \left(\frac{1}{2} \frac{c}{1+c} L M_{T}(l)\right) d J(l) \tag{A.14}
\end{equation*}
$$

The larger $c$ is, the more weight is given to alternatives for which break magnitude $\delta$ is large. We consider two cases $c \rightarrow 0$ and $c \rightarrow \infty$ which is as follows, suitably normalized.

$$
\begin{gathered}
\lim _{c \rightarrow 0} 2\left(\operatorname{Exp}-L M_{T c}-1\right) / c=\int L M_{T}(l) d J(l) \\
\lim _{c \rightarrow \infty} \log \left((1+c)^{p / 2} \operatorname{Exp}-L M_{T c}\right)=\log \int \exp \left(\frac{1}{2} L M_{T}(l)\right) d J(l)
\end{gathered}
$$

The directional break test statistics are similar to $\operatorname{Exp}-L R_{T}$ under Gaussian disturbances and $c \rightarrow 0$. If we assume $J(l)$ is an uniform weight function over $[\varepsilon, 1-\varepsilon], \varepsilon>0$ then under the null,

$$
\begin{aligned}
\int_{\varepsilon}^{1-\varepsilon} L M_{T}^{*}(l) d l & \Rightarrow \int_{\varepsilon}^{1-\varepsilon} \frac{(W(l)-l W(1))^{\prime}(W(l)-l W(1))}{l(1-l)} d l \\
\log \int_{\varepsilon}^{1-\varepsilon} \exp \left(\frac{1}{2} L M_{T}^{*}(l)\right) d l & \Rightarrow \log \int_{\varepsilon}^{1-\varepsilon} \exp \left(\frac{1}{2} \frac{(W(l)-l W(1))^{\prime}(W(l)-l W(1))}{l(1-l)}\right) d l .
\end{aligned}
$$

From our model (2.1) the test statistic $L M_{T}^{*}(l)$ is the test statistic when the break location $l$ is known.

$$
L M_{T}^{*}(l)=\frac{1}{l(1-l)}\left(T^{-1 / 2} \sum_{t=[T l]+1}^{T} X_{t} \hat{u}_{t}\right)^{\prime} \hat{\Omega}^{-1}\left(T^{-1 / 2} \sum_{t=[T l]+1}^{T} X_{t} \hat{u}_{t}\right) .
$$

where $\hat{u}_{t}$ is the residual under the null (no break). Under the local alternative $\delta=T^{-1 / 2} d$ and $c \rightarrow 0$,

$$
\int_{\varepsilon}^{1-\varepsilon} L M_{T}^{*}(l) d l \Rightarrow \int_{\varepsilon}^{1-\varepsilon} \frac{1}{l(1-l)}\left[W(l)-l W(1)+\left(\min \left\{l, r_{0}\right\}-l r_{0}\right) \Omega^{-1 / 2} \Sigma_{X} d\right]^{\otimes 2} d l
$$

where $r_{0}$ is the true break location and $\Omega$ is the long-run variance of $\left\{X_{t} u_{t}\right\}$.
Refer to Elliott and Müller (2006) for the construction and asymptotic distribution of the "quasi Local Level" test statistic $\widehat{q L L}$.

## A.2.2 Proofs for Chapter 2

## Proof of Lemma 5

Proof. Let $R=\left[I_{k} \mathbf{0}\right]^{\prime}$ be a $((k+m) \times k)$ matrix where $I_{k}$ is an identity matrix with dimension $k$ and $\mathbf{0}$ a $(k \times m)$ matrix of zeros. Then we have $R^{\prime} Q^{\prime} Q=X^{\prime} Q$. From the definition of $M$ and $l=t / T$ we have

$$
\begin{aligned}
T^{-1 / 2}(X-X(t))^{\prime} M U= & T^{-1 / 2}(X-X(t))^{\prime} U-(X-X(t))^{\prime} Q\left(Q^{\prime} Q\right)^{-1} T^{-1 / 2} Q^{\prime} U \\
= & T^{-1 / 2}(X-X(t))^{\prime} U-T^{-1} R^{\prime}(Q-Q(t))^{\prime} Q\left(T^{-1} Q^{\prime} Q\right)^{-1} T^{-1 / 2} Q^{\prime} U \\
= & T^{-1 / 2}(X-X(t))^{\prime} U-(1-l) R^{\prime} \Sigma_{Q} \Sigma_{Q}^{-1} T^{-1 / 2} Q^{\prime} U \\
& \quad+\left[(1-l) R^{\prime} \Sigma_{Q} \Sigma_{Q}^{-1}-T^{-1} R^{\prime}(Q-Q(t))^{\prime} Q\left(T^{-1} Q^{\prime} Q\right)^{-1}\right] T^{-1 / 2} Q^{\prime} U \\
= & T^{-1 / 2}(X-X(t))^{\prime} U-(1-l) R^{\prime} \Sigma_{Q} \Sigma_{Q}^{-1} T^{-1 / 2} Q^{\prime} U+o_{p}(1)
\end{aligned}
$$

where the last equation is from $(1-l) R^{\prime} \Sigma_{Q} \Sigma_{Q}^{-1}-T^{-1} R^{\prime}(Q-Q(t))^{\prime} Q\left(T^{-1} Q^{\prime} Q\right)^{-1} \xrightarrow{p} 0$ by Condition 1(iv) and $T^{-1 / 2} Q^{\prime} U=\left[\begin{array}{c}T^{-1 / 2} X^{\prime} U \\ T^{-1 / 2} Z^{\prime} U\end{array}\right]=O_{p}(1)$ from Condition 1 (ii) and (iii). Hence we have

$$
\begin{aligned}
T^{-1 / 2}(X-X(t))^{\prime} M U & =T^{-1 / 2}(X-X(t))^{\prime} U-(1-l) T^{-1 / 2} X^{\prime} U+o_{p}(1) \\
& \Rightarrow \Omega^{1 / 2}(W(1)-W(l))-(1-l) \Omega^{1 / 2} W(1) \\
& =-\Omega^{1 / 2}(W(l)-l W(1))
\end{aligned}
$$

We can ignore the minus sign because the distribution $W(l)-l W(1)$ is symmetric.

