Essays on Nonparametric and Dynamic Time-Seies Econometrics

Shih-Tang Hwu

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

Doctor of Philosophy

University of Washington

2018

Reading Committee:
Yanqin Fan, Co-Chair
Chang-Jin Kim, Co-Chair
Jing Tao

Program Authorized to Offer Degree:

Department of Economics

University of Washington

Abstract

Essays on Nonparametric and Dynamic Time-Seies Econometrics

Shih-Tang Hwu

Co-Chairs of the Supervisor Committee:

Yanqin Fan

Chang-Jin Kim

Department of Economics

This dissertation explores important macroeconomics issues based on both classical and Bayesian Econometrics tools developed. One goal of the first chapter of the dissertation is to develop identification conditions and algorithm for estimating Markov-switching models without imposing distribution assumptions. Since the seminal work of Hamilton (1989), the basic Markov-switching model has been extended in various ways. Without a single exception, estimation of the aforementioned models and the other Markov-switching models in the literature has relied upon parametric assumptions on the distribution of the error terms. Most applications of Markov-switching models in the literature assume normally distributed error terms, with rare exceptions like Dueker (1997) who proposes a model of stock returns in which the innovation comes from a Student-t distribution. The question then would be: what if a normal log-likelihood is maximized but the normality assumption is violated? Based on simulation studies, we find that maximum likelihood estimation could lead to sizable bias in the parameter estimates and poor inferences about regime probabilities

when the normality assumption is violated, even for a sample size as large as 5,000. We approximate the unknown distribution of the error term by the Dirichlet process mixture of normals, in which the number of mixtures is treated as a parameter to estimate. In doing so, we pay a special attention to identification of the model. We apply the proposed model to the growth of postwar U.S. industrial production index in order to investigate its regime-switching dynamics. Our univariate model can effectively control for the irregular components that is not related to business conditions. This leads to sharp and accurate inferences on recession probabilities just like the dynamic factor models of Kim and Yoo (1995), Chauvet (1998), and Kim and Nelson (1998) do.

The second chapter of the dissertation investigates the relationships between innovations to trend inflation and inflation-gap in a univariate unobserved components model with with Markov-switching volatility. Building on the work of Stock and Watson (2007), we empirically shows that a negative correlation between innovations to trend inflation and the inflation gap, when it is combined with time-varying inflation gap persistence, plays an important role in the dynamics of postwar US inflation. A negative correlation between trend inflation and the markup shock may be an important source of their negative correlation. Like the time-varying VAR models of Cogley and Sbordone (2008) and Ascari and Sbordone (2014), our model results in smooth trend inflation, from which inflation persistently deviates during the Great inflation period. Furthermore, our model provides superior out-of-sample forecasts than Stock and Watson's (2007) unobserved components model with stochastic volatility or than Atkeson and Ohanian's (2001) random walk model does.

One goal of the last chapter of the dissertation is to develop estimation methods in linear regression model with endogenous variables but only weak instrument variables. The proposed methods exploit the time-varying volatility of the endogenous variables. We show that the proposed estimators are consistent and asymptotically normally distributed. We also show that the proposed methods have much better power compare with the existing weak instrument robust test through simulations. Another goal of the last chapter is to investigate the magnitude of elasticity of intertemporal substitution (EIS), which is one of the most important parameters in applied macroeconomics and finance. Yogo (2004) applies the existing weak instrument robust test to estimate EIS and find 22 out of 33 confidence interval to be $(-\infty, \infty)$, which is very uninformative. We apply proposed approach to estimate the EIS using the data employed by Yogo (2004). Confidence intervals based on proposed methods are much tighter than those constructed by weak instrument robust tests and its value is generally close to 0.

Table of Contents

List of Tables

List of Figures

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my advisors: Professor Yanqin Fan, and Chang-Jin Kim. I owe my deepest gratitude to them for their generous guidance and endless support. This dissertation would never have been accomplished without their invaluable advice.

My parents, Ko-Chung Ho and Zong-Bin Hwu, always uphold and inspire me. Without them, my pursuit of this advanced degree would never have been started. Their constant love have sustained me throughout my life. My wife, Sara Sung, constantly support and encourage me. She is always with me when I need her the most. This dissertation is lovingly dedicated to my parents, my wife, and my son, Joseph.

Chapter 1

Markov-switching models with Unknown Error Distributions [1]

1.1 Introduction

Since the seminal work of Hamilton (1989), the basic Markov-switching model has been extended in various ways. For example, Diebold et al. (1994) and Filardo (1994) extend the model to allow the transition probabilities governing the Markov process to be a function of exogenous or predetermined variables. Kim (1994) extends the model to the state-space representation of general dynamic linear models, which includes autoregressive moving average processes, unobserved components models, dynamic factor models, etc. Chib (1998) introduces a structural break model with multiple change-points by constraining the transition probabilities of the Markov-switching model so that the state variable can either stay at the current value or jump to the next higher value. More recently, Fox et al. (2011), Song (2014), and Bauwens et al. (2017) introduce infinite hidden Markov models and generalize the finite-state Markov switching model of Hamilton (1989) to an infinite number of states. Their models integrate the regime switching and structural break dynamics in a unified Bayesian framework. For these models, the number of states is possibly infinite and is determined when estimating the model.

Without a single exception, estimation of the aforementioned models and the other Markov-switching models in the literature has relied upon parametric assumptions on the distribution of the error terms. Most applications of Markov-switching models in the literature assume normally distributed error terms, with rare exceptions like Dueker (1997) who proposes a model of stock returns in which the innovation comes from a Student-t distribution. The question then would be: what if a normal log-likelihood is maximized but the

[1] This chapter is based on a joint work with Chang-Jin Kim.

normality assumption is violated? We performed a simulation study in order to investigate the finite sample performance of the maximum likelihood estimation of Markov-switching models when a normal log-likelihood is maximized but the normality assumption is violated. We find that maximum likelihood estimation could lead to sizable bias in the parameter estimates and poor inferences about regime probabilities when the normality assumption is violated, even for a sample size as large as 5,000. [2]

In Bayesian semi-parametric econometrics, approximating an unknown distribution based on a mixture of normals is popular as surveyed in Marin et al. (2005). There are two alternative models for achieving the goal. They are: i) the finite mixture normals model in which the number of states is fixed, and ii) the Dirichlet process mixture normals model in which the number of states is treated as a random variable. Kim et al. (1998) and Omori et al. (2007) demonstrate the usefulness of the finite mixture of normals in approximating the log chi-square distribution in stochastic volatility models; and Alexander and Lazar (2006) employ it to approximate the unknown error distribution in a GARCH model. More recently, Jensen and Maheu (2013) apply the Dirichlet process mixture of normals to a multivariate GARCH model; Jensen and Maheu (2010, 2014) apply it to deal with unknown error distributions in stochastic volatility models; and Jin and Maheu (2016) apply it for Bayesian semi-parametric modeling of realized covariance matrices.

The goal of this paper is to present a Bayesian approach to estimating Markov-switching models without imposing a priori parametric assumption on the distribution of the error term. We implement the Dirichlet process mixture normals model in order to approximate the unknown and potentially non-normal error distribution. We note that, in order to allow for an asymmetric, as well as fat-tailed, error distribution within a Markov-switching model, two identification issues need to be addressed.

The first identification issue is associated with the Markov-switching intercept and heteroskedasticity. Specifically, in a model with Markov-switching in intercept and heteroskedasticity, if we employ mixture of normals with unknown mean and variance to approximate the unknown error distributions, we will need to normalize the unconditional mean and variance of the mixture normals to be 0 and 1, respectively, in order to identify the Markov-switching

[2] We deal with this issue in Section 1.2.

intercept and heteroskedasticity. Although implementing this normalization restriction is relatively straightforward under finite number of mixture of normal, it is not the case under Dirichlet process mixture of normal. We propose a simple and efficient way to overcome this problem. The second identification issue is associated with distinguishing the Markov-switching and mixture of normals parts. In particular, our proposed model is a model combined Markov switching mixture with iid mixture of normal. [3] On the one hand, since the iid mixture is a Markov-switching mixture without persistence in the state variable, when there is no additional restriction on the dynamics of the Markov-switching mixture, we may identify some of the mixture of normals to be a Markov-switching mixture. On the other hand, when the persistence of Markov-switching mixture is close to iid mixture, one may identify some of the Markov-switching mixture as mixture of normals. To avoid this identification problem, we explicitly derive the persistent conditions of Markov-switching model and propose a straightforward way to impose the conditions by employing proper priors. See Section 1.3 for detailed discussion on these two identification issues.

We apply the proposed model to the growth of postwar U.S. Industrial Production index covering the period January 1947-January 2017. We demonstrate that a model with a normality assumption performs poorly in identifying the NBER reference cycles. The null hypothesis of normality for the error term is rejected at a 5% significance level. However, the proposed univariate model can effectively control for the irregular components that are not related to business conditions. This leads to sharp and accurate inferences on recession probabilities just like the dynamic factor models of Kim and Yoo (1995), Chauvet (1998), and Kim and Nelson (1998) do. Furthermore, the null of normality is not rejected for the standardized error term that is obtained conditional on the mixing indicator variable.

The rest of this paper is organized as follows. In Section 1.2, we motivate our paper by exploring the finite sample properties of the quasi-maximum likelihood estimation of Markov-switching models. We discuss our model specifications with special attention to identification issues in Section 1.3. In Section 1.4, we present a Markov Chain Monte Carlo (MCMC) algorithm for estimating the proposed model. We present two model comparison

[3] See Fruhwirth-Schnatter (2006) for detailed discussion on Markov-switching and iid mixture of normal models.

methods in Section 1.5. Section 1.6 provides an empirical application of the proposed model, and Section 1.7 concludes the paper.

1.2. Pitfalls in Assuming Normality in Markov-Switching Model: Finite Sample Performance Based on Simulation Study

In this section, we investigate finite sample performance of the maximum likelihood estimation of Markov-switching models when a normal log-likelihood is maximized but the normality assumption is violated. For this purpose, we consider the following model with Markov-switching mean and variance:

$$\begin{aligned} y_t = \beta_{S_t} + h_{S_t}\varepsilon_t, \quad \varepsilon_t \sim i.i.d(0,1), \quad S_t = 1,2, \\ t = 1,2,...,T, \end{aligned} \tag{1.1}$$

where S_t is a 2-state Markov-switching process with transition probabilities

$$Pr[S_t = 1|S_{t-1} = 1] = p_{11}, \quad Pr[S_t = 2|S_{t-1} = 2] = p_{22}. \tag{1.2}$$

We consider the following four alternative distributions for the error term ε_t, two of which are symmetric and the other two are asymmetric:

Case #1

$$\varepsilon_t \sim i.i.d. \;\; N(0,1)$$

Case #2

$$\varepsilon_t \sim \frac{u_t}{\sqrt{\nu/(\nu-2)}}, \quad u_t \sim i.i.d. \;\; t-distribution \;\; with \;\; d.f. = \nu$$

Case #3

$$\varepsilon_t = \frac{ln(u_t^2) - E(ln(u_t^2))}{\sqrt{(var(ln(u_t^2))}}, \quad u_t \sim i.i.d. \;\; N(0,1),$$

where $E(\ln v_t^2) = -1.2704$, $var(\ln v_t^2) = \pi^2/2$.

Case #4

$$\varepsilon_t \ |D_t \sim i.i.d. \ \ N(\mu_{D_t}, \sigma^2_{D_t}), \ D_t = 1, 2, 3,$$
$$Pr[D_t = 1] = w_1, \ Pr[D_t = 2] = w_1, \ Pr[D_t = 1] = w_3$$

For each of the above four cases, we generate 1,000 sets of data. For each data set generated, we estimate the model in equations (1.1) and (1.2) by maximizing a normal log-likelihood. While the normality assumption is satisfied for Case #1, it is violated for the other three cases. We consider two alternative sample sizes: $T = 500$ and $T = 5000$. The parameter values we assign are given below:

$$\beta_1 = -0.5, \quad \beta_2 = 1; \quad h_1 = 2, \quad h_2 = 1; \quad p_{11} = 0.9, \quad p_{22} = 0.95;$$
$$\nu = 5;$$
$$\mu_1 = 0.72, \quad \mu_2 = 0, \quad \mu_3 = -1.8; \quad \sigma_1^2 = 0.025, \quad \sigma_2^2 = 0.2, \quad \sigma_3^2 = 0.1;$$
$$w_1 = 0.5, \quad w_2 = 0.3, \quad w_3 = 0.2$$

For each of the above four cases and for each parameter, Table 1.1 reports the mean of 1,000 point estimates, as well as the root mean squared error (RMSE) of the estimates from the true value. For case #1, in which we have normally distributed error term, the mean parameter estimates are very close to their true values for both sample sizes. For the other cases in which the error term is not normally distributed, the mean parameter estimates are far from their true values. Note that the mean parameter estimates are almost identical when $T = 500$ or $T = 5,000$, suggesting the bias in these parameter estimates may not just be a small sample phenomenon. When we compare the results among Cases #2, #3, and #4, the bias is smallest for Case #2, in which the error term is non-normal but symmetrically distributed.

In order to investigate how inferences on regime probabilities are affected by the violation of the normality assumption, we conduct another simulation study. When generating data, we consider the same data generating processes as given above, except that we generate S_t, $t = 1, 2, ..., T$, only once and fix them in repeated sampling. The sample size we consider is $T = 500$. For each data set generated in this way, we estimate the model in equations (1.1) and (1.2) by maximizing a normal log-likelihood and then calculate smoothed probabilities conditional on estimated parameters. Figure 1.1 plots the average smoothed probabilities of high-mean regime for each case. The shaded areas represent the true high-mean regime. Case #1 with the normal error term provides us with the sharpest regime inferences. However, as the distribution of the error term deviates from normality, inferences about regime probabilities deteriorate a lot especially for Cases #3 and #4, in which the error terms are asymmetrically distributed.

The simulation study in this section clearly demonstrates the pitfalls of estimating Markov-switching models by maximizing a normal log-likelihood when the normality assumption is violated. When normality assumption is violated, the resulting estimators have poor finite sample performance in parameter estimations and inferences about regime probabilities. In the next two sections, we introduce a Bayesian approach to estimating Markov-switching models with unknown and potentially non-normal error distributions.

1.3. Model Specifications and Identification Issues

1.3.1. Basic Model Specifications

We consider the following Markov-switching regression model:

Specification #1

$$y_t = \beta_0 + \beta_1 S_t + x_t^*, \tag{1.3}$$

$$\phi(L)x_t^* = h_{W_t}\varepsilon_t, \quad \varepsilon_t \sim i.i.d.(0,1), \tag{1.4}$$

$${h_1}^2 < {h_2}^2 < \ldots < {h_J}^2, \tag{1.5}$$

$$\varepsilon_t \sim i.i.d.(0,1), \tag{1.6}$$

where the distribution of ε_t is unknown; $S_t \in \{0,1\}$ is two-state Markov-switching variable; $W_t \in \{1,2,...,J\}$ is $J-$state Markov-switching variable; and $\phi(L) = \Big(1 - \phi_1 L - \ldots - \phi_k L^k\Big)$.

$$Pr[S_t = j|S_{t-1} = i] = p_{s,ij},\ \ i,j = 0,1; \quad Pr[W_t = j|W_{t-1} = i] = p_{w,ij},\ \ \ i,j = 1,2,...,J \tag{1.7}$$

We can approximate the distribution of ε_t by the following mixture of normals: [4]

$$\varepsilon_t | D_t \ \sim i.i.d. \ \ N(\mu^*_{D_t}, \sigma^{*2}_{D_t}), \quad D_t = 1, 2, ..., M, \tag{1.8}$$

where D_t is the mixing indicator variable with the following mixing probabilities:

$$Pr[D_t = m] = p_{d,m}, \quad m = 1, 2, ..., M \tag{1.9}$$

As the unconditional expectation and variance of ε_t are 0 and 1, respectively, we have the following restrictions on the conditional means and variances of ε^*_t:

$$\sum_{m=1}^{M} \mu^*_m w_m = 0; \quad and \quad \sum_{m=1}^{M} (\sigma^{*2}_m + \mu^{*2}_m) w_m = 1. \tag{1.10}$$

Bayesian inference of the above model with restrictions in equation (1.10) does not seem to be very straightforward, especially when we employ Dirichlet process mixture of normals. In order to circumvent the difficulties associated with imposing these restrictions, we consider an alternative representation of the model. By defining $x_t = \beta_0 + x^*_t$, we can rewrite equations (1.3) and (1.4) as:

Specification #2

$$y_t = \beta_1 S_t + x_t, \tag{1.3$'$}$$

$$\phi(L) x_t = g_{W_t} u_t, \quad u_t | W_t \sim i.i.d. \left(\frac{\phi(1)\beta_0}{g_{W_t}}, h_1^2 \right), \tag{1.4$'$}$$

where $u_t = \phi(1)\beta_0 + h_1 \varepsilon_t$, and $g_{W_t} = h_{W_t}/h_1$ with $g_1 = 1$.
Conditional on the mixture indicator variable D_t and W_t, the distribution of u_t is specified as:

$$u_t | D_t, W_t \ \sim i.i.d. \ \ N\left(\frac{\mu_{D_t}}{g_{W_t}}, \sigma^2_{D_t} \right), \quad D_t = 1, ..., M, \quad W_t = 1, \ldots, J \tag{1.11}$$

with $p_{d,m}$ referring to the mixture probability in equation (1.9). In this case, the mean and variance of u_t without conditioning on D_t are unconstrained.

[4] We allow for potential asymmetry in the distribution of ε_t. Note that in case $\mu^*_m = 0$ for all m, the distribution is ε^*_t is symmetric.

Based on the mean and variance of mixture of normals, we can recover β_0 and h_1 as the following,

$$\beta_0 = \frac{1}{\phi(1)} \sum_{m=1}^{M} \mu_m p_{d,m} \tag{1.12}$$

$$h_1^2 = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{g_{W_t}} - \frac{1}{T} \sum_{t=1}^{T} \frac{1}{g_{W_t}} \right)^2 \left(\sum_{m=1}^{M} (\mu_m - \bar{\mu})^2 p_{d,m} - \bar{\mu}^2 \right) + \bar{\sigma}^2 \tag{1.13}$$

where $\bar{\mu} = \sum_{m=1}^{M} \mu_m^2 p_{d,m}$ $\bar{\sigma}^2 = \sum_{m=1}^{M} \sigma_m^2 p_{d,m}$

As the parameter of interest can be recovered easily from Specification #2, the Markov Chain Monte Carlo (MCMC) algorithm presented in Section 1.4 is based on it. To let the model in equations (1.3')-(1.4') consistent with the inequality constrain in equation (1.5), we specify g_{W_t} in equation (1.4')

$$g_j^2 = g_{j-1}^2(1 + \bar{g}_j), \quad j = 2, 3, ..., J; \; g_1^2 = 1, \tag{1.14}$$

where $\bar{g}_j > 0$ and $(1 + \bar{g}_j) \sim IG(.,.)_{1[1+\bar{g}_j>1]}$.

The above specification implies the following inequality restriction:

$$g_J^2 > g_{J-1}^2 > \ldots > 1. \tag{1.15}$$

Furthermore, the above specification allows us to employ independent priors for $(1+\bar{g}_j)$, see Section 1.4 for discussion on MCMC algorithm. In the next subsection, we discuss the prior specification of mixture of normals.

1.3.2. Bayesian Modeling of the Finite Mixture of Normals and the Dirichlet Process Mixture of Normals

The Dirichlet process mixture of normals that we employ in this paper builds on the finite mixture of normals. In order to help understand the Dirichlet process mixture of normals and its relation to the finite mixture of normals, we review both models in this section. When the total number of mixtures, M, is fixed and pre-specified, we have the following specification for finite mixture of normals:

Finite Mixture of Normals

$$
\begin{aligned}
u_t|D_t, W_t \;&\sim\; i.i.d.\;\; N\left(\frac{\mu_{D_t}}{g_{W_t}}, \sigma^2_{D_t}\right),\;\; D_t = 1, 2, ..., M, \\
(p_{d,1}, p_{d,2}, \ldots, p_{d,M}) \;&\sim\; Dirichlet\left(\frac{\alpha}{M}, \ldots, \frac{\alpha}{M}\right), \\
\left(\mu_m, \sigma^2_m\right) \;&\sim\; G_0, \qquad m = 1, 2, ..., M, \\
G_0 \;&\equiv\; N\left(\lambda_0, \psi_0 \sigma^2_m\right) IG\left(\frac{\delta_0}{2}, \frac{\nu_0}{2}\right),
\end{aligned}
\tag{1.16}
$$

where $p_{d,m}$ is the mixing probability in equation (1.9) and G_0, the joint prior distribution of (μ_m, σ^2_m), is assumed to be Normal-Inverse-Gamma. The α parameter can be either fixed or random.

For the above finite mixture of normals, the prior probability of D_t conditional on $\tilde{D}_{\neq t}$ can be derived as: [5]

$$
\begin{aligned}
&Pr[D_t = m|\tilde{D}_{\neq t}, \alpha] = \frac{T_{m,\neq t} + \frac{\alpha}{M}}{T - 1 + \alpha}, \quad m = 1, 2, ..., M, \\
&(with \sum_{m=1}^{M} Pr[D_t = m|\tilde{D}_{\neq t}, \alpha] = 1)
\end{aligned}
\tag{1.17}
$$

where $\tilde{D}_{\neq t} = [\, D_1 \;\; \ldots \;\; D_{t-1} \;\; D_{t+1} \;\; \ldots \;\; D_T \,]'$ is the collection of mixing indicators excluding D_t; and $T_{m,\neq t}$ is the total number of observations that belong to class m in a sample that excludes period t. An important thing to note is that the above probabilities always add up to 1. With this background, we are now ready to discuss the Dirichlet process mixture of normals and its properties.

As suggested by Neal (2000), Gorur and Rasmussen (2010), and others, the limit of the model in equation (1.17) as $M \to \infty$ is equivalent to the Dirichlet process mixture of normals. A formal specification for the Dirichlet process mixture of normals is given below:

Dirichlet Process Mixture of Normals

[5] Proof of equation (1.17) is given in Appendix 1.A.

$$
\begin{aligned}
u_t|D_t, W_t \ &\sim \ i.i.d. \ \ N\left(\frac{\mu_{D_t}}{g_{W_t}}, \sigma^2_{D_t}\right), \ D_t = 1, 2, ..., M, \\
\left(\mu_m, \sigma^2_m\right) \ &\sim \ G, \qquad m = 1, 2, ..., M, \\
G \mid G_0, \alpha \ &\sim DP\left(\alpha, G_0\right) \\
G_0 \ &\equiv \ N\left(\lambda_0, \psi_0 \sigma^2_m\right) IG\left(\frac{\delta_0}{2}, \frac{\nu_0}{2}\right),
\end{aligned} \tag{1.18}
$$

where $DP(.,.)$ refers to the Dirichlet process; G_0 and α are referred to as the base distribution and the concentration parameter, respectively.

Here, M is a random variable, potentially infinite, that is to be estimated. Note that in the case of the finite mixture of normals, the joint distribution of (μ_m, σ^2_m) is given by G_0, and thus, $G \equiv G_0$. In the case of the Dirichlet process mixture of normals, however, the joint distribution of (μ_m, σ^2_m) is a random distribution generated by a Dirichlet process with based distribution G_0 and the concentration parameter α. [6]

The prior probability of D_t conditional on $\tilde{D}_{\neq t}$ can be obtained by taking the limit $M \rightarrow \infty$ for equation (1.17), as given below:

$$
\begin{aligned}
&Pr[D_t = m|\tilde{D}_{\neq t}, \alpha] = \frac{T_{m,\neq t}}{T - 1 + \alpha}, \ m = 1, 2, ..., M^*_{\neq t}, \\
&(with \ \sum_{m=1}^{M} Pr[D_t = m|\tilde{D}_{\neq t}, \alpha] < 1)
\end{aligned} \tag{1.19}
$$

where $T_{m,\neq t}$ is is defined earlier and $M^*_{\neq t}$ is the total number of distinctive classes (or mixtures) realized in the sample that excludes period t.

Unlike the case of the finite mixture of normals in equation (1.17), the above probabilities do not add up to 1, suggesting that there always exists non-zero probability that an observation at period t belongs to a new class that does not belong to the existing $M^*_{\neq t}$ classes. This probability is given below:

$$
\begin{aligned}
Pr[D_t = M^*_{\neq t} + 1|\tilde{D}_{\neq t}, \alpha] &= 1 - \sum_{m=1}^{M^*_{\neq t}} Pr[D_t = m|\tilde{D}_{\neq t}, M^*_{\neq t}] \\
&= \frac{\alpha}{T - 1 + \alpha},
\end{aligned} \tag{1.20}
$$

[6] That is, the Dirichlet process provides a random distribution over distributions on infinite sample spaces. The hierarchical models in which the Dirichlet process is used as a prior over the distribution of the parameters are referred to as the Dirichlet process mixture model.

which suggests that, if α is larger, the prior mean of M is higher with less concentrated distribution for G in equation (1.18).

The α parameter can be either fixed or random. In case α is treated as random, its conjugate prior is the Gamma distribution, given below:

$$\alpha \sim Gamma(a, b), \quad a > 0, \; b > 0. \tag{1.21}$$

1.3.3. Identification of Markov-switching Regimes and Mixture of Normals

As we mentioned in the introduction, one may not be able to identify the Markov-switching regimes, S_t and W_t, and Mixture of Normals, D_t, without additional restriction on the dynamic of Markov-switching regimes. On the one hand, since the i.i.d. mixture is a Markov-switching mixture without persistence in the state variable, when there is no additional restriction on the dynamics of the Markov-switching mixture, we may identify some of the mixture of normals to be a Markov-switching mixture. On the other hand, when the persistence of Markov-switching mixture is close to i.i.d. mixture, one may identify some of the Markov-switching mixture as mixture of normals.

This problem can be related to the case of summation of two independent time-series. Consider two independent time-series, a_t and b_t, where $a_t \sim N(0, \sigma_a^2)$, and $b_t \sim N(0, \sigma_b^2)$. Suppose we can only observe the summation of these two time series, $c_t = a_t + b_t$. In the case where both a_t and b_t are independent over time, one cannot identify σ_a^2 and σ_b^2 separately, but can only identify $\sigma_c^2 = \sigma_a^2 + \sigma_b^2$.

In contrast, suppose a_t is independent over time, and b_t is an AR(1) process,

$$b_t = \phi b_{t-1} + b_t^*, \quad b_t^* \sim i.i.d.N\left(0, \left(1 - \phi^2\right) \sigma_b^2\right)$$

where b_t^* is independent with a_t.

In this case, one can rewrite the model $c_t = a_t + b_t$ into an ARMA(1,1) model,

$$c_t = \phi c_{t-1} + e_t - \theta e_{t-1}, \quad e_t \sim i.i.d.N(0, \sigma_e^2), \tag{1.22}$$

We can identify the three parameters in the ARMA(1,1) model. We can also show that θ and σ_e^2 are functions of σ_a^2 and σ_b^2.

$$(1+\theta^2)\sigma_e^2 = (1+\phi^2)\sigma_a^2 + (1-\phi^2)\sigma_b^2$$

$$\theta\sigma_e^2 = \phi\sigma_a^2$$

Thus, by estimating the model in equation (1.22), we can identify the σ_a^2, σ_b^2, and ϕ.

The above example is the well known result in unobserved components model. It suggests that combination of two different time-series can be identified by imposing different dynamics. Follow the same logic, we derive the condition that guarantee the Markov-switching mixture to be persistent. Consider a general $J-$state first-order Markov-switching mixture, W_t, we can write it in the following vector autoregressive form,

$$\begin{bmatrix} W_{1,t} \\ W_{2,t} \\ \vdots \\ W_{J,t} \end{bmatrix} = \begin{bmatrix} p_{w,11} & p_{w,21} & \cdots & p_{w,J1} \\ p_{w,12} & p_{w,22} & \cdots & p_{w,J2} \\ \vdots & & & \\ p_{w,1J} & p_{w,2N} & \cdots & p_{w,JJ} \end{bmatrix} \begin{bmatrix} W_{1,t-1} \\ W_{2,t-1} \\ \vdots \\ W_{J,t-1} \end{bmatrix} + \begin{bmatrix} W^*_{1,t-1} \\ W^*_{2,t-1} \\ \vdots \\ W^*_{J,t-1} \end{bmatrix}, \qquad (1.23)$$
$$\Rightarrow \widetilde{W}_t = P\widetilde{W}_{t-1} + v_t$$

where $W_{j,t} = 1$ if $W_t = j$. $v_t = \widetilde{W}_t - P\widetilde{W}_{t-1}$, which is a martingale difference sequence.

Conventional Markov-switching model usually implies that the Markov-switching regimes are positively dependent on past regimes. For the above vector autoregressive model to be positively dependent, the transition probability matrix, P, needs to only have either strictly positive real eigenvalues or complex eigenvalue with strictly positive real part. [7] Proposition 1.1 below provide the condition that guarantees the Markov-switching mixture to be persistent.

Proposition 1.1. If $p_{W,ii} > 0.5$, $i = 1, \ldots, J$, then the eigenvalues of P can only be positive real number or complex eigenvalue with positive real part.

The proof of proposition 1.1 is in Appendix 1.B. Proposition 1.1 suggests if the diagonal term in the transition probability matrix, P, is larger than 0.5, the Markov-switching mixture

[7] For the persistence condition of vector autoregressive model, readers are referred to Hamilton (1994).

are positively dependent on past regimes. This condition allows us to successfully distinguish the Markov-switching mixture and iid mixture of normal. We impose this condition in practice through proper prior distribution. In particular, we impose the following priors for $p_{s,ij},\ i,j=0,1$ and $p_{w,ij},\ i,j=1,2,...,J$:

$$p_{s,11} \sim Beta(\alpha_{s,11}, \beta_{s,11})_{1[p_{s,11}>0.5]}, \quad p_{s,22} \sim Beta(\alpha_{s,22}, \beta_{s,22})_{1[p_{s,22}>0.5]},$$

$$[p_{w,i1}, p_{w,i2}, \ldots, p_{w,iJ}] \sim Dirichlet(\alpha_{w,i1}, \alpha_{w,i2}, \ldots, \alpha_{w,iJ})_{1[p_{w,ii}>0.5]},$$

In other words, we impose truncated beta distribution and truncated Dirichlet distribution as the prior distribution for $p_{s,ij},\ i,j=0,1$ and $p_{w,ij},\ i,j=1,2,...,J$, respectively.

1.4. Estimation of the Model

We denote $\tilde{\theta}_1$ as a vector that contains all the parameters associated with the Markov-switching regression equation in (1.3')-(1.4'), as given below:

$$\tilde{\theta}_1 = [\,\beta_1 \quad \tilde{\phi}' \quad \tilde{g}^{2'} \quad \tilde{p}'\,]',$$

where $\tilde{\phi} = [\,\phi_1 \quad \phi_2 \quad \ldots \quad \phi_k\,]'$; $\tilde{g}^2 = [\,g_2^2 \quad g_3^2 \quad \ldots \quad g_J^2\,]'$; and $\tilde{p}_s$and $\tilde{p}_w$ are vectors that contain the transition probabilities of S_t and W_t, respectively.

For the parameters associated with the Dirichlet process mixture of normals for u_t, we define

$$\tilde{\theta}_2 = [\,\tilde{\mu}' \quad \tilde{\sigma}^{2'} \quad \alpha \quad M\,]',$$

where $\tilde{\mu} = [\,\mu_1 \quad \ldots \quad \mu_M\,]'$; $\tilde{\sigma}^{2'} = [\,\sigma_1^2 \quad \ldots \quad \sigma_M^2\,]'$.

Then, the hierarchical nature of our model allows us to decompose the posterior distribution of our interest as follows:

$$\begin{aligned} f(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{S}_T, \tilde{W}_T, \tilde{D}_T|\tilde{Y}_T) &\propto f(\tilde{\theta}_2, \tilde{D}_T|\tilde{\theta}_1, \tilde{S}_T, \tilde{Y}_T) f(\tilde{\theta}_1, \tilde{S}_T|\tilde{Y}_T) \\ &= f(\tilde{\theta}_2, \tilde{D}_T|\tilde{\varepsilon}_T) f(\tilde{\theta}_1, \tilde{S}_T, \tilde{W}_T|\tilde{Y}_T), \end{aligned} \tag{1.24}$$

where $\tilde{S}_T = [\,S_1 \quad \ldots \quad S_T\,]'$, $\tilde{W}_T = [\,W_1 \quad \ldots \quad W_T\,]'$, $\tilde{D}_T = [\,D_1 \quad \ldots \quad D_T\,]'$, $\tilde{Y}_T = [\,Y_1 \quad \ldots \quad Y_T\,]'$, and $\tilde{u}_T = [\,u_1 \quad \ldots \quad u_T\,]'$. Equation (1.24) suggests that the MCMC algorithm consists of the following two steps:

Step 1: Draw the variates for the Markov-switching regression model conditional on mixture of Normals and data $\tilde{Y}_T$. That is, draw $\tilde{\theta}_1$ and $\tilde{S}_T$ conditional on $\tilde{\theta}_2$, $\tilde{D}_T$, and data.

Step 2: Draw the variates associated with the mixture of normals conditional on the error term for the Markov-switching regression equation in (1.3')-(1.4'). That is, draw $\tilde{\theta}_2$ and $\tilde{D}_T$ conditional on $\tilde{u}_T$.

1.4.1. Drawing Variates Associated with Markov-switching Regression Equation Conditional on the Mixture of Normals

Equation (1.11) implies that

$$u_t = \frac{\mu_{D_t}}{g_{W_t}} + \sigma_{D_t} v_t, \quad v_t \sim i.i.d. \;\; N(0,1), \tag{1.25}$$

and thus, by substituting this into equations (1.3')-(1.4') and rearranging terms, we obtain

$$y_t^* = \beta_1 S_t^* + (1 - \phi(L)) x_t^* + g_{w_t} v_t, \tag{1.26}$$

where $y_t^* = \frac{y_t - \mu_{D_t}}{\sigma_{D_t}}$; $S_t^* = \frac{S_t}{\sigma_{D_t}}$; and $x_t^* = \frac{x_t}{\sigma_{D_t}}$.

Based on equation (1.26), we can draw $[\,\beta_1 \quad \tilde{\phi}' \quad \tilde{g}^{2'} \quad \tilde{p}'\,]'$, $\tilde{S}_T$, and $\tilde{W}_T$ in the following sequence:

i) Draw $\tilde{S}_T$ conditional on β_1, $\tilde{\phi}'$, $\tilde{g}^{2'}$, $\tilde{Y}_T^* = [\,y_1^* \quad \ldots \quad y_T^*\,]'$, and $\tilde{X}_T^* = [\,x_1^* \quad \ldots \quad x_T^*\,]'$.
ii) Draw β_1 conditional on $\tilde{\phi}'$, $\tilde{g}^{2'}$, $\tilde{Y}_T^*$, $\tilde{S}_T^* = [\,S_1^* \quad \ldots \quad S_T^*\,]'$, and $\tilde{X}_T^*$.
iii) Draw $\tilde{g}^2$ conditional on β_1, $\tilde{\phi}'$, $\tilde{Y}_T^*$, $\tilde{S}_T^*$, and $\tilde{X}_T^*$.
iv) Draw $\tilde{\phi}'$ conditional on β_1, $\tilde{g}^{2'}$, $\tilde{Y}_T^*$, $\tilde{S}_T^*$, and $\tilde{X}_T^*$.