## Proof of Theorem 6

Proof. From Lemma 5, consistency of estimators $\hat{\Sigma}_{X}, \hat{\Omega}$, and continuous mapping theorem we have

$$
\begin{aligned}
& \left(\frac{\bar{d}^{\prime} \bar{d}}{\bar{\delta}^{\prime} \hat{\Sigma}_{X}^{-1 / 2} \hat{\Omega} \hat{\Sigma}_{X}^{-1 / 2} \bar{\delta}}\right)^{1 / 2} \bar{\delta}^{\prime} \hat{\Sigma}_{X}^{-1 / 2}\left(T^{-1 / 2} \sum_{s=t+1}^{T} \hat{v}_{s}\right) \\
& \Rightarrow\left(\frac{\bar{d}^{\prime} \bar{d}}{\bar{\delta}^{\prime} \Sigma_{X}^{-1 / 2} \Omega \Sigma_{X}^{-1 / 2} \bar{\delta}}\right)^{1 / 2} \bar{\delta}^{\prime} \Sigma_{X}^{-1 / 2} \Omega^{1 / 2}(W(l)-l W(1)) \\
& \quad=\left(\bar{d}^{\prime} \bar{d}\right)^{1 / 2} B B(l)
\end{aligned}
$$

where the last equation is from $\bar{\delta}^{\prime} \Sigma_{X}^{-1 / 2} \Omega^{1 / 2}(W(l)-l W(1))=\left(\bar{\delta}^{\prime} \Sigma_{X}^{-1 / 2} \Omega \Sigma_{X}^{-1 / 2} \bar{\delta}\right)^{1 / 2} B B(l)$ where $B B(l)$ is a univariate Brownian Bridge. The second term inside the exponential function of $d b_{T}^{a}$, which is $\frac{1}{2} \bar{d}^{\prime} \bar{d}\left(\frac{t}{T}\right)\left(1-\frac{t}{T}\right)$, is asymptotically equivalent to the second term of the likelihood ratio test statistic in (2.9):

$$
\begin{aligned}
& \frac{1}{2}(\bar{d} \circ \bar{\delta})^{\prime} \Sigma_{X}^{-1 / 2} T^{-1}(X-X(t))^{\prime} M(X-X(t)) \Sigma_{X}^{-1 / 2}(\bar{d} \circ \bar{\delta}) \\
= & \frac{1}{2}(\bar{d} \circ \bar{\delta})^{\prime} \Sigma_{X}^{-1 / 2}\left[T^{-1}(X-X(t))^{\prime}(X-X(t))\right. \\
& \left.\quad-T^{-1} R^{\prime}(Q-Q(t))^{\prime} Q\left(T^{-1} Q^{\prime} Q\right)^{-1} T^{-1} Q^{\prime}(Q-Q(t)) R\right] \Sigma_{X}^{-1 / 2}(\bar{d} \circ \bar{\delta}) \\
\rightarrow & \frac{1}{2}(\bar{d} \circ \bar{\delta})^{\prime} \Sigma_{X}^{-1 / 2}\left[(1-l) \Sigma_{X}-(1-l)^{2} R^{\prime} \Sigma_{Q} \Sigma_{Q}^{-1} \Sigma_{Q} R\right] \Sigma_{X}^{-1 / 2}(\bar{d} \circ \bar{\delta}) \\
= & \frac{1}{2} \bar{d}^{\prime} \bar{d} l(1-l),
\end{aligned}
$$

where the last equation is from $(\bar{d} \circ \bar{\delta})^{\prime}(\bar{d} \circ \bar{\delta})=\bar{d}^{\prime} \bar{d}$.
For $d b_{T}^{c}$ the application of the continuous mapping theorem yields the distribution

$$
d b_{T}^{c} \Rightarrow \sqrt{12} \int_{0}^{1} B B(l) d l=\sqrt{12}\left[\int_{0}^{1} W_{1}(l) d l-\frac{1}{2} W_{1}(1)\right]
$$

where $W_{1}(l)$ is a univariate Brownian Motion. The mean and variance of the distribution can be
computed from

$$
\binom{\int_{0}^{1} W_{1}(l) d l}{W_{1}(1)} \sim N\left(\binom{0}{0},\left(\begin{array}{cc}
\frac{1}{3} & \frac{1}{2} \\
\frac{1}{2} & 1
\end{array}\right)\right)
$$

The mean is zero and the variance is $\frac{1}{3}+\frac{1}{4}-2 \cdot \frac{1}{2} \cdot \frac{1}{2}=\frac{1}{12}$, thus the null limit distribution of $d b_{T}^{c}$ is a standard normal distribution.

## Proof of Theorem 7

Proof. Under the alternative we have $M Y=M\left(X-X\left(\tau_{0}\right)\right) \delta+M U$, and if $\delta=T^{-1 / 2} \sigma \Sigma_{X}^{-1 / 2} d$ then

$$
T^{-1 / 2} \sum_{s=t+1}^{T} \hat{v}_{s}=T^{-1 / 2}(X-X(t))^{\prime} M U+T^{-1} \sigma(X-X(t))^{\prime} M\left(X-X\left(\tau_{0}\right)\right) \Sigma_{X}^{-1 / 2} d
$$

The first term is the same as under the null hypothesis. Consider the second term:

$$
\begin{aligned}
& T^{-1}(X-X(t))^{\prime} M\left(X-X\left(\tau_{0}\right)\right) \\
= & T^{-1}(X-X(t))^{\prime}\left(X-X\left(\tau_{0}\right)\right)-T^{-1} R^{\prime}(Q-Q(t))^{\prime} Q\left(T^{-1} Q^{\prime} Q\right)^{-1} T^{-1} Q^{\prime}\left(Q-Q\left(\tau_{0}\right)\right) R
\end{aligned}
$$

From Condition 1(iv),

$$
\begin{gathered}
T^{-1}(X-X(t))^{\prime}\left(X-X\left(\tau_{0}\right)\right) \xrightarrow{p}\left(1-\max \left\{l, r_{0}\right\}\right) \Sigma_{X} \\
T^{-1} R^{\prime}(Q-Q(t))^{\prime} Q\left(T^{-1} Q^{\prime} Q\right)^{-1} T^{-1} Q^{\prime}\left(Q-Q\left(\tau_{0}\right)\right) R \xrightarrow{p}(1-l)\left(1-r_{0}\right) R^{\prime} \Sigma_{Q} \Sigma_{Q}^{-1} \Sigma_{Q} R
\end{gathered}
$$

Hence, by continuous mapping theorem the second term converges as follows.

$$
\begin{aligned}
& T^{-1} \sigma(X-X(t))^{\prime} M\left(X-X\left(\tau_{0}\right)\right) \Sigma_{X}^{-1 / 2} d \\
\xrightarrow{p} & \sigma\left[\left(1-\max \left\{l, r_{0}\right\}\right) \Sigma_{X}-(1-l)\left(1-r_{0}\right) \Sigma_{X}\right] \Sigma_{X}^{-1 / 2} d \\
= & \sigma\left(\min \left\{l, r_{0}\right\}-l r_{0}\right) \Sigma_{X}^{1 / 2} d
\end{aligned}
$$

Again, by continuous mapping theorem we have

$$
\begin{aligned}
& \left(\frac{\bar{d}^{\prime} \bar{d}}{\bar{\delta}^{\prime} \hat{\Sigma}_{X}^{-1 / 2} \hat{\Omega} \hat{\Sigma}_{X}^{-1 / 2} \bar{\delta}}\right)^{1 / 2} \bar{\delta}^{\prime} \hat{\Sigma}_{X}^{-1 / 2}\left(T^{-1 / 2} \sum_{s=t+1}^{T} \hat{v}_{s}\right) \\
\Rightarrow & \frac{\sigma\left(\bar{d}^{\prime} \bar{d}\right)^{1 / 2}\left(\min \left\{l, r_{0}\right\}-l r_{0}\right)}{\left(\bar{\delta}^{\prime} \Sigma_{X}^{-1 / 2} \Omega \Sigma_{X}^{-1 / 2} \bar{\delta}\right)^{1 / 2}} \bar{\delta}^{\prime} d+\left(\bar{d}^{\prime} \bar{d}\right)^{1 / 2} B B(l)
\end{aligned}
$$