Equation (1.26) is a standard Markov-switching model with Markov-switching intercept and heteroskedasticity. Thus, drawing $\tilde{S}_T$ and $\tilde{\theta}_1$ from the appropriate full conditional distributions is standard.

1.4.2. Drawing Variates Associated with the Mixture of Normals Conditional on $\tilde{u}_T$

Conditional on $\tilde{\beta}$, $\tilde{h}^2$, $\tilde{S}_T$, and data $\tilde{Y}_T$, we can calculate the error term ε_t in equation (1.11) as

$$u_t = \frac{y_t - \beta_1 S_t - (1 - \phi(L))x_t}{h_{S_t}}, \quad t = 1, 2, ..., T. \tag{1.27}$$

Then, based on equation (1.25), we can draw the variates associated with mixture of normals in the following sequence:

i) Conditional on $\tilde{\mu}$, $\tilde{\sigma}^2$, and $\tilde{u}_T$, draw $\tilde{D}_T$ and α for the Dirichlet process mixture of normals. The total number of mixtures M is generated as a byproduct of generating $\tilde{D}_T$.

ii) Conditional on $\tilde{\sigma}^2$, $\tilde{D}_T$, M, and $\tilde{u}_T$, draw $\tilde{\mu}$.

iii) Conditional on $\tilde{\mu}$, $\tilde{D}_T$, M, and $\tilde{u}_T$, draw $\tilde{\sigma}^{2\prime}$

Drawing $\tilde{\mu}$ and $\tilde{\sigma}^2$ from appropriate full conditional posterior distributions derived based on equation (1.25) is standard. We thus focus on drawing $\tilde{D}_T$ and α in what follows. [8]

1.4.2.1. Drawing $\tilde{D}_T$ Conditional on α

If the total number of mixtures, M, were fixed as in the case of the finite mixture of normals, it would be straightforward to generate D_t based on the following full conditional distribution of D_t:

$$f(D_t|\tilde{\mu}, \tilde{\sigma}^2, \tilde{D}_{\neq t}, \varepsilon_t) \propto f(D_t|\tilde{D}_{\neq t}, \alpha) f(u_t|\tilde{\mu}, \tilde{\sigma}^2, D_t), \quad D_t = 1, 2, ..., M, \tag{1.28}$$

where $\tilde{D}_{\neq t}$ is collection of mixing indicators in the sample excluding D_t; $f(D_t|\tilde{D}_{\neq t}, \alpha)$ is the prior probability in equation (1.17); and $f(u_t|\tilde{\mu}, \tilde{\sigma}^2, D_t = m) = \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left[-\frac{(u_t - \mu_m/g_{W_t})^2}{2\sigma_m^2}\right]$. That is, we could draw D_t based on the following probabilities:

[8] This section is largely based on the works of West et al. (1994), Escobar and West (1995), and Neal (2000).

$$Pr[D_t = m|u_t, \tilde{\mu}, \tilde{\sigma}^2, \tilde{D}_{\neq t}] = \frac{Pr[D_t = m|\tilde{D}_{\neq t}]f(u_t|\tilde{\mu}, \tilde{\sigma}^2, D_t = m)}{\sum_{m=1}^{M} Pr[D_t = m|\tilde{D}_{\neq t}]f(u_t|\mu_m, \sigma_m^2, D_t = m)}, \ m = 1, 2, ..., M. \tag{1.29}$$

For the Dirichlet process mixture of normals, in which M is a random variable, Neal (2000) suggests that equation (1.28) should be replaced by:

$$f(D_t|\tilde{\mu}, \tilde{\sigma}^2, D_{\neq t}, \alpha, \varepsilon_t) \propto f(D_t|\alpha, \tilde{D}_{\neq t})f(\varepsilon_t|\tilde{\mu}, \tilde{\sigma}^2, D_t), \quad D_t = 1, \ldots, M^*_{\neq t}, M^*_{\neq t} + 1, \tag{1.30}$$

where $M^*_{\neq t}$ is the number of distinctive classes (or mixtures) in the sample that exclude period t; and $f(D_t|\tilde{D}_{\neq t}, \alpha)$ is the prior probability given in equation (1.19) or (1.20). Here, when $D_t = M^*_{\neq t} + 1$, it means that period t belongs to a new class that does not exist in $\tilde{D}_{\neq t}$. Given equation (1.30), we can then generate D_t using the following probabilities:

$$Pr[D_t = m|\tilde{\mu}, \tilde{\sigma}^2, D_{\neq t}, \alpha, u_t] = \frac{Pr[D_t = m|\tilde{D}_{\neq t}, \alpha]f(u_t|\tilde{\mu}, \tilde{\sigma}^2, D_t)}{\sum_{m=1}^{M^*_{\neq t}+1} Pr[D_t = m|\tilde{D}_{\neq t}, \alpha]f(u_t|\tilde{\mu}, \tilde{\sigma}^2, D_t)}, \tag{1.31}$$

$$m = 1, 2, ..., M^*_{\neq t}, M^*_{\neq t} + 1.$$

Depending on whether D_t belongs to the existing class ($m = 1, 2, ...,$ *or* $M^*_{\neq t}$) or a new class ($m = M^*_{\neq t} + 1$), we have the following two conditional densities for u_t:

$$f(u_t|\tilde{\mu}, \tilde{\sigma}^2, D_t = m) = f_N(u_t|\mu_m, \sigma_m^2), \quad for \ m = 1, 2, ..., M^*_{\neq t}; \tag{1.32}$$

$$f(u_t|\tilde{\mu}, \tilde{\sigma}^2, D_t = M^*_{\neq t} + 1) = \int\int f_N(u_t|\mu_{M^*_{\neq t}+1}, \sigma^2_{M^*_{\neq t}+1})G_0(\mu_{M^*_{\neq t}+1}, \sigma^2_{M^*_{\neq t}+1})d\mu_{M^*_{\neq t}+1}d\sigma^2_{M^*_{\neq t}+1}, \tag{1.33}$$

where $f_N(\cdot|\mu_j, \sigma_j^2)$ refers to a normal density function with mean μ_j and variance σ_j^2. The intuition for the integral in equation (1.33) is that, when period t belongs to a new class of normal with unknown mean and variance, we evaluate the density of u_t by taking average of the densities for all possible values of mean and variance generated from the base distribution

G_0. This integral can be evaluated by Monte Carlo simulation as suggested by West et al. (1994). [9]

By denoting $\tilde{D}_T$ as a collection of the mixing indicators (or class indicators) generated from the previous iteration of the MCMC, we can generate D_t by repeating the following steps sequentially for $t = 1, 2, ..., T$, starting with $t = 1$:

i) Count the total number of distinctive classes in $\tilde{D}_{\neq t}$ and set it as $M^*_{\neq t}$.

ii) Generate D_t according to the probabilities in equation (1.31), and replace the $t-$th element of $\tilde{D}_T$ with the generated D_t.

iii) If D_t is generated to be $M^*_{\neq t} + 1$, it means that period t belongs to a new class that does not exists in $\tilde{D}_{\neq t}$. In this case, we have to generate intermediate values for the mean ($\mu_{M^*_{\neq t}+1}$) and variance ($\sigma^2_{M^*_{\neq t}+1}$) that are associated with this new class. They can be generated from the following posterior distributions:

$$\sigma^2_{M^*_{\neq t}+1} \,|u_t \;\sim IG\left(\frac{1+d_0}{2}, \frac{v_0 + (\varepsilon_t - \lambda_0/g_{W_t})^2/(1+\psi_0/g^2_{W_t})}{2}\right), \tag{1.34}$$

$$\mu_{M^*_{\neq t}+1}|\sigma^2_{M^*_{\neq t}+1}, u_t \sim N\left(\frac{\lambda_0 + \psi_0 u_t/g_{W_t}}{1+\psi_0/g^2_{W_t}}, \frac{\psi_0}{1+\psi_0/g^2_{W_t}}\sigma^2_{M^*_{\neq t}+1}\right), \tag{1.35}$$

which can be easily derived given the joint prior G_0 for $(\mu_{M^*_{\neq t}+1}, \sigma^2_{M^*_{\neq t}+1})$ in equation (1.18) and a single observation u_t.

iv) We discard those mixture indices which no observation belongs to them. We then calculate the number of distinct mixture in the rest mixtures as M and redefine the rest mixture indices as $1, 2, \ldots, M$. That is, we assign $m = 1$ to a specific mixture and $m = 2$ to another specific mixture, etc.

v) Set t=t+1, and go to i).

[9] The integral in equation (1.33) can be approximated by

$$\int\int f_N(u_t|\mu_{M^*_{\neq t}+1}, \sigma^2_{M^*_{\neq t}+1})G_0(\mu_{M^*_{\neq t}+1}, \sigma^2_{M^*_{\neq t}+1})d\mu_{M^*_{\neq t}+1}d\sigma^2_{M^*_{\neq t}+1} \approx \frac{1}{R}\sum_{i=1}^{R} f_N(u_t|\mu_i/g_{W_t}, \sigma^2_i),$$

where μ_i and σ^2_i are drawn from the base distribution G_0 in equation (1.18) and R is large enough. Alternatively, Escobar and West (1995) analytically derive that this integral results in a density function for a scaled and shifted Student's t-distribution.

At the end of the iteration, we have a new set of $\tilde{D}_T$. The number of distinctive classes in $\tilde{D}_T$ is the realized M or the realized total number of mixtures.

1.4.2.2. Drawing α conditional on $\tilde{D}_T$, and thus, on M

Drawing α conditional on $\tilde{D}_T$ is equivalent to drawing α conditional on M, the total number of mixtures or classes in the sample. [10] In this section, we explain an algorithm for generating α as proposed by Escobar and West (1995).

Given the prior distribution of α in equation (1.21), the prior density is:

$$f(\alpha) \propto \alpha^{a-1} \exp(-\alpha b), \tag{1.36}$$

and as derived by Antoniak (1974), the likelihood for M is

$$f(M|\alpha) \propto \alpha^M \frac{\Gamma(\alpha)}{\Gamma(\alpha + T)}, \tag{1.37}$$

where $\Gamma(.)$ refers to the Gamma function and T is the sample size. Thus, Escobar and West (1995) derive the posterior density of α as: [11]

$$\begin{aligned} f(\alpha|M) &\propto f(\alpha) f(M|\alpha) \\ &\propto \alpha^{a+M-2} \exp(-\alpha b)(\alpha + T) \int_0^1 x^\alpha (1-x)^{T-1} dx, \end{aligned} \tag{1.38}$$

which implies that the posterior distribution of α is the marginal distribution obtained from a joint distribution of α and a continuous quantity η such that

[10] Note that the posterior distribution of α depends only on M, for given $\tilde{D}_T$.

[11] Note that gamma functions in equation (1.37) can be written as

$$\frac{\Gamma(\alpha)}{\Gamma(\alpha + T)} = \frac{(\alpha + T)\beta(\alpha + 1, T)}{\alpha \Gamma(T)},$$

where $\beta(.,.)$ refers to the beta function, and

$$\beta(\alpha + 1, T) = \int_0^1 x^\alpha (1-x)^{T-1} dx$$

$$f(\alpha, \eta|M) \propto \alpha^{a+M-1} \exp(-\alpha b)(\alpha + T)\eta^{\alpha}(1-\eta)^{T-1}, \; 0 < \eta < 1. \tag{1.39}$$

As shown in Appendix 1.C, Escobar and West(1995) further derive the conditional posterior densities $f(\eta|\alpha, M)$ and $f(\alpha|\eta, M)$, and show that

$$\eta|\alpha, M \;\sim Beta(\alpha + 1, T) \tag{1.40}$$

and

$$\alpha|\eta, M \;\sim r_{\eta}G(a + M, b - \ln(\eta)) + (1 - r_{\eta})G(a + M - 1, b - \ln(\eta)), \tag{1.41}$$

where the latter is a mixture of two Gamma distributions with $r_{\eta}/(1 - r_{\eta}) = (a + M - 1)/\{T[b - \ln(\eta)]\}$.

Thus, the following two-step algorithm can be employed to draw α:

i) Conditional on α generated in the previous iteration of the Gibbs sampling, draw an intermediate random variable η from the distribution given in equation (1.40).

ii) Conditional on η, draw α from the distribution given in equation (1.41).

1.5. Model Comparison Methods

We consider formal model comparisons in Bayesian framework. For the in sample model comparison, we consider the Watanabe-Alkaike Information Criterion proposed by Watanabe (2010). For the out-of-sample model comparison, we consider the predictive likelihood.

1.5.1. Watanabe-Alkaike Information Criterion

For in sample model comparison, Spiegelhalter et al. (2002) first proposed using Deviance Information Criterion (DIC) for complex hierarchical models. Although DIC can be calculated in a straightforward way, it has been criticized in many different perspectives. First, the penalty term is not invariant to reparameterization. For example, we would obtain a (slightly) different penalty term (and hence DIC) if we parameterized in terms of

h_{S_t} or $\log(h_{S_t})$, even if the priors on each were mathematically equivalent. Furthermore, Celeux et al. (2006) showed well in the context of mixture models, DIC is not based on a universal principle that could lead to a procedure that was both computationally practical and generically applicable. Therefore, we adopt another model comparison criterion, the Watanabe-Akaike information criterion (WAIC) (Watanabe (2010)). WAIC is defined as the following,

$$WAIC = -2\sum_{t=1}^{T} E_{\Psi|Y_t}\left\{\log\left[f(Y_t|\Psi)\right]\right\} + 2\left\{\sum_{t=1}^{T}\log E\left[f(Y_t|\Psi)\right] - \sum_{t=1}^{T} E_{\Psi|Y_t}\left\{\log\left[f(Y_t|\Psi)\right]\right\}\right\}$$

The first term in WAIC represents the model fit, and the second term in WAIC represents the penalty. Thus, WAIC prefers the model with a small value. In practice, the WAIC can be calculated by replacing the expectations by averages over the R posterior draws Ψ_R as

$$\sum_{t=1}^{T} E_{\Psi|Y_t}\left\{\log\left[f(Y_t|\Psi)\right]\right\} = \sum_{t=1}^{T}\left[\frac{1}{R}\sum_{j=1}^{R}\log f(Y_t|\Psi_j)\right]$$

$$\sum_{t=1}^{T}\log E\left[f(Y_t|\Psi)\right] = \sum_{t=1}^{T}\log\left[\frac{1}{R}\sum_{j=1}^{R} f(Y_t|\Psi_j)\right]$$

where R is the number of MCMC simulations.

1.5.2. Predictive Density and Likelihood

For out-of-sample model comparison, Gelfand and Mukhopadhyay (1995) discuss the predictive likelihood of linear functionals for Dirichlet process mixture models. Following their findings, the one-step-ahead predictive posterior density for the proposed model in Section 1.3 is

$$\begin{aligned} f\left(y_{t+1}|I_t\right) &= \int f\left(y_{t+1}|\tilde{\theta}_1, \tilde{\theta}_2, S_{t+1}, W_{t+1}, \alpha\right) f\left(\tilde{\theta}_1, \tilde{\theta}_2, S_{t+1}, W_{t+1}, \alpha|I_t\right) d\tilde{\theta}_1 d\tilde{\theta}_2 dS_{t+1} dW_{t+1} d\alpha \\ &\approx \frac{1}{R}\sum_{i=1}^{R} f\left(y_{t+1}|\tilde{\theta}_1^{(r)}, \tilde{\theta}_2^{(r)}, S_{t+1}^{(r)}, W_{t+1}^{(r)}, I_t\right) \end{aligned}$$

where the conditional density

$$\begin{aligned} f\left(y_{t+1}|\tilde{\theta}_1^{(r)}, \tilde{\theta}_2^{(r)}, S_{t+1}^{(r)}, W_{t+1}^{(r)}, I_t\right) =& \frac{\alpha^{(r)}}{\alpha^{(r)}+T}\int f\left(y_{t+1}|\tilde{\theta}_1^{(r)}, \mu, \sigma^2, S_{t+1}^{(r)}, W_{t+1}^{(r)}, I_t\right) G_0(\mu,\sigma^2) d\mu d\sigma^2 \\ &+ \sum_{m=1}^{M^{(r)}} \frac{T_m^{(r)}}{\alpha^{(r)}+T} f_N\left(y_{t+1}|\tilde{\theta}_1^{(r)}, \mu_m^{(r)}, \sigma_m^{2\ (r)}, S_{t+1}^{(r)}, W_{t+1}^{(r)}, I_t\right) \end{aligned}$$

and $S_{t+1}^{(r)}$ and $W_{t+1}^{(r)}$ are randomly drawn from the transition probabilities, $Pr\left(S_{t+1}^{(r)}|S_t^{(r)}\right)$ and $Pr\left(W_{t+1}^{(r)}|W_t^{(r)}\right)$, respectively.

1.6. An Application to the Growth of Postwar U.S. Industrial Production Index [1947M1-2017M1]

1.6.1. Specification for an Empirical Model

We consider the following univariate Markov-switching model for the growth of industrial production index (y_t):

$$\begin{aligned} y_t &= \beta_{0,C_t} + \beta_{1,C_t} S_t + x_t^* \\ x_t^* &= \phi x_{t-1}^* + h_{W_t}\varepsilon_t, \quad \varepsilon_t \sim i.i.d.(0,1), \\ S_t &= 0,1, \quad W_t = 1,2,3, \\ \beta_{1,C_t} &< 0, \quad \forall t, \end{aligned} \tag{1.42}$$

β_{0,C_t} is the mean growth rate during boom and $\beta_{0,C_t} + \beta_{1,C_t}$ is the mean growth rate during recession. We have a boom when $S_t = 0$ and we have a recession when $S_t = 1$. The distribution of the error term u_t is potentially non-normal, and it is approximated by the Dirichlet process mixture of normals in equation (1.18).

The model in equation (1.42) is an extension of the model in equations (1.3')-(1.4') with $J = 3$. We adopt this extension because it would be unreasonable to assume that the regime-specific mean growth rates during boom or recession are constant in a sample that covers the entire postwar period. While Eo and Kim (2016) propose to specify the regime-specific mean growth rates of real GDP to be random walks, we assume that there are two structural breaks in the regime-specific mean growth rates. We thus specify β_{0,C_t} and β_{1,C_t} as follows: [12]

$$\beta_{j,C_t} = \delta_{j,1} + \delta_{j,2}\ C_{2,t} + \delta_{j,3}\ C_{3,t}, \quad j = 0,1, \tag{1.44}$$

where $C_{j,t} = 1$ if $C_t = j$, and 0, otherwise. We impose the following restrictions in order to

[12] The inequalities in equation (1.44) is based on Kim and Nelson (1999), who suggest that the difference between the mean growth rates during recessions and booms has been decreasing.

incorporate the observation that the gap between the mean growth rates during boom and recession has been decreasing (Kim and Nelson (1999)):

$$\delta_{0,3} < \delta_{0,2} < 0; \quad \delta_{1,3} > \delta_{1,2} > 0. \tag{1.45}$$

By defining the transition probabilities for C_t as $p_{C,ij} = Pr[C_t = j|C_{t-1} = i]$, the following restrictions are employed to model two structural breaks in the sample:

$$p_{C,13} = p_{C,21} = p_{C,31} = p_{C,32} = 0 \textit{ and } p_{C,33} = 1. \tag{1.46}$$

Note that β_1 is the mean growth rate during boom after the second structural break and $\beta_1 + \beta_2$ is the mean growth rate during recession after the second structural break.

We can rewrite the above model as the following equation:

$$y_t = (\delta_{0,2}C_{2,t} + \delta_{0,3}C_{3,t}) + (\delta_{1,1} + \delta_{1,2}C_{2,t} + \delta_{1,3}C_{3,t})S_t + x_t, \tag{1.47}$$

$$x_t = \phi x_{t-1} + g_{W_t}u_t, \quad u_t \sim i.i.d.\left(\frac{(1-\phi)\delta_{0,1}}{g_{W_t}}, h_1^2\right), \tag{1.48}$$

One can see that the model in equations (1.47)-(1.48) has very similar form as the model in equations (1.3')-(1.4'), except the new model allows for two structural breaks in the long-run mean growth rate. Note that the model specified above implies a time-varying long-run mean growth rate, which can be estimated by:

$$\begin{aligned} \tau_t = {} & \delta_{0,0} + \delta_{0,1}Pr[C_t = 1|\tilde{Y}_T] + \delta_{0,2}Pr[C_t = 2|\tilde{Y}_T] \\ & + \Big(\delta_{1,0} + \delta_{1,1}Pr[C_t = 1|\tilde{Y}_T] + \delta_{1,2}Pr[C_t = 2|\tilde{Y}_T]\Big) \times Pr[S_t = 1], \end{aligned} \tag{1.49}$$

where $Pr[S_t = 1]$ is the steady-state probability that $S_t = 1$, which is given below:

$$Pr[S_t = 1] = \frac{1 - p_{s,11}}{2 - p_{s,11} - p_{s,22}}.$$

1.6.2. Specification of Prior Distributions

We estimate the model by employing the normalization introduced in *Specification #2* in Section 1.3, and set $\delta_{0,1} = 0$ and $g_1^2 = 1$. We then employ the estimation procedure in

Section 1.4. The original parameters $\delta_{0,1}$, and h_i^2, $i = 1, 2, 3$, are recovered as discussed in Section 1.3. The priors that we employ are described below:

$$[\delta_{0,2} \;\; \delta_{0,3} \;\; \delta_{1,1} \;\; \delta_{1,2} \;\; \delta_{1,3}]' \sim N\left([-0.5 \;\; -1 \;\; -1.5 \;\; 0.5 \;\; 1]', 0.5I_5\right)_{1[\delta_{1,1}<0, \delta_{0,3}<\delta_{0,2}<0, \delta_{1,3}>\delta_{1,2}>0]},$$

$$\phi \sim N(0.3, 0.5)_{1[|\phi|<1]}$$

$$g_2^2 = g_1^2(1+\bar{g}_2), \;\; (1+\bar{g}_2) \sim IG(1,2)_{1[1+\bar{g}_2>1]}$$

$$g_3^2 = g_2^2(1+\bar{g}_3), \;\; (1+\bar{g}_3) \sim IG(1,2)_{1[1+\bar{g}_3>1]}$$

$$p_{s,00} \sim Beta(0.9, 0.1)_{1[p_{s,00}>0.5]}, \quad p_{s,11} \sim Beta(0.8, 0.2)_{1[p_{s,11}>0.5]},$$

$$p_{C,11} \sim Beta(24.9, 0.1)_{1[p_{C,11}>0.5]}, \quad p_{C,22} \sim Beta(19.9, 0.1)_{1[p_{C,22}>0.5]},$$

$$[p_{w,11}, p_{w,12}, p_{w,13}] \sim Dirichlet(0.9, 0.05, 0.05)_{1[p_{w,11}>0.5]},$$

$$[p_{w,21}, p_{w,22}, p_{w,23}] \sim Dirichlet(0.05, 0.9, 0.05)_{1[p_{w,22}>0.5]},$$

$$[p_{w,31}, p_{w,32}, p_{w,33}] \sim Dirichlet(0.05, 0.05, 0.9)_{1[p_{w,33}>0.5]},$$

$$(\mu_m, \sigma_m^2) \sim \;\; G_0 \;\; \equiv \;\; N(2, \sigma_m^2) IG\,(2, 1)\,, \;\; m = 1, 2, \ldots,$$

$$\alpha \sim Gamma(1, 2).$$

For model comparison purpose, we consider two alternative cases about the assumption on error distributions. In the first case, ε_t is assumed to be normally distributed. In this case, β_1 and h_1^2 are estimated directly. We employ the following priors for these parameters:

$$\delta_{0,1} \sim N(2, 0.5), \quad h_1^2 \sim IG(2, 1).$$

In the second case, ε_t is assumed to be Student-t distributed with unknown degree of freedom, ν. [13] β_1 and h_1^2 are also estimated directly in this case, and we employ the same priors as in the first case. For the degree of freedom parameter (ν), we employ the following prior distribution,

$$\nu \sim \chi^2(3)$$

[13] For model with t distribution, we follow Geweke's (1993) specification. In particular,

and ν is draw from its posterior distribution by Metropolis-Hasting Algorithm.

1.6.3. Empirical Results

Data employed are seasonally-adjusted postwar U.S. industrial production index. Data are obtained from the Federal Reserve Bank of St. Louis economic database (FRED), and the sample covers the period 1947M1-2017M1. Figure 1.2 depicts the data. We obtain 150,000 MCMC draws and discard the first 50,000 to avoid the effect of the initial values. All the inferences are based on the remaining 100,000 draws.

Table 1.2 reports the posterior moments of the parameters obtained under the normality assumption for the error term. When we performed a normality test for the error term for this case, however, the null was rejected at a 5% significance level. This provides a justification for employing the proposed model, in which we approximate the unknown error term with the Dirichlet process mixture of normals.

For the proposed model, the posterior mean for the total number of mixtures is slightly higher than 3, as shown in Table 1.3. The null hypothesis of normality is not rejected for the standardized error term estimated conditional on the mixing indicator variable. [14] These results suggest that the Dirichlet process mixture normals model reasonably well approximates the unknown distribution of the error term. Furthermore, Both in sample and out-of-sample Bayesian model selection criteria, WAIC and predictive likelihood, strongly prefers the proposed model.

---

model in equation (1.42) with ε_t follows a independent t distribution can be written as,

$$\begin{aligned} y_t =&\beta_{0,C_t} + \beta_{1,C_t} S_t + x_t^* \\ &x_t^* = \phi x_{t-1}^* + h_{W_t}\varepsilon_t, \quad \varepsilon_t \sim i.i.d.N(0, \frac{1}{\lambda_t}), \\ &\lambda_t \sim i.i.d.\Gamma(\frac{\nu}{2}, \frac{\nu}{2}) \\ &S_t = 0,1, \ W_t = 1,2,3 \\ &\beta_{1,C_t} < 0, \quad \forall t, \end{aligned}$$

[14] To calculate the Jarque-Bera test statistic for the normality test, we use the posterior mean of the standardized error term $(\frac{u_t - \mu_{D_t}/g_{W_t}}{\sigma_{D_t}})$ from equation (1.11), for $t = 1, 2, ..., T$.

Figures 1.3.A, 1.3.B, and 1.3.C depict the posterior probabilities of recession for the three models. The shaded areas represent the NBER recessions. Estimates of turning points from the proposed model are much sharper and agree much more closely with the NBER reference cycles than the estimates from a model with normally distributed errors do.

Figure 1.4 depicts the time-varying volatility for the IP growth rate estimated from the proposed model. Note that we model the volatility process as a 3-state Markov-switching process. It seems that the high and the medium volatility regimes are mostly focused on the period prior to the mid 1980s. However, in most of the post-1984 period, the low volatility regime dominates except for a few episodes of high or medium volatility. Finally, Figure 1.5 depicts the posterior mean of the long-run mean growth rates estimated based on equation (1.49). It demonstrates a pattern for steadily decreasing long-run mean growth rate, which is consistent with Stock and Watson (2012) and Eo and Kim (2016).

1.6.4. Sensitivity Analysis of the Prior Distribution

As we mentioned in the previous sections, one important identification condition we impose in this paper is that Markov-switching mixture is sufficiently persistent. We employ truncated Beta distribution to impose this condition on the business cycle dynamics. In other words, our model achieves identification through proper prior distribution. It will be interesting to see whether the empirical results in Section 1.6.3 depends on the tightness of the prior distribution. In particular, we are interested to see whether the posterior probabilities of recession, based on the proposed model as well as the model with Gaussian or Student-t distributed errors, will be affected when we employ relatively diffused prior on the transition probability of S_t and W_t.

We consider the following alternative priors on the transition probability of S_t,

$$p_{s,00} \sim Beta(0.18, 0.02)_{1[p_{s,00}>0.5]}, \quad p_{s,11} \sim Beta(0.16, 0.04)_{1[p_{s,11}>0.5]},$$

$$[p_{w,11}, p_{w,12}, p_{w,13}] \sim Dirichlet(0.18, 0.01, 0.01)_{1[p_{w,11}>0.5]},$$

$$[p_{w,21}, p_{w,22}, p_{w,23}] \sim Dirichlet(0.01, 0.18, 0.01)_{1[p_{w,22}>0.5]},$$

$$[p_{w,31}, p_{w,32}, p_{w,33}] \sim Dirichlet(0.01, 0.01, 0.18)_{1[p_{w,33}>0.5]},$$

and all the other prior distributions maintain the same as in Section 1.6.3. Figure 1.3.A, 1.3.B, and 1.3.C present the posterior probabilities of recession based on Gaussian, Student-t, and Dirichlet process mixture error terms, respectively. We can see that, the the posterior probabilities of recession based on proposed model are more robust than the other two alternative models. This also demonstrate that the proposed model is able to exploit the data information more efficiently.

1.7. Concluding Remarks

In their dynamic factor models of business cycle, Kim and Yoo (1996), Chauvet (1998), and Kim and Nelson (1998) assume that each individual coincident variable consists of an idiosyncratic component and a common factor component, which is subject to Markov switching mean. They estimate their models either by the maximum likelihood estimation or by the Bayesian method, under the assumption of normally distributed shocks. They all show that their estimates of turning points are much sharper and agree much more closely with the NBER reference cycles than the estimates from a univariate Markov switching model do. The intuition is that the idiosyncratic components in these multivariate models, which consist of irregular components and outliers, are averaged out across individual series.

However, even in case the common factor component is normally distributed, the existence of irregular components and outliers in individual series makes the error term in a univariate model to deviate from normality. This is the main reason why our univariate Markov-switching model of the postwar industrial production index results in poor inferences on recession probabilities under a normality assumption. By modeling the error term as the Dirichlet process mixture of normals, we can effectively control for the irregular component that is not related to the business conditions. This leads to sharp and accurate inferences on recession probabilities just like the dynamic factor models do.

Appendix 1.A. Derivation of Equation (1.17)

Given the prior for $(w_1, w_2, \ldots, w_M)$ in equation (1.16), the marginal distribution of w_m is

$$w_m \sim Beta(\frac{\alpha}{M}, \frac{\alpha}{M}(M-1)), \tag{1.A.1}$$

with the following density function:

$$f(w_m) \propto w_m^{\frac{\alpha}{M}-1}(1-w_m)^{\frac{\alpha}{M}(M-1)-1}. \tag{1.A.2}$$

The likelihood of $\tilde{D}_{\neq t}$ given w_m can be expressed as:

$$f(\tilde{D}_{\neq t}|w_m) \propto w_m^{T_{m,\neq t}}(1-w_m)^{T-1-T_{m,\neq t}}, \tag{1.A.3}$$

where $T_{m,\neq t}$ denotes the total number of observations that belong to the $m-th$ class in a sample that excludes period t.

By combining equations (1.A.2) and (1.A.3), we have:

$$\begin{aligned} f(w_m|\tilde{D}_{\neq t}) &\propto f(w_m)Pr(\tilde{D}_{\neq t}|w_m) \\ &= w_m^{T_{m,\neq t}+\frac{\alpha}{M}-1}(1-w_m)^{T-1-T_{m,\neq t}+\frac{\alpha}{M}(M-1)-1}, \end{aligned} \tag{1.A.4}$$

which suggests that

$$w_m|\tilde{D}_{\neq t} \;\sim\; Beta(T_{m,\neq t}+\frac{\alpha}{M}, T-1-T_{m,\neq t}+\frac{\alpha}{M}(M-1)). \tag{1.A.5}$$

From equation (1.A.5), we can derive the following probability of interest in equation (1.17):

$$\begin{aligned} Pr[D_t = m|\tilde{D}_{\neq t}) &= E(w_m|\tilde{D}_{\neq t}) \\ &= \frac{T_{m,\neq t}+\frac{\alpha}{M}}{T-1+\alpha}. \end{aligned} \tag{1.A.6}$$

Appendix 1.B. Proof of Proposition 1.1

The transition probabilities matrix, P, in equation (1.23) has the following two properties by definition: (i) $p_{w,ij} \geq 0$ for all $1 < i,j < J$ and $\sum_{j=1}^{J} p_{w,ij} = 1, \forall i$. Given this two properties, it is easy to see that if $p_{w,ii} > 0.5, i = 1, \ldots, J$, then $p_{w,ij} > \sum_{j=1,j\neq i}^{J} |p_{w,ij}|$. Define $R_i = p_{w,ii} - \sum_{j=1,j\neq i}^{J} |p_{w,ij}|$, where $R_i > 0$ for all i by construction.

Let $sign_{ij} = \frac{p_{w,ij}}{|p_{w,ij}|}$ be the sign of $p_{w,ij}$. Then we can easily check the following quadratic for all $x \in R^J$

$$x'Px = \sum_{i=1}^{J} R_i x_i^2 + \sum_{i=1}^{J} \sum_{j>i} |p_{w,ij}|(x_i + sign_{ij} x_j)^2 \tag{1.B.1}$$

All the summands in equation (1.B.1) are nonnegative as R_i is positive when $p_{w,ii} > 0.5, i = 1, \ldots, J$. Furthermore, if $x \neq 0$ then $x_i \neq 0$ for some i, so $x'Px \geq R_i x_i^2 > 0$ for all $x \in R^J$.

Define the eigenvalues of P as $\mu + i\nu$ and the corresponding eigenvectors as $x + iy$, where i denote $\sqrt{-1}$. By the definition of eigenvalues and eigenvectors,

$$\begin{aligned}
&(P - \mu + i\nu)(x + iy) \\
\Rightarrow &\begin{cases} (P - \mu)x + \nu y = 0 \\ (P - \mu)y - \nu x = 0 \end{cases} \\
\Rightarrow &x'(P - \mu)x + y'(P - \mu)y = \nu(y'x - x'y) = 0 \\
\Rightarrow &\mu = \frac{x'Px + y'Py}{x'x + y'y}
\end{aligned} \tag{1.B.2}$$

As $x'Px > 0$ for all $x \in R^J$, one can easily see that μ in equation (1.B.2) is always positive. Thus, we can conclude that when $p_{w,ii} > 0$

Appendix 1.C. Derivation of Equations (1.40) and (1.41)

Conditional on α, equation (1.39) results in

$$f(\eta|\alpha, M) \propto \eta^{\alpha}(1-\eta)^{T-1}, \quad 0 < \eta < 1, \tag{1.C.1}$$

which suggests that

$$\eta \mid \alpha, M \ \sim Beta(\alpha + 1, T). \tag{1.C.2}$$

Thus, as in Escobar and West (1995), the conditional density of α given η and M can be derived as:

$$
\begin{aligned}
&f(\alpha|\eta, M) \\
&\quad \propto \alpha^{a+M-1}\exp\{-\alpha[b-\ln(\eta)]\} + T\alpha^{a+M-2}\exp\{-\alpha[b-\ln(\eta)]\} \\
&\quad = \frac{\Gamma(a+M)}{[b-\ln(\eta)]^{a+M}}G(a+M, b-\ln(\eta)) + T\frac{\Gamma(a+M-1)}{[b-\ln(\eta)]^{a+M-1}}G(a+M-1, b-\ln(\eta)) \\
&\quad \propto (a+M-1)G(a+M, b-\ln(\eta)) + T[b-\ln(\eta)]G(a+M-1, b-\ln(\eta)),
\end{aligned} \tag{1.C.3}
$$

which can be written as the following mixture of two Gamma distributions:

$$
(\alpha|\eta, M) \sim r_\eta G(a+M, b-\ln(\eta)) + (1-r_\eta)G(a+M-1, b-\ln(\eta)), \tag{1.C.4}
$$

where $r_\eta/(1-r_\eta) = (a+M-1)/\{T[b-\ln(\eta)]\}$.