Under conditional homoskedasticity, we have $\sigma\left(\bar{d}^{\prime} \bar{d}\right)^{1 / 2} /\left(\bar{\delta}^{\prime} \Sigma_{X}^{-1 / 2} \Omega \Sigma_{X}^{-1 / 2} \bar{\delta}\right)^{1 / 2}=\left(\bar{d}^{\prime} \bar{d} /\left(\bar{\delta}^{\prime} \bar{\delta}\right)\right)^{1 / 2}$ $=\bar{d}_{1}$ under $\bar{d}=\bar{d}_{1} 1_{k}$ and the local alternative asymptotic distribution of the test statistics reduces to the distribution inside the probability function of (2.15), (2.16) and (2.17).

## A. 3 Appendix for Chapter 3

We show that our method is a special case of smoothing the direct forecast estimates themselves. Denote $\hat{\beta}_{L S, h}$ a $(n \times 1)$ vector as the horizon $h$ direct regression least-square estimates. Suppose we want to smooth out the estimates by taking a weighted average of the previous $h-1$ and $h$ to obtain a smoothed forecast, $\hat{\beta}_{h}=\omega_{h} \hat{\beta}_{L S, h}+\left(1-\omega_{h}\right) \hat{\beta}_{L S, h-1}$. Let $\hat{\beta}=\left(\hat{\beta}_{1}^{\prime}, \ldots, \hat{\beta}_{H}^{\prime}\right)^{\prime}$ and $\hat{\beta}_{L S}=\left(\hat{\beta}_{L S, 1}^{\prime}, \ldots, \hat{\beta}_{L S, H}^{\prime}\right)^{\prime}$. Then we can express the smoothing procedure in a linear function of
direct forecasts $\hat{\boldsymbol{\beta}}=\mathbf{A} \hat{\boldsymbol{\beta}}_{L S}$, where $\mathbf{A}$ is a $(n H \times n H)$ matrix.

$$
\left[\begin{array}{c}
\hat{\beta}_{1} \\
\vdots \\
\hat{\beta}_{H}
\end{array}\right]=\underbrace{\left[\begin{array}{ccccc}
\mathbf{I}_{n} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\
\left(1-\omega_{2}\right) \mathbf{I}_{n} & \omega_{2} \mathbf{I}_{n} & \cdots & \mathbf{0} & \mathbf{0} \\
\vdots & \vdots & \cdots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \left(1-\omega_{H}\right) \mathbf{I}_{n} & \omega_{H} \mathbf{I}_{n}
\end{array}\right]}_{\mathbf{A}}\left[\begin{array}{c}
\hat{\beta}_{L S, 1} \\
\vdots \\
\hat{\beta}_{L S, H}
\end{array}\right]
$$

The smoothed direct forecasts can also be expressed as a linear function of $\hat{\beta}_{L S}$ with a coefficient matrix. In section 2.2 we showed the smoothed direct forecasts are

$$
\begin{align*}
\hat{\beta} & =\left[T^{-1} \sum_{t}\left(\hat{\mathbf{V}}^{-1} \otimes \mathbf{x}_{t} \mathbf{x}_{t}^{\prime}\right)+\mathbf{M}^{-1}\right]^{-1}\left[T^{-1} \sum_{t}\left(\hat{\mathbf{V}}^{-1} \otimes \mathbf{x}_{t}\right) \mathbf{Y}_{t}\right] \\
& =\left[T^{-1} \mathbf{X}_{v}^{\prime} \mathbf{X}_{v}+\mathbf{M}^{-1}\right]^{-1}\left[T^{-1} \sum_{t}\left(\hat{\mathbf{V}}^{-1} \otimes \mathbf{x}_{t}\right) \mathbf{Y}_{t}\right] \tag{A.15}
\end{align*}
$$

where $\mathbf{X}_{v}^{\prime} \mathbf{X}_{v}=\hat{\mathbf{V}}^{-1} \otimes \mathbf{X}^{\prime} \mathbf{X}$. Since $\mathbf{M}^{-1}$ is a positive definite matrix by construction, $\mathbf{M}^{-1}=\mathbf{P}^{\prime} \Lambda \mathbf{P}$ by triangular factorization. If we use the matrix inversion lemma on the first term of $\hat{\beta}$, then

$$
\begin{aligned}
&\left(T^{-1} \mathbf{X}_{v}^{\prime} \mathbf{X}_{v}+\mathbf{M}^{-1}\right)^{-1}=\left(T^{-1} \mathbf{X}_{v}^{\prime} \mathbf{X}_{v}\right)^{-1} \\
&-\left(T^{-1} \mathbf{X}_{v}^{\prime} \mathbf{X}_{v}\right)^{-1} \mathbf{P}^{\prime}\left(\Lambda+\mathbf{P}\left(T^{-1} \mathbf{X}_{v}^{\prime} \mathbf{X}_{v}\right)^{-1} \mathbf{P}^{\prime}\right)^{-1} \mathbf{P}\left(T^{-1} \mathbf{X}_{v}^{\prime} \mathbf{X}_{v}\right)^{-1}
\end{aligned}
$$

If we plug this into (A.15),

$$
\begin{aligned}
\hat{\beta}=\left(T^{-1} \mathbf{X}_{v}^{\prime} \mathbf{X}_{v}\right)^{-1} & \left(T^{-1} \sum_{t}\left(\hat{\mathbf{V}}^{-1} \otimes \mathbf{x}_{t}\right) \mathbf{Y}_{t}\right)-\left(T^{-1} \mathbf{X}_{v}^{\prime} \mathbf{X}_{v}\right)^{-1} \mathbf{P}^{\prime} \\
& \times\left(\Lambda^{-1}+\mathbf{P}\left(T^{-1} \mathbf{X}_{v}^{\prime} \mathbf{X}_{v}\right)^{-1} \mathbf{P}^{\prime}\right)^{-1} \mathbf{P}\left(T^{-1} \mathbf{X}_{v}^{\prime} \mathbf{X}_{v}\right)^{-1}\left(T^{-1} \sum_{t}\left(\hat{\mathbf{V}}^{-1} \otimes \mathbf{x}_{t}\right) \mathbf{Y}_{t}\right) .
\end{aligned}
$$

The first line is equivalent to the least-square estimates of the direct forecast regression since the diagonal matrix $\hat{\mathbf{V}}^{-1}$ is cancelled out and gives us

$$
\hat{\beta}_{L S}=\left(T^{-1} \sum_{t} \mathbf{I}_{H} \otimes \mathbf{x}_{t} \mathbf{x}_{t}^{\prime}\right)^{-1}\left(T^{-1} \sum_{t}\left(\mathbf{I}_{H} \otimes \mathbf{x}_{t}\right) \mathbf{Y}_{t}\right) .
$$