References

[1] Alexander, C., E. Lazar, 2006, "Normal Mixture GARCH(1,1): Applications to Exchange Rate Modeling," *Journal of Applied Econometrics*, 21, 307-336.

[2] Antoniak, C. E., 1974, "Mixtures of Dirichlet Processes With Applications to Nonparametric Problems," *The Annals of Statistics*, 2, 1152-1174.

[3] Bauwens, L., J.-F. Carpantier, A. Dufays, 2017, "Autoregressive Moving Average Infinite Hidden Markov-Switching Models," *Journal of Business and Economic Statistics*, 35, 162-182.

[4] Chauvet, M., 1998, "An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switches," *International Economic Review*, 39, 969-996.

[5] Chib, S., 1998, "Estimation and Comparison of Multiple Change-Point Models," *Journal of Econometrics*, 86, 221-241.

[6] Diebold, F. X., J.-H. Lee, and G. C. Weinbach, 1994, "Regime Switching with Time-Varying Transition Probabilities," In C. Hargreaves, Ed.: *Nonstationary Time Series Analysis and Cointegration*, 283-302. Oxford University Press.

[7] Dueker, M., 1997, "Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility," *Journal of Business and Economic Statistics*, 15, 26-34.

[8] Eo, Y. and C-J. Kim, 2016, "Markov-Switching Models with Evolving Regime-Specific Parameters: Are Postwar Booms and Recessions All Alike?," *Review of Economics and Statistics*, 98, 940-949.

[9] Escobar, M. D., and M. West, 1995, "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577-588.

[10] Filardo, A. J., 1994, "Business Cycle Phases and Their Transitional Dynamics," *Journal of Business and Economic Statistics*, 12, 299-308.

[11] Fox, E., E. Sudderth, M. Jordan, and A. Willsky, 2011, "A Sticky HDP-HMM with

Application to Speaker Diarization," *Annals of Applied Statistics*, 5, 1020-1056.

[12] Frhwirth-Schnatter, S., 2006, "Finite Mixture and Markov Switching Models. Springer Series in Statistics," Springer, New York.

[13] Geweke, J., 1993, "Bayesian Treatment of the Independent Student-t Linear Model," *Journal of Applied Econometrics*, 8, S19-S40.

[14] Gorur, D., and C. E. Rasmussen, 2010, "Dirichlet Process Gaussian Mixture Models: Choice of the Base Distribution," *Journal of Computer Science and Technology*, 25, 615-626.

[15] Hamilton, J. D., 1989, "A New Approach to the Economic Analysis of Nonstationary Time Series and The Business Cycle," *Econometrica*, 57, 357-384.

[16] Hamilton, J. D., 1994 "Time Series Analysis," Princeton University Press, Princeton. NJ.

[17] Jensen, M. and J. Maheu, 2010, "Bayesian Semiparametric Stochastic Volatility Modeling," *Journal of Econometrics*, 157, 306-316.

[18] Jensen, M. and J. Maheu, 2013, "Bayesian Semiparametric Multivariate GARCH Modeling," *Journal of Econometrics*, 176, 3-17.

[19] Jensen, M, and J. Maheu, 2014, "Estimating a Semiparametric Asymmetric Stochastic Volatility Model with a Dirichlet Process Mixture," *Journal of Econometrics*, 178, 523-538.

[20] Jin, X. and J. Maheu, 2016, "Bayesian Semiparametric Modeling of Realized Covariance Matrices," *Journal of Econometrics*, 192, 19-39.

[21] Kim, C-J., 1994, "Dynamic Linear Models with Markov Switching," *Journal of Econometrics*, 60, 1-22.

[22] Kim, C-J. and C. Nelson, 1998, "Business Cycle Turning Points, A New Coincident Index, and Tests of Duration Dependence Based on A Dynamic Factor Model with Regime-Switching," *Review of Economics and Statistics*, 80, 188-201.

[23] Kim, C-J. and C. Nelson, 1999, "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, 608-616.

[24] Kim, S., N. Shephard, and S. Chib, 1998, "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361-393.

[25] Kim, M.-J. and J.-S. Yoo, 1995, "New Index of Coincident Indicators: A Multivariate Markov Switching Factor Model Approach," *Journal of Monetary Economics*, 36, 607-630.

[26] Koop, G., 2003, *Bayesian Econometrics*, New York: Wiley

[27] Marin, J.-M., K. Mengersen, C.P. Robert, 2005, "Bayesian Modeling and Inference on Mixtures of Distributions," *Handbook of Statistics*, 25, 459-507.

[28] Neal, R. M., 2000, "Markov Chain Sampling Methods for Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, 9, 249-265.

[29] Omori, Y., S. Chib, N. Shephard, and J. Nakajima, 2007, "Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference," *Journal of Econometrics*, 140, 425-449.

[30] Song, Y., 2014, "Modeling Regime Switching and Structural Breaks with an Infinite Hidden Markov Model," *Journal of Applied Econometrics*, 29, 825-842.

[31] Stock, J. and M. Watson, 2012, "Disentangling the Channels of the 2007-09 Recession," *Brookings Papers on Economic Activity*, 81-156.

[32] Watanabe, S., 2010, "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory," *Journal of Machine Learning Research*, 11, 3571-3594.

[33] West, M., P. Muller, and M. D. Escobar, 1994, "Hierarchical Priors and Mixture Models, With Application in Regression and Density Estimation," in *Aspects of Uncertainty: A Tribute to D.V. Lindley*, eds. P. R. Freeman and A. F. M. Smith, New York: Wiley, 363-386.

[34] White, H., 1994, *Estimation, Inference, and Specification Analysis*, New York: Cambridge University Press.

Table 1.1. Quasi Maximum Likelihood Estimation of Markov-switching Models: Monte Carlo Experiment

$T = 500$

	True	Case #1	Case #2	Case #3	Case #4
$\delta_{0,1}$	-0.5	-0.519 (0.232)	-0.360 (0.288)	-1.031 (0.634)	-1.148 (1.092)
β_2	1	0.999 (0.067)	1.044 (0.676)	1.141 (0.175)	1.324 (0.397)
h_1	2	1.990 (0.128)	2.089 (0.330)	1.982 (0.256)	1.679 (0.451)
h_2	1	1.004 (0.004)	0.891 (0.109)	0.726 (0.275)	0.495 (0.505)
p_{11}	0.9	0.900 (0.042)	0.884 (0.057)	0.651 (0.279)	0.497 (0.426)
p_{22}	0.95	0.950 (0.019)	0.933 (0.045)	0.854 (0.110)	0.702 (0.274)

$T = 5000$

	True	Case #1	Case #2	Case #3	Case #4
β_1	-0.5	-0.505 (0.069)	-0.339 (0.176)	-1.042 (0.551)	-0.907 (0.415)
β_2	1	1.000 (0.020)	0.989 (0.023)	1.148 (0.151)	1.391 (0.392)
h_1	2	1.997 (0.039)	2.081 (0.107)	1.958 (0.086)	1.707 (0.295)
h_2	1	1.000 (0.001)	0.872 (0.128)	0.718 (0.282)	0.431 (0.569)
p_{11}	0.9	0.899 (0.012)	0.882 (0.023)	0.601 (0.303)	0.462 (0.439)
p_{22}	0.95	0.950 (0.006)	0.930 (0.022)	0.833 (0.119)	0.658 (0.293)

Note:

1. This table reports quasi maximum likelihood estimation results under different error distributions. Each cell contains the average of the 1,000 point estimates for each parameter and the root mean squared error of the estimates from the true value (in parentheses).
2. Case #1: normal distribution; Case #2: t-distribution; Case #3: χ^2 distribution; Case #4: mixture of 3 normals.

Table 1.2. Bayesian Inference of a Model under Normality Assumption [Log Difference of the U.S. Industrial Production Index, 1947M1-2017M1]

Parameter	Mean	SD	Median	90% HPDI
$\delta_{0,1}$	1.178	0.453	1.154	[0.505,1.961]
$\delta_{0,2}$	-0.540	0.344	-0.487	[-1.184,-0.096]
$\delta_{0,3}$	-0.321	0.201	-0.260	[-0.736,-0.099]
$\delta_{1,1}$	-1.049	0.394	-1.018	[-1.752,-0.455]
$\delta_{1,2}$	0.456	0.327	0.394	[0.042,1.087]
$\delta_{1,3}$	0.327	0.226	0.289	[0.033,0.765]
ϕ	0.257	0.063	0.259	[0.150,0.353]
h_1^2	0.493	0.025	0.492	[0.452,0.535]
g_2^2	1.911	0.270	1.908	[1.471,2.356]
g_3^2	2.282	0.383	2.256	[1.709,2.928]
$p_{s,00}$	0.886	0.120	0.940	[0.598,0.981]
$p_{s,11}$	0.939	0.063	0.953	[0.839,1.000]
$p_{w,11}$	0.961	0.023	0.966	[0.919,0.989]
$p_{w,12}$	0.027	0.030	0.022	[0.000,0.079]
$p_{w,21}$	0.098	0.077	0.086	[0.000,0.244]
$p_{w,22}$	0.851	0.087	0.867	[0.678,0.959]
$p_{w,32}$	0.228	0.114	0.216	[0.056,0.434]
$p_{w,33}$	0.745	0.104	0.757	[0.557,0.898]
$p_{c,11}$	0.980	0.023	0.988	[0.932,0.999]
$p_{c,22}$	0.993	0.014	0.997	[0.968,0.999]

Table 1.2. (Continued).

WAIC	1769
LPL	-480.7
JB	9.046 (0.008)

Note:

1. Out of 150,000 MCMC draws, the first 50,000 are discarded and inferences are based on the remaining 100,000 draws.
2. SD refers to standard deviation.
3. HPDI refers to a highest posterior density interval.
4. WAIC refers to the Watanabe-Akaike Information Criterion.
5. LPL refers to the log of predictive likelihood.
6. JB refers to the Jarque-Bera test statistic for a normality test. In the parenthesis is the p-value.

Table 1.3. Bayesian Inference of a Model under Student-t Distribution Assumption [Log Difference of the U.S. Industrial Production Index, 1947M1-2017M1]

Parameter	Mean	SD	Median	90% HPDI
$\delta_{0,1}$	1.130	0.465	1.106	[0.428,1.941]
$\delta_{0,2}$	-0.503	0.321	-0.457	[-1.100,-0.081]
$\delta_{0,3}$	-0.319	0.204	-0.258	[-0.739,-0.084]
$\delta_{1,1}$	-0.992	0.403	-0.962	[-1.716,-0.398]
$\delta_{1,2}$	0.442	0.313	0.391	[0.039,1.033]
$\delta_{1,3}$	0.307	0.227	0.262	[0.026,0.735]
ϕ	0.274	0.060	0.279	[0.167,0.365]
h_1^2	0.494	0.025	0.493	[0.454,0.536]
g_2^2	1.971	0.273	1.980	[1.512,2.406]
g_3^2	2.265	0.394	2.237	[1.689,2.920]
$p_{s,00}$	0.856	0.138	0.923	[0.558,0.989]
$p_{s,11}$	0.935	0.079	0.957	[0.789,1.000]
$p_{w,11}$	0.955	0.024	0.959	[0.914,0.987]
$p_{w,12}$	0.038	0.029	0.038	[0.0001,0.085]
$p_{w,21}$	0.117	0.073	0.106	[0.014,0.254]
$p_{w,22}$	0.826	0.086	0.839	[0.660,0.942]
$p_{w,32}$	0.245	0.112	0.235	[0.079,0.448]
$p_{w,33}$	0.742	0.108	0.752	[0.546,0.904]
$p_{c,11}$	0.980	0.022	0.988	[0.935,0.999]
$p_{c,22}$	0.992	0.016	0.997	[0.964,0.999]

Table 1.3. (Continued).

ν	293.9 (57.5)
WAIC	1759
LPL	-471.2
Accept Probability	0.376

Note:

1. Out of 150,000 MCMC draws, the first 50,000 are discarded and inferences are based on the remaining 100,000 draws.
2. SD refers to standard deviation.
3. HPDI refers to a highest posterior density interval.
4. ν refers the posterior average of degree of freedom, standard deviation is reported in parenthesis.
5. WAIC refers to the Watanabe-Akaike Information Criterion.
6. LPL refers to the log of predictive likelihood.
7. Accept probability refers to the acceptance probability of Metropolis Hasting algorithm for ν.

Table 1.4. Bayesian Inference of a Model with Unknown Error Distribution [Log Difference of the U.S. Industrial Production Index, 1947M1-2017M1]

Parameter	Mean	SD	Median	90% HPDI
$\delta_{0,1}$	0.971	0.498	0.743	[0.507,2.014]
$\delta_{0,2}$	-0.450	0.412	-0.262	[-1.350,-0.090]
$\delta_{0,3}$	-0.236	0.132	-0.216	[-0.478,-0.087]
$\delta_{1,1}$	-1.402	0.292	-1.365	[-1.965,-0.999]
$\delta_{1,2}$	0.536	0.287	0.545	[0.072,0.993]
$\delta_{1,3}$	0.413	0.248	0.373	[0.068,0.870]
ϕ	0.140	0.049	0.139	[0.062,0.223]
h_1	0.456	0.031	0.454	[0.407,0.509]
g_2	1.764	0.305	1.731	[1.336,2.322]
g_3	2.504	0.463	2.514	[1.686,3.256]
$p_{s,00}$	0.954	0.026	0.960	[0.911,0.981]
$p_{s,11}$	0.887	0.046	0.889	[0.807,0.961]
$p_{w,11}$	0.967	0.023	0.972	[0.923,0.992]
$p_{w,12}$	0.010	0.024	0.001	[0.000,0.065]
$p_{w,21}$	0.110	0.147	0.066	[0.000,0.408]
$p_{w,22}$	0.862	0.159	0.915	[0.530,0.967]
$p_{w,32}$	0.221	0.127	0.212	[0.001,0.439]
$p_{w,33}$	0.746	0.110	0.764	[0.549,0.886]
$p_{C,11}$	0.984	0.025	0.994	[0.931,0.999]
$p_{C,22}$	0.997	0.003	0.998	[0.991,0.999]

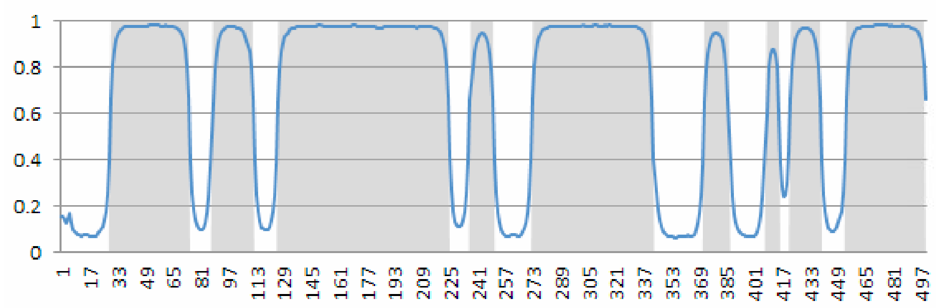
Table 1.4. (Continued).

M	3.349 (1.272)
WAIC	1549
LPL	-362.3
JB	1.218 (0.544)
Acceptance Probability 1	0.315
Acceptance Probability 2	0.474

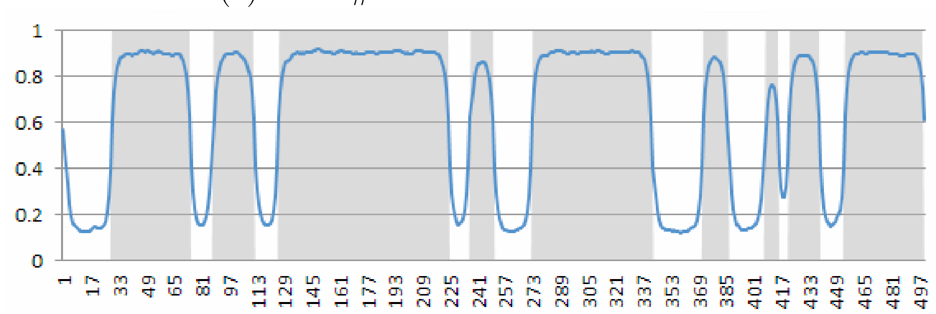
Notes:

1. Out of 150,000 MCMC draws, the first 50,000 are discarded and inferences are based on the remaining 100,000 draws.
2. SD refers to standard deviation.
3. HPDI refers to a highest posterior density interval.
4. M refers to the posterior average number of non-empty mixtures, standard deviation is reported in parenthesis.
5. WAIC refers to the Watanabe-Akaike Information Criterion.
6. LPL refers to the log of predictive likelihood.
7. JB refers to the Jarque-Bera test statistic for a normality test. In the parenthesis is the p-value.
8. Acceptance Probability 1 refers to the acceptance probability of Metropolis Hasting algorithm for h_1^2; Acceptance Probability 2 refers to the acceptance probability of the Metropolis Hasting algorithm for h_2^2.

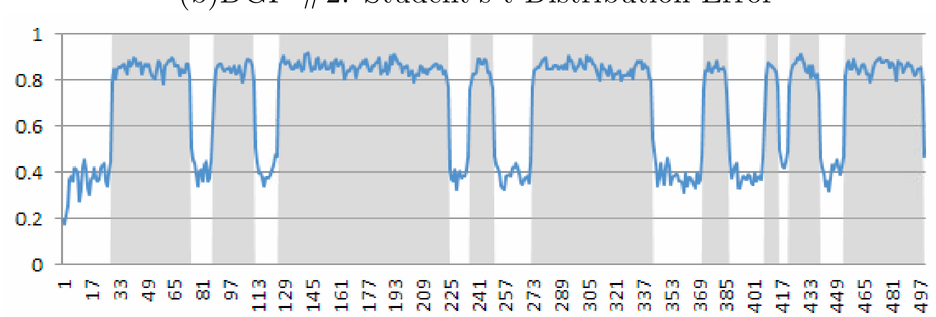
Figure 1.1. Smoothed Probabilities of Regime 2 based on Quasi-Maximum Likelihood Estimation under Different Error Distributions [T=500].



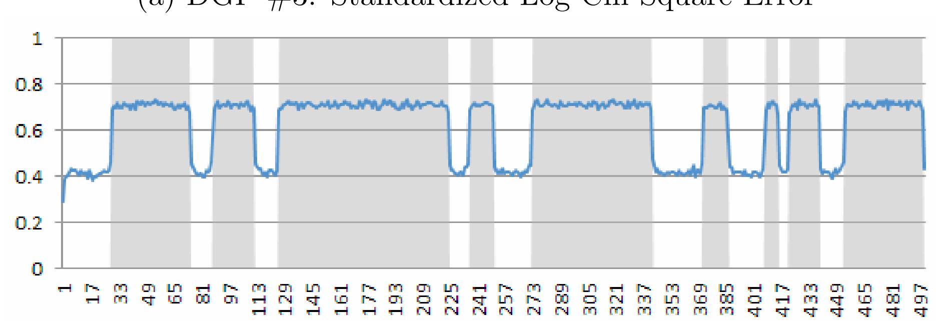
(a) DGP #1: Standard Normal Error



(b)DGP #2: Student's t Distribution Error



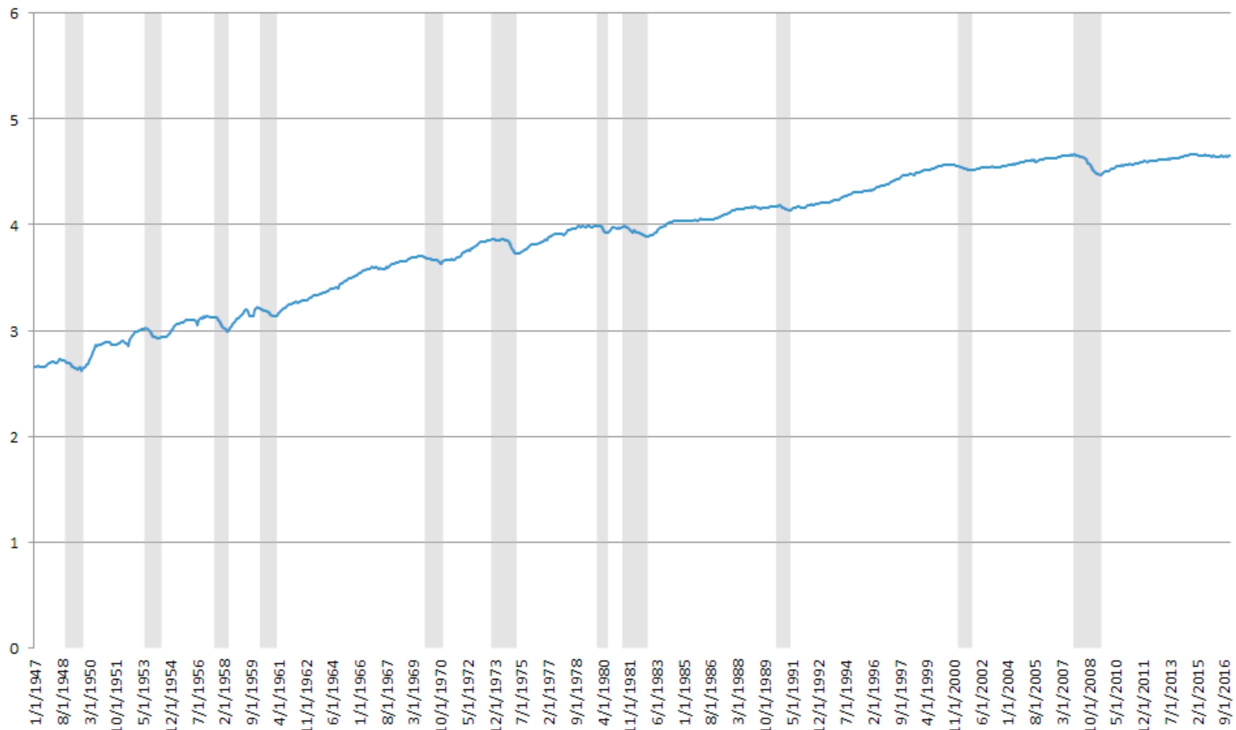
(a) DGP #3: Standardized Log Chi-Square Error



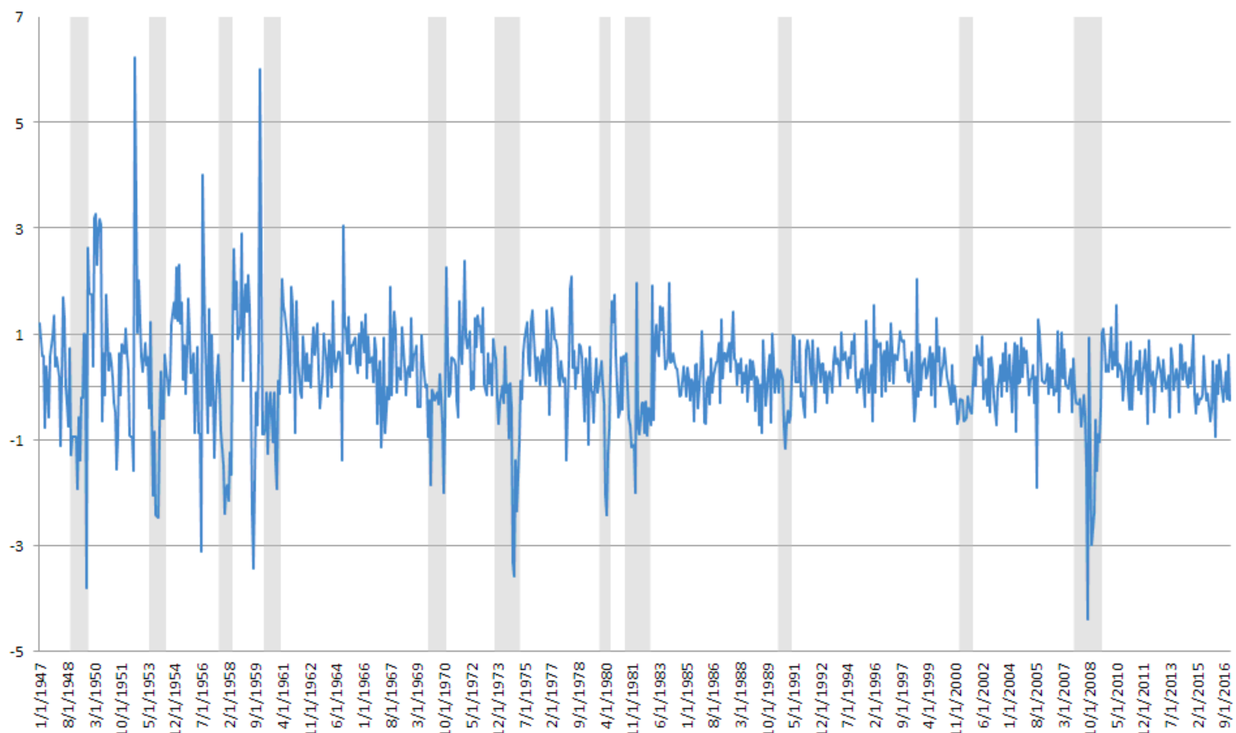
(b)DGP #4: Mixture Normal Error

Note: The shaded area denotes the data periods associated with regime 2.

Figure 1.2. U.S. Industrial Production (IP) Index and Its Growth Rate [1947M1 - 2017M1]



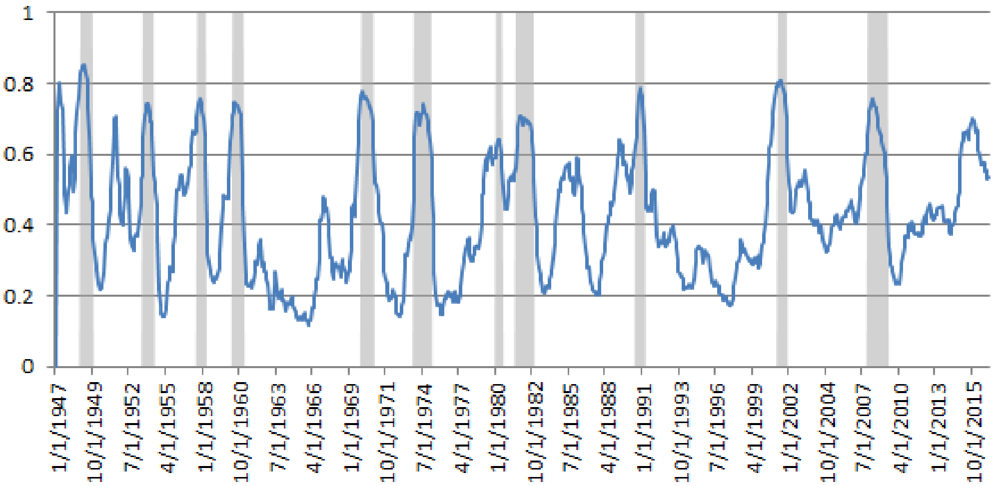
(a) Logarithm of U.S. Industrial Production (IP)



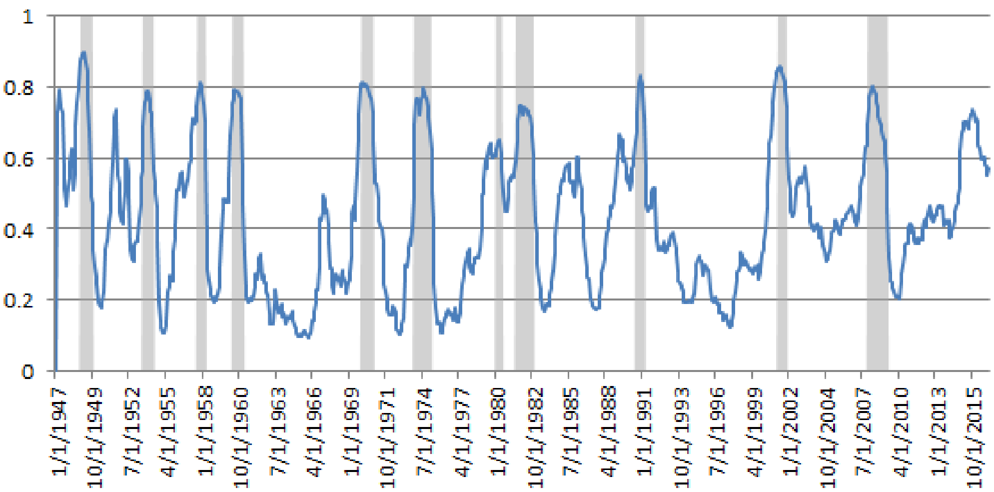
(b) IP growth

Note: The shaded area denotes the NBER recession date.

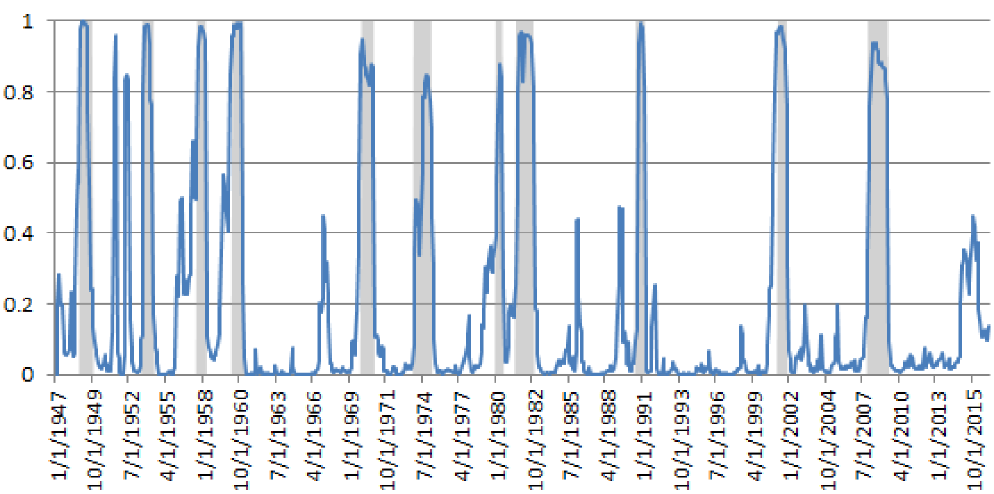
Figure 1.3. Posterior Probabilities of Recession



(a) For the Model with Normality Assumption.



(b) For the Model with Student-t Distribution Assumption.



(c) For the Model with Dirichlet Process Mixture of Normals.