The last two terms of the second line also gives us least-square estimates. Therefore the smoothed direct forecast estimates can be expressed as a linear function of direct forecast estimates.

$$
\hat{\beta}=\left[\mathbf{I}_{n H}-\left(T^{-1} \mathbf{X}_{v}^{\prime} \mathbf{X}_{v}\right)^{-1} \mathbf{P}^{\prime}\left(\Lambda^{-1}+\mathbf{P}\left(T^{-1} \mathbf{X}_{v}^{\prime} \mathbf{X}_{v}\right)^{-1} \mathbf{P}^{\prime}\right)^{-1} \mathbf{P}\right] \hat{\beta}_{L S} .
$$

## Forecast accuracy test

We describe West's (1996) forecast accuracy test of two competing models. The competing models are described in section 3.2, which are the direct forecast and smoothed direct forecast.

For the smoothed direct forecast, "asymptotic irrelevance" does not hold and we need to take into account estimation uncertainty. Asymptotic irrelevance means that one conducts inference by applying standard results to the mean of the loss function of interest, treating parameter estimation as irrelevant (West, 1996). Asymptotic irrelevance holds when predictors are uncorrelated with the prediction error if the object of interest is the out-of-sample MSFE. Because smoothed direct forecasts of all horizons are estimated jointly, we need to derive a joint distribution of the horizon $h$ direct forecast and smoothed direct forecasts of all horizons in order to compare the MSFE at horizon $h$.

Assumptions 1-4 from West (1996) are maintained and the same notations are used to derive the joint distribution of direct and smoothed forecasts. Let $E f_{t+h}:=E\left[\left(y_{t+h}-\mathbf{x}_{t}^{\prime} \gamma_{h}\right)\right]$ be the forecast measure of interest, which is the MSFE at horizon $h$. Then the estimate is $\hat{f}_{t+h}:=\left(\hat{u}_{h, t+h}^{2}, \hat{v}_{1, t+1}^{2}, \hat{v}_{2, t+2}^{2}, \hat{v}_{3, t+3}^{2}, \ldots, \hat{v}_{H, t+H}^{2}\right)^{\prime}$ where $\hat{u}_{h, t+h}=y_{t+h}-\mathbf{x}_{t}^{\prime} \hat{\gamma}_{h t}$ is the direct forecast residual and $\hat{v}_{h, t+h}=y_{t+h}-\mathbf{x}_{t}^{\prime} \hat{\beta}_{h t}$ is the smoothed direct forecast residual, each from horizon $h$
regression. Estimators $\hat{\gamma}_{h t}$ and $\hat{\beta}_{h t}$ satisfies $\hat{\gamma}_{t}-\gamma^{*}=\mathbf{B}(t) \mathbf{H}(t)$ and $\hat{\beta}_{t}-\beta^{*}=\tilde{\mathbf{B}}(t) \tilde{\mathbf{H}}(t)$, where

$$
\begin{aligned}
& \mathbf{H}(t)=t^{-1} \sum_{j=1}^{t} h_{j}\left(\gamma^{*}\right)=t^{-1} \sum_{j=1}^{t}\left(\mathbf{y}_{j}-\left(\mathbf{I}_{H} \otimes \mathbf{x}_{j}\right)^{\prime} \gamma^{*}\right) \otimes \mathbf{x}_{j} \\
& \tilde{\mathbf{H}}(t)=t^{-1} \sum_{j=1}^{t} \tilde{h}_{j}\left(\gamma^{*}\right)=t^{-1} \sum_{j=1}^{t} \hat{\mathbf{V}}^{-1}\left(\mathbf{y}_{j}-\left(\mathbf{I}_{H} \otimes \mathbf{x}_{j}\right)^{\prime} \beta^{*}\right) \otimes \mathbf{x}_{j}-\mathbf{M}^{-1} \beta^{*}
\end{aligned}
$$

and $\mathbf{y}_{t}=\left(y_{t+1}, \ldots, y_{t+H}\right)$. Hence expectations of $h_{j}\left(\gamma^{*}\right)$ and $\tilde{h}_{j}\left(\beta^{*}\right)$ are orthogonality conditions for each estimator: $E\left[h_{j}\left(\gamma^{*}\right)\right]=0, E\left[\tilde{h}_{j}\left(\beta^{*}\right)\right]=0$. Denote the horizon $h$ component of $h_{j}\left(\gamma^{*}\right)$ as $h_{j, h}\left(\gamma^{*}\right) \equiv \mathbf{x}_{j} u_{j+h}$.

Let $\bar{f}=P^{-1} \sum_{t=R}^{T} \hat{f}_{t+h}, \theta=\left(\gamma_{h}, \beta_{1}, \beta_{2}, \ldots, \beta_{H}\right), f_{t \theta}:=\frac{\partial f_{t+h}}{\partial \theta}, h_{j}(\theta):=\left(h_{j, h}\left(\gamma^{*}\right), \tilde{h}_{j}\left(\beta^{*}\right)\right)$, and $\mathbf{F}:=E\left[f_{t \theta}\right]$. Below are terms we need to derive the distribution of $\left(\bar{f}-E f_{t}\right)$.

$$
\begin{aligned}
\Gamma_{f h}(j) & =E\left(f_{t}-E f_{t}\right) h_{t-j}^{\prime}, S_{f h}=\sum_{j=-\infty}^{\infty} \Gamma_{f h}(j), \\
\Gamma_{h h}(j) & =E\left(h_{t} h_{t-j}^{\prime}\right), S_{h h}=\sum_{j=-\infty}^{\infty} \Gamma_{h h}(j), \\
\Gamma_{f f}(j) & =E\left[\left(f_{t}-E f_{t}\right)\left(f_{t}-E f_{t}\right)^{\prime}\right], S_{f f}=\sum_{j=-\infty}^{\infty} \Gamma_{f f}(j), \\
\quad \Pi & =1-\pi^{-1} \ln (1+\pi), \quad \lim _{T \rightarrow \infty}(P / R)=\pi, 0 \leq \pi<\infty,
\end{aligned}
$$

Theorem 4.1 of West (1996) implies that

$$
\begin{aligned}
P^{1 / 2}\left(\bar{f}-E f_{t}\right) & \stackrel{A}{\sim} N(\mathbf{0}, \Omega) \\
\Omega & =S_{f f}+\Pi\left(\mathbf{F B} S_{f h}^{\prime}+S_{f h} \mathbf{B}^{\prime} \mathbf{F}\right)+2 \Pi \mathbf{F B} S_{h h} \mathbf{B}^{\prime} \mathbf{F}^{\prime} .
\end{aligned}
$$

A consistent estimator of $\Omega$ is obtained by replacing the parameters to the sample analogue. The
test statistic for the null hypothesis $E\left[u_{h, t+h}^{2}\right]=E\left[v_{h, t+h}^{2}\right]$ has a chi-squared limit distribution,

$$
\frac{P(\alpha \bar{f})^{2}}{\alpha^{\prime} \hat{\Omega} \alpha} \stackrel{A}{\sim} \chi^{2}(1)
$$

where $\alpha=(1,0, \cdots, 0,-1,0, \cdots 1)$ has -1 at the $(h+1)$ th element.