Figure 1.4. Time-Varying Volatility for the IP Series: Proposed Model

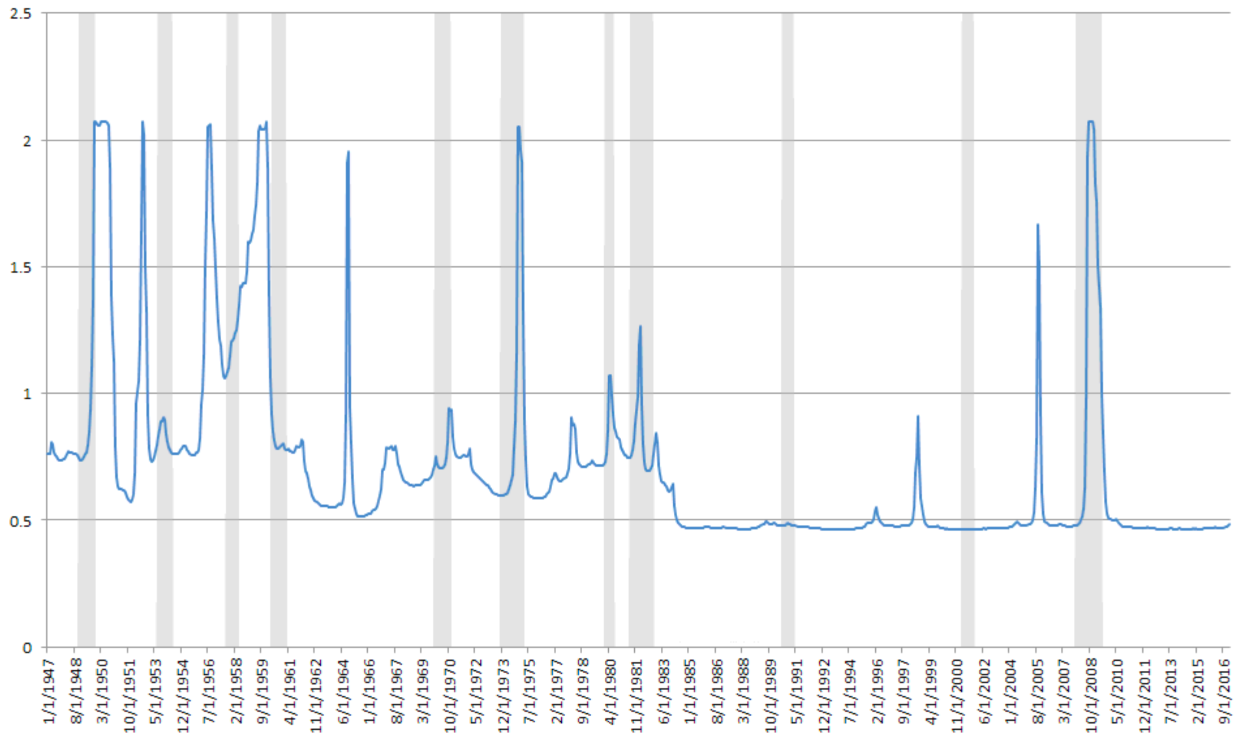


Figure 1.5. Time-varying Long-Run Mean Growth Rate: Proposed Model

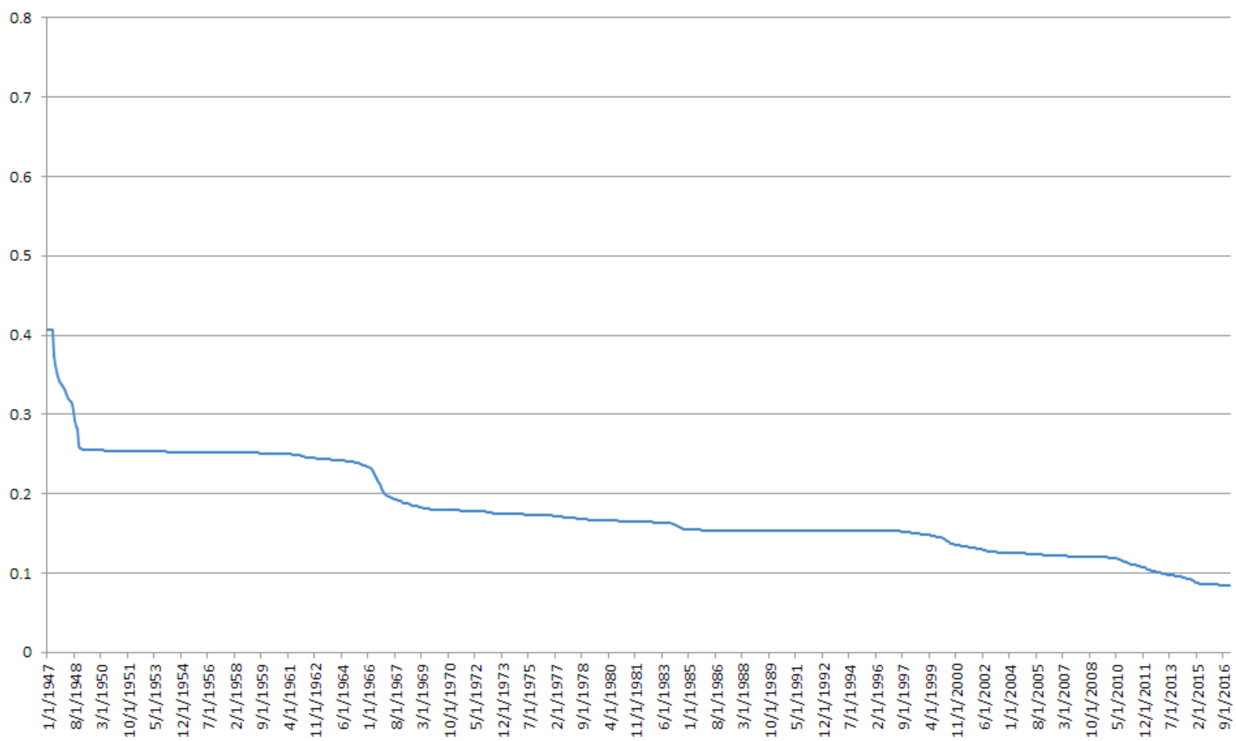
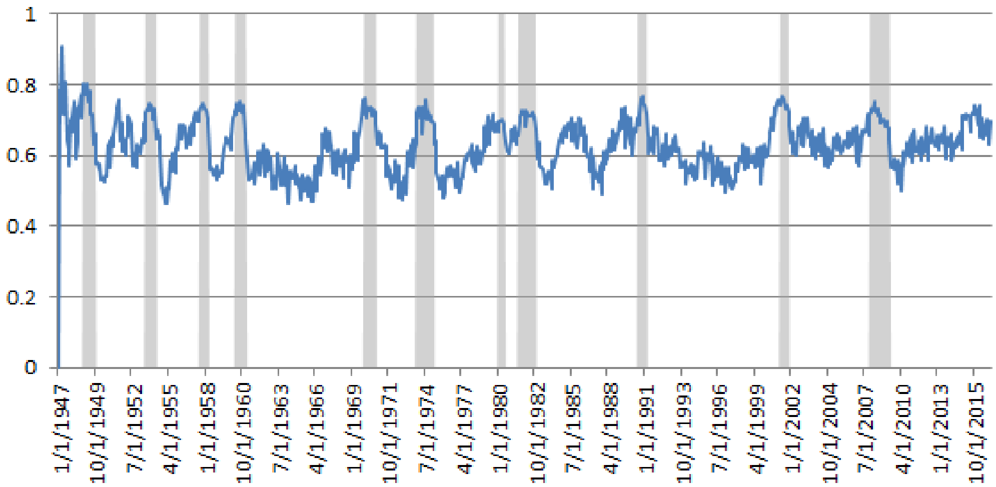
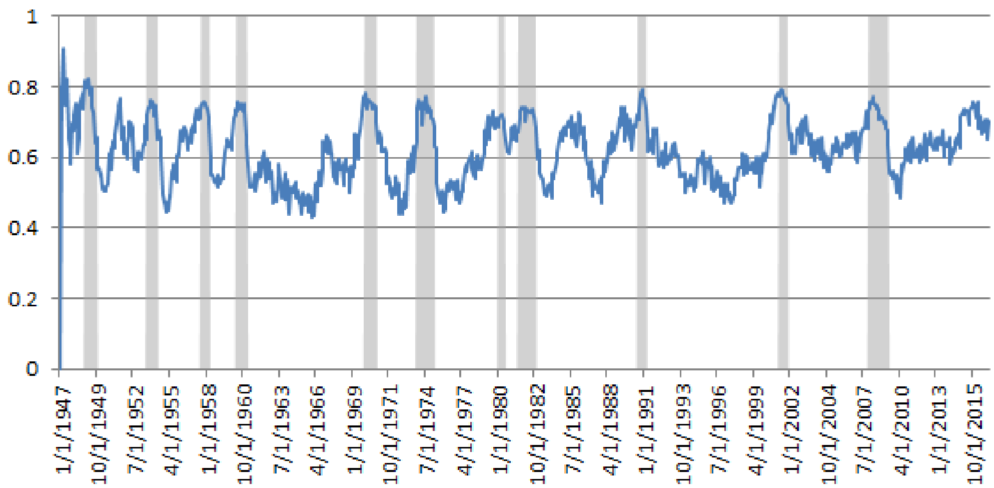


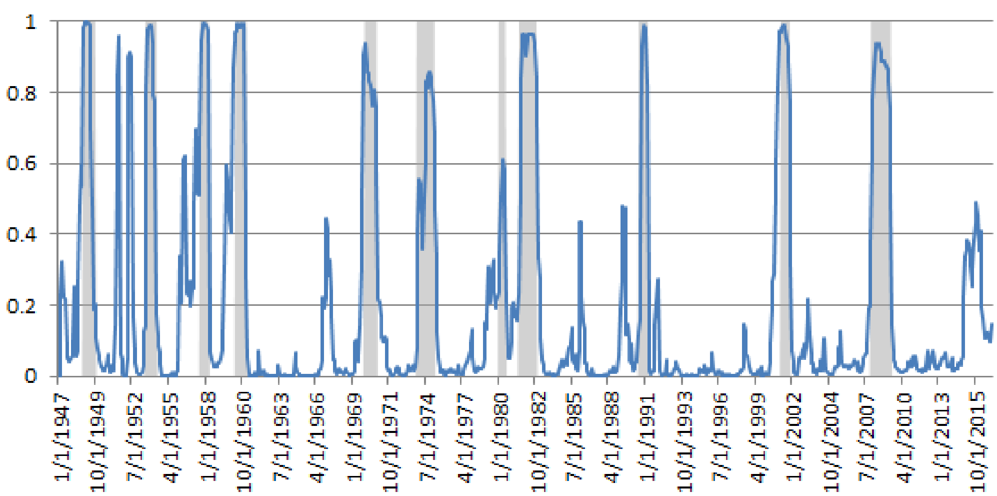
Figure 1.6. Posterior Probabilities of Recession: with Diffused Prior on Transition Probabilities



(a) For the Model with Normality Assumption.



(b) For the Model with Student-t Distribution Assumption.



(c) For the Model with Dirichlet Process Mixture of Normals.

Chapter 2

Estimating Trend Inflation Based on Unobserved Components Models: Is It Correlated with the Inflation Gap? [1]

2.1. Introduction

Trend inflation, which is usually defined as the long-run inflation expectation, is one of the key issues in both theoretical and empirical Macroeconomics. As the movements of trend inflation are sometimes attributed to shifts in monetary policy, modeling and estimating trend inflation is important in the study of monetary policy (see, for example, Bernanke (2007) and Mishkin (2007)). Furthermore, an accurate estimate of trend inflation can also serve as a useful centering point in constructing inflation forecasts at different horizons.

In their seminal work, Stock and Watson (2007) estimate trend inflation by employing a univariate unobserved components model, in which the volatilities of trend inflation and inflation gap (cyclical component of inflation) are subject to stochastic volatility processes. Under the assumption that trend inflation and inflation gap are not correlated with each other, they show that their estimated trend inflation is very volatile and it closely tracks the actual inflation. Correspondingly, they show that their estimate of the inflation gap, which is a priori assumed to be serially uncorrelated, is small in magnitude with relatively small volatility throughout the sample. Recently, Stock and Watson (2016) extend their earlier work on the univariate unobserved components model to a multivariate framework, in which the common persistent and the transitory factors are subject to stochastic volatility. They show that, even though uncertainty about trend inflation is substantially reduced by employing disaggregated data within a multivariate unobserved components model, the resulting estimate of trend inflation is similar to univariate estimate of trend inflation based on aggregate data.

[1] This chapter is based on a joint work with Chang-Jin Kim.

In the meantime, Cogley and Sbordone (2008) and Ascari and Sbordone (2014) estimate trend inflation by employing time-varying parameter vector autoregressive (VAR) models, in which the coefficients are assumed to follow random walk processes. [2] Unlike Stock and Watson (2007, 2016), however, they do not impose a priori assumption of uncorrelated trend inflation and inflation gap. Their empirical results are in contrast to those in Stock and Watson (2007, 2016), in the sense that their estimated trend inflation is much smoother than that reported in Stock and Watson (2007, 2016). Furthermore, unlike estimate of Stock and Watson (2007, 2016) which a priori assume serially uncorrelated inflation gap, their estimate of inflation gap suggests that there were more persistent deviations of inflation from trend in the 1960-1983 period than in the post-1983 period.

In this paper, we explicitly show that nonzero correlation between innovations to trend inflation and inflation gap is an important feature of the postwar U.S. inflation dynamics. [3] This is done within a univariate unobserved components model, in which the stochastic volatilities for trend inflation and the inflation gap in Stock and Watson's (2007) model are approximated by Markov-switching volatilities. These approximations allow us to easily incorporate nonzero correlation between innovations to the two unobserved components of inflation. Our model also allows us to easily incorporate other important features of postwar U.S. inflation not delivered in Stock and Watson's (2007) model. These additional features are: i) regime-switching inflation gap persistence and ii) association between inflation and inflation uncertainty.

Main findings of this paper can be summarized as follows. First, innovations to trend inflation and inflation gap are negatively correlated. Second, there were persistent deviations of inflation from its trend component in the 1970s, while the persistence in inflation gap almost disappears since the mid-1980s. Cogley et al. (2010) also find similar result within a time-varying parameter VAR framework. Third, there exist positive and statistically significant correlation between inflation and inflation uncertainty. Fourth, our estimated trend inflation is much smoother than that estimated by Stock and Watson (2007), and it has

[2] They define trend inflation as the long-run conditional expectation of inflation from their VAR model, in the spirit of Beveridge and Nelson (1981).

[3] As illustrate by Morley et al. (2003) in an unobserved components model for the log of real GDP, imposing a zero restriction on the correlation between the cyclical and the stochastic trend components could lead to misleading trend-cycle decompositions.

the same general pattern as the trend inflation estimated by Cogley and Sbordone (2008) or Ascari and Sbordone (2014). Lastly, but not the least important, our model results in smaller mean squared prediction errors than Stock and Watson's (2007) unobserved components model with stochastic volatility or the random walk model of Atkeson and Ohanian (2001) in pseudo out-of-sample forecast exercises.

The rest of this paper is organized as follows. In Section 2.2, we review important features of postwar U.S. inflation discussed in the literature. Then, we present a theoretical background for a nonzero correlation between innovations to trend inflation and inflation gap. In Section 2.3, we present a univariate unobserved components model of inflation with Markov-switching volatility, in which all the features of inflation reviewed in Section 2.2 are incorporated. Section 2.4 presents our empirical results. Section 2.5 compares the pseudo out-of-sample predictability of our model with those of Stock and Watson's (2007) model and Atkeson and Ohanian's (2001) random walk model. Section 2.6 concludes the paper.

2.2. Important Features of Postwar U.S. Inflation Dynamics

2.2.1. Review of Empirical Literature on Inflation Dynamics

Time-Varying Volatilities for Trend Inflation and Inflation Gap
(Stock and Watson, 2007)

Stock and Watson (2007) estimate the following unobserved components model of inflation with stochastic volatility (hereafter, UCSV model):

$$\pi_t = \tau_t + z_t, \tag{2.1}$$

$$\tau_t = \tau_{t-1} + \sigma_{\varepsilon,t}\varepsilon_t, \tag{2.2}$$

$$z_t = \sigma_{\eta,t}\eta_t, \tag{2.3}$$

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim i.i.d.N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), \tag{2.4}$$

where

$$\begin{aligned} \ln \sigma^2_{\eta,t} &= \ln \sigma^2_{\eta,t-1} + \nu_{\eta,t}, \quad \nu_{\eta,t} \sim i.i.d.N(0, \sigma^2_{\nu,\eta}), \\ \ln \sigma^2_{\varepsilon,t} &= \ln \sigma^2_{\varepsilon,t-1} + \nu_{\varepsilon,t}, \quad \nu_{\varepsilon,t} \sim i.i.d.N(0, \sigma^2_{\nu,\varepsilon}), \end{aligned} \tag{2.5}$$

where τ_t is trend inflation or the permanent component of inflation; z_t is the inflation gap or the transitory component of inflation; $\ln \sigma^2_{\eta,t}$ and $\ln \sigma^2_{\varepsilon,t}$ are the stochastic volatilities; and ε_t, η_t, $\nu_{\eta,t}$ and $\nu_{\varepsilon,t}$ are independent of one another. As replicated in Figure 2.1, they report empirical evidence of substantial changes over time in the volatility of trend inflation, while the volatility of the inflation gap is relatively small and stable. Furthermore, as shown in upper panel of Figure 2.2, their estimate of trend inflation closely tracks actual inflation, resulting in inflation gap which is small in magnitude and noisy throughout the whole sample period. Kim (1993), Kang et al. (2009), Stock and Watson (2016), and Mertens (2016) also demonstrate similar results within the unobserved components model framework, in which innovations to trend inflation and the inflation gap are assumed to be independent.

Time-Varying Inflation Gap Persistence

Consider the following generalization of the transitory inflation or inflation gap in equation (2.3):

$$\begin{aligned} z_t =& \phi_t z_{t-1} + \sigma_{\eta,t}\eta_t, \\ &|\phi_t| < 1, \end{aligned} \tag{2.6}$$

where the persistence of inflation gap ϕ_t potentially varies over time.

While Stock and Watson (2007, 2016) assume that inflation gap persistence (ϕ_t) is 0 for all t, researchers who estimate trend inflation and the inflation gap based on VAR models with time-varying parameters report evidence in favor of time-varying persistence of inflation gap. For example, Cogley and Sargent (2001, 2005), Cogley and Sbordone (2008), Cogley et al. (2010), and Ascari and Sbordone (2014) all report that ϕ_t was high until the mid-1980s and has decreased considerably since the mid-1980s. Furthermore, Conrad and Eife (2012) derive inflation gap persistence as a function of the policy weights in the central bank's Taylor rule, and they present empirical evidence showing a decline in the inflation gap persistence in the mid-1980s.

<u>*Association Between Inflation and Inflation Uncertainty*</u>

Friedman (1977) suggests that understanding the costs of inflation requires us to understand the link between inflation and its uncertainty. Following Friedman (1997), Ball (1992) theoretically shows that a rise in inflation can raise uncertainty about future inflation within a model of monetary policy, in which there exists information asymmetry and the public faces uncertainty about the preference of the policymaker. [4] Cukierman and Meltzer (1986) discuss the channel through which increased inflation uncertainty can have a positive effect on the level of inflation.

On the empirical side, Ball and Cecchetti (1990) show that the association between inflation and its uncertainty may differ between short- and long-run horizons. In particular, they show that a positive correlation between inflation and inflation uncertainty is more compelling as the horizon considered increases. Within an unobserved components model of U.S. postwar inflation, Kim (1993) finds evidence of a positive relationship between inflation and the uncertainty associated with trend inflation. Bhar and Hamori (2004) find the similar results for G7 countries. Bredin and Fountas (2006) also report positive relationship between inflation and its long-run uncertainty in European countries [5]. These empirical results suggest that equation (2.1) can be extended as follows:

$$\pi_t = \tau_t + z_t + f(\sigma^2_{\eta,t}, \sigma^2_{\epsilon,t}), \tag{2.7}$$

where $f(\sigma^2_{\eta,t}, \sigma^2_{\epsilon,t})$ is a function of variances for the innovations to trend inflation and the

[4] As Ball (1992) states, when inflation is high, for example, policymakers face a dilemma: they would like to disinflate, but fear the resulting recession.

[5] In the earlier literature that does not distinguish between long run and short run inflation, empirical results seem to be conflicting. For example, within the GARCH (Generalized Autoregressive Conditional Heteroscedasticity) framework, Engle (1983) and Cosimano and Jansen (1988) find little evidence of correlation between inflation and its uncertainty, while Grier and Perry (1998) find inflation has a significantly positive effect on its uncertainty in G7 countries.

inflation gap.

2.2.2 An Additional Issue to Be Investigated: Correlation Between Innovations to Trend Inflation and Inflation Gap

As in Woodford (2008) and Goodfriend and King (2012), let us consider the following New Keynesian Phillips Curves with time-varying trend inflation,

$$\pi_t - \tau_t = \beta E(\pi_{t+1} - \tau_{t+1}|I_t) + \kappa x_t + \zeta_t, \tag{2.8}$$

where τ_t is the trend inflation; β denotes a subjective discount factor; x_t is the output gap; and I_t refers to information up to t. Here, ζ_t denotes the mark-up shock, which is potentially serially correlated. It determines firms' pricing power in the New Keynesian literature (see, e.g. Steinsson (2003), and Ireland (2007)). By iterating equation (2.8) in forward direction and rearranging terms, we have:

$$\pi_t = \tau_t + \kappa \sum_{j=0}^{\infty} \beta^j E(x_{t+j}|I_t) + \sum_{j=0}^{\infty} \beta^j E(\zeta_{t+j}|I_t). \tag{2.9}$$

From equation (2.9), note that transitory inflation or the inflation gap is given by:

$$z_t = \kappa \sum_{j=0}^{\infty} \beta^j E(x_{t+j}|I_t) + \sum_{j=0}^{\infty} \beta^j E(\zeta_{t+j}|I_t), \tag{2.10}$$

where the second term on the right-hand-side represents the present value of conditional expectations of future markup shocks.

In the meantime, Bénabou (1992a,b) theoretically prove that trend inflation is negatively correlated with the markup shock. Bénabou's theoretical result has been empirically supported by researchers including Gali and Gertler (1999), Banerjee et al. (2001), and Head et al. (2010). These suggest that shocks to trend inflation and inflation gap may be negatively correlated. To investigate this possibility, we extend equation (2.4) in the following way:

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim i.i.d.N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \quad |\rho| < 1. \tag{2.11}$$

2.3. An Unobserved Components Model of Inflation with Markov-Switching and Correlated Shocks

2.3.1. Model Specification

In this section, we present an unobserved components model of inflation, in which we incorporate all the features discussed in Section 2.2. To begin with, we consider the following model with i) a nonzero correlation between innovations to trend inflation (τ_t) and inflation gap (z_t) and ii) time-varying persistence (ϕ_t) of inflation gap:

$$\begin{aligned} \pi_t &= \tau_t + z_t, \\ \tau_t &= \tau_{t-1} + \sigma_{\varepsilon,t}\varepsilon_t, \\ z_t &= \phi_t z_{t-1} + \sigma_{\eta,t}\eta_t, \end{aligned} \tag{2.12}$$

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim i.i.d.N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$

We assume that time-varying persistence of inflation gap (ϕ_t) and the volatility processes ($\sigma_{\varepsilon,t}$ and $\sigma_{\eta,t}$) for the postwar U.S. data can be approximated by independent Markov-switching processes, as described below:

$$\begin{aligned} \sigma_{\varepsilon,t} &= Q_0 + Q_1 S_{1,t}, \quad Q_0, Q_1 > 0, \\ \sigma_{\eta,t} &= h_0 + h_1 S_{2,t}, \quad h_0, h_1 > 0, \\ \phi_t &= (1 - S_{3,t})\phi_0 + S_{3,t}\phi_1, \quad |\phi_0|, |\phi_1| < 1, \end{aligned} \tag{2.13}$$

where each of the discrete latent variables $S_{1,t}$, $S_{2,t}$ and $S_{3,t}$ evolves according to a two-state first-order Markov-switching process with the following transition probabilities:

$$\begin{aligned} Pr[S_{1,t} = i | S_{1,t-1} = j] &= p_{1,ij}, \quad i, j = 0, 1, \\ Pr[S_{2,t} = i | S_{2,t-1} = j] &= p_{2,ij}, \quad i, j = 0, 1, \\ Pr[S_{3,t} = i | S_{3,t-1} = j] &= p_{3,ij}, \quad i, j = 0, 1. \end{aligned} \tag{2.14}$$

Note that the above model with $\rho = 0$ and $\phi_t = 0$ for all t is directly comparable to Stock and Watson's (2007) unobserved components stochastic volatility (UCSV) model given in equations (2.1)-(2.5). As will be shown in Section 2.4, when we estimate both our model with these restrictions and Stock and Watson's (2007) UCSV model using the postwar data, estimated volatilities and estimated trend inflation from these two models are qualitatively very similar. However, unlike the UCSV model, the above model with Markov-switching volatility allows us to identify the correlation (ρ) between innovations to trend inflation and inflation gap. [6] Furthermore, our model also allows us to easily incorporate the other features of inflation dynamics reviewed in Section 2.1.

Our most general model incorporates potentially nonzero relationship between inflation and its uncertainty. In line with equation (2.7), this is done by replacing the first equation of the model presented in equation (2.12) by: [7]

$$\pi_t = \tau_t + z_t + \alpha_1 S_{1,t} + \alpha_2 S_{2,t} + \alpha_3 S_{1,t} S_{2,t}, \tag{2.15}$$

where $f(\sigma^2_{\eta,t}, \sigma^2_{\epsilon,t})$ in equation (2.7) is replaced by $\alpha_1 S_{1,t} + \alpha_2 S_{2,t} + \alpha_3 S_{1,t} S_{2,t}$, following Kim (1993). In our model, the variance of inflation has 4 regimes: i) when $S_{1,t} = 0$ and $S_{2,t} = 0$, we are in regime 1; ii) when $S_{1,t} = 0$ and $S_{2,t} = 1$, we are in regime 2; i) when $S_{1,t} = 1$ and $S_{2,t} = 0$, we are in regime 3; and i) when $S_{1,t} = 1$ and $S_{2,t} = 1$, we are in regime 4. Thus, α_1 captures a shift in the level of inflation during regime 2 relative to regime 1; α_2 captures a shift in the level of inflation during regime 3 relative to regime 1; and $\alpha_1 + \alpha_2 + \alpha_3$ captures a shift in the level of inflation during regime 4 relative to regime 1.

2.3.2. Estimation of the Model

In order to estimate our model using the maximum likelihood estimation, we first cast it to the following state-space model:

Measurement Equation

[6] For identification of the model, readers are referred to Appendix 2.A, where Morley et al.'s (2003) results on identifiability of unobserved components models with homoscedastic shocks are extended to the case of Markov-switching variances.

[7] Kim (1993) considers a similar unobserved components model with Markov-switching variances. However, he assumes that $\phi_t = 0$ for all t and that $\rho = 0$, as in Stock and Watson (2007).

$$\pi_t = [\,1 \quad 1\,] \begin{bmatrix} \tau_t \\ z_t \end{bmatrix} + \alpha_1 S_{1,t} + \alpha_2 S_{2,t} + \alpha_3 S_{1,t} S_{2,t}, \tag{2.16}$$

$$\left(\ \pi_t = H\beta_t + \alpha_1 S_{1,t} + \alpha_2 S_{2,t} + \alpha_3 S_{1,t} S_{2,t}\ \right).$$

<u>Transition Equation</u>

$$\begin{bmatrix} \tau_t \\ z_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \phi_{S_{3,t}} \end{bmatrix} \begin{bmatrix} \tau_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \sigma_{\varepsilon,S_{1,t}} & 0 \\ 0 & \sigma_{\eta,S_{2t}} \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}, \qquad \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim i.i.d.N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \tag{2.17}$$

$$\Big(\ \beta_t = F_{S_{3,t}} \beta_{t-1} + R_{S_{1,t},S_{2,t}} V_t, \quad V_t \sim i.i.d.N(0,\Omega)\ \Big).$$

The model is estimated via the maximum likelihood estimation method based on Kim's (1994) approximate Kalman filter. For details of the maximum likelihood estimation procedure, readers are referred to Appendix 2.B.

2.4. Empirical Results

We employ quarterly data on the personal consumption expenditure deflator for core items (PCE-core). The inflation rate is calculated as annualized quarterly percentage change in price index. To demonstrate the importance of each feature of inflation discussed in Section 2.2, we consider the following 4 versions of our model that differ in the assumptions employed,

<u>**Model 1**</u>: $\rho = 0$, $\phi_t = 0$ for all t, $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

<u>**Model 2**</u>: $\phi_t = 0$ for all t, $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

<u>**Model 3**</u>: $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

<u>**Model 4**</u>: Unrestricted Model.

Note that Model 1 is directly comparable to Stock and Watson's (2007) UCSV model. Model 2 allows for potential nonzero correlation between innovations to trend inflation and the inflation gap in Model 1. Model 3 incorporates the time-varying inflation gap persistence into

Model 2. Model 4 is our most general model and it incorporates potential nonzero relationship between the level of inflation and uncertainty associated with the two components of inflation.

Table 2.1 reports estimation results for the above four models. For Model 1, estimates of the transition probabilities show that the volatility of trend inflation is more persistent ($\hat{p}_{1,00} + \hat{p}_{1,11} - 1 = 0.956$) than that of the inflation gap ($\hat{p}_{2,00} + \hat{p}_{2,11} - 1 = 0.844$), and this pattern is maintained for the other models as well. In Figure 2.1, we depict the volatilities of trend inflation and the inflation gap estimated from Stock and Watson's (2007) model with stochastic volatility as well as those from our benchmark model (Model 1) with Markov-switching volatility. Note that volatilities estimated from our benchmark model reasonably well approximate those from Stock and Watson's (2007) model. Furthermore, as depicted in Figure 2.2, the estimates of trend inflation obtained from these two models are very similar. These results provide us with justifications for replacing stochastic volatility in Stock and Watson's (2007) model with Markov-switching volatilities. An advantage of employing Markov-switching volatilities in an unobserved component model of inflation is that it allows us to easily incorporate important features of inflation discussed in Section 2.2.

Focusing on estimates of the correlation (ρ) between innovations to trend inflation and the inflation gap for Models 2, 3, and 4, they range between -0.684 and -0.907, and they are all statistically significant at the 1% level. These results shed light on potential drawback of Stock and Watson's (2007) UCSV model, in which ρ is constrained to be 0. In Figure 2.3, we depict estimates of trend inflation volatility and the inflation gap volatility from Model 2. The volatilities of both trend inflation and inflation gap are in general estimated to be higher than those from Model 1. However, an increase in the inflation gap volatility during the period covering mid-1970s$\sim$mid-1980s is much more pronounced for Model 2. As shown in Figure 2.4, trend inflation estimated from Model 2 does not seem to be very different from that from Model 1. That is, estimation of trend inflation seems to remain almost intact even when ρ is estimated to be negative and statistically significant.

A negative correlation between trend inflation and inflation gap is consistent with the literature of the relationship between trend inflation and markup. Intuitively, assuming trend inflation and inflation gap are uncorrelated is equivalent to assuming inflation gap is merely an exogenous shock to inflation.

Concerning the estimation of the inflation gap persistence (ϕ_t) from Models 3 and 4, they exhibit two distinct regimes. As depicted in Figure 2.5, both models exhibit high inflation gap persistence (with $\hat{\phi}_0 = 0.879$ for Model 3 and 0.868 for Model 4) for the period covering mid-1960s∼mid 1980s and almost zero inflation gap persistence for the rest of the period. These patterns of time-varying inflation gap persistence from Model 3 or 4 are very close to those in Cogley et al. (2010) and Ascari and Sbordone (2014), whose results are based on VAR models with time-varying parameters.

In order to investigate the effects of incorporating time-varying inflation gap persistence on volatility measures, we depict trend inflation volatilities and the inflation gap volatilities from Model 3 and Model 4 in Figure 2.6. Volatility measures from these two models have very similar patterns. However, they are very different from those from Model 1 or 2. For Models 3 or 4, the volatility of trend inflation is estimated to be much lower than that from Model 1 or 2. Furthermore, unlike for Model 1 or 2, high and volatile inflation during the 1970s is associated with a surge in the volatility of inflation gap for Model 3 or 4. These results suggest that high and volatile inflation in the 1970s is in large part due to transitory shocks, which may be a monetary phenomenon. This result is consistent with Clarida et al. (2000), who suggest that high and volatile inflation combined with volatile economic activity in the 1970s were engendered by loose monetary policy. [8]

Figure 2.7 depicts estimates of trend inflation from Models 3 and 4. Trend inflation estimated from Model 3 or 4 is very different from that estimated from Model 1, Model 2, or Stock and Watson's (2007) UCSV model. In particular, it is much smoother than the actual inflation rate, while trend inflation from our benchmark model (Model 1) or Stock and Watson's (2007) model depicted in Figure 2.2 very closely tracks actual inflation. Our results based on Model 3 or 4 are in line with Cogley and Sbordone (2008) and Ascari and Sbordone (2014), who also report estimates of trend inflation which are much smoother than the actual inflation rate, based on VAR models with time-varying parameters.

Our discussion so far seems to suggest that the smoothness of trend inflation from Model

[8] Barsky and Kilian (2002, 2004), De Long (1997), Friedman (1975), Mckinnon (1982), and Kilian (2008), among others, also suggest that high and volatile inflation rate in the 1970s may not have been related to permanent shocks.

3 or 4 results mainly from allowing for time-varying inflation gap persistence. However, as depicted in Figure 2.8, when we estimate trend inflation based on Model 3 or 4 with the constraint $\rho = 0$, it turns out to be very close to that from Model 1. Furthermore, as shown in Figure 2.9, the inflation gap persistence was estimated to be close to zero throughout the sample for Model 3. For Model 4, it was estimated to be considerably lower than when ρ was not constrained to be 0. Thus, conditional on $\rho = 0$, allowing for time-varying inflation gap persistence does seem to be redundant. We conjecture that this is the reason why Stock and Watson (2007) and Kim (1993) specified the transitory component of inflation as serially uncorrelated in their unobserved components models of inflation, under the maintained hypothesis $\rho = 0$. We conclude that time-varying inflation gap persistence plays a meaningful role in the estimation of trend inflation, only when it is combined with nonzero correlationship between innovations to trend inflation and the inflation gap.

The last column of Table 2.1 reports estimation results for our most general model, which incorporates nonzero association between inflation uncertainty and the level of inflation in Model 3. The estimate for the α_1 parameter, which measures the relationship between inflation and its uncertainty associated with trend inflation (i.e., long-run uncertainty), is positive and statistically significant at the 1% significance level. The estimate for the α_2 parameter, which measures the relation between inflation and its uncertainty associated with inflation gap (i.e., short-run uncertainty), is statistically significant at the 10% significance level. We conclude that a positive association between inflation and its long-run uncertainty is much more compelling than that between inflation and its short-run uncertainty. These results are in line with earlier empirical evidence presented by Ball and Cecchetti (1990), Kim (1993), and Bhar and Hamori (2004), among others. As depicted in the lower panel of Figure 2.7, allowing for association between the level of inflation and the volatilities of the two components of inflation results in smoother trend inflation than that from Model 3.

2.5. Pseudo Out-of-Sample Forecasting Performance

In this section, we evaluate the pseudo out-of-sample forecasting performance of our models against Stock and Watson's (2007) UCSV model and Atkeson and Ohanian's (2001) random walk model (AO model, hereafter). Following Stock and Watson (2007), we define the inflation over the next h quarters as

$$\pi_{t+h}^{h} \equiv \sum_{i=t+1}^{t+h} \pi_i / h,$$

and the forecasting horizon we consider are: $h = 4, 8, 12$ quarters.

For forecasts made at date t, estimation of each model was performed using only data available through date t. The forecasts are recursive, so that forecasts at date t are based on all the data from 1959Q1 through date t. The first forecasts are based on model estimates obtained using data from 1959Q1 through 1989Q4.

For the random-walk model of Atkeson and Ohanian (2001), the average h-quarter rate of inflation is forecasted as the average of inflation over the previous h quarters. The forecast error for the AO model is given by:

<u>*AO Model*</u>

$$u_{t+h|t}^{AO} = \pi_{t+h}^{h} - \frac{1}{h}(\pi_t + \ldots + \pi_{t-h}).$$

For Stock and Watson's (2007) UCSV model and for Models 1 and 2 represented in this paper, inflation forecasts are purely determined by the random walk component. Thus, forecast errors for these models can be calculated as:

<u>*UCSV Model*</u>

$$u_{t+h|t}^{SW} = \pi_{t+h}^{h} - \tau_{t|t}$$

<u>*Models 1 and 2*</u>

$$u_{t+h|t}^{j} = \pi_{t+h}^{h} - \tau_{t|t}, \quad j = 1, 2,$$

where $\tau_{t|t}$ denotes the filtered estimate of the trend inflation conditional on information up to t. For Models 3 and 4 of ours, forecast errors can be calculated as:

<u>*Models 3 and 4*</u>

$$u_{t+h|t}^{j} = \pi_{t+h}^{h} - \frac{1}{h}\sum_{k=1}^{h} \Big(HF^k \beta_{t|t} + E(\alpha_1 S_{1,t} + \alpha_2 S_{2,t} + \alpha_3 S_{1,t} S_{2,t} | I_t) \Big), \quad j = 3, 4,$$

where H and F are defined in equations (2.16) and (2.17), and $\beta_{t|t}$ refers to the expectation of β_t $(= [\,\tau_t \quad z_t\,]')$ conditional on information up to date t.

Table 2.2 reports root mean squared errors (RMSEs) obtained from each model. The results can be summarized as follows. First, Stock and Watson's (2007) UCSV model has smaller RMSEs than Atkeson and Ohanian's (2001) random-walk model or our benchmark model (Model 1) for all the forecasting horizons considered. Second, however, RMSE's from Model 2, 3, or 4 in this paper are in general smaller than those from the AO model, the UCSV model, or Model 1. These results suggest that allowing for nonzero correlation (ρ) between innovations to trend inflation and inflation gap improves pseudo out-of-sample forecasting performance.

Table 2.3 reports the Diebold-Mariano test statistics for testing the null hypothesis that mean squared errors for two competing models are the same. Mean squared errors from Model 2 or 3 are in general smaller than those from the AO model, the UCSV model, or Model 1, even though the differences are not statistically significant. However, our most general model (Model 4) beats the UCSV model and the random walk model at the 10% significance level for forecasting horizons 4 and 8 quarters. Furthermore, it outperforms our benchmark model (Model 1) at the 1% significance level for forecasting horizons 4 and 8 quarters.

Our results in this section suggest that nonzero correlation between trend inflation and inflation gap plays an important role in improving the out-of-sample forecasting performance when it is combined with i) time-varying inflation gap persistence and ii) the association between inflation uncertainty and the level of inflation.

2.6. Conclusion

Stock and Watson (2007) propose and estimate an unobserved components model with stochastic volatility, in which they assume i) inflation gap is serially uncorrelated; and ii) innovations to trend inflation and inflation gap are uncorrelated. Trend inflation estimated from their model is volatile and it closely tracks inflation throughout the sample, leaving no room for persistent deviations of inflation from this trend. On the contrary, trend inflation estimated from a time-varying parameter VAR model (e.g., Cogley and Sbordone, 2008, or Ascari and Sbordone, 2014) is smooth and inflation deviates persistently from this trend during the Great inflation period. This paper provides some insights on why the two alternative measures of trend inflation are so different.

Within the framework of a univariate unobserved components model with Markov-switching volatility, this paper empirically shows that a negative correlation between innovations to trend inflation and the inflation gap, when it is combined with time-varying inflation gap persistence, plays an important role in the dynamics of postwar US inflation. This provides indirect evidence supporting Bénabou (1992a,b), Gali and Gertler (1999), Banerjee et al. (2001), Head et al. (2010), and Head et al. (2010), who suggest the markup shock, which is positively correlated with the inflation gap, is negatively correlated with trend inflation. When association between the inflation and its uncertainty is additionally incorporated into the model, trend inflation is estimated to be much smoother than that in Stock and Watson (2007). Besides, our model provides superior out-of-sample forecasts than Stock and Watson's (2007) unobserved components model with stochastic volatility or Atkeson and Ohanian's (2001) random walk model.

Appendix 2.A. Identification of the Unobserved Components Model with Markov-Switching Heteroscedasticity

By Granger's Lemma (Granger and Newbold, 1986), our model in equations (2.12)-(2.13) can be rewritten as the following reduced-form ARIMA(1,1,1) model:

$$\begin{aligned}(1-\phi_{S_{3,t}}L)\Delta\pi_t = &(1-\theta_{S_{1,t},S_{2,t},S_{3,t}}L)e_t \\ &+(1-\phi_{S_{3,t}}L)(1-L)(\alpha_1 S_{1,t}+\alpha_2 S_{2,t}+\alpha_3 S_{1,t}S_{2,t}), \\ &e_t|S_{1,t},S_{2,t} \sim N(0,\sigma^2_{e,S_{1,t},S_{2,t},S_{3,t}}),\end{aligned} \qquad (2.A.1)$$

where $\Delta\pi_t = \pi_t - \pi_{t-1}$. By matching the variance and autocovariance in the structural model in equations (2.12)-(2.13) and the reduced-form model in equation (2.A.1), we have the following equations

$$\begin{aligned}&\sigma^2_{e,S_{1,t},S_{2,t},S_{3,t}} + \theta^2_{S_{1,t-1},S_{2,t-1},S_{3,t-1}}\sigma^2_{e,S_{1,t-1},S_{2,t-1},S_{3,t-1}} \\ =&\sigma^2_{\varepsilon,S_{1,t}} + \phi^2_{S_{3,t-1}}\sigma^2_{\varepsilon,S_{1,t-1}} + \sigma^2_{\eta,S_{2,t}} + \sigma^2_{\eta,S_{2,t-1}} + 2\rho\left(\sigma_{\varepsilon,S_{1,t}}\sigma_{\eta,S_{2,t}} + 2\phi_{S_{3,t-1}}\sigma_{\varepsilon,S_{1,t-1}}\sigma_{\eta,S_{2,t-1}}\right), \\ &\theta_{S_{1,t-1},S_{2,t-1},S_{3,t-1}}\sigma^2_{e,S_{1,t-1},S_{2,t-1},S_{3,t-1}} \\ =&\phi_{S_{3,t-1}}\sigma^2_{\varepsilon,S_{1,t-1}} - \sigma^2_{\eta,S_{2,t-1}} - \left(1+\phi_{S_{3,t-1}}\right)\rho\sigma_{\varepsilon,S_{1,t-1}}\sigma_{\eta,S_{2,t-1}}\end{aligned}$$

Our identification strategy exploits the fact that the reduced-form parameters can be expressed as functions of the structure parameters and that the number of reduced-form parameters increases exponentially as the number of regimes for the volatility processes increase. [9] It is easy to see that we have more moment conditions above than the number of parameters in the covariance matrix of structural model, thus the structural parameters are identified. With the same logic, one can show that all the models (Models 1,2,3, and 4) considered in Section 2.4 are identified.

Appendix 2.B. Maximum Likelihood Estimation of the Model

We rewrite the state-space model in section 2.3 in the following way:

[9] Rigobon (2003) discusses a related identification strategy in a simultaneous equations model.

<u>*Measurement Equation*</u>

$$\pi_t = H\beta_t + \mu_{S_t}, \tag{2.B.1}$$

<u>*Transition Equation*</u>

$$\beta_t = F_{S_t}\beta_{t-1} + R_{S_t}V_t, \quad V_t \sim N(0, \Omega), \tag{2.B.2}$$

where $\mu_{S_t} = \alpha_1 S_{1,t} + \alpha_2 S_{2,t} + \alpha_3 S_{1,t}S_{2,t}$; $F_{S_t} = F_{S_{3,t}}$; $R_{S_t} = R_{S_{1,t},S_{2,t}}$; and

$$S_t = \begin{cases} 1,\ if & S_{1,t} = 0, S_{2,t} = 0, S_{3,t} = 0 \\ 2,\ if & S_{1,t} = 0, S_{2,t} = 0, S_{3,t} = 1 \\ 3,\ if & S_{1,t} = 0, S_{2,t} = 1, S_{3,t} = 0 \\ 4,\ if & S_{1,t} = 0, S_{2,t} = 1, S_{3,t} = 1 \\ 5,\ if & S_{1,t} = 1, S_{2,t} = 0, S_{3,t} = 0 \\ 6,\ if & S_{1,t} = 1, S_{2,t} = 0, S_{3,t} = 1 \\ 7,\ if & S_{1,t} = 1, S_{2,t} = 1, S_{3,t} = 0 \\ 8,\ if & S_{1,t} = 1, S_{2,t} = 1, S_{3,t} = 1. \end{cases}$$

As $S_{1,t}$, $S_{2,t}$, and $S_{3,t}$ are independent of one another, the 8×8 matrix of transition probabilities for S_t is given by:

$$\widetilde{P} = \widetilde{P}_1 \otimes \widetilde{P}_2 \otimes \widetilde{P}_3, \tag{2.B.3}$$

where the (i, j)th element of $\widetilde{P}$ is the probability of $S_t = j$ conditional on $S_{t-1} = i$; $\otimes$ denote the Kroncker product; and

$$\widetilde{P}_1 = \begin{bmatrix} p_{1,00} & p_{1,10} \\ p_{1,01} & p_{1,11} \end{bmatrix}, \quad \widetilde{P}_2 = \begin{bmatrix} p_{2,00} & p_{2,10} \\ p_{2,01} & p_{2,11} \end{bmatrix}, \quad \widetilde{P}_3 = \begin{bmatrix} p_{3,00} & p_{3,10} \\ p_{3,01} & p_{3,11} \end{bmatrix}.$$

Conditional on $S_{t-1} = i$ and $S_t = j$, $i, j = 1, 2, \ldots, 8$, the Kalman filter is given by:

$$\beta^{(i,j)}_{t|t-1} = F_j \beta^{(i)}_{t-1|t-1}, \tag{2.B.4}$$

$$P^{(i,j)}_{t|t-1} = F_j P^{(i)}_{t-1|t-1} F'_j + R_j Q_j R'_j, \tag{2.B.5}$$

$$\eta^{(i,j)}_{t|t-1} = \pi_t - H\beta^{(i,j)}_{t-1|t-1} - \mu_j, \tag{2.B.6}$$

$$f^{(i,j)}_{t|t-1} = HP^{(i,j)}_{t|t-1}H', \tag{2.B.7}$$

$$\beta^{(i,j)}_{t|t} = \beta^{(i,j)}_{t|t-1} + P^{(i,j)}_{t|t-1}H'\left[f^{(i,j)}_{t|t-1}\right]^{-1}\eta^{(i,j)}_{t|t-1}, \tag{2.B.8}$$

$$P^{(i,j)}_{t|t} = \left\{I_2 - P^{(i,j)}_{t|t-1}H'\left[f^{(i,j)}_{t|t-1}\right]^{-1}H\right\}P^{(i,j)}_{t|t-1}, \tag{2.B.9}$$

where $\beta^{(i,j)}_{t-1|t-1}$ is the expectation of β_{t-1} conditional on $S_{t-1} = i$, $S_t = j$, and information up to $t-1$ (I_{t-1}); $P^{(i,j)}_{t|t-1}$ is the variance covariance matrix of β_t conditional on $S_{t-1} = i$, $S_t = j$, and I_{t-1}; $\eta^{(i,j)}_{t|t-1}$ is the prediction error of π_t conditional on $S_{t-1} = i$, $S_t = j$, and I_{t-1}; and $f^{(i,j)}_{t|t-1}$ is the variance π_t conditional on $S_{t-1} = i$ and $S_t = j$, and I_{t-1}.

Each iteration of the above Kalman filter produces an 8-fold increase in the number of cases to consider. It is necessary to introduce some approximations to make the above Kalman filter operable. We follow Kim (1994) in 'collapsing' the 8×8 posterior moments ($\beta^{(i,j)}_{t|t}$ and $P^{(i,j)}_{t|t}$) into 8×1 posterior moments ($\beta^{(j)}_{t|t}$ and $P^{(j)}_{t|t}$), as given below:

$$\beta^{(j)}_{t|t} = \sum_{i=1}^{8} \frac{Pr(S_t = j, S_{t-1} = i|I_t)}{Pr(S_t = j|I_t)}\beta^{i,j}_{t|t}, \tag{2.B.10}$$

$$P^{(j)}_{t|t} = \sum_{i=1}^{8} \frac{Pr(S_t = j, S_{t-1} = i|I_t)}{Pr(S_t = j|I_t)}\left[P^{i,j}_{t|t} + (\beta^{j}_{t|t} - \beta^{i,j}_{t|t})(\beta^{j}_{t|t} - \beta^{i,j}_{t|t})'\right], \tag{2.B.11}$$

$$j = 1, 2, ..., 8,$$

where $\beta^{(j)}_{t|t}$, for example, is a weighted average of $\beta^{(i,j)}_{t|t}$, with the weights being $\frac{Pr(S_t=j,S_{t-1}=i|I_t)}{Pr(S_t=j|I_t)}$, $i, i = 1, 2, \ldots, 8$.

To complete the above the Kalman filter, we need to calculate $Pr(S_t = j, S_{t-1} = i|I_t)$ and other probability terms. This is done by adopting the Hamilton (1989) filter, as given below:

$$Pr(S_t = j, S_{t-1} = i|I_t) = \frac{f(\pi_t|S_t = j, S_{t-1} = i, I_{t-1})Pr(S_t = j, S_{t-1} = i|I_{t-1})}{f(\pi_t|I_{t-1})}, \tag{2.B.12}$$

where

$$f(\pi_t|S_t = j, S_{t-1} = i, I_{t-1}) = \frac{1}{\sqrt{2\pi f^{i,j}_{t|t-1}}}\exp\left\{-\frac{(\eta^{i,j}_{t|t-1})^2}{2f^{i,j}_{t|t-1}}\right\}, \tag{2.B.13}$$

$$f(\pi_t|I_{t-1}) = \sum_{i=1}^{8}\sum_{j=1}^{8} f(\pi_t|S_t = j, S_{t-1} = i, I_{t-1}) Pr(S_t = j, S_{t-1} = i|I_{t-1}), \tag{2.B.14}$$

$$Pr(S_t = j, S_{t-1} = i|\Pi_{t-1}) = Pr(S_t = j|S_{t-1} = i) Pr(S_{t-1} = i|I_{t-1}), \tag{2.B.15}$$

and

$$Pr(S_t = j|I_t) = \sum_{i=1}^{8} Pr(S_t = j, S_{t-1} = i|I_t). \tag{2.B.16}$$

Conditional on initial values for the moments of the state vector β_t and the regime probabilities, we can run the above Kalman filter to obtain the following log likelihood function:

$$\log(f(\pi_T, \pi_{T-1}, \ldots)) = \sum_{t=1}^{T} \log(f(\pi_t|I_{t-1})), \tag{2.B.17}$$

which can be numerically maximized with respect to the parameters of the model.

References

[1] Atkeson, A. and L.E. Ohanian, 2001, Are Phillips Curves Useful for Forecasting Inflation? *Quarterly Review*, 25, 2-11.

[2] Ascari, G. and A.M. Sbordone, 2014, The Macroeconomics of Trend Inflation, *Journal of Economic Literature*, 52, 679-739.

[3] Ball, L., 1992, Why does high inflation raise inflation uncertainty?, *Journal of Monetary Economics*, 29, 371-388.

[4] Ball, L. and S.G. Cecchetti, 1990, Inflation and Uncertainty at Short and Long Horizons, *Brookings Papers on Economic Activity*, 1, 215-254.

[5] Banerjee, A., L. Cockerell and B. Russell, 2001, An I(2) analysis of Inflation and the Markup, *Journal of Applied Econometrics*, 16, 221-240.

[6] Barsky, R.B. and L. Kilian, 2002, "Do We Really Know that Oil Caused the Great Stagflation? A Monetary Alternative," *NBER Macroeconomics Annual 2001*, 137-198.

[7] Barsky, R.B. and L. Kilian, 2004, "Oil and the Macroeconomy since the 1970s," *Journal of Economic Perspectives*, 18, 115-134.

[8] Bénabou, R., 1992a, Inflation and Efficiency in Search Markets, *Review of Economic Studies*, 59, 299-329.

[9] Bénabou, R., 1992b, Inflation and Markups: Theories and Evidence from the Retail Trade Sector, *European Economic Review*, 36, 566-574.

[10] Bernanke, B.S., 2007, Inflation Expectations and Inflation Forecasting, Speech at NBER Monetary Economics Workshop.

[11] Beveridge, S. and C.R. Nelson, 1981, A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the "Business Cycle", *Journal of Monetary Economics*, 7, 151-174.

[12] Bhar, R. and S. Hamori, 2004, The Link between Inflation and Inflation Uncertainty:

Evidence from G7 Countries, *Empirical Economics*, 29, 825-853.

[13] Bredin, D. and S. Fountas, 2006, Inflation, Inflation Uncertainty, and Markov Regime Switching Heteroskedasticity: Evidence from European Countries, Working paper.

[14] Clarida, R., J. Galí and M. Gertler, 2000, "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics,* 115, 147-180.

[15] Cogley, T., G.E. Primiceri and T.J. Sargent, 2010, Inflation-Gap Persistence in the US, *American Economic Journal: Macroeconomics*, 2, 43-69.

[16] Cogley, T. and T.J. Sargent, 2001, Evolving Post World War II U.S. Inflation Dynamics, *NBER Macroeconomics Annual*, 16, 331-373.

[17] Cogley, T. and T.J. Sargent, 2005, Drifts and Volatilities: Monetary Policies and Outcomes in the Post War U.S., *Review of Economic Dynamics*, 8, 262-302.

[18] Cogley, T. and A.M. Sbordone, 2008, Trend Inflation, Indexation, and Inflation Persistence in the New Keynesian Phillips Curve, *American Economic Review*, 98, 2101-2126.

[19] Conrad, C. and T. Eife, 2012, Explaining Inflation-Gap Persistence by a Time-Varying Taylor Rule, *Journal of Macroeconomics*, 34, 419-428.

[20] Cosimano, T.F. and D.W. Jansen, 1988, Estimates of the Variance of U.S. Inflation based upon the ARCH Model: Comment, *Journal of Money, Credit, and Banking*, 20, 409-421.

[21] Cukierman, A. and A. Meltzer, 1986, A Theory of Ambiguity, Credibility, and Inflation under Discretion and Asymmetric Information, *Econometrica*, 54, 1099-1128.

[22] De Long, B.J., 1997, "America's peacetime inflation: The 1970s," In *Reducing Inflation: Motivation and Strategy,* C. Romer and D. Romer (eds). Chicago: University of Chicago Press.

[23] Engle, R., 1983, Estimates of the Variance of U.S. Inflation based upon the ARCH Model, *Journal of Money, Credit, and Banking*, 15, 286-301.

[24] Friedman, M., 1975, "Perspective on Inflation," *Newsweek*, June 24, p. 73.

[25] Friedman, M., 1977, "Nobel lecture: Inflation and Unemployment," *Journal of Political Economy*, 85, 451-472.

[26] Galí, J. and M. Gertler, 1999, Inflation Dynamics: A Structural Econometric Analysis, *Journal of Monetary Economics*, 44, 195-222.

[27] Goodfriend, M. and R.G. King, 2012, "The Great Inflation Drift." In *The Great Inflation: The Rebirth of Modern Central Banking*, edited by Michael D. Bordo and Athanasios Orphanides, 181?16. Chicago, IL: University of Chicago Press.

[28] Grier, K. and M. Perry, 1998, On Inflation and Inflation Uncertainty in the G7 Countries, *Journal of International Money and Finance*, 17, 671-689.

[29] Granger, C.W.J. and P. Newbold, 1986, Forecasting Economic Time Series, 2nd edition. (Academic Press, Orlando).

[30] Head, A., A. Kumar and B. Lapham, 2010, Market Power, Price Adjustment, and Inflation, *International Economic Review*, 51, 73-98.

[31] Ireland, P.N., 2007, Changes in the Federal Reserve's Inflation Target: Causes and Consequences, *Journal of Money, Credit, and Banking*, 39, 1851-1882.

[32] Kang, Kyu Ho, C.J. Kim and J. Morley, 2009, Changes in U.S. Inflation Persistence, *Studies in Nonlinear Dynamics and Econometrics*, 13, 1-21.

[33] Kilian, L., 2008, "The Economic Effects of Energy Price Shocks," *Journal of Economic Literature,* 46, 871-909.

[34] Kim, C-J, 1993, Unobserved-Component Time-Series Models with Markov-Switching Heteroskedasticity: Changes in Regime and the Link between Inflation Rates and Inflation Uncertainty, *Journal of Business and Economic Statistics*, 11, 341-349.

[35] Kim, C-J, 1994, Dynamic Linear Models with Markov-Switching, *Journal of Econometrics*, 60, 1-22.

[36] McKinnon, R.I., 1982, "Currency Substitute and Instability in the World Dollar Standard," *American Economic Review,* 72, 320-333.

[37] Mertens, Elmar, 2016, Measuring the Level and Uncertainty of Trend Inflation, *Review of Economic and Statistics*, 98, 950-967.

[38] Mishkin, F.S., 2007, Inflation Dynamics. *International Finance*, 10, 317?34.

[39] Morley, James C., C.R. Nelson and E. Zivot, 2003, Why Are the Beveridge-Nelson and Unobserved-Components Decompositions of GDP so Different? *Review of Economics and Statistics*, 85, 235-243.

[40] Rigobon, R. 2003, "Identification through Heteroskedasticity," *Review of Economic and Statistics*, 85, 777-792.

[41] Steinsson, Jón, 2003, Optimal Monetary Policy in an Economy with Inflation Persistence, *Journal of Monetary Economics*, 50, 1425-1456.

[42] Stock, J. and M. Watson, 2007, Why Has Inflation Become Harder to Forecast? *Journal of Money, Credit, and Banking*, 39, 3-33.

[43] Stock, J. and M. Watson, 2016, Core Inflation and Trend Inflation, *Review of Economic and Statistics*, 98, 770-784.

[44] Woodford, M., 2008, "How Important Is Money in the Conduct of Monetary Policy?" *Journal of Money, Credit, and Banking*, 40, 1561-1598.

Table 2.1. Estimation of Models 1-4.

$$\pi_t = \tau_t + z_t + \alpha_1 S_{1,t} + \alpha_2 S_{2,t} + \alpha_3 S_{1,t} S_{2,t}$$
$$\tau_t = \tau_{t-1} + (Q_0 + Q_1 S_{1,t})\varepsilon_t$$
$$z_t = \phi_{S_{3,t}} z_{t-1} + (h_0 + h_1 S_{2,t})\eta_t$$
$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

	Model 1	**Model 2**	**Model 3**	**Model 4**
p1,00	0.979 (0.014)***	0.977 (0.015)***	0.985 (0.022)***	0.971 (0.031)***
p1,11	0.931 (0.045)***	0.907 (0.057)***	0.933 (0.086)***	0.872 (0.101)***
p2,00	0.982 (0.017)***	0.989 (0.017)***	0.995 (0.006)***	0.994 (0.006)***
p2,11	0.954 (0.041)***	0.956 (0.043)***	0.978 (0.018)***	0.979 (0.018)***
p3,00	-	-	0.970 (0.029)***	0.990 (0.014)***
p3,11	-	-	0.992 (0.009)***	0.995 (0.007)***
h_0	0.329 (0.044)***	0.426 (0.089)***	0.500 (0.052)***	0.477 (0.059)***
h_1	0.328 (0.139)**	0.496 (0.164)***	0.781 (0.122)***	0.778 (0.169)***
Q_0	0.175 (0.039)***	0.193 (0.058)***	0.166 (0.051)***	0.129 (0.053)***
Q_1	0.891 (0.162)***	1.085 (0.326)***	0.437 (0.178)**	0.391 (0.194)**
ϕ_0	-	-	0.879 (0.104)***	0.868 (0.114)***
ϕ_1	-	-	0.093 (0.153)	0.013 (0.138)
α_1	-	-	-	0.940 (0.180)***
α_2	-	-	-	1.172 (0.701)*
α_3	-	-	-	-0.466 (3.083)
ρ	-	-0.694 (0.188)***	-0.883 (0.105)***	-0.907 (0.075)***
$Ln(L)$	-238.662	-236.406	-235.390	-232.755

Table 2.1. (Continued).

Note:

1. In the parentheses are standard errors of parameter estimates.
2. The superscripts '*', '**', and '***' indicate significance at the 10%, 5%, and 1% level, respectively.
3. $Ln(L)$ denotes the log likelihood value.

Table 2.2. Root Mean Squared Errors for Pseudo Out-of-Sample Forecasting: Out-of-Sample Period: 1990Q1-2016Q1

	Forecasting Horizon		
	4 Quarters	8 Quarters	12 Quarters
AO Model	0.4077	0.4848	0.5964
UCSV Model	0.4007	0.4179	0.4681
Model 1	0.4251	0.4346	0.4888
Model 2	0.3742	0.3942	0.4548
Model 3	0.3914	0.4093	0.4711
Model 4	0.3449	0.3515	0.4060

Note:

1. AO model refers to a random walk model proposed by Atkeson and Ohanian (2001); and UCSV model refers to Stock and Watson's (2007) unobserved components model with stochastic volatility.
2. Root mean squared forecast errors are obtained from the pseudo out-of-sample inflation forecast, $\pi^h_{t+h} \equiv \sum_{i=t+1}^{t+h} \pi_i/h$. The forecasting period starts at 1990Q1 with expanding windows.

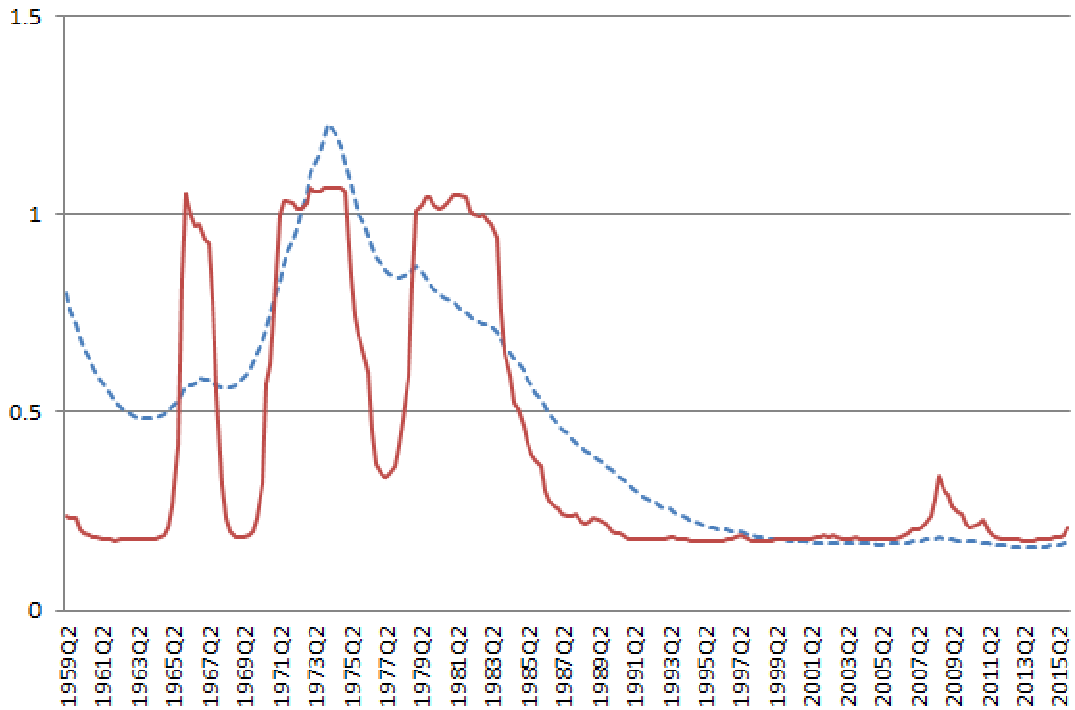
Table 2.3. Diebold and Mariano's (1995) Test Statistics for Evaluating Pseudo Out-of-Sample Forecasting Performance

	Forecasting Horizon		
	4 Quarters	8 Quarters	12 Quarters
Model 2 vs. AO Model	1.42	1.31	1.19
	(0.16)	(0.19)	(0.23)
Model 3 vs. AO Model	0.77	1.15	1.08
	(0.44)	(0.25)	(0.28)
Model 4 vs. AO Model	1.74	1.86	1.43
	(0.08)*	(0.06)*	(0.15)
Model 2 vs. UCSV Model	1.49	1.55	0.57
	(0.14)	(0.12)	(0.57)
Model 3 vs. UCSV Model	0.11	0.73	-0.11
	(0.91)	(0.47)	(0.91)
Model 4 vs. UCSV Model	1.68	1.84	0.55
	(0.09)*	(0.07)*	(0.58)
Model 2 vs. Model 1	2.33	1.59	1.24
	(0.02)**	(0.11)	(0.22)
Model 3 vs. Model 1	1.89	0.78	1.42
	(0.06)*	(0.44)	(0.16)
Model 4 vs. Model 1	2.45	2.65	0.87
	(0.01)***	(0.01)***	(0.38)

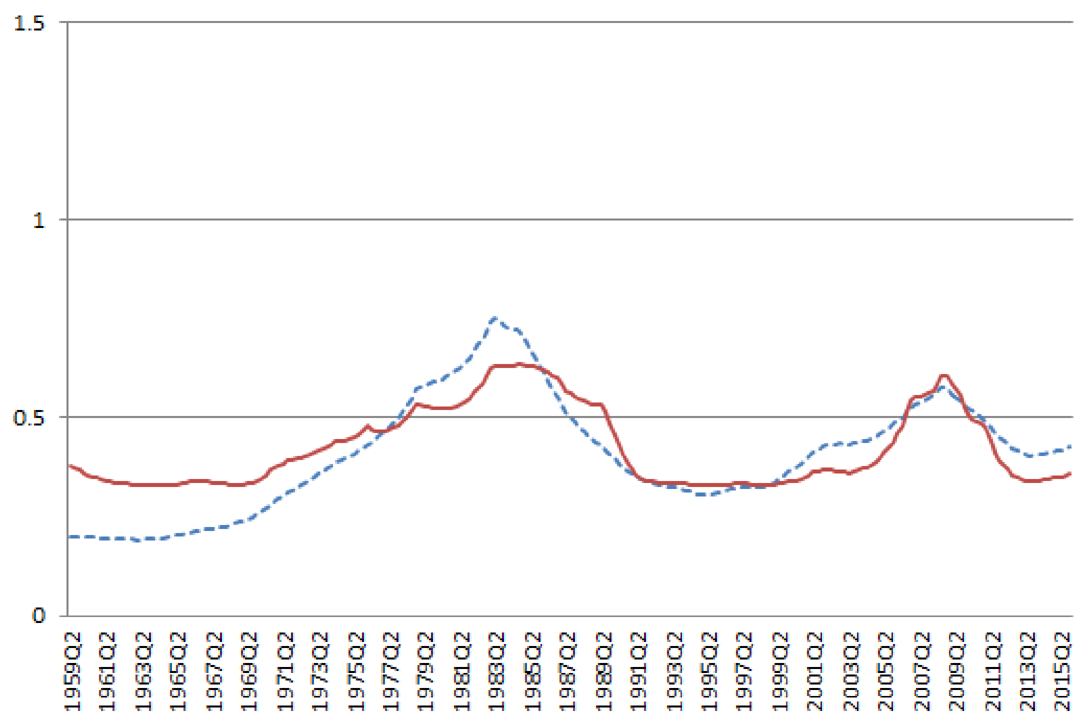
Note:

1. AO model refers to a random walk model proposed by Atkeson and Ohanian (2001); and UCSV model refers to Stock and Watson's (2007) unobserved components model with stochastic volatility.
2. A negative value for the test statistic means that the mean squared error for Model 2,3, or 4 is larger than that for the competing model (AO model, UCSV model or Model 1).
3. Reported are the Diebold-Mariano test statistics. In the parentheses are their corresponding p-values under the null of equal mean squared errors.
4. The superscripts '*', '**', and '***' indicate significance at the 10%, 5%, and 1% level, respectively.

Figure 2.1. Measures of Trend Inflation Volatility and Inflation Gap Volatility: Stock and Watson's (2007) UCSV model vs. Model 1 (Our Benchmark Model)



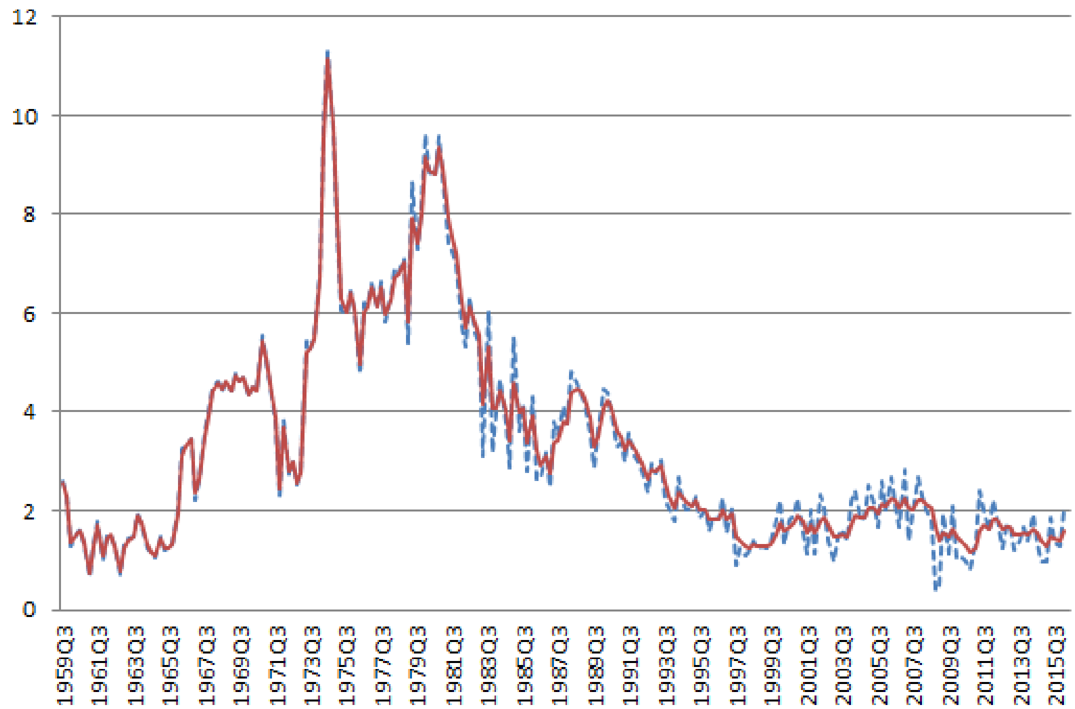
(a) Trend Inflation Volatilities



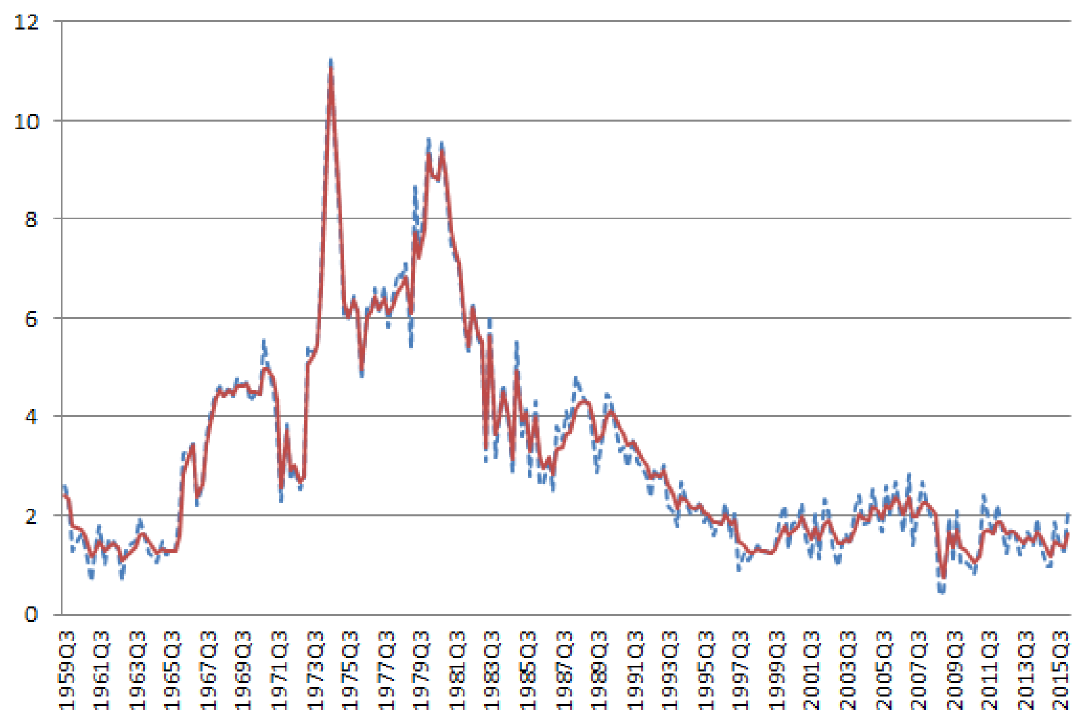
(b) Inflation Gap Volatilities

Note: The dotted lines are for Stock and Watson's (2007) UCSV model; the solid lines are for Model 1 (Our Benchmark Model).