## Data description

Table A.9: Category (A) Income, Output, Sales, Capacity Utilization

| Series | Trans. | Sample period | Description |
| :--- | :--- | :--- | :--- |
| IPNB50030.M | $\Delta \mathrm{Ln}$ | $1939: 1-2016: 7$ | Industrial Production Index - Final <br> Products \& Nonindust Supplies <br> (2012=100, NSA) |
| IPNB50002.M | $\Delta \mathrm{Ln}$ | $1939: 1-2016: 7$ | Industrial Production Index - Final <br> Products (2012=100, NSA) |
| IPNB51000.M | $\Delta \mathrm{Ln}$ | $1939: 1-2016: 7$ | Industrial Production Index - Consumer <br> goods (2012=100, NSA) <br> Industrial Production Index - Durable <br> consumer goods (2012=100, NSA) |
| IPNB51100.M | $\Delta \mathrm{Ln}$ | $1947: 1-2016: 7$ | Industrial Production Index - |
| IPNB51200.M | $\Delta \mathrm{Ln}$ | $1947: 1-2016: 7$ | Nondurable consumer goods <br> (2012=100, NSA) |
| IPNB52100.M | $\Delta \mathrm{Ln}$ | $1947: 1-2016: 7$ | Industrial Production Index - Business <br> equipment (2012=100, NSA) |
| IPNB53000.M | $\Delta \mathrm{Ln}$ | $1939: 1-2016: 7$ | Industrial Production Index - Materials <br> (2012=100, NSA) |
| IPNB53100.M | $\Delta \mathrm{Ln}$ | $1947: 1-2016: 7$ | Industrial Production Index - Durable <br> goods materials (2012=100, NSA) |
| IPNB53200.M | $\Delta \mathrm{Ln}$ | $1954: 1-2016: 7$ | Industrial Production Index - <br> Nondurable goods materials <br> (2012=100, NSA) |
| IPSB00004.M | $\Delta \mathrm{Ln}$ | $1919: 1-2016: 7$ | Industrial Production Index - <br> Manufacturing (SIC) (2012=100, SA) |
| IPSGMFD.M | $\Delta \mathrm{Ln}$ | $1972: 1-2016: 7$ | Industrial Production Index - Durable <br> manufacturing (NAICS) (2012=100, |
| IPSGMFN.M | $\Delta \mathrm{Ln}$ | $1972: 1-2016: 7$ | SA) <br> Industrial Production Index - <br> Nondurable manufacturing (NAICS) <br> (2012=100, SA) |

Table A.9: Category (A) Income, Output, Sales, Capacity Utilization; continued from previous page.

| Series | Trans. | Sample period | Description |
| :---: | :---: | :---: | :---: |
| IPSG21.M | $\Delta \mathrm{Ln}$ | 1972:1-2016:7 | Industrial Production Index - Mining (2012=100, SA) |
| IPSG2211A2.M | $\Delta \mathrm{Ln}$ | 1972:1-2016:7 | Industrial Production Index - Electricity and gas utilities (2012=100, SA) |
| UTLB50001.M | Lev | 1967:1-2016:7 | $\begin{aligned} & \text { Capacity Utilization - Total Index (\%, } \\ & \text { SA) } \end{aligned}$ |
| UTLB00004.M | Lev | 1948:1-2016:7 | Capacity Utilization - Manufacturing (SIC) $(\%$, SA) |
| UTLGMFD.M | Lev | 1967:1-2016:7 | Capacity Utilization - Durable manufacturing (NAICS) (\%, SA) |
| UTLGMFN.M | Lev | 1967:1-2016:7 | Capacity Utilization - Nondurable manufacturing (NAICS) (\%, SA) |
| UTLG21.M | Lev | 1967:1-2016:7 | Capacity Utilization - Mining (\%, SA) |
| UTLG335.M | Lev | 1972:1-2016:7 | Capacity Utilization - Electrical equip., Appliance \& component ( $\%$, SA) |
| YP.M | $\Delta \mathrm{Ln}$ | 1959:1-2016:6 | Nominal Personal Income (Bil \$, SAAR |
| A0M051.M | $\Delta \mathrm{Ln}$ | 1959:1-2016:6 | Personal Income less Transfer <br> Payments (Bil/Chained 2009 \$, SAAR |
| CXFAE.M | $\Delta \mathrm{Ln}$ | 1959:1-2016:6 | Personal Consumption Expenditures excluding Food and Energy (Bil \$, SAAR) |
| CD.M | $\Delta \mathrm{Ln}$ | 1959:1-2016:6 | Personal Consumption Expenditures Durable goods (Bil \$, SAAR) |
| CN.M | $\Delta \mathrm{Ln}$ | 1959:1-2016:6 | Personal Consumption Expenditures - <br> Nondurable goods (Bil \$, SAAR) |
| CSV.M | $\Delta \mathrm{Ln}$ | 1959:1-2016:6 | Personal Consumption Expenditures - <br> Services (Bil \$, SAAR) |

Table A.10: Category (B) Employment and Unemployment

| Series | Trans. | Sample period | Description |
| :--- | :--- | :--- | :--- |
| EPNEGP.M | $\Delta \mathrm{Ln}$ | $1947: 1-2016: 7$ | Employment of prod and nonsup on <br> nonfarm payrolls - Goods (thous., SA) |
| EPNET.M | $\Delta \mathrm{Ln}$ | $1964: 1-2016: 7$ | Employment of prod and nonsup on <br> nonfarm payrolls - Total, private <br> (thous., SA) |
| EPNECON.M | $\Delta \mathrm{Ln}$ | $1947: 1-2016: 7$ | Employment of prod and nonsup on <br> nonfarm payrolls - Construction <br> (thous., SA) |
| EPNEPSP.M | $\Delta \mathrm{Ln}$ | $1964: 1-2016: 7$ | Employment of prod and nonsup on <br> nonfarm payrolls - Private service pro- <br> viding (thous., SA) |
| EPNETTU.M | $\Delta \mathrm{Ln}$ | $1964: 1-2016: 7$ | Employment of prod and nonsup on <br> nonfarm payrolls - Trade, trans. and <br> utilities (thous., SA) |
| EPNEWST.M | $\Delta \mathrm{Ln}$ | $1972: 1-2016: 7$ | Employment of prod and nonsup on <br> nonfarm payrolls - Wholesale trade <br> (thous., SA) |
| EPNERET.M | $\Delta \mathrm{Ln}$ | $1972: 1-2016: 7$ | Employment of prod and nonsup on <br> nonfarm payrolls - Retail trade (thous., |
| EPNEFIN.M | $\Delta \mathrm{Ln}$ | $1964: 1-2016: 7$ | SA) <br> Employment of prod and nonsup on <br> nonfarm payrolls - Financial activities <br> (thous., SA) |
| EPNEOTS.M | $\Delta \mathrm{Ln}$ | $1964: 1-2016: 7$ | Employment of prod and nonsup on <br> nonfarm payrolls - Other services <br> (thous., SA) |
| EG.M | $\Delta \mathrm{Ln}$ | $1939: 1-2016: 7$ | Employment all employees - Govern- <br> ment (thous., SA) |