Figure 2.2. Measures of Trend Inflation: Stock and Watson's (2007) UCSV model vs. Model 1



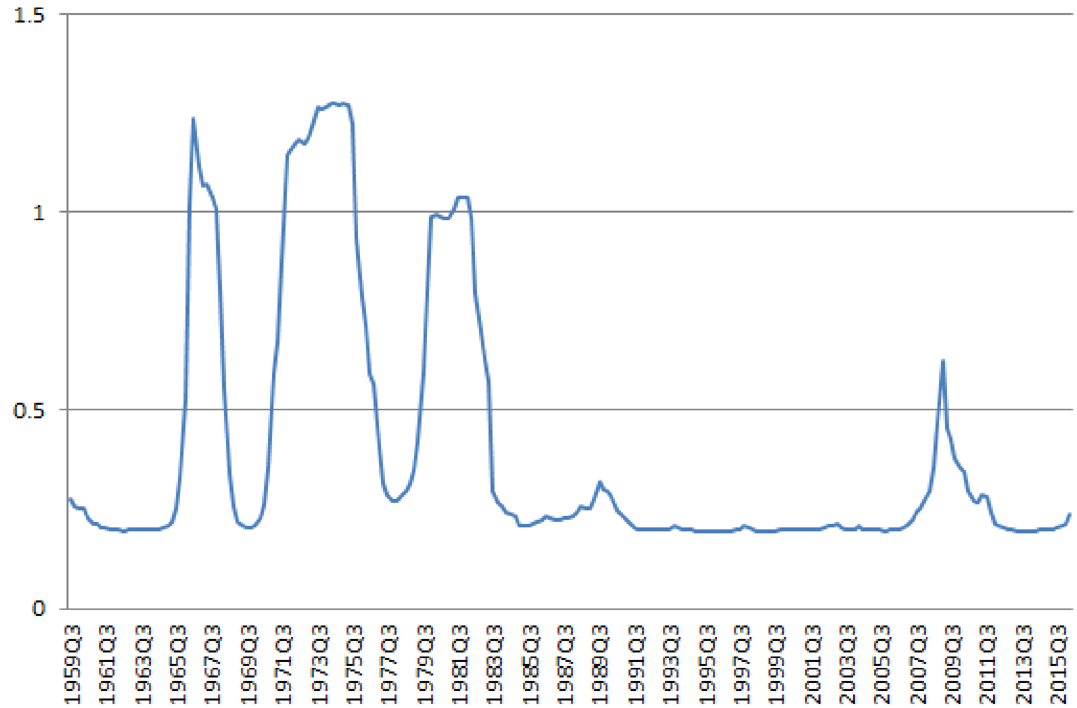
(a) Stock and Watson's (2007) UCSV Model



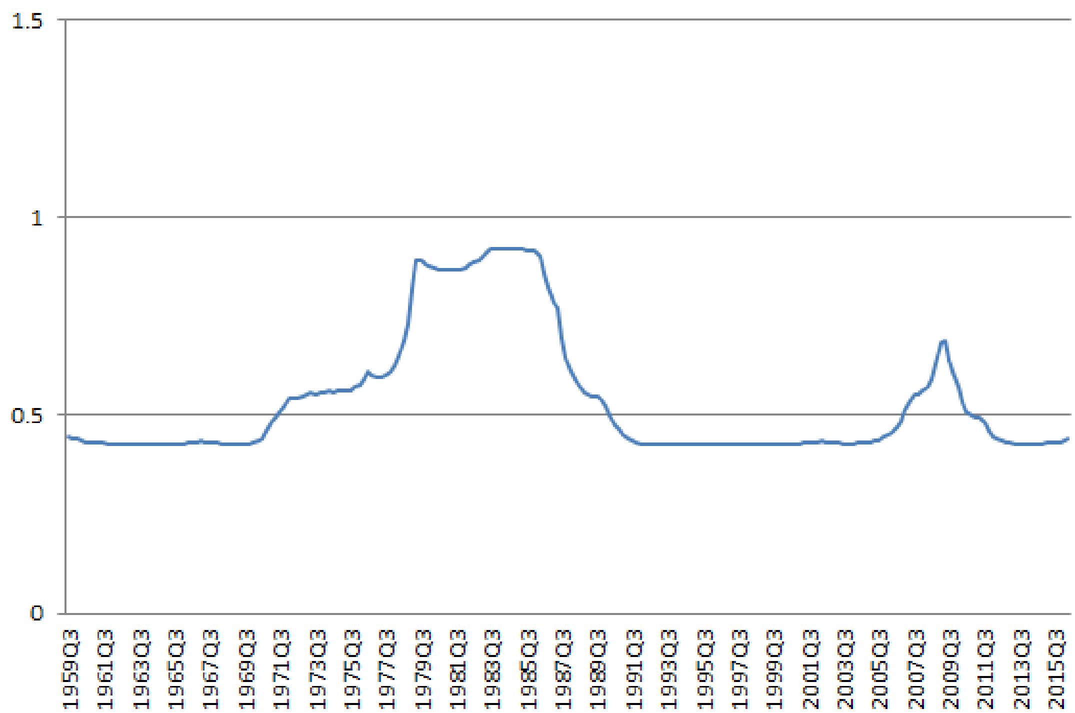
(b) Model 1 (Our Benchmark Model)

Note: The dotted lines are for actual inflation; the solid lines are for trend inflation.

Figure 2.3. Measures of Trend Inflation Volatility and Inflation Gap Volatility: Model 2

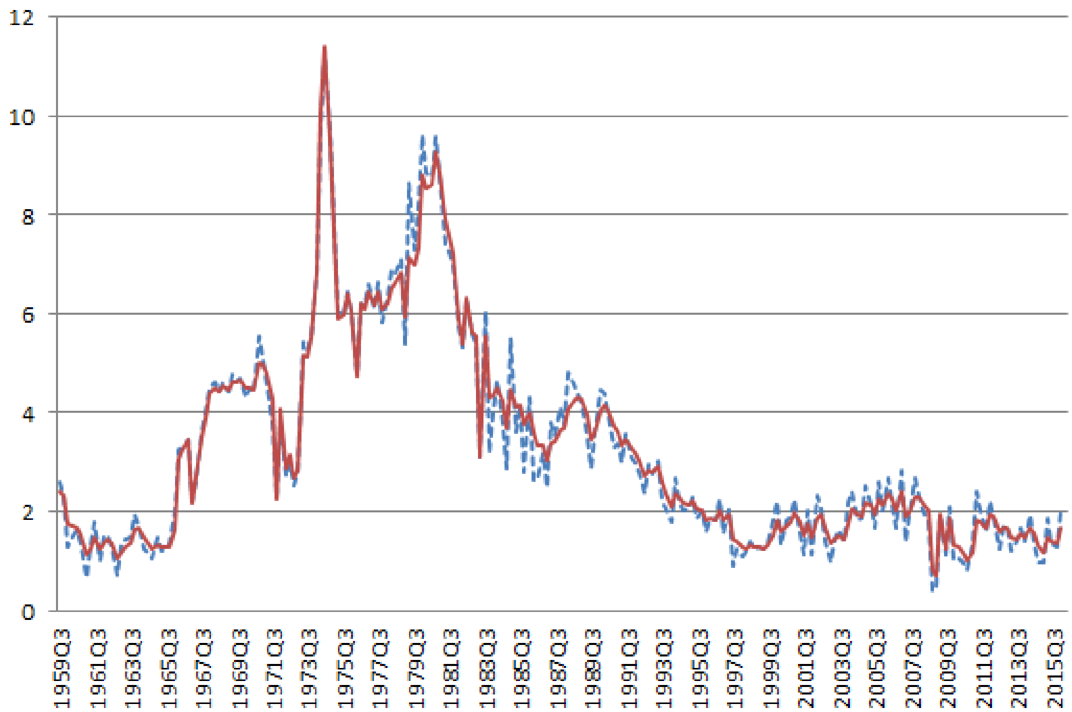


(a) Trend Inflation Volatilities



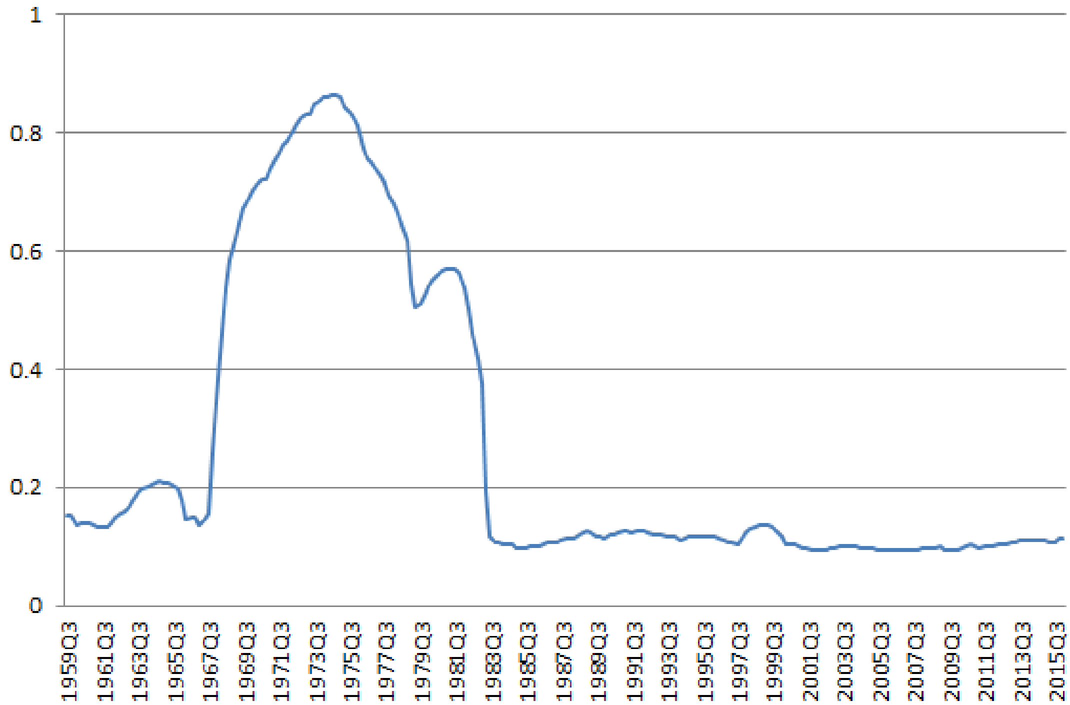
(b) Inflation Gap Volatilities

Figure 2.4. Measures of Trend Inflation: Model 2

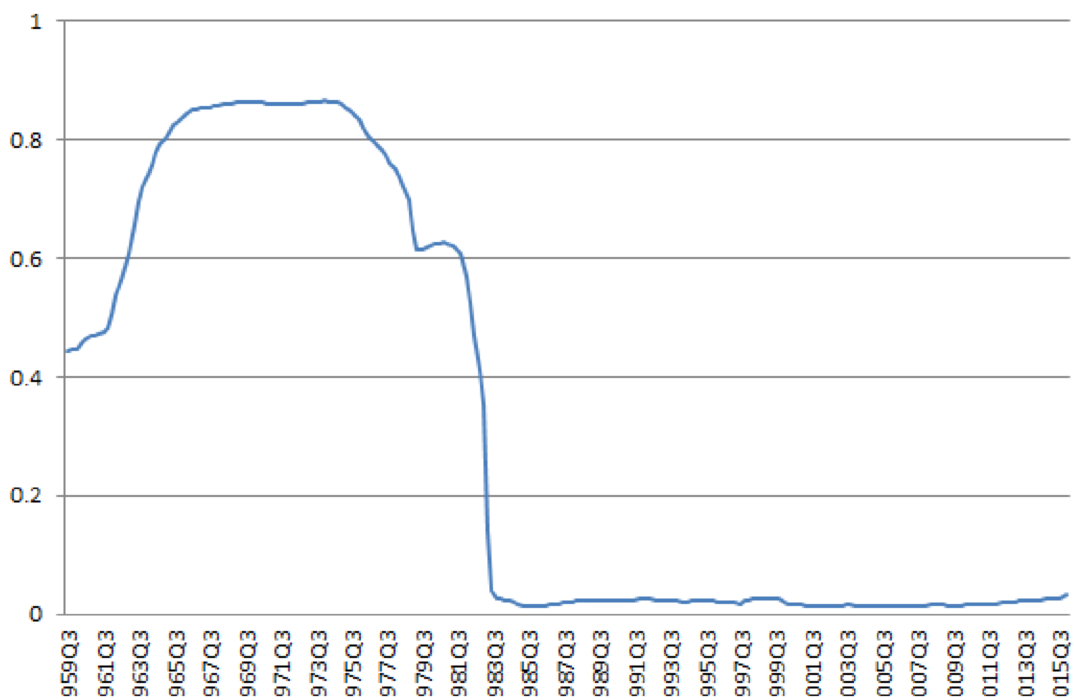


Note: The dotted lines are for actual inflation; the solid lines are for trend inflation.

Figure 2.5. Time-Varying Inflation Gap Persistence: Model 3 vs. Model 4

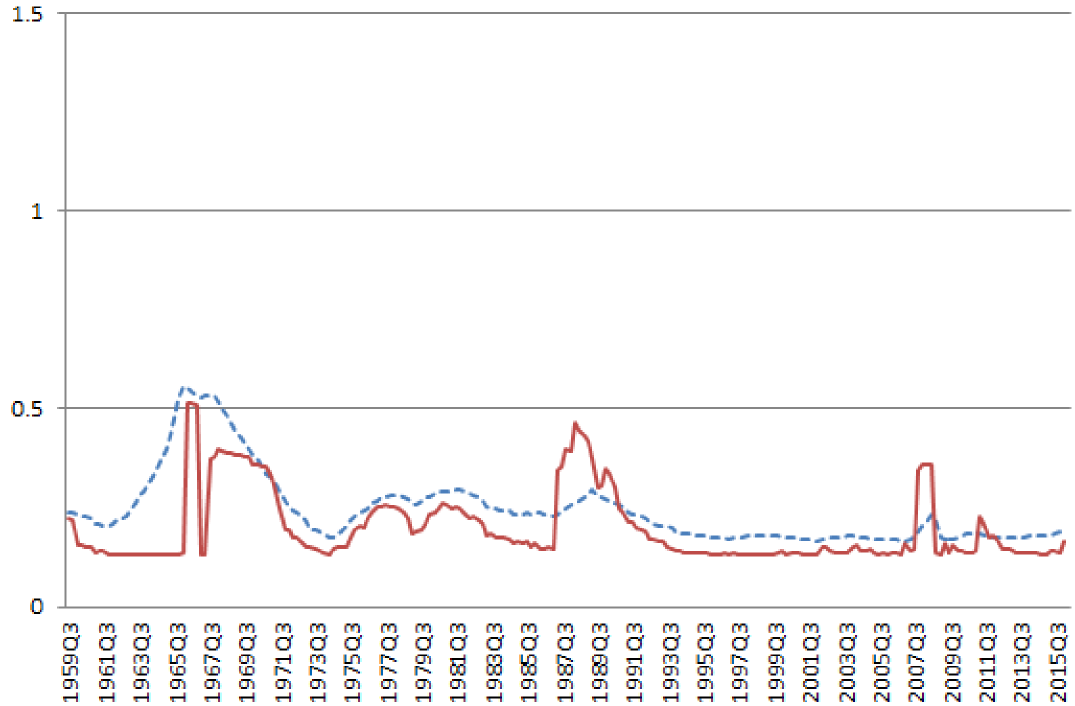


(a) Model 3

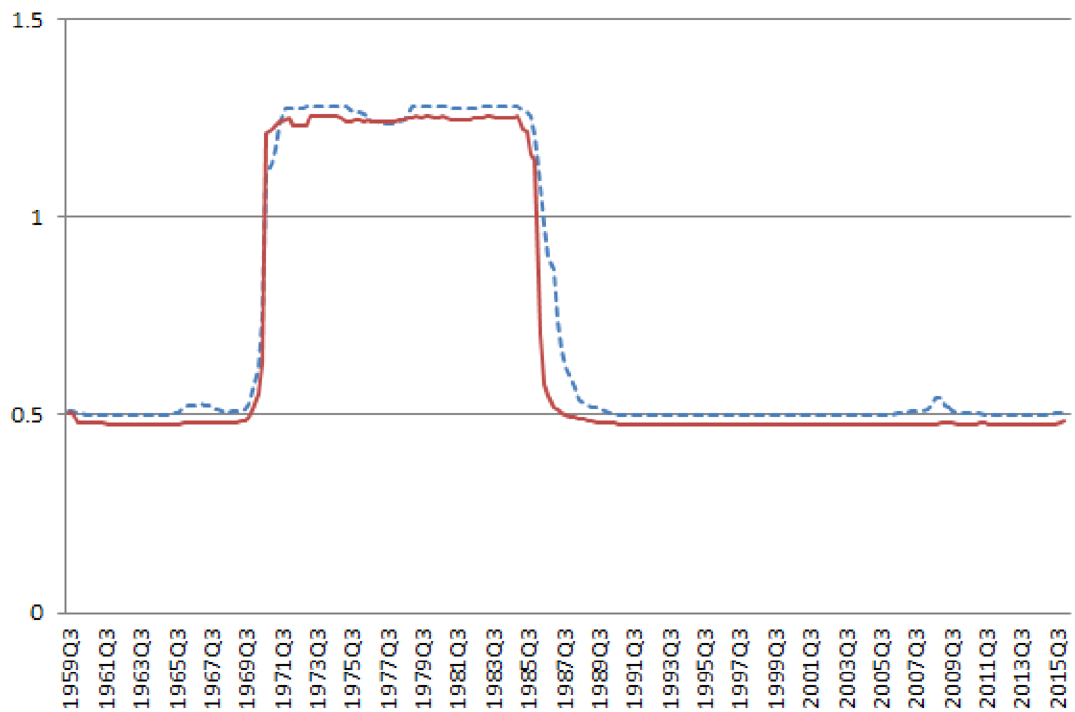


(b) Model 4

Figure 2.6. Measures of Trend Inflation Volatility and Inflation Gap Volatility: Model 3 vs. Model 4



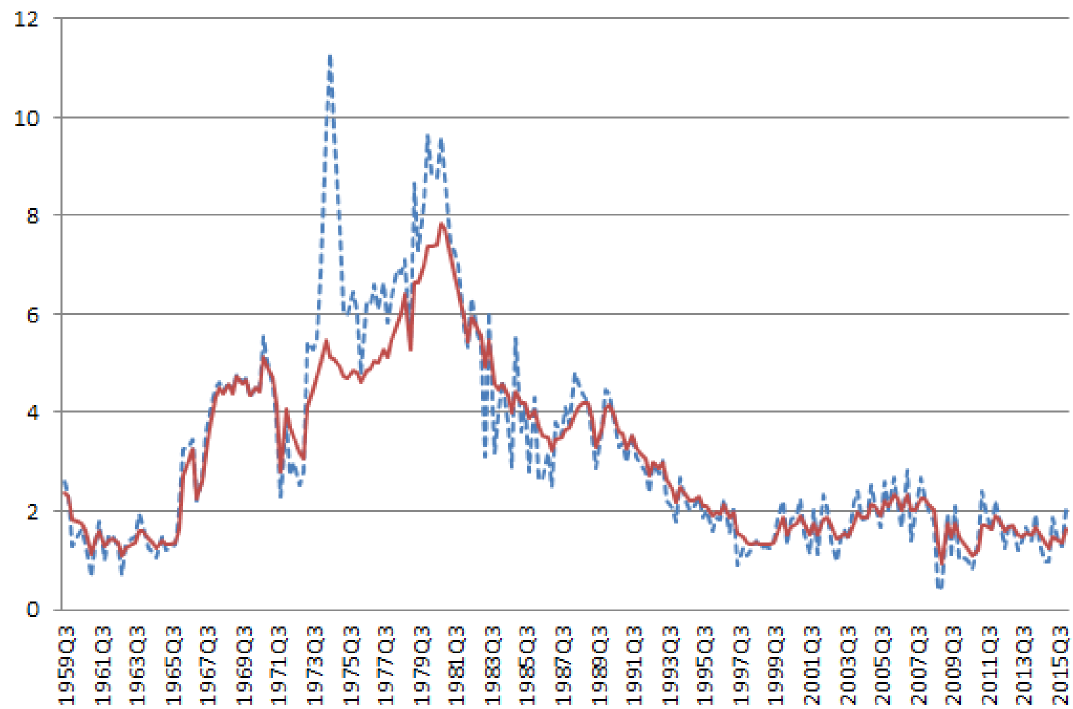
(a) Trend Inflation Volatilities



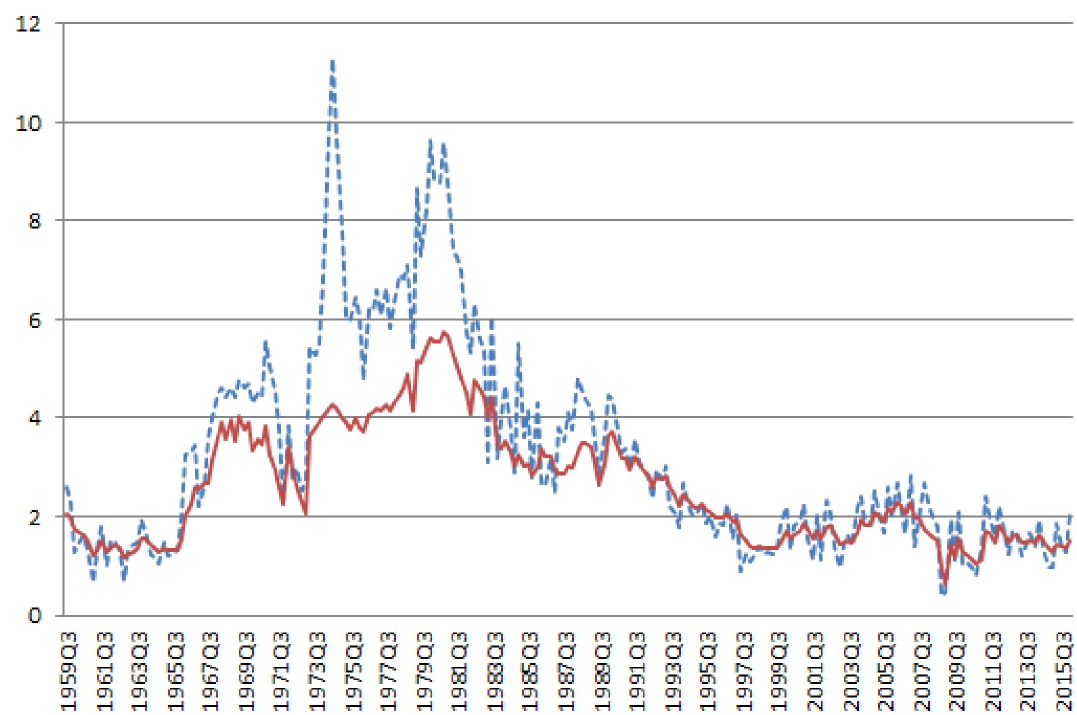
(b) Inflation Gap Volatilities

Note: The dotted lines are for volatilities from Model 3; the solid lines are for volatilities from Model 4.

Figure 2.7. Measures of Trend Inflation: Model 3 vs. Model 4



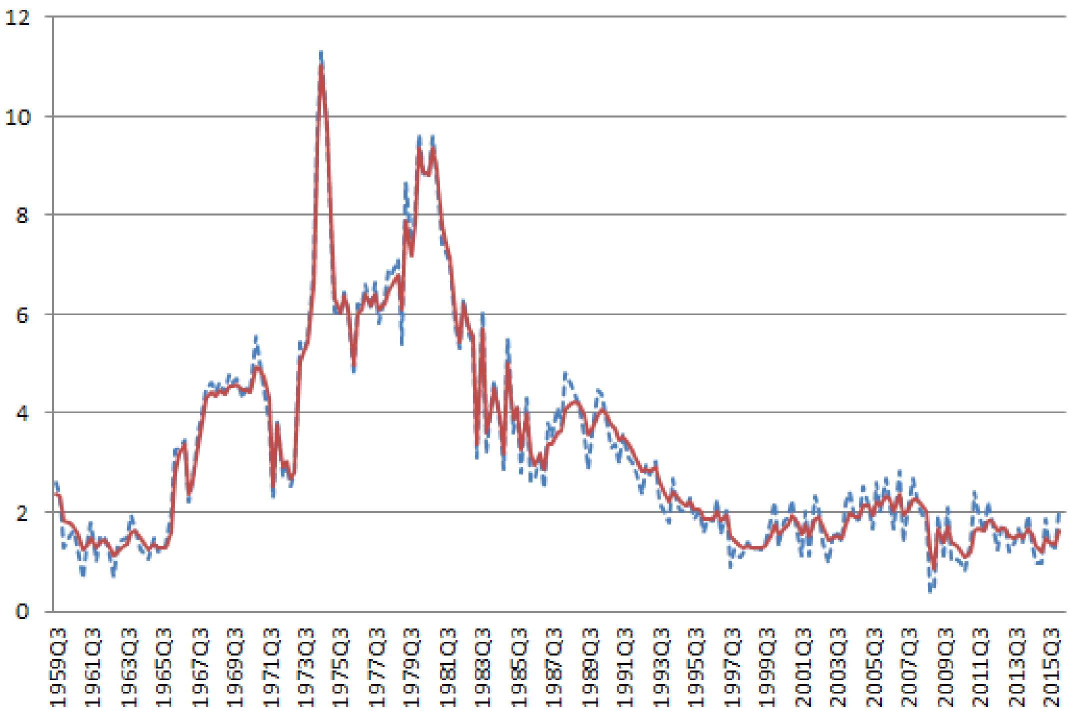
(a) Model 3



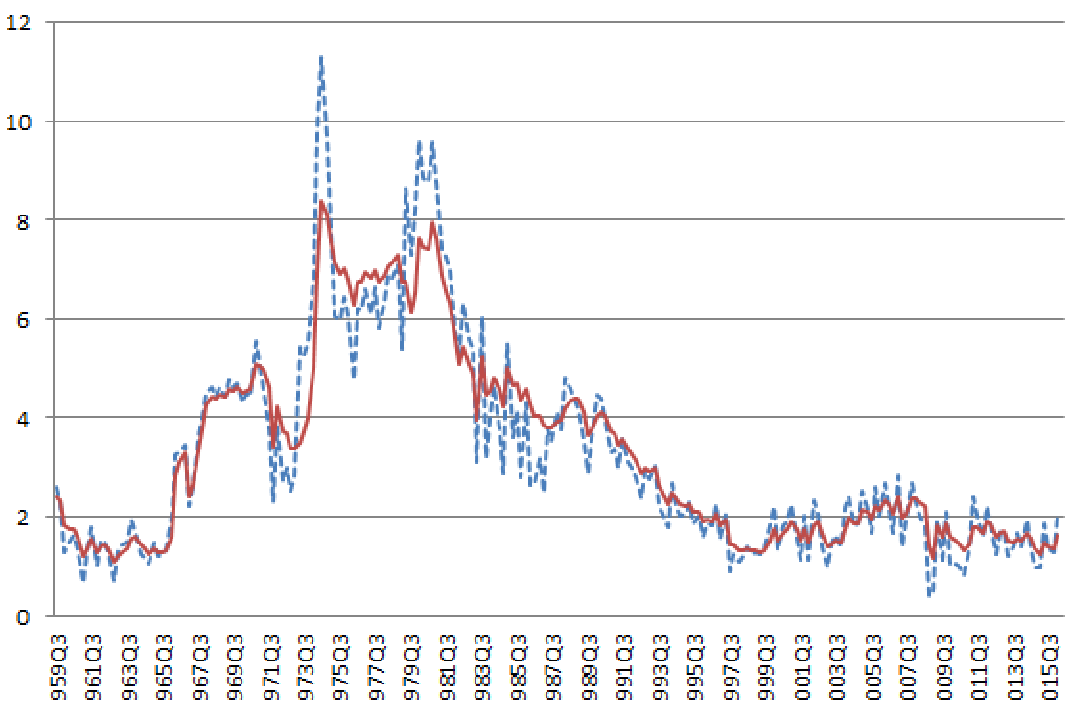
(b) Model 4

Note: The dotted lines are for actual inflation; the solid lines are for trend inflation.

Figure 2.8. Measures of Trend Inflation: Model 3 with $\rho = 0$ vs. Model 4 with $\rho = 0$



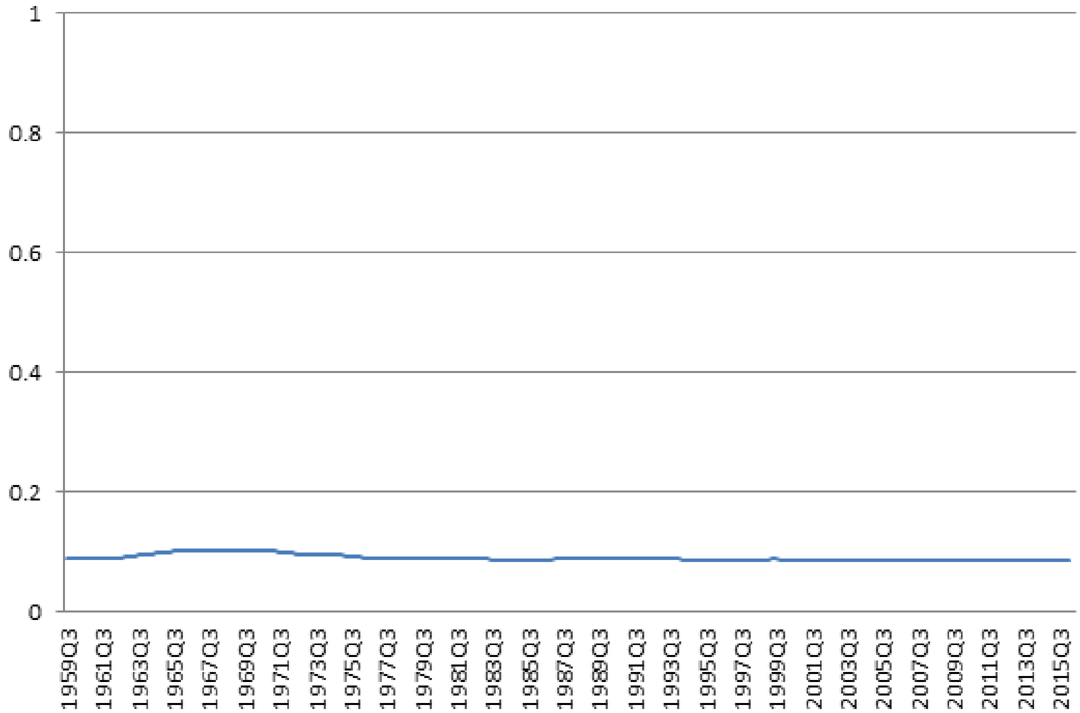
(a) Model 3



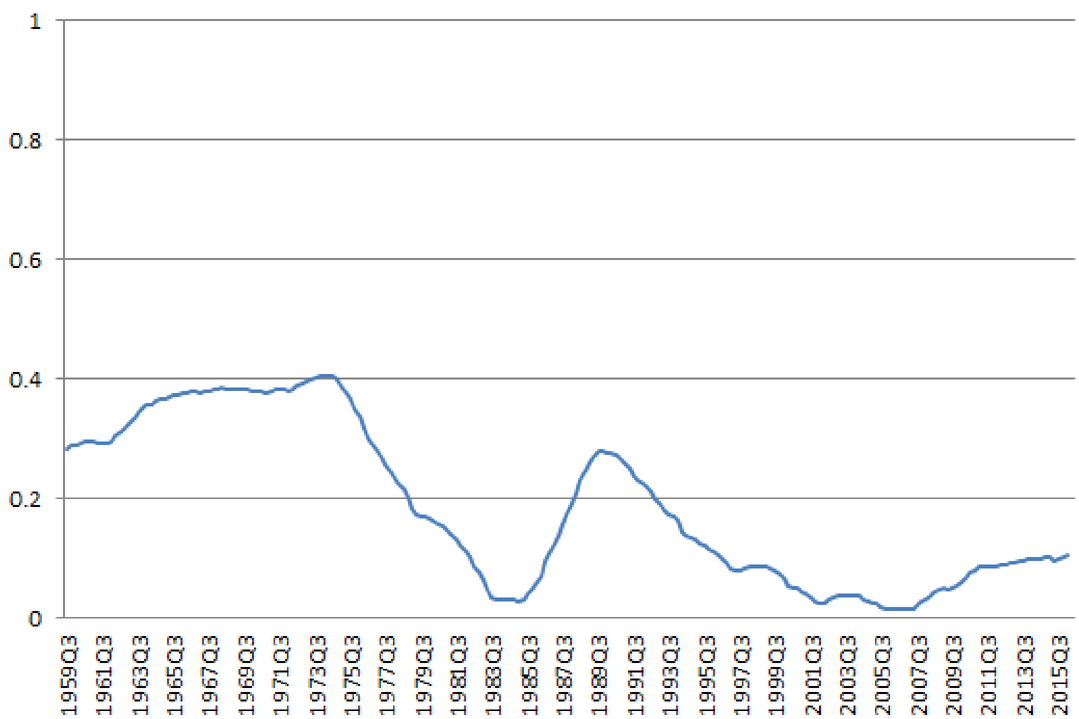
(b) Model 4

Note: The dotted lines are for actual inflation; the solid lines are for trend inflation.

Figure 2.9. Time-Varying Inflation Gap Persistence: Model 3 with $\rho = 0$ vs. Model 4 with $\rho = 0$



(a) Model 3



(b) Model 4

Chapter 3

Estimating Elasticity of Intertemporal Substitution when Instruments are Weak: Identification Through Time-Varying Volatility

3.1 Introduction

The elasticity of intertemporal substitution (EIS) in consumption is a parameter of central importance in macroeconomics and finance. In a basic model of the effects of monetary policy, the EIS is the parameter that relates current and expected future real interest rates to the current level of aggregate demand in the inter-temporal relation. In the consumption and portfolio choice problem of an infinite-lived investor with Epstein-Zin (1989) preferences, the EIS is the key parameter in the optimal consumption rule.

EIS can be defined by the log-linearized Euler equation based on the Epstein-Zin utility function (Epstein and Zin, 1991). As the rate of return and the consumption levels are simultaneously determined, OLS does not provide consistent estimates. Therefore, numerous papers have applied the standard instrument variable (IV) method to estimate EIS (e.g. Hall (1988) and Campbell (2003)). The general finding in the literature is that the EIS is small (Hall, 1988). For instance, Campbell (2003, table 9) reports a 95% confidence interval of $[-0.14, 0.28]$ for EIS, using quarterly U.S. data (1947-1998) on nondurable consumption and T-bill returns. However, a high value for the EIS is a precondition that the long-run risk model of Bansal and Yaron (2004) - a successful model for explaining the risk premium and other key asset markets phenomena - has to satisfy. As pointed out by Neely, Roy, and Whiteman (2001), Campbell (2003), and Yogo (2004), weak instruments could be one of the reason for this inconsistency in the estimation of EIS as asset returns are notoriously difficult to predict.

In linear IV regression, weak instruments arise when the instruments are weakly correlated with the endogenous variables. Rothenberg (1984) and Nelson and Startz (1990) show in their seminal work that the exact finite sample distribution of two stage least square (2SLS) estimator depends on the strength of the instruments. Staiger and Stock (1997) extend results in Rothenberg (1984) and Nelson and Startz (1990) to more general framework and show that the asymptotic distributions of generalized method of moment (GMM) and IV statistics are in general nonnormal. Furthermore, the standard GMM and 2SLS point estimates, hypothesis tests, and confidence intervals are all unreliable. In order to make valid inference, several weak instrument robust tests have been proposed, including Anderson-Rubin (AR) (Anderson and Rubin, 1949), Lagrange multiplier (LM) (Kleibergen, 2002), and conditional likelihood ratio (CLR) (Moreira, 2003).

Yogo (2004) applies the above three robust tests to estimate EIS of eleven developed countries and reports quite uninformative confidence interval for EIS based on stock return. For example, based on CLR test, Yogo (2004) reports 95% confidence intervals of $[-\infty, \infty]$ for EIS for eight out of eleven countries, using quarterly data on stock returns. These uninformative confidence intervals indicate the weakness of weak instrument robust test - its power performance is usually poor, especially when the instruments are really weak. Recently, Andrews (2014) and Andrews and Guggenberger (2015) propose several more powerful weak instrument robust tests and revisit the estimation problem of EIS in Yogo (2004). Although they successfully demonstrate that their methods are more powerful than AR, LM, and CLR tests in artificial data, but none of them report reasonable confidence intervals for EIS using stock returns.