Table A.10: Category (B) Employment and Unemployment; continued from previous page.

| Series | Trans. | Sample period | Description |
| :---: | :---: | :---: | :---: |
| EPNEND.M | $\Delta \mathrm{Ln}$ | 1939:1-2016:7 | Employment of prod and nonsup on nonfarm payrolls - Nondurable goods (thous., SA) |
| EPNEMF.M | $\Delta \mathrm{Ln}$ | 1939:1-2016:7 | Employment of prod and nonsup on nonfarm payrolls - Manufacturing (thous., SA) |
| EPNED.M | $\Delta \mathrm{Ln}$ | 1939:1-2016:7 | Employment of prod and nonsup on nonfarm payrolls - Durable goods (thous., SA) |
| EPNEML.M | $\Delta \mathrm{Ln}$ | 1947:1-2016:7 | Employment of prod and nonsup on nonfarm payrolls - Mining and logging (thous., SA) |
| U15@26WZ.M | Lev | 1948:1-2016:7 | Unemployment level of duration: 15-26 wks (thous, SA) |
| U15\&WZ.M | Lev | 1948:1-2016:7 | Unemployment level of duration: 15 wks+ (thous, SA) |
| U@5WZ.M | Lev | 1948:1-2016:7 | Unemployment level of duration: less than 5 wks (thous, SA) |
| U5@14WZ.M | Lev | 1948:1-2016:7 | Unemployment level of duration: 5-14 wks (thous, SA) |
| WEEKSU.M | Lev | 1948:1-2016:7 | Unemployed average duration in weeks (weeks, SA) |
| RDUTTTOD5U.M | Lev | 1948:1-2016:7 | Unemployment rate: 15 wks and over (\%, NSA) |
| HOPMF.M | Lev | 1956:1-2016:7 | Avg. weekly overtime hrs. of prod. and nonsup. employees: mfg., overtime (SA) |
| HPEAP.M | Dif | 1964:1-2016:7 | Avg. weekly overtime hrs. of prod. and nonsup. employees: Total private (SA) |
| HPMF.M | Lev | 1939:1-2016:7 | Avg. weekly overtime hrs. of prod. and nonsup. employees: manufacturing (SA) |
| LCZ.M | $\Delta \mathrm{Ln}$ | 1948:1-2016:7 | Civilian labor force: 16 years and over (thous., SA) |

Table A.11: Category (C) Construction, Inventories and Orders

| Series | Trans. | Sample period | Description |
| :--- | :--- | :--- | :--- |
| HUSTSSO.M | Ln | $1959: 1-2016: 7$ | Housing starts: South (mil.u., SAAR) |
| HUSTSWT.M | Ln | $1959: 1-2016: 7$ | Housing starts: West (mil.u., SAAR) |
| HUSTSNC.M | Ln | $1959: 1-2016: 7$ | Housing starts: Midwest (mil.u., <br> SAAR) |
| HUSTSNE.M | Ln | $1959: 1-2016: 7$ | Housing starts: Northeast (mil.u., <br> SAAR) |
| UHS.M | Ln | $1947: 1-2016: 7$ | Housing starts: Total (thous.u., SAAR) |
| HU1NOFFERZ.M | Ln | $1963: 1-2016: 6$ | New 1-family houses for sale at end of <br> month (thous.u., SA) |
| HU1NSOLDZ.M | Ln | $1963: 1-2016: 6$ | New 1-family houses for sale during <br> month (thous.u., SA) |
| HUATZNSTNS.M | Ln | $1968: 1-2016: 7$ | Housing authorized but not started at <br> end of period (thous.u., NSA) |
| SHPMH.M | Ln | $1959: 1-2016: 6$ | Mobile homes: manufactures' <br> shipments (thous.u., SAAR) |
| RISMAT.M | $\Delta$ Ln | $1959: 1-2016: 7$ | Ratio of mfg. and trade: inventory or <br> sales (SAAR NIPA) |
| PMI.M | Lev | $1948: 1-2016: 7$ | Purchasing managers' index (SA) |
| PMP.M | Lev | $1948: 1-2016: 7$ | Napm production index (\%) |
| JDIFFO@NAPMZ.M | LLn | $1948: 1-2016: 7$ | New orders index (\%, SA) |

Table A.12: Category (D) Interest Rate and Asset Prices

| Series | Trans. | Sample period | Description |
| :---: | :---: | :---: | :---: |
| RMGFCM@10NS.M | Diff | 1953:4-2016:7 | Interest rate: U.S. treasury const maturities, $10-\mathrm{yr}$ (\% per ann., NSA) |
| RMFEDFUNDNS.M | Diff | 1954:7-2016:7 | Interest rate: federal funds (effective) (\% per ann., NSA) |
| RMGFCM@1NS.M | Diff | 1953:4-2016:7 | Interest rate: U.S. treasury const maturities, 1-yr (\% per ann., NSA) |
| RMGFCM@5NS.M | Diff | 1953:4-2016:7 | Interest rate: U.S. treasury const maturities, 5 -yr (\% per ann., NSA) |
| ALCIBL00.M | $\Delta \mathrm{Ln}$ | 1959:1-2016:7 | Commerical and industiral loans outstanding in 2009 dollars (bci) |
| JNYSE02Z.M | $\Delta \mathrm{Ln}$ | 1966:1-2016:7 | Nyse common stock price index: composite $(12 / 31 / 02=5000)$ |
| JS\&PNS.M | $\Delta \mathrm{Ln}$ | 1901:1-2016:7 | S\&P's common stock price index: composite (1941-43=10) |
| JS\&PINDNS.M | $\Delta \mathrm{Ln}$ | 1901:1-2016:7 | S\&P's common stock price index: industrial (1941-43=10) |
| MNY2@00.M | $\Delta \mathrm{Ln}$ | 1959:1-2016:6 | Money supply, m2 in 2009 dollars (bci) |
| RMMTGNS.M | $\Delta \mathrm{Ln}$ | 1963:1-2016:6 | Effective interest rate: conventional home mtge loans (\%) |
| RXC146\%USNS.M | $\Delta \mathrm{Ln}$ | 1967:11-2016:7 | Foreign exchange rate: <br> Switzerland (swiss franc per U.S. \$) |
| RXC158\%USNS.M | $\Delta \mathrm{Ln}$ | 1967:11-2016:7 | Foreign exchange rate: Japan (yen per U.S. \$) |
| RXC156\%USNS.M | $\Delta \mathrm{Ln}$ | 1956:1-2016:7 | Foreign exchange rate: <br> Canada (canadian \$ per U.S. \$) |

## Bibliography

R. Alquist, L. Kilian, and R. J. Vigfusson. Forecasting the price of oil. In G. Elliott and A. Timmermann, editors, Handbook of Economic Forecasting, volume 2A, chapter 8. Elsevier, 2013.
T. Amemiya. Advanced Econometrics. Harvard University Press, 1985.
D. W. K. Andrews. Laws of large numbers for dependent nonidentically distributed random variables. Econometric Theory, 4:458-466, 1988.
D. W. K. Andrews. Tests for parameter instability and structural change with unknown change point. Econometrica, 61:821-856, 1993.
D. W. K. Andrews and W. Ploberger. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica, 62(6):1383-1414, 1994.
J. Bai. Least squares estimation of a shift in linear processes. Journal of Time Series Analysis, 15 (5):453-472, 1994.
J. Bai. Estimation of a change point in multiple regression models. The Review of Economics and Statistics, 79:551-563, 1997.
J. Bai and P. Perron. Estimating and testing linear models with multiple structural changes. Econometrica, 66:47-78, 1998.
J. Bai and P. Perron. Computation and analysis of multiple structural changes. Journal of Applied Econometrics, 18:1-22, 2003.
J. Bai, R. L. Lumsdaine, and J.H. Stock. Testing for and dating common breaks in multivariate time series. Review of Economic Studies, 65:395-432, 1998.
R. J. Bhansali. Approximating the distribution of the ml estimate of the change-point in a sequence of independent r.v.'s. Annals of the Institute of Statistical Mathematics, 48(3):577-602, 1996.
P. K. Bhattacharya. Maximum likelihood estimation of a change-point in the distribution of independent random variables. Journal of Multivariate Analysis, 23:183-208, 1987.
Z. W. Birnbaum and Marshall A. W. Some multivariate chebyshev inequalitites with extensions to continuous parameter processes. Annals of Mathematical Statistics, 32:687-703, 1961.
R. L. Brown, J. Durbin, and J. M. Evans. Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Society: Series B, 37(2):149-163, 1978.
P. Bühlmann. Sieve bootstrap for time series. Bernoulli, 3:123-148, 1997.
D. Card, A. Mas, and J. Rothstein. Tipping and the dynamics of segregation. The Quarterly Journal of Economics, 123(1):177-218, 2008.
A. Casini and P. Perron. Continuous record laplace-based inference about the break date in structural models. Unpublished manuscript, Department of Economics, Boston University, 2017.
P. Chang and S. Sakata. Estimation of impulse response functions using long autoregression. The Econometrics Journal, 10(2):453-469, 2007.
T. T. L. Chong. Structural change in AR(1) models. Econometric Theory, 17(1):87-155, 2001.
G. C. Chow. Tests of equality between sets of coefficients in two linear regressions. Econometrica, 28:591-605, 1960.
D. Cox. Prediction by exponentially weighted moving averages and related methods. Journal of Royal Statistical Society, Series B, 23(2):412-422, 1961.
E. Dvir and K. S. Rogoff. Three epochs of oil. In NBER Working Paper Series, number 14927, Cambridge, MA, 2009. National Bureau of Economic Research.
G. Elliott and U. K. Müller. Efficient tests for general persistent time variation in regression coefficients. Review of Economic Studies, 73:907-940, 2006.
G. Elliott and U. K. Müller. Confidence sets for the date of a single break in linear time series regression. Journal of Econometrics, 141:1196-1218, 2007.
Y. Eo and J. Morley. Likelihood-ratio-based confidence sets for the timing of structural breaks. Quantitative Economics, 6:463-497, 2015.
R. González-Val and M. Marcén. Unilateral divorce versus child custody and child support in the u.s. Journal of Economic Behavior \& Organization, 81:613-643, 2012.
R. Hájek and A. Rényi. Generalization of an inequality of kolmogorov. Acta Mathematica Hungarica, 6:281-283, 1955.
J. D. Hamilton. What is an oil shock? Journal of Econometrics, 113:363-398, 2003.
B. E. Hansen. Strong laws for dependent heterogeneous processes. Econometric Theory, 7(2): 213-221, 1991.
B. E. Hansen. The new econometrics of structural change: Dating breaks in u.s. labor productivity. Journal of Economic Perspective, 15(4):117-128, 2001.
D. Hinkley. Inference about the change point in a sequence of random variables. Biometrika, 57: 1-17, 1970.
C. Ing. Multistep prediction in autoregressive processes. Econometric Theory, 19(2):254-279, 2003.
H. Jeffreys. An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 186(1007): 453-461, 1946.
L. Jiang, X. Wang, and S. Yu. In-fill asymptotic theory for structural break point in autoregression: A unified theory. Singapore Management University, School of Economics, 2017.
L. Jiang, X. Wang, and S. Yu. New distribution theory for the estimation of structural break. Journal of Econometrics, 205:156-176, 2018.

Ò. Jordà. Estimation and inference of impulse responses by local projections. American Economic Review, 95(1):161-182, 2005.
L. Kilian. Exogenous oil supply shocks: How big are they and how much do they matter for the u.s. economy? Review of Economics and Statistics, 90(2):216-240, 2008.
C. J. Kim and C. R. Nelson. Has the u.s. economy become more stable? a bayesian approach based on a markov-switching model of the business cycle. Review of Economics and Statistics, 81(4):608-616, 1999.
E. Kurozumi and P. Tuvaandorj. Model selection criteria in multivariate models with multiple structural changes. Journal of Econometrics, 164(2):218-238, 2011.
R. Y. Liu. Bootstrap procedures under some non-i.i.d. models. Annals of Statistics, 16:1696-1708, 1988.
M. Marcellino, J. Stock, and M. Watson. A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. Journal of Econometrics, 136:499-526, 2006.
M.M. McConnell and G. Perez-Quiros. Output fluctuations in the united states: What has changed since the early 1980's? American Economic Review, 90(5):1464-1476, 2000.
D. L. McLeish. A maximal inequality and dependent strong laws. The Annals of Probability, 5: 829-839, 1975.
K. A. Mork. Oil and the macroeconomy when prices go up and down. Journal of Political Economy, 91:740-744, 1989.
W. Newey and K. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55(3):703-708, 1987.
W. K. Newey and K. D. West. Automatic lag selection in covariance matrix estimation. Review of Economic Studies, 61(4):631-653, 1994.
J. Nyblom. Testing for the constancy of parameters over time. Journal of the American Statistical Association, 84:223-230, 1989.
M. T. Owyang, V. A. Ramey, and S. Zubairy. Are government spending multipliers greater during periods of slack? evidence from twentieth-century historical data. American Economic Review, 103(3):129-134, 2013.
B. S. Paye and A. Timmermann. Instability of return prediction models. Journal of Empirical Finance, 13:274-315, 2006.
P. Perron. The great crash, the oil-price shock, and the unit-root hypothesis. Econometrica, 57(6): 1361-1401, 1989.
W. Ploberger and W. Krämer. The CUSUM test with OLS residuals. Econometrica, 60:271-285, 1992.
W. Ploberger, W. Krämer, and K. Kontrus. A new test for structural stability in the linear regression model. Journal of Econometrics, 40(2):307-318, 1989.
Z. Qu and P. Perron. Estimating and testing structural changes in multivariate regressions. Econometrica, 75(2):459-502, 2007.
R. Quandt. The estimations of the parameters of a linear regression system obeying two separate regimes. Journal of the American Statistical Association, 53:873-880, 1958.
R. Quandt. Tests of the hypothesis that a linear regression system obeys two separate regimes. Journal of the American Statistical Association, 55:324-330, 1960.
R. J. Shiller. A distributed lag estimator derived from smoothness priors. Econometrica, 41(4): 775-788, 1973.
J. H. Stock. Unit roots, structural breaks and trends. In R. F. Engle and D. L. McFadden, editors, Handbook of Econometrics, volume 4, chapter 46. Elsevier, 1994.
J. H. Stock and M. W. Watson. Has the business cycle changed, and why? NBER Macroeconomics Annual, 17:159-218, 2002.
A. Weiss. Multi-step estimation and forecasting in dynamic models. Journal of Econometrics, 48 (1-2):135-149, 1991.
K. D. West. Asymptotic inference about predictive ability. Econometrica, 64(5):1067-1084, 1996.
P. A. Wolak. Testing inequality constraints in linear econometric models. Journal of Econometrics, 41:205-235, 1989.
Y. C. Yao. Approximating the distribution of the ml estimate of the change-point in a sequence of independent r.v.'s. The Annals of Statistics, 3:1321-1328, 1987.