In linear regression framework, it is well known that the instrument variable approach is equivalent to OLS regression that includes additional regressors to control for endogeneity. Commonly, the additional regressors are the reduced form residuals for the endogenous variables. This method is known as control function (CF) approach. For example, in the case of a linear regression, a two-step estimation procedure based on the CF approach proceeds as follows. In the first step, the reduced-form residuals for the endogenous regressors are estimated. In the second step, these residuals are included in the primary equation as additional regressors. Conditional on these additional regressors, the new error term is orthogonal to the endogenous variables. However, the residuals of the reduced-form equations are (near) multi-collinear with the endogenous variables when instruments are weak and thus the identification of the primary equation failed.

In this paper, we propose semiparametric CF approaches to avoid the multi-collinearity problem in the conventional CF approach when the instruments are arbitrarily weak. Our identification strategy is through time-varying volatility. We show that the proposed estimators are $\sqrt{n}$ consistent and asymptotically normally distributed. Monte Carlo experiments show that the proposed methods can estimate the coefficients of endogenous variables precisely in finite sample. The simulation results also show that the proposed methods have good performance on both size and power. Given the tremendous empirical support of the time-varying volatility in the stock returns (e.g. Schwert (1989), Kim, Shephard, and Chib (1998)), we apply the propose methods to estimate EIS based on stock return. The estimated confidence intervals for EIS for all the eleven countries in Yogo (2004) are consistent with the general finding in the literature.

The rest of the paper is organized as follows. Section 3.2 reviews the literature on both weak instruments and EIS. Section 3.3 presents our model specification and the identification conditions. Section 3.4 discusses the estimation and inference under two different model specification. Section 3.5 reports the finite sample performance of the proposed method, compare with the weak instrument robust test in the literature. We apply the proposed methods to estimate the EIS based on the data employed by Campbell (2003) and Yogo (2004) in Section 3.6. Section 3.7 is the conclusion.

3.2 Literature Review

3.2.1. Weak Instrument

IV regression is one of the most widely applied method in practical work to consistently estimate the parameter of interest when there exist endogenous variable in the linear regression model. The typical requirements for the validity of the IV regression are twofold: the instruments are required to be (i) exogenous (not correlated with the error term) and (ii) relevant. The second requirement implies that Z should be correlated with the endogenous variables. The problem of weak instruments arise when the instruments in linear IV regression are weakly correlated with the included endogenous variables. As shown in Rothenberg (1984) and Nelson and Startz (1990), 2SLS estimator has significant bias and is poorly approximated by a normal distribution when IVs are weak and the degree of endogeneity is medium to strong. Staiger and Stock (1997) extend results in Rothenberg (1984) and Nelson and Startz (1990) to more general framework and show that the asymptotic distributions of generalized method of moment (GMM) and IV statistics are in general nonnormal.

There are numerous examples of weak IVs in the empirical literature. Besides the estimation of elasticity of inetrtemporal substitution we studied in this paper, another classic example is from labor economics of Angrist and Krueger (1991) IV regression of wages on the endogenous variable years of education and additional covariates. Dummies for quarter of birth (with and without interactions with exogenous variables) are used as IVs for years of education. The argument is that quarter of birth is related to years of education via mandatory school laws for children aged sixteen and lower. When the relationship is weak, this leads to weak IVs. A notable feature of this application is that weak instrument issues arise despite the fact that Angrist and Krueger (1991) use a 5% Census sample with hundreds of thousands of observations. Evidently, weak instruments should not be thought of as merely a small-sample problem, and the difficulties associated with weak instruments can arise even if the sample size is very large.

Since weak IV arises when instruments in linear IV regression are weakly correlated with the included endogenous variables. A small first stage F statistic (or, equivalently, a low partial R^2) provides evidence that IVs are weak. Stock and Yogo (2005) develop formal tests based on the F statistic for the null hypothesis: the bias of 2SLS is greater than 10% of the bias based on OLS. The F test rejects the null of weak IVs at the 5% level if $F > 10.3$. Analogous tests when the null hypothesis is specified in terms of the limited information maximum likelihood (LIML) estimator or Fuller's (1977) modification of LIML are provided in Stock and Yogo (2005). These tests have different (smaller) critical values. An alternative test for the detection of weak IVs based on reverse regressions is given by Hahn and Hausman (2002). Unfortunately, this test has very low power and is not recommended, at least for the purpose of detecting weak instruments, see Hausman, Stock, and Yogo (2005).

As the conventional GMM and 2SLS estimate might lead to significant bias under weak

IV, an alternative approach is to apply tests which are fully robust to the weak IV asymptotically. By this we mean that the designed tests have asymptotically correct probability under weak IV asymptotics (instruments are weakly correlated with the included endogenous variables). The confidence intervals (CI) are then obtained by inverting the test, this approach was first advocated by Dufour (1997) and Staiger and Stock (1997).

The first test employed specifically to deal with weak IVs is the AR test, see Dufour (1997) and Staiger and Stock (1997). The power of AR test is good when we only have 1 endogenous variable. However, AR test has low power when we have more than 1 endogenous variables. Since that, the literature has sought more powerful tests than the AR test that are robust to weak IVs. Kleibergen (2002) and Moreira (2007) independently introduce an LM test whose null rejection rate is robust to weak IVs. The power of the LM test often is better than that of the AR test when the number of endogenous variables is larger than 1. Moreira (2003) propose the conditional CLR test, which is a more sophisticated version of Wang and Zivots (1998) likelihood ratio test in which a critical value function replaces a constant to achieve the desired confidence level. The CLR test has higher power than the Wang-Zivot likelihood ratio test. Andrews, Moreira, and Stock (2006) developed the asymptotic power envelop of the weak IV robust test. Among the above three robust tests, they show that the CLR test is (essentially) on the weak IV asymptotic power envelop under homoskedastic errors and thus recommend CLR test for iid homoskedastic error models. For heteroskedastic/autocorrelated error, the CLR test statistic can be replaced by a heteroskedasticity-robust version, HR-CLR, or a heteroskedasticity and autocorrelation-robust version, HAR-CLR, see Andrews, Moreira, and Stock (2007).

There are some other researches focus on many weak IV asymptotics. Such asymptotics are designed for the case in which the IVs are weak and the number of IVs, k, is relatively large compared to the sample size n. Chao and Swanson (2005) consider asymptotics in which $k \to \infty$, and $n \to \infty$. Under the weak IV asymptotics proposed in Staiger and Stock (1997), they show that as the k grows to infinity faster enough, the 2SLS estimator and the jackknife IV estimator are generally consistent. Since the estimation of consumption CAPM usually only contains fairly amount of instruments, we decide not to discuss this case in this paper. For more detailed survey for weak instrument, see Stock, James, and Wright (2000), Andrews and Stock (2005), Mikushava (2010).

3.2.2. Elasticity of Intertemporal Substitution

The EIS is considered as one of the main behavioral parameters in macroeconomics and financial economics. The magnitude of the EIS is central for policy analysis and for a host of economic issues including: (i) The value of the EIS determines the consumption saving decisions, since it measures the sensitivity of changes in the expected consumption growth rate in response to changes in the expected return on the portfolio for a typical stockholder; (ii) The effectiveness of fiscal and monetary policies depends on the level of the EIS. Specifically, the higher the value of the EIS, the less effective fiscal policy, and the higher the value of the EIS the more effective monetary policy in increasing output (Hall, 1988); (iii) The EIS plays a key role in fitting the data in a real business cycle. The value of the EIS is a central determinant of the level and volatility of interest rates over the business cycle.

Although the magnitude of EIS is very important, there is no consensus about its value. Some articles support the hypothesis of a low EIS (Hall, 1988; Campbell and Viceira, 1999; Campbell, 2003; and Yogo, 2004), whereas others support the hypothesis of a high EIS (Hansen and Singleton, 1982; Attanasio and Weber, 1989; Vissing-Jorgensen, 2002; and Bansal, Kiku, and Yaron, 2007). Havranek (2014) examines 2,375 estimates of the EIS reported in 169 published studies. He reports that the mean estimate of the EIS is about 0.5, which is rather in favor of the low EIS hypothesis. Yet, a high value for the EIS is a precondition that the long-run risk model of Bansal and Yaron (2004) - a successful model for explaining the risk premium and other key asset markets phenomena - has to satisfy. When working with this type of model, we should assume that the EIS is more than one, which conflicts with most empirical evidence.

To obtain a precise estimate of EIS, one needs to develop methods which are not only valid under weak IVs, but also exploit sample information more efficiently. This motivates our research.

3.3. Model Specifications and Identification

We consider the following regression model,

$$y_t = x_t'\beta_0 + u_t, \quad E(x_t u_t) \neq 0, \tag{3.1}$$

$$x_t = (I_p \otimes z_{t-1}')\delta_0 + v_t \tag{3.2}$$

where y_t is the dependent variable. x_t and z_t are the $p \times 1$ endogenous variables and $q \times 1$ instrumental variables, $q \geq p$. Note that this model specification accounts for other strictly exogenous variables (include constant term), which have been "partial out" of the specification (Hahn and Hausman, 2005). Our main interest is to identify $\beta_0 \in B$, where B is a compact subset of R^p.

Let F_t is the σ-field to which y_t, x_t and z_t are adapted, define

$$\sigma_{ut}^2 = Var(u_t|F_{t-1}), \quad u_t^* = \frac{u_t}{\sigma_{ut}},$$
$$\Omega_{vt} = Var(v_t|F_{t-1}), \quad v_t^* = \Omega_{vt}^{-\frac{1}{2}} v_t$$

We assume the following mixing conditions throughout the paper

Assumption M:

u_t^* and v_{it}^* are strong mixing (α-mixing) martingale difference process with unit conditional variance, a.s., for all t, with filter F_t, where v_{it}^* is the i^{th} element in v_{it}^*. There exist $\alpha_1 > 1$, $\alpha_{2i} > 1$, $M_1 > 0$ and $M_{2i} > 0$, such that $\sup_t E u_t^{*4\alpha_1} < M_1 < \infty$ and $\sup_t E v_{it}^{*\,4\alpha_{2i}} < M_{2i} < \infty$, for $i = 1, \ldots, p$

Assumption M states that both u_t^* and v_t^* has finite fourth moment. By this assumption and Lyapunov's inequality, $E\left(|u_t^*|^{4\xi}\right)$ and $E\left(|v_{it}^*|^{4\xi}\right)$ exists for all $\xi \leq \min(\alpha_1, \alpha_{21}, \ldots, \alpha_{2p})$, as do all expectations involving up to any four combinations of u_t^* and v_{it}^*, for $i = 1, \ldots, p$. Note that by construction $E(u_t^*|F_{t-1}) = E(v_{it}^*|F_{t-1}) = 1$.

The model in equation (3.1) cannot be directly estimated by the least square method due to the endogeneity. The two most popular methods to solve the problem are 2SLS and control function approach. In linear models, it is well known that these two methods give algebraically identical results when instruments are strong. However, as discussed in the previous section, the conventional 2SLS method suffers from weak instruments. In contrast, we show that the control function approach proposes in this paper suffers less from the weak instruments problem. Follow the control function approach, we can project the structure equation residual, u_t, on the reduced-form error terms, v_t, and use it as additional regressors. In particular, we can rewrite the model in equations (3.1)-(3.2) as,

$$\begin{aligned} y_t &= x_t'\beta_0 + \left(E(v_t^2|F_{t-1})^{-1}v_t\right)' E(u_tv_t|F_{t-1}) + w_t \\ &= x_t'\beta_0 + \left(E(v_t^2|F_{t-1})^{-\frac{1}{2}}v_t\right)' E(u_tv_t^*|F_{t-1}) + w_t \end{aligned} \tag{3.3}$$

where

$$w_t = u_t - \left(E(v_t^2|F_{t-1})^{-\frac{1}{2}}v_t\right)' E(u_tv_t^*|F_{t-1})$$

Lemma 3.1. Under assumption M, we can show that

$$(i) E(w_t|F_{t-1}) = 0$$

$$(ii) E(x_tw_t|F_{t-1}) = 0$$

$$(iii) E(v_tw_t|F_{t-1}) = 0$$

$$(iv) E(v_t^*w_t|F_{t-1}) = 0$$

The proof of Lemma 3.1 is given in the Appendix. Lemma 3.1 shows that the new residual, w_t, is a martingale difference sequence and is uncorrelated with x_t, v_t, and the standardized v_t^*. Thus the endogeneity in the reduced form equation (3.3) is controlled. However, when the instruments are weak, the model in equation (3.3) is not always identifiable. We discuss the identification conditions for equation (3.3) under weak instruments in the following subsection.

3.3.1. Identification under Weak Instruments

In this subsection, we discuss the identification conditions for the control function model in equation (3.3) under weak instruments. As mentioned in the previous section, the control function approach is algebraically identical to the 2SLS when instruments are strong. Therefore, we focus on the weak instrument case. In particular, we maintain the following weak instruments assumption throughout the paper.

Assumption IV:

(i) $\frac{1}{T}\sum_{t=2}^{T} z_{t-1}u_t \to_p E\left(z_{t-1}u_t\right) = \mathbf{0}_q, \quad \frac{1}{T}\sum_{t=2}^{T} z_{t-1}\otimes v_t \to_p E\left(z_{t-1}\otimes v_t\right) = \mathbf{0}_{p\times q}.$

(ii) $\sup_t (z_t z_t') < \infty$.

(iii) $\delta_0 = \frac{C}{T^\nu}$, where $\nu \geq \frac{1}{2}$, $C = [c_{11}, c_{12}, \ldots, c_{1q}, c_{21}, \ldots, c_{2q}, \ldots, c_{pq}]'$, and c_{ij} is finite constant, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, q$.

Assumption IV(i) states that the instruments satisfy the conventional orthogonality condition. IV(ii) states that the instruments have finite second moment for all t. Both are standard in the literature. IV(iii) states the partial correlations between endogenous variable and the instrument variables converge to 0. This assumption has been widely assumed in the weak instrument literature and is the same as Staiger and Stock (1997) when $\nu = \frac{1}{2}$. Formally, IV(iii) induces a triangular array structure, but we drop the additional affix T in the arguments when it introduces no confusion.

To clarify identification conditions, we rewrite the model in equation (3.3) into moment condition form. First, literature usually assume $E(v_t^2|F_{t-1}) = \Omega_v$ and $E(u_t v_t^*|F_{t-1}) = \gamma_0$, where Ω_v and γ_0 are constant for $t = 1, \ldots, T$, $\gamma_0 = (\gamma_{10}, \ldots, \gamma_{p0})$, $\gamma_{j0} = E(u_t v_{jt}^*|F_{t-1})$. In this case, $v_t^* = \Omega_v^{-\frac{1}{2}} v_t$ for $t = 1, \ldots, T$, and equation (3.3) becomes

$$E\begin{bmatrix} x_t\left(y_t - x_t'\beta_0 - v_t^{*\prime}\gamma_0\right) \\ v_t^*\left(y_t - x_t'\beta_0 - v_t^{*\prime}\gamma_0\right)\end{bmatrix} = 0 \tag{3.4}$$

When instruments are weak, $\delta \to 0$ and $x_t \to v_t$ as $T \to \infty$, thus x_t and v_t are asymptotically collinear. Since $v_t^* = \Omega_v^{-1} v_t$ for $t = 1, \ldots, T$, x_t and v_t^* are also asymptotically collinear. As a consequence, the first p moment conditions and the $p + 1^{th}$ to $2p^{th}$ moment conditions in equation (3.4) are asymptotically the same, so we only have p moment conditions,

$$E\left[\, x_t\left(y_t - x_t'\beta_0 - v_t^{*\prime}\gamma_0\right)\right] = 0 \tag{3.5}$$

We have $2p$ unknown parameters, (β_0, γ_0), but only p moment conditions. Furthermore, rank condition is not satisfied since x_t and v_t^* are asymptotically collinear. Therefore, the identification failed in this case.

The above analysis shows the weak instrument problem in control function approach stems from the collinearity between regressors when both $E(v_t^2|F_{t-1})$ and $E(u_t v_t^*|F_{t-1})$ are constant over time. In contrast, when we don't assume $E(v_t^2|F_{t-1})$ and $E(u_t v_t^*|F_{t-1})$ are constant over time, equation (3.3) can be rewritten as the following moment conditions,

$$E\begin{bmatrix} x_t\left(y_t - x_t'\beta_0 - v_t^{*\prime}E(u_t v_t^*|F_{t-1})\right) \\ v_t^*\left(y_t - x_t'\beta_0 - v_t^{*\prime}E(u_t v_t^*|F_{t-1})\right)\end{bmatrix} = 0 \tag{3.6}$$

Under assumption IV and M, v_t, Ω_{vt}, and v_t^* can be identified from the reduced-form equation (3.2). However, u_t cannot be identified separately without identifying β_0, this makes the model in equation (3.6) unidentified without restriction on the functional form of $E(u_t v_t^*|F_{t-1})$.

To see this more clearly, we can define $u_t(\beta^*) = y_t - x_t\beta^*$, for any given $\beta^* \in B$. Then the model in equation (3.6) can be rewritten as,

$$E\begin{bmatrix} x_t\left(y_t - x_t'\beta^* - v_t^{*\prime}E(u_t(\beta^*)v_t^*|F_{t-1})\right) \\ v_t^*\left(y_t - x_t'\beta^* - v_t^{*\prime}E(u_t(\beta^*)v_t^*|F_{t-1})\right)\end{bmatrix} = 0 \tag{3.7}$$

By the law of iterated expectation, the above moment conditions can be expressed as,

$$\begin{aligned} &E\begin{bmatrix} E\left(x_t\left(y_t - x_t'\beta^* - v_t^{*\prime}E(u_t(\beta^*)v_t^*|F_{t-1})\right)|F_{t-1})\right) \\ E\left(v_t^*\left(y_t - x_t'\beta^* - v_t^{*\prime}E(u_t(\beta^*)v_t^*|F_{t-1})\right)|F_{t-1})\right)\end{bmatrix} \\ \to &E\begin{bmatrix} E\left(v_t\left(u_t(\beta^*) - v_t^{*\prime}E(u_t(\beta^*)v_t^*|F_{t-1})\right)|F_{t-1})\right) \\ E\left(v_t^*\left(u_t(\beta^*) - v_t^{*\prime}E(u_t(\beta^*)v_t^*|F_{t-1})\right)|F_{t-1})\right)\end{bmatrix} = 0 \end{aligned} \tag{3.8}$$

where the convergence result comes from IV(3). As the equality in equation (3.8) holds for any arbitrary $\beta^* \in B$, the moment conditions model in equation (3.6) can't be identified.

In summary, the model in equation (3.3) cannot be identified under assumption IV when we assume both $E(v_t^2|F_{t-1})$ and $E(u_t v_t^*|F_{t-1})$ are constant over time, or when there is no restriction on the functional form of $E(u_t v_t^*|F_{t-1})$. Since Ω_{vt}, and v_t^* can be identified from the reduced-form equation under assumption IV and M, the time varying variance of v_t can be easily taken into account. Therefore, we focus on the functional form of $E(u_t v_t^*|F_{t-1})$ which can allow us to identify the parameters in equation (3.3). In the following, we proposed two different specifications of $E(u_t v_t^*|F_{t-1})$ for this purpose,

Case 1: $E(u_t v_{it}^*|F_{t-1}) = \gamma_{i0}$ where γ_{i0} are constant for all t, $i = 1, \ldots, p$.

In this case, we have the same moment conditions as equation (3.4). Nevertheless, when Ω_{vt} is time varying, x_t and v_t^* are not asymptotically collinear. Therefore, the convergence results in equation (3.5) won't happen and we have $2p$ moment conditions. Meanwhile, rank condition is also satisfied since x_t and v_t^* are not asymptotically collinear. Thus, the model in equation (3.3) can be identified under this specification. Note that constant $E(u_t v_{it}^* | F_{t-1})$ is often assumed in the literature since the model is then easier to compute. Case 1 have same advantage as the model can be estimated by simple OLS method, but it exploits more information from data since it accounts for the time varying variance in v_t. We propose another specification to relax the constant assumption in the next case.

Case 2: $E(u_t v_{it}^* | F_{t-1}) = \alpha_{00_i} + \alpha_{10_i} u_{t-1} v_{it-1}^*$ where α_{00_i} and α_{10_i} are constant for all t, $i = 1, \ldots, p$.

This specification is in the same spirit as Engle (2002), where we assume the covariance between u_t and v_t^* follows an autoregressive conditional heteroskedasticity (ARCH) process [1]. In this case, the moment conditions in equation (3.6) can be rewritten as

$$E\begin{bmatrix} x_t\left(y_t - x_t'\beta_0 - \sum_{i=1}^{p} v_{it}^*(\alpha_{00_i} + \alpha_{10_i} u_{t-1} v_{it-1}^*)\right) \\ v_t^*\left(y_t - x_t'\beta_0 - \sum_{i=1}^{p} v_{it}^*(\alpha_{00_i} + \alpha_{10_i} u_{t-1} v_{it-1}^*)\right) \end{bmatrix} = 0$$

As in case 1, the convergence results in equation (3.5) won't happen and rank condition is satisfied as Ω_{vt} is time varying. Thus, the model in equation (3.3) can be identified under this specification.

In the next section, we discuss the estimation and inference of the model in equation (3.3) under the above two different specifications.

3.4. Model Estimation and Inference

3.4.1. Case 1

[1] We focus on ARCH(1) specification here for simplicity. The model can be easily extended to ARCH(k), where $k > 1$. Furthermore, if we assume that both u_t and v_t follow normal distribution, then we can extend the specification in case 2 to generalized autoregressive conditional heteroskedasticity (GARCH) process.

In this case, the model in equation (3.3) becomes a linear model with constant coefficient

$$y_t = x_t'\beta_0 + v_t^{*\prime}\gamma_0 + w_t \tag{3.9}$$

where w_t is uncorrelated with x_t and v_t^* and a martingale difference sequence as we show in Lemma 3.1.

To address the time varying variance in v_t, we make the following assumption throughout this section.

Assumption V

(i)

$$\Omega_{vt} \equiv \Omega\left(\frac{t}{T}\right) = \begin{bmatrix} h_{11}\left(\frac{t}{T}\right) & h_{12}\left(\frac{t}{T}\right) & \dots & h_{1p}\left(\frac{t}{T}\right) \\ h_{21}\left(\frac{t}{T}\right) & h_{22}\left(\frac{t}{T}\right) & \dots & h_{2p}\left(\frac{t}{T}\right) \\ \vdots & \vdots & & \vdots \\ h_{p1}\left(\frac{t}{T}\right) & h_{p2}\left(\frac{t}{T}\right) & \dots & h_{pp}\left(\frac{t}{T}\right) \end{bmatrix}$$

where $h_{ij}(\cdot) = h_{ji}(\cdot)$, $h_{ij}(\cdot)$ are non-stochastic, measurable, uniformly bounded on the interval $(0,1]$, and satisfy Lipschitz condition, for all $i,j = 1,\dots,p$. $\inf_t h_{ii}\left(\frac{t}{T}\right) > 0$, for all $i = 1,\dots,p$.

(ii) $\Omega\left(\frac{t}{T}\right)$ is not constant and nonsingular over t.

Assumption V is in the same spirit as Cavaliere (2004), Phillips and Xu (2005), and Xu and Phillips (2008). Note that under Assumption V, the function $h_{ij}(\cdot)$ is integrable on the interval [0,1] up to any finite order, $i = 1,\dots,p$, $j = 1,\dots,p$. For brevity, we write $\int_0^1 h_{ij}^m(r)dr$ as $\int_0^1 h_{ij}^m(r)$ for any finite positive integer m. It is straightforward to see that x_t and v_t^* are not asymptotically collinear under assumption V(2). We focus on the time varying unconditional variance of v_t in assumption V since the endogenous variable in our motivating example is stock returns, which has tremendous amount of evidence about its time varying unconditional variance (see e.g. Kim, Shephard, and Chib (1998), Omori, Chib, Shephard, and Nakajima (2007)). Feng (2004) and Engle and Rangel (2008) discuss how to combine the time varying unconditional variance with GARCH model with more assumptions on the error term, u_t and v_t.

Define $s_t = (x_t', v_t^{*\prime})$, the estimator $(\widehat{\beta}, \widehat{\gamma})$ can be expressed as,

$$\begin{pmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{pmatrix} = (\sum_{t=1}^{T} s_t' s_t)^{-1} \sum_{t=1}^{T} s_t' y_t$$

Theorem 3.1. Under assumption IV, M, and V, assume Ω_{vt} and δ are known,

$$\sqrt{T} \begin{pmatrix} \widehat{\beta} - \beta_0 \\ \widehat{\gamma} - \gamma_0 \end{pmatrix} \to dN(0, \Gamma_1^{-1} \Lambda_1 \Gamma_1^{-1})$$

where

$$\Gamma_1 = \begin{pmatrix} \int \Omega & \int \Omega^{\frac{1}{2}} \\ \int \Omega^{\frac{1}{2}} & 1 \end{pmatrix}, \quad \Lambda_1 = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E\left(w_t^2 s_t' s_t\right)$$

The proof of Theorem 3.1 is given in the Appendix. Theorem 3.1 shows that the estimators are $\sqrt{n}$ consistent and asymptotic normally distributed when the true value of δ and Ω_{vt} are known. It also demonstrates the problem of weak instruments in the absence of time-varying variance. In particular, when $\Omega_{vt} = \overline{\Omega}$,

$$\Gamma_1 = \begin{pmatrix} \overline{\Omega} & \overline{\Omega}^{\frac{1}{2}} \\ \overline{\Omega}^{\frac{1}{2}} & 1 \end{pmatrix}$$

which is a singular matrix. Thus the asymptotic variance covariance matrix cannot be computed.

The OLS estimators in Theorem 3.1 are infeasible in practice, since the true value of δ and Ω_{vt} are unknown. To produce a feasible procedure, we propose to use a kernel-based estimator that can consistently estimate Ω_{vt} and v_t^*. In order to have uniform convergence rate throughout the support of $h_{ij}(\cdot)$, $i, j = 1, \ldots, p$, we propose to use the generalized kernel proposed by Muller (1991).

Let $\widehat{v}_t = x_t - (I_p \otimes z_t)\widehat{\delta}$ be the first stage OLS residuals. Also let $b_{ij} \equiv b_{ij}(N) \to 0$ denote a bandwidth and let $K(\cdot)_{b_{ij}}$ denote a univariate generalized kernel function with the properties $K_{b_{ij}}(u, t) = 0$ if $u > t$ or $u < 1 - t$; for all $t \in [0, 1]$,

$$h^{-(j+1)} \int_{t-1}^{t} u^j K_{b_{ij}}(u, t) du = \begin{cases} 1, & if j = 0 \\ 0, & if 1 \leq j \leq r - 1 \end{cases}$$

We call $K_{b_{ij}}(\cdot, \cdot)$ a univariate generalized kernel function of order r.

The following example is taken from Muller (1991). Define

$$M_{\theta,r}([a_1,a_2]) = \left\{ g \in Lip([a_1,a_2]), \int_{a_1}^{a_2} x^j g(x)dx = \left\{ \begin{array}{rr} 1, & if j=0 \\ 0, & if 1 \le j \le r-1 \end{array} \right\} \right.$$

where $Lip([a_1,a_2])$ denotes the space of Lipschitz continuous functions on $[a_1,a_2]$. Define $K_+(\cdot,\cdot)$ and $K_-(\cdot,\cdot)$ as follows:

(i) The support of $K_+(x,q')$ is $[-1,q'] \times [0,1]$ and the support of $K_-(\cdot,\cdot)$ is $[-q',1] \times [0,1]$.

(ii) $K_+(\cdot,q') \in M_{\theta,r}([-1,q'])$ and $K_-(\cdot,q') \in M_{\theta,r}([-q',1])$. We note that $K_+(\cdot,1') = K_-(\cdot,1) = K(\cdot) \in M_{\theta,r}([-1,1])$. Now let

$$K_h(u,t) = \begin{cases} K_+(u,1), & if h \le t \le 1-h \\ K_+(\frac{u}{h},\frac{t}{h}), & if 0 \le t \le h \\ K_-(\frac{u}{h},\frac{1-t}{h}), & if 1-h \le t \le 1 \end{cases}$$

Then we can show that $K_h(\cdot,\cdot)$ is a generalized kernel function of order r.

Assume the bandwidth parameter b_{ij} satisfies $O(T^{-\frac{1}{2}}) < b_{ij} < O(T^{-\frac{1}{5}}) \forall i,j = 1,2,\ldots,p$. The estimator $\widehat{v}_t^*$ can be expressed as

$$\widehat{v}_t^* = \widehat{\Omega}_{vt}^{-\frac{1}{2}} \widehat{v}_t$$

where

$$\widehat{\Omega}_{vt} = \begin{pmatrix} \widehat{h}_{11}\left(\frac{t}{T}\right) & \widehat{h}_{12}\left(\frac{t}{T}\right) & \ldots & \widehat{h}_{1p}\left(\frac{t}{T}\right) \\ \widehat{h}_{21}\left(\frac{t}{T}\right) & \widehat{h}_{22}\left(\frac{t}{T}\right) & \ldots & \widehat{h}_{2p}\left(\frac{t}{T}\right) \\ \vdots & \vdots & & \vdots \\ \widehat{h}_{p1}(tT) & \widehat{h}_{p2}\left(\frac{t}{T}\right) & \ldots & \widehat{h}_{pp}\left(\frac{t}{T}\right) \end{pmatrix}$$

$$\widehat{h}_{ij}\left(\frac{t}{T}\right) = \sum_{\tau=1}^{T} w_{t\tau ij} \widehat{v}_{i\tau} \widehat{v}_{j\tau}, \quad w_{t\tau ij} = \left(\sum_{\tau=1}^{T} K_{b_{ij}}\left(\frac{t-\tau}{T}\right) \right)^{-1} K_{b_{ij}}\left(\frac{t-\tau}{T}\right)$$

Under the above generalized kernel estimator, follow Feng (2004) and Feng and Yu (2006), we can have the following convergence results

Lemma 3.2.

$$(i)\ \widehat{v}_t - v_t = O_p\left(T^{-\frac{1}{2}}\right)$$
$$(ii)\ \widehat{h}_{ij}(r) - h^*_{ij}(r) = O_p(b_{ij}^2) + O_p\left(Tb_{ij}{}^{-\frac{1}{2}}\right),\ \ r \in [0,1]$$
$$(iii)\ \widehat{v}^*_t - v^*_t = o_p\left(T^{-\frac{1}{4}}\right)$$

The proof of Lemma 3.2 follows the proof in Feng (2004). Lemma 3.2 states that the time-varying volatility can be estimated precisely. Lemma 3.2 (ii) is the conventional convergence result in nonparametric estimation. Lemma 3.2 (iii) states is the under smoothing results from choosing the convergence rate of bandwidth appropriately. This condition is used for Corollary 3.2 below.

Therefore, a feasible estimator $(\widetilde{\beta}, \widetilde{\gamma})$ can be obtained by the following three step approach.

Step 1: Estimate the instrument equation by OLS, obtain a consistent estimators of v_t, $\widehat{v}_t$.

Step 2: Apply the generalized kernel estimator above to estimate the time varying variance covariance matrix Ω_{vt}, and calculate $\widehat{v}^*_t$ as $\widehat{\Omega}_{vt}\widehat{v}_t$

Step 3: Replace v^*_t in equation (3.9) with $\widehat{v}^*_t$ and estimate the new model by OLS.

Follow Lemma 5.1 in Newey (1994), the feasible approach can consistently estimate β and γ. To make correct inference on β, however, we need to take into account the generated regressors problem due to the estimation of δ and σ_{vt} (Pagan (1984), Newey (1994)). Nevertheless, Theorem 3.1 still provides a useful result for testing endogeneity. Note that in this case, $\gamma_i = \rho_i \sigma_u$, thus when there exist no endogeneity, $\rho_i = \gamma_i = 0$.

Corollary 3.1. Under the same assumptions in Theorem 3.1, a test for endogeneity can be established as

$$H_0 : \gamma_0 = 0$$

where the test statistics $\sqrt{T}\widehat{\gamma}'(\Gamma_1^{-1}\Lambda_1\Gamma_1^{-1})_*^{-1}\widehat{\gamma} \rightarrow_a \chi^2_p$, $(\Gamma_1^{-1}\Lambda_1\Gamma_1^{-1})_*$ is the $p \times p$ bottom right block of the $\Gamma_1^{-1}\Lambda_1\Gamma_1^{-1}$ in Theorem 3.1.

The proposed test in corollary 3.1 does not suffer from the generated regressors problem. This is because when $\rho = \gamma = 0$, the model in equation (3.1) can be directly estimated by OLS and thus no extra uncertainty is introduced. [2] This can be seen as an extended version of the Hausman-Wu test (Hausman, 1978; Wu, 1973) under weak instrument. Note that, in the absence of weak instrument and time-varying variance, Wu's (1973) approach is to test for the significance of γ in equation (3.9), and the result is equivalent to the Hausman test.

Corollary 3.2. The feasible estimator $(\widetilde{\beta}, \widetilde{\gamma})$ has the following asymptotic distribution

$$\sqrt{T}\begin{pmatrix} \widetilde{\beta} - \beta_0 \\ \widetilde{\gamma} - \gamma_0 \end{pmatrix} \to_d N(0, \Gamma_1^{-1}\Lambda_1^{*}\Gamma_1^{-1})$$

where

$$\left(\begin{array}{c} \Lambda_1^{*} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=2}^{T} E\left(w_t s_t' - G_\delta M^{-1} z_{t-1} v_t + \sum_{i=1}^{p}\sum_{i\le j}^{p} D_{i,j,t}\left[v_{it}v_{jt} - h_{ij}\left(\frac{t}{T}\right) - G_\delta M^{-1} z_{t-1} v_t\right]\right) \\ \times\left(w_t s_t' - G_\delta M^{-1} z_{t-1} v_t + \sum_{i=1}^{p}\sum_{i\le j}^{p} D_{i,j,t}\left[v_{it}v_{jt} - h_{ij}\left(\frac{t}{T}\right) - G_\delta M^{-1} z_{t-1} v_t\right]\right)' \\ G_\delta = \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T} E\left(\frac{\partial (y_t - x_t\beta_0 - \gamma_0 v_t^{*}) s_t'}{\partial \delta}\right) \\ M = \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T} E(z_{t-1} z_{t-1}') \\ D_{i,j,t} = E\left(\frac{\partial\left(y_t - x_t\beta_0 - \gamma_0 v_t^{*}\right) s_t'}{\partial h_{ij}\left(\frac{t}{T}\right)}\right) \end{array}\right)$$

3.4.2 Case 2

In this case, the model in equation (3.9) can be rewritten as,

$$y_t = x_t'\beta_0 + \sum_{i=1}^{p} v_{it}^{*}(\alpha_{00_i} + \alpha_{10_i} u_{t-1} v_{it-1}^{*}) + w_t \tag{3.10}$$

Unlike case 1, the model in equation (3.10) can't be estimated by OLS method, since u_t can't be identified separately without identifying β. Nevertheless, under the specification in equation (3.10), one can calculate $u_t(\beta) = y_t - x_t\beta$ given any fixed value of β, for $t = 1, \ldots, T$. Therefore, one can estimate β and u_t simultaneously. Let $\alpha_{00} =$

[2] This is similar as in the control function literature (see e.g. Kim (2004) and Kim (2010)).