[^0]:    ${ }^{1}$ Also known as the maximum a-posteriori probability (MAP) estimator.
    ${ }^{2}$ In-fill/continuous record asymptotics derives the limit distribution by assuming the time span is fixed with a shrinking sampling interval under a continuous time approximation model.

[^1]:    ${ }^{3}$ The prior distribution (1.6) is equivalent to a Beta distribution with shape parameters $(3 / 2,3 / 2)$.
    ${ }^{4}$ Consistency has been proved by Yao (1987).

[^2]:    ${ }^{5}$ This is equivalent to Jeffreys (1946) prior of $\delta$ conditional on $\rho$. Jeffreys rule is to use the square root of the determinant of the Fisher information matrix as a uniform prior. This does not correspond to our motivation of using the Fisher information as an informative prior.

[^3]:    ${ }^{6}$ Ornstein-Uhlenbeck process is a continuous time analogue of a discrete time AR(1) model

[^4]:    ${ }^{7}$ If weight function is $\omega_{k}=(k / T(1-k / T))^{\gamma}$, Assumption 2 is satisfied if $0 \leq \gamma \leq 1 / 2$ for an arbitrary small $\alpha$ in Assumption 1(i). For $\gamma \in\{1 / 8,1 / 4,3 / 8\}$ the results (omitted due to space constraints) do not change qualitatively; the probability at boundaries decrease compared to the finite sample distribution of LS. Because $\omega(\rho)=(\rho(1-\rho))^{\gamma} \rightarrow 1$ as $\gamma \rightarrow 0$, the difference between the two estimators finite sample behavior shrinks when $\gamma$ is close to zero.

[^5]:    ${ }^{8}$ For consistency of the break point estimator in a $\operatorname{AR}(1)$ model, we use $\Omega_{k}=\omega_{k}^{2} I_{q}$ where $\omega_{k}$ is a function of $k / T$ only.

[^6]:    ${ }^{9}$ For sample period 1952:7-2003:12 the mean excess return of the FTSE All-share index is 0.5949 and standard deviation is 5.4890. At $t=1975: 1$ the excess return $\operatorname{Ret}_{t}=0.4556$ and at $t=1975: 2$ we have $\operatorname{Ret}_{t}=53.2187$, so the

[^7]:    ${ }^{10}$ Data were downloaded from http://econ.ucsd.edu/~jhamilton

[^8]:    ${ }^{11}$ Sub-samples with start dates 1957Q3 to 1961Q2 has matrix $\left(X^{\prime} X\right)$ that are near singular for nonlinear oil price measures.

[^9]:    ${ }^{12}$ Data were downnloaded from https://sites.google.com/site/lkilian2019/research/data-sets
    ${ }^{13}$ Kilian (2008) does not estimate for structural breaks but considers OLS regression of real GDP growth on a constant, four lags of the dependent variable and eight lags of the exogenous oil shock measure to see the effect of oil shock to real GDP. The break date estimates using eight lags of exogenous oil shock measure are 1981Q2 for both methods, similar to the estimate using four lags.

[^10]:    ${ }^{1}$ For critical values of $d b_{T}^{b}$ in Table 2.1 we assumed weights $w_{t}=T^{-1} \exp \left[\frac{1}{2} \bar{d}^{\prime} \bar{d}\left(\frac{t}{T}\right)\left(1-\frac{t}{T}\right)-\frac{1}{2} \bar{d}^{\prime} \bar{d}(0.1)(0.9)\right]$ so that critical values are not too large at $1 \%$ significance level. It essentially does not change the test statistic because the additional term of weights is a constant.

[^11]:    ${ }^{2}$ The mean of annualized growth rate from 1947 Q 1 to 1984 Q 1 is 3.5 percent and the standard deviation is 4.7 percent. From 1984Q1 to 2017Q2 the mean is 2.6 percent and the standard deviation is 2.4 percent.

[^12]:    ${ }^{1}$ Data are obtained from the EIA website: https://www.eia.gov/petroleum/data.php. The West Texas Intermediate (WTI) monthly price for crude oil are calculated by EIA from daily data by taking an unweighted average of the daily closing spot prices for a given product over the specified time period. For the U.S. refiners' acquisition cost (RAC) of crude oil, the U.S. is defined as the 50 states, the District of Columbia, Puerto Rico, the Virgin Islands, and all American territories and possessions. Values reflect the PAD District in which the crude oil is intended to be refined.
    ${ }^{2}$ There is evidence of a major structural change in the distribution of pre- and post-1973 real oil prices; Alquist, Kilian, and Vigfusson (2013), Dvir and Rogoff (2009). Hence we use data the of post-1973 real oil price only.
    ${ }^{3}$ Smoothed direct forecasts are jointly estimated for horizon 1 to 60 months, hence the validation sample is trimmed by 60 months from the end of the data set.
    ${ }^{4} \mathrm{AR}(p)$ model is estimated using in-sample data (1986.1-2001.12 for WTI price and 1974.1-2001.12 for RAC price) and then 500 replication of out-of-sample pseudodata are generated.

[^13]:    ${ }^{5}$ Because the inverse of the variance estimates have extremely small values relative to the penalty term, the matrix $\mathbf{M}^{*}$ from (3.7) is close to singular.