$(\alpha_{00_1}, \dots, \alpha_{00_p})$, $\alpha_{10} = (\alpha_{10_1}, \dots, \alpha_{10_p})$, and $\theta = (\beta', \alpha'_{00}, \alpha'_{10})'$, and define $g(y_t, x_t, v_t^*; \theta) \equiv \left[y_t - x_t'\beta - \sum_{i=1}^{p} v_{it}^*(\alpha_{0_i} + \alpha_{1_i} u_{t-1}(\beta) v_{it-1}^*)\right]^2$ the estimators can be expressed as,

$$\widehat{\theta} = \arg\min_{\theta} \frac{1}{T} \sum_{t=1}^{T} g(y_t, x_t, v_t^*; \theta) \tag{3.11}$$

Theorem 3.2. Under assumption M, IV, and V

$$\widehat{\beta} = \beta_0 + o_p(1)$$
$$\widehat{\alpha}_0 = \alpha_{00} + o_p(1)$$
$$\widehat{\alpha}_1 = \alpha_{01} + o_p(1)$$

The proof of Theorem 3.2 is given in Appendix. Theorem 3.2 shows that the proposed estimator in equation (3.11) can consistently estimate the model parameters. The following theorem shows its asymptotic distribution,

Theorem 3.3 Under assumption M, IV, and V

$$\sqrt{T} \begin{pmatrix} \widehat{\beta} - \beta_0 \\ \widehat{\alpha}_0 - \alpha_{00} \\ \widehat{\alpha}_1 - \alpha_{01} \end{pmatrix} \to_d N(0, \Gamma_2^{-1} \Lambda_2 \Gamma_2^{-1})$$

where

$$\Gamma_2 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} E\left(\frac{\partial^2 g(y_t, x_t, v_t^*; \theta)}{\partial\theta\partial\theta'} \right) \Bigg|_{\theta=\theta_0},$$
$$\Lambda_2 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} E\left(\frac{\partial g(y_t, x_t, v_t^*; \theta)}{\partial\theta} \frac{\partial g(y_t, x_t, v_t^*; \theta)}{\partial\theta'} \right) \Bigg|_{\theta=\theta_0}$$

Same as case 1, the estimators in equation (3.11) are infeasible in practice, and the feasible estimator can be obtained by the same three step approach in case 1, except we minimize the quadratic function in equation (3.11) in step 3. The extended Hausman-Wu Test in this case can also be performed as follows,

Corollary 3.3 Under the same assumptions in Theorem 3.1, a test for endogeneity can be established as

$$H_0 : \alpha_0 = \alpha_1 = 0$$

where the test statistics $\sqrt{T}\widehat{\gamma}'(\Gamma_2^{-1}\Lambda_2\Gamma_2^{-1})_*^{-1}\widehat{\gamma} \to_a \chi_p^2$, $(\Gamma_2^{-1}\Lambda_2\Gamma_2^{-1})_*$ is the $2p \times 2p$ bottom right block of the $\Gamma_2^{-1}\Lambda_2\Gamma_2^{-1}$ in Theorem 3.3.

Follow Newey (1994), the asymptotic distribution of the estimator in case 2 can be derived as,

Corollary 3.4 The feasible estimator $(\widetilde{\beta}, \widetilde{\alpha}_0, \widetilde{\alpha}_1)$ has the following asymptotic distribution,

$$\sqrt{T}\begin{pmatrix} \widetilde{\beta} - \beta_0 \\ \widetilde{\alpha}_0 - \alpha_{00} \\ \widetilde{\alpha}_1 - \alpha_{01} \end{pmatrix} \to_d N(0, \Gamma_2^{-1}\Lambda_2^*\Gamma_2^{-1})$$

where

$$\Lambda_1^* = \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T} E\left(\frac{\partial g(y_t, x_t, v_t^*; \theta)}{\partial \theta} - G_\delta M^{-1} z_{t-1} v_t + \sum_{i=1}^{p}\sum_{i\le j}^{p} D_{i,j,t}\left[v_{it}v_{jt} - h_{ij}\left(\frac{t}{T}\right) - G_\delta M^{-1} z_{t-1} v_t\right]\right)$$
$$\times \left(\frac{\partial g(y_t, x_t, v_t^*; \theta)}{\partial \theta} - G_\delta M^{-1} z_{t-1} v_t + \sum_{i=1}^{p}\sum_{i\le j}^{p} D_{i,j,t}\left[v_{it}v_{jt} - h_{ij}\left(\frac{t}{T}\right) - G_\delta M^{-1} z_{t-1} v_t\right]\right)'$$

$$G_\delta = \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T} E\left(\frac{\partial (y_t - x_t\beta_0 - \gamma_0 v_t^*) s_t'}{\partial \delta}\right)$$

$$M = \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T} E(z_{t-1}z_{t-1}')$$

$$D_{i,j,t} = E\left(\frac{\partial\left(y_t - x_t\beta_0 - \gamma_0 v_t^*\right) s_t'}{\partial h_{ij}\left(\frac{t}{T}\right)}\right)$$

3.5. Monte Carlo Simulation

In this section, we analyze the finite sample performance of the proposed methods. We consider two different data generating process (DGP) in the Monte Carlo experiments to fit with our case 1 and 2,

DGP 1

$$y_t = x_t\beta + u_t,$$

$$x_t = z_t\delta + v_t,$$

$$Corr(v_t, u_t) = \rho$$

$$\sigma_{vt}^2 = \sigma_{v0} + \left(\frac{T-t}{T}\right)\sigma_{v1}$$

where $T = 200$, $\sigma_{v0} = 0.1$, $\sigma_{v1} = 5$, and $\delta = 0.01/\sqrt{T}$. $u_t^* \sim N(0,1)$, $v_t^* \sim N(0,1)$.

DGP 2

$$y_t = x_t\beta + u_t,$$

$$x_t = z_t\delta + v_t,$$

$$Cov(v_t^*, u_t) = \alpha_0 + \alpha_1 u_{t-1} v_{t-1}^*$$

$$\sigma_{vt}^2 = \sigma_{v0} + \left(\frac{T-t}{T}\right)\sigma_{v1}$$

where $T = 200$, $\sigma_{v0} = 0.1$, $\sigma_{v1} = 5$, and $\delta = 0.01/\sqrt{T}$. $u_t^* \sim N(0,1)$, $v_t^* \sim N(0,1)$.

First, Table 3.1 reports estimation results under each DGP. The upper panel of column 2 shows the estimation results under DGP 1 when $\rho = 0.9$, whereas the bottom panel corresponds to $\rho = 0.5$. The third Column presents the estimation results under GDP 2 when $\alpha_0 = 0.5$, $\alpha_1 = 0.1$ (upper panel), and $\alpha_0 = 0.05$, $\alpha_1 = 0.3$ (lower panel). The results shows that the proposed methods can estimate the parameter of interest, β, precisely in all different cases.

Second, follow the weak instrument literature, we focus on the finite sample power of the test, $H_0 : \beta = 0$, against fixed alternative. Figure 3.1 presents the power curve under DGP 1. Solid line is the power curve for the proposed method in case 1. The power curve of the proposed method goes to 1 very fast, while the power curves for the three robust test are almost flat line. Figure 3.2 presents the power curve under DGP 2, which shows the same pattern as in Figure 3.1. Thus we conclude that the proposed methods perform much better than the weak instrument robust test when there exist time vary volatility in the endogenous variable.

3.6. Elasticity of Intertemporal Substitution: Consumption CAPM

The elasticity of intertemporal substitution (EIS) in consumption is a parameter of central importance in macroeconomics and finance. Let δ be the subjective discount factor and γ be the coefficient of relative risk aversion. The Epstein-Zin (1989, 1991) objective function is defined recursively by

$$U_t = \left[(1-\delta)C_t^{(1-1/\psi)} + \delta\left(E_t U_{t+1}^{1-\gamma}\right)^{(1-1/\psi)(1-\gamma)}\right]^{1/(1-1/\psi)} \tag{3.12}$$

where C_t is consumption at time t, $E_t(\cdot)$ is the conditional expectation on information up to t. The representative household maximizes the objective function in equation (3.12), subject to the intertemporal budget constraint

$$W_{t+l} = (1 + R_{W,t+l})(W_t - C_t) \tag{3.13}$$

where W_{t+1} is the household's wealth and $1 + R_{W,t+1}$ is the gross real return on the portfolio of all invested wealth at $t+1$. Epstein and Zin (1991) show that equations (3.12) and (3.13) together imply an Euler equation of the form

$$E_t\left[\left(\delta\left(\frac{C_{t+1}}{C_t}\right)^{-1/\psi}\right)^{\theta}\left(\frac{1}{1+R_{W,t+1}}\right)^{1-\theta}(1+R_{i,t+1})\right] = 1$$

where $\theta = (1-\gamma)/(1-1/\psi)$, $1 + R_{i,t+1}$ is the gross real return on asset i.

Given a vector of instruments, the parameters δ, γ, and ψ can be estimated by GMM through the above nonlinear Euler equation. This is the approach taken by Hansen and Singleton (1982) and Epstein and Zin (1991). As noted by Epstein and Zin (1991), the difficulty with this approach is that it requires knowledge of returns on the wealth portfolio, which includes returns on human capital. Hence, Roll's (1977) critique of the testability of CAPM applies.

Alternatively, the above nonlinear Euler equation can be log linearized as

$$E_t \Delta c_{t+1} = \tau_{it} + \psi E_t r_{i,t+1} \tag{3.14}$$

where

$$\Delta c_{t+1} = \log\left(\frac{C_{t+1}}{C_t}\right)$$
$$r_{i,t+1} = \log(1 + R_{i,t+1})$$
$$\tau_{it} = \psi\left(\log\delta + \frac{1-\theta}{2}Var(r_{w,t+1}|F_{t-1})\right) + \frac{\theta}{2\psi}Var(\Delta c_{t+1}|F_{t-1})$$
$$+\frac{\psi}{2}Var(r_{i,t+1}|F_{t-1}) - \theta Cov(r_{i,t+1}, \Delta c_t|F_{t-1})$$

To estimate the EIS, one typically replaces the expectations in equation (3.14) by the actual observations and uses the following regression equation,

$$\Delta c_{t+1} = \tau_i + \psi r_{i,t+1} + u_{i,t+1} \tag{3.15}$$

where Δc_{t+1} is the consumption growth at time $t+1$, $r_{i,t+1}$ is the real return on asset i at time $t+1$. In general, two types of assets are usually considered, risk free asset (treasury bonds) and the risky asset (aggregate stock return). The error $u_{i,t+1}$, which is linear in the innovation to consumption growth and asset return, is correlated with the regressors $r_{i,t+1}$. Thus, one usually need instrument variables to estimate equation (3.15). Given a vector of instruments Z_t is uncorrelated with the error, we have the moment restriction, $E(Z_t u_{i,t+1}) = 0$, so that ψ can be estimated by the following reduced form equations,

$$r_{i,t+1} = z_t' \delta + v_{i,t+1} \tag{3.16}$$

In the finance literature, equation (3.16) is also called predictive regression, where we use lagged variable, z_t, to predict return on assets, $r_{i,t+1}$. However, as noted in Cochrane (2008), returns on assets, especially aggregate stock return, are notoriously hard to predict, which implies z_t are only weakly correlated with $r_{i,t+1}$. Therefore, we will have weak instrument problem when we estimate the model in equations (3.15)-(3.16), this is consistent with the statement in Neely, Roy, and Whiteman (2001) and Campbell (2003), and Yogo (2004).

In order to overcome the weak instrument problem, Yogo (2004) apply the weak instrument robust test to estimate the model in equations (3.15)-(3.16). The author considered eleven developed countries in Campbell (2003)'s data set. The instruments he used are the nominal interest rate, inflation, consumption growth rate, and log dividend-price ratio. To avoid the time aggregation problem (Hall, 1988), all instruments are lagged twice. The last three column in Table 3.3 shows Yogo's (2004) estimation results (Yogo (2004), table 5). For 22 out of 33 confidence intervals, the three widely used weak instrument robust tests can only report real line as confidence interval.

We propose to estimate model in equations (3.15)-(3.16) by the CF approach. Notice that there exist strong empirical support of the time-varying unconditional variance of the aggregate stock return, so equation (3.15) fit into our framework perfectly. For comparison purpose, we use the same data set and instruments as Yogo (2004). Figures 3.3-3.13 show the aggregate stock return and the estimated volatility based on the generalized kernel estimator. These figures show that the generalized kernel estimator can capture the volatility clustering (Mandelbrot, 1963) quite well.

Table 3.1 shows the point estimates obtained by the CF approach and the conventional GMM approach. For most of the countries, GMM estimator are larger than the CF estimator. This is consistent with the weak instrument literature. In particular, when instruments are weak, the GMM or 2SLS estimator will bias toward the OLS estimator. In contrast, the proposed method is robust to weak instrument and thus the point estimate will be consistent. Table 3.2 shows the confidence intervals obtained by the CF approach and the three weak instrument robust tests. The AR, LM, and CLR tests all report very uninformative confidence intervals for most of the countries as in Yogo (2004). This is consistent with our argument that the power of the weak instrument robust test strongly depends on the strength of the instruments. When the instruments are really weak, one can hardly reject any null hypothesis. In contrast, the confidence interval obtained by CF approach is much tighter and also consistent with the findings in Hall (1988).

3.7. Concluding Remarks

We propose a feasible CF approach to deal with the weak instrument problem in this paper. We show that the estimator is root-n consistent and is asymptotic normal. We also propose to use wild bootstrap to estimate the asymptotic variance of the estimator. Simulation studies show that the CF approach has significant power advantage compare with the weak instrument robust test. We apply the CF approach to estimate EIS as in Yogo (2004) and obtain a much tighter confidence interval.

Appendix 3.A.

3.A.1. Proof of Lemma 3.1

By construction, $w_t \equiv u_t - v_t^{*\prime}\gamma_t = u_t - v_t^{*\prime}\rho\sigma_{ut}$.

For (i), it follows immediately from assumption M.

For (ii),

$$\begin{aligned} E(x_t w_t|F_{t-1}) &= E\left(\left((I_p \otimes z_{t-1}')\delta_0 + v_t\right)\left(u_t - v_t^{*\prime}\rho\sigma_{ut}\right)|F_{t-1}\right) \\ &= E\left(v_t\left(u_t - v_t^{*\prime}\rho\sigma_{ut}\right)|F_{t-1}\right) \\ &= E\left(v_t u_t|F_{t-1}\right) - E\left(v_t v_t^{*\prime}|F_{t-1}\right)\rho\sigma_{ut} \\ &= \left(\Omega_{vt}^{-\frac{1}{2}} - \Omega_{vt}^{-\frac{1}{2}} E\left(v_t^* v_t^{*\prime}|F_{t-1}\right)\right)\rho\sigma_{ut} \\ &= 0 \end{aligned}$$

where the second equation follows by assumption M, and the last follows $E\left(v_t^* v_t^{*\prime}|F_{t-1}\right) = I_p$ by construction.

Proofs for (iii) and (iv) follows the same logic as (ii) and thus are omitted.

3.A.2. Proof of Theorem 3.1

The proof of Theorem 3.1 relies on a set of Lemma which we present below.

Lemma 3.A.1:

$$\frac{1}{T}\sum_{t=1}^{T} s_t' s_t \rightarrow_p \begin{pmatrix} \int\Omega & \int\Omega^{\frac{1}{2}} \\ \int\Omega^{\frac{1}{2}} & 1 \end{pmatrix}$$

Proof:

It suffices to show,

(i) $\frac{1}{T}\sum_{t=1}^{T} x_t x_t' \rightarrow_p \int\Omega$.

(ii) $\frac{1}{T}\sum_{t=1}^{T} x_t v_t^{*\prime} \rightarrow_p \int\Omega^{\frac{1}{2}}$.

For (i)

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} x_t x_t' = \frac{1}{T}\sum_{t=1}^{T}(I_p \otimes z_{t-1}')\delta\delta'(I_p \otimes z_{t-1}')' + \frac{1}{T}\sum_{t=1}^{T}(I_p \otimes z_{t-1}')\delta(\Omega^{\frac{1}{2}} v_t^*)' \\ + \frac{1}{T}\sum_{t=1}^{T}(\Omega^{\frac{1}{2}} v_t^*)\delta'(I_p \otimes z_{t-1}')' + \frac{1}{T}\sum_{t=1}^{T}(\Omega^{\frac{1}{2}} v_t^*)(\Omega^{\frac{1}{2}} v_t^*)' \end{aligned} \tag{3.A.1}$$

For the first term, IV(ii) and (iii) ensures the following convergence result,

$$\frac{1}{T}\sum_{t=1}^{T}(I_p \otimes z'_{t-1})\delta\delta'(I_p \otimes z'_{t-1})' = \frac{1}{T^2}\sum_{t=1}^{T}(I_p \otimes z'_{t-1})CC'(I_p \otimes z'_{t-1})' \to_p 0 \qquad (3.A.2)$$

For the second and third term, follow the orthogonality conditions in IV(i),

$$\frac{1}{T}\sum_{t=1}^{T}(I_p \otimes z'_{t-1})\delta(\Omega^{\frac{1}{2}}v_t^*)' \to_p 0 \qquad (3.A.3)$$

$$\frac{1}{T}\sum_{t=1}^{T}(\Omega^{\frac{1}{2}}v_t^*)\delta'(I_p \otimes z'_{t-1})' \to_p 0 \qquad (3.A.4)$$

For the fourth term, by the law of large number for martingale difference sequence and assumption V,

$$\frac{1}{T}\sum_{t=1}^{T}(\Omega^{\frac{1}{2}}v_t^*)(\Omega^{\frac{1}{2}}v_t^*)' \to_p \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E\left((\Omega^{\frac{1}{2}}v_t^*)(\Omega^{\frac{1}{2}}v_t^*)'\right) = \int \Omega \qquad (3.A.5)$$

Thus the desired result can be obtained by combining the results in equations (3.A.1) to (3.A.5).

For (ii),

$$\frac{1}{T}\sum_{t=1}^{T}x_t v_t^{*\prime} = \frac{1}{T}\sum_{t=1}^{T}\left(I_p \otimes z'_{t-1}\right)v_t^{*\prime} + \frac{1}{T}\sum_{t=1}^{T}\Omega^{\frac{1}{2}}v_t^* v_t^{*\prime} \qquad (3.A.6)$$

The first term converge to 0 follow the same argument in (i). For the second term

$$\frac{1}{T}\sum_{t=1}^{T}\Omega^{\frac{1}{2}}v_t^* v_t^{*\prime} \to_p \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E\left(\Omega^{\frac{1}{2}}v_t^* v_t^{*\prime}\right) = \int \Omega^{\frac{1}{2}}$$

Lemma 3.A.2:

(i) $\frac{1}{T}\sum_{t=1}^{T}\left(v_t w_t^2 v_t' - E\left(v_t w_t^2 v_t'\right)\right) \to_p 0$

(ii) $\frac{1}{T}\sum_{t=1}^{T}\left(v_t w_t^2 v_t^{*\prime} - E\left(v_t w_t^2 v_t^{*}\right)\right) \to_p 0$

(iii) $\frac{1}{T}\sum_{t=1}^{T}\left(v_t^* w_t^2 v_t^{*\prime} - E\left(v_t^* w_t^2 v_t^{*\prime}\right)\right) \to_p 0$

Proof:

For (i)

$$v_t w_t^2 v_t' = \begin{bmatrix} v_{1t}^2 w_t^2 & v_{1t}v_{2t}w_t^2 & \dots & v_{1t}v_{pt}w_t^2 \\ v_{2t}v_{1t}w_t^2 & v_{2t}^2 w_t^2 & \dots & v_{2t}v_{pt}w_t^2 \\ \vdots & \vdots & \dots & \vdots \\ v_{pt}v_{1t}w_t^2 & v_{pt}v_{2t}w_t^2 & \dots & v_{pt}^2 w_t^2 \end{bmatrix}$$

Note that v_{it} and w_t are α-mixing by Theorem 14.1 in Davidson (1994), thus $v_{it}v_{jt}$ and w_t^2 are mixing and therefore near-epoch dependent in L^2 norm on the α-mixing sequence $\{v_t^*, u_t^*\}$. So by Theorem 17.9 in Davidson (1994), $v_{it}v_{jt}w_t^2$ is near-epoch dependent in L^1 norm on $\{v_t^*, u_t^*\}$. Let $\alpha = \min\{\alpha_1, \alpha_{21}, \dots, \alpha_{2p}\}$, we have,

$$E\left[v_{it}v_{jt}w_t^2\right]^\alpha \le Ev_{it}^{4\alpha}Ev_{jt}^{4\alpha}Ew_t^{4\alpha} < \infty$$

by assumption M and Cauchy-Schwartz inequality. Thus by the law of large numbers for L^1-mixingales (Andrews, 1988), we have the desired results. Proofs for (ii) and (iii) follows the same logic and thus are omitted.

Lemma 3.A.3:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_t' w_t \to_d N(0, \Omega_2)$$

Proof:

Follow Cramer-Wold theorem, it suffices to show,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \lambda' s_t' w_t \to_d N(0, \lambda'\Omega_2\lambda) \tag{3.A.7}$$

for every fixed $2p\times 1$ vector $\lambda = (\lambda_1, \lambda_2)'$, where $\lambda_1 = (\lambda_{11}, \lambda_{12}, \dots, \lambda_{1p})$, $\lambda_2 = (\lambda_{21}, \lambda_{22}, \dots, \lambda_{2p})$.■

First,

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\lambda' s_t' w_t &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\lambda_1' x_t w_t + \lambda_2' v_t^* w_t) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(\lambda_1'(I_p\otimes z_{t-1}')\delta w_t + \lambda_1' v_t w_t + \lambda_2' v_t^* w_t\Big) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\lambda_1' v_t w_t + \lambda_2' v_t^* w_t) + o_p(1) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\lambda' s_t^{*\prime} w_t + o_p(1)
\end{aligned}
$$

where $s_t^* \equiv (v_t, v_t^*)$, and the third equation follows from assumption IV(i) and (iii).

Second, note that $\{\lambda' s_t^{*\prime} w_t, F_t\}$ is a martingale difference sequence by Lemma 3.1. Then the following convergence results follows Lemma 3.A.2

$$\frac{1}{T}\sum_{t=1}^{T}\lambda' w_t^2 s_t^{*\prime} s_t^* \lambda \to_p \lambda'\Lambda\lambda$$

Moreover, we have

$$
\begin{aligned}
E(\lambda' s_t^* w_t)^{2\alpha} &= \|\lambda_1' v_t w_t + \lambda_2' v_t^* w_t\|_{2\alpha}^{2\alpha} \\
&\leq (\sum_{i=1}^{p}\|\lambda_{1i} v_{ti} w_t\|_{2\alpha} + \sum_{i=1}^{p}\|\lambda_{2i} v_{ti}^* w_t\|_{2\alpha})^{2\alpha}
\end{aligned}
\tag{3.A.8}
$$

where the inequality follow by Minkowski inequality.

Furthermore, we can show that each component in equation (3.A.8) are finite,

$$\|\lambda_{1i} v_{ti} w_t\|_{2\alpha} = |\lambda_{1i}|(E(v_{it} w_t)^{2\alpha})^{\frac{1}{2\alpha}} \leq |\lambda_{1i}|(E v_{it}^{4\alpha} E w_t^{4\alpha})^{\frac{1}{4\alpha}} < \infty$$

$$\|\lambda_{2i} v_{ti}^* w_t\|_{2\alpha} = |\lambda_{2i}|(E(v_{it}^* w_t)^{2\alpha})^{\frac{1}{2\alpha}} \leq |\lambda_{2i}|(E v_{it}^{*\,4\alpha} E w_t^{4\alpha})^{\frac{1}{4\alpha}} < \infty$$

where the inequality follow by Cauchy-Schwartz inequality.

So we can conclude that equation (3.A.8) is finite, by the central limit theorem for the martingale differences, equation (3.A.7) holds and then we have the desired results.

The proof of Theorem 3.1 follows Lemma 3.A.1-3.A.3.

3.A.3. Proof of Theorem 3.2

Let $\theta = (\beta', \alpha_0', \alpha_1')'$, we have the following result uniformly in θ,

$$
\begin{aligned}
Q_n(\theta) \equiv & \frac{1}{T}\sum_{t=1}^{T}\left(y_t - x_t'\beta - \sum_{i=1}^{p} v_{it}^*(\alpha_{0_i} + \alpha_{1_i} u_{t-1}(\beta) v_{it-1}^*)\right)^2 \\
= & \frac{1}{T}\sum_{t=1}^{T}\left(u_t + x_t'(\beta_0 - \beta) - \sum_{i=1}^{p} v_{it}^*(\alpha_{0_i} + \alpha_{1_i}(u_{t-1} + x_{t-1}'(\beta_0 - \beta)) v_{it-1}^*)\right)^2 \\
= & \frac{1}{T}\sum_{t=1}^{T}\left(u_t + v_t'(\beta_0 - \beta) - \sum_{i=1}^{p} v_{it}^*(\alpha_{0_i} + \alpha_{1_i}(u_{t-1} + v_{t-1}'(\beta_0 - \beta)) v_{it-1}^*)\right)^2 + o_p(1)
\end{aligned}
$$

where the last equation follows assumption IV(iii).

$Q_n(\theta)$ is continuous everywhere in θ. We also have the following boundedness result,

$$
\begin{aligned}
& E\left(u_t + v_t'(\beta_0 - \beta) - v_t^{*\prime}\rho\left(\sigma_{ut}^2 + Cov(u_t, v_t)'(\beta_0 - \beta) + (\beta_0 - \beta)'\Omega_{vt}(\beta_0 - \beta)\right)^{\frac{1}{2}}\right)^{4\alpha} \\
= & \left\| u_t + v_t'(\beta_0 - \beta) - v_t^{*\prime}\rho\left(\sigma_{ut}^2 + Cov(u_t, v_t)'(\beta_0 - \beta) + (\beta_0 - \beta)'\Omega_{vt}(\beta_0 - \beta)\right)^{\frac{1}{2}}\right\|_{4\alpha}^{4\alpha} \\
\leq & \left(\|u_t\|_{4\alpha} + \|v_t'(\beta_0 - \beta)\|_{4\alpha} + \left\| v_t^{*\prime}\rho\left(\sigma_{ut}^2 + Cov(u_t, v_t)'(\beta_0 - \beta) + (\beta_0 - \beta)'\Omega_{vt}(\beta_0 - \beta)\right)^{\frac{1}{2}}\right\|_{4\alpha}\right)^{4\alpha} \\
< & \infty
\end{aligned}
$$

where the inequality follow by Minkowski and the finiteness follows assumption M and IV.

Define $Q_n^*(\theta) = Q_n(\theta) - Q_n(\theta_0)$, where $\theta_0 = (\beta_0', \alpha_{00}', \alpha_{10}')'$. Note that u_t and v_t are α mixing by assumption M, thus $\left(u_t + v_t'(\beta_0 - \beta) - \sum_{i=1}^{p} v_{it}^*(\alpha_{00_i} + \alpha_{10_i}(u_{t-1} - v_{t-1}'(\beta - \beta_0)) v_{it-1}^*)\right)$ is near-epoch dependent in L^1 norm on the α-mixing sequence $\{v_t^*, u_t^*\}$. By the law of large numbers for L^1-mixingales (Andrews, 1988) and the boundedness result above, we can show that $Q_n^*(\theta) \to pE(Q_n^*(\theta))$ uniformly in θ. Consistency then follows if the minimum is unique.

It can be shown that

$$
\begin{aligned}
E\left(\widehat{Q}_n^*(\theta)\right) = {} & E\left[\frac{1}{T}\sum_{t=2}^{T}\left(y_t - x_t'\beta - \sum_{i=1}^{p} v_{it}^*(\alpha_{0_i} + \alpha_{1_i}(u_{t-1} + v_{t-1}'(\beta_0-\beta))v_{it-1}^*)\right)^2\right] \\
& - E\left[\frac{1}{T}\sum_{t=2}^{T}\left(y_t - x_t'\beta_0 - \sum_{i=1}^{p} v_{it}^*(\alpha_{00_i} + \alpha_{10_i}u_{t-1}v_{it-1}^*)\right)^2\right] \\
= {} & E\left[\frac{1}{T}\sum_{t=2}^{T}\left(x_t'\beta + \sum_{i=1}^{p} v_{it}^*(\alpha_{0_i} + \alpha_{1_i}(u_{t-1} + v_{t-1}'(\beta_0-\beta))v_{it-1}^*)\right)^2\right] \\
& - E\left[\frac{1}{T}\sum_{t=2}^{T} 2y_t\left(x_t'\beta + \sum_{i=1}^{p} v_{it}^*(\alpha_{0_i} + \alpha_{1_i}(u_{t-1} + v_{t-1}'(\beta_0-\beta))v_{it-1}^*)\right)\right] \\
& - E\left[\frac{1}{T}\sum_{t=2}^{T}\left(x_t'\beta + \sum_{i=1}^{p} v_{it}^*(\alpha_{00_i} + \alpha_{10_i}u_{t-1}v_{it-1}^*)\right)^2\right] \\
& + E\left[\frac{1}{T}\sum_{t=2}^{T} 2y_t\left(x_t'\beta + \sum_{i=1}^{p} v_{it}^*(\alpha_{00_i} + \alpha_{10_i}u_{t-1}v_{it-1}^*)\right)\right] \\
= {} & E\Bigg\{\frac{1}{T}\sum_{t=2}^{T}\Bigg[x_t'(\beta_0-\beta) + \sum_{i=1}^{p} v_{it}^*\left((\alpha_{00_i}-\alpha_{0_i}) + (\alpha_{10_i}-\alpha_{1_i})u_{t-1}v_{it-1}^*)\right) \\
& \qquad + \sum_{i=1}^{p}\alpha_{1_i}v_{i,t-1}v_{t-1}'(\beta_0-\beta)\Bigg]^2\Bigg\}
\end{aligned}
$$

Under assumption V, x has full rank and thus the above moment is uniquely minimized at $\theta = \theta_0$.

References

[1] Anderson, T. W., and Rubin, H. (1949), "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Annals of Mathematical Statistics* 20, 46-63.

[2] Andrews, I. (2016), "Conditional Linear Combination Tests for Weakly Identified Models," *Econometrica*, forthcoming.

[3] Andrews, D.W.K. and Guggenberger, P. (2015), "Identification- and Singularity-Robust Inference for Moment Condition," Cowles Foundation Discussion Papers 1978, Cowles Foundation for Research in Economics, Yale University.

[4] Andrews, D. W. K., Moreira, M. J. and Stock, J. H. (2006), "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74: 715-752.

[5] Andrews DWK, Moreira M. J., Stock J.H (2007), "Performance of Conditional Wald Tests in IV Regression with Weak Instruments," *Journal of Econometrics*, 139, 116-132.

[6] Angrist J.D. and Krueger, A.B. (1991), "Does Compulsory School Attendance Affect Schooling And Earnings?" *Quarterly Journal of Economics*, 106, 979-1014.

[7] Attanasio, O.P. and G. Weber (1989) "Intertemporal Substitution, Risk Aversion and the Euler Equation for Consumption," *Economic Journal*, 99, 59-73.

[8] Bansal, R., Kiku, D., and Yaron, A. (2007), "Risk For the Long Run: Estimation and Inference," Working paper.

[9] Bansal, R. and A. Yaron (2004), "Risks for the long run: A potential resolution of asset pricing puzzles," *Journal of Finance*, 59, 1481-1509.

[10] Campbell, J. Y. (2003), "Consumption-based asset pricing," *Handbook of the Economics of Finance*.

[11] Campbell J. Y., and Viceira, L. M. (1999), "Consumption and portfolio decisions when expected returns are time varying," *Quarterly Journal of Economics*, 114(2), 433-495.

[12] Cavaliere, G. (2004), "Unit root tests under time-varying variance shifts," *Econometric Reviews*, 23, 259-292.

[13] Chao, J., and N. Swanson (2007), "Alternative Specifications of the Bias and MSE of the IV Estimator under Weak Identification with an Application to Bias Correction,"

Journal of Econometrics, 137, 515-555.

[14] Dufour, J.M. (1997): "Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365-1387.

[15] Epstein, L. G., and Zin, S. E. (1989), "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoreical framework," *Econometrica*, 57(4), 937-969.

[16] Epstein, L. G., and Zin, S. E. (1991), "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: an Empirical Analysis," *Journal of Political Economy*, 99, 263-286.

[17] Engle, R. (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models," *Journal of Business and Economics Statistics*, 20, 339-350.

[18] Engle R. and J.G. Rangel (2008), "The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes," *Review of Financial Studies*, 21, 1187-1222.

[19] Fuller, W. A. (1977), "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, 939-953.

[20] Hahn, J., and Hausman, J.A. (2002), "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70, 163-189.

[21] Hall, R.E. (1988), "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 96, 339-357.

[22] Hansen, L.P. and K.J. Singleton (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269-1286.

[23] Hausman, J. (1978), "Specification tests in econometrics," *Econometrica* 46, 1251-1271.

[24] Hausman, J., J. H. Stock, and M. Yogo (2005), "Asymptotic Properties of the Hahn-Hausman Test for Weak Instruments," *Economics Letters*, 89, 333-342.

[25] Kleibergen, F. (2002), "Pivotal statistics for testing structural parameters in instrumental variables regression," *Econometrica*, 70(5), 1781-1803.

[26] Kim, S., Shephard, N., and Chib, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361-393.

[27] Moreira, M. J. (2003), "A conditional Likelihood ratio test for structure models," *Econo-*

metrica, 71(4), 1027-1048.

[28] Moreira, M. J. (2007), "Test with Correct size when instruments can be arbitrarily weak," *Journal of Econometrics*, 152, 131-140.

[29] Muller, H-G, (1991), "Smooth optimum kernel estimators near endpoints," *Biometrika*, 78, 521-530.

[30] Neely, C. J., Roy, A., and Whiteman, C. H. (2001), "Risk aversion versus intertemporal substitution: a case study of identification failure in the intertemporal consumption capital asset prcing model," *Journal of Business and Economic Statistics*, 19(4), 395-403.

[31] Nelson, C. and R. Startz (1990), "Some further results on the exact small sample properties of the instrumental variable estimator," *Econometrica*, 58, 967-976.

[32] Newey, W. K. (1994), "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349-1382.

[33] Omori, Y., Chib, S., Shephard, N., and Nakajima, J., (2007), "Stochastic volatility with leverage: Fast and efficient likelihood inference," *Journal of Econometrics*, 140, 425-449.

[34] Pagan, A. (1984), 11Econometric issues in the analysis of regressions with generated regressors," *International Economic Review* 25(1), 221-247.

[35] Phillips, P.C.B., Xu, K.-L., (2006), "Inference in Autoregression under Heteroskedasticity," *Journal of Time Series Analysis*, 27, 289-308.

[36] Roll, R. (1977), "A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory," *Journal of Financial Economics*, 4, 129-176.

[37] Rothenberg, T. J. (1984), "Approximating the distributions of econometric estimators of test statistics," in *Handbook of Econometrics*, Vol. II, ed. by Z Griliches and M.D. Intriligator. Amsterdam: North-Holland, 881-936.

[38] Schwert, G.W. (1989), "Why does Stock Market Volatility Change Over Time?," *Journal of Finance*, 5, 1115-1153.

[39] Staiger, D. and Stock, J. H. (1997), "Instrumental variables regression with weak instruments," *Econometrica* 65, 557-586.

[40] Stock, J. H., Wright, J. H., and Yogo, M. (2002), "A survey of weak instruments and weak identification in generalized method of moments," *Journal of Business and Economic*

Statistics, 20(4), 518-529.

[41] Stock, J.H., and Yogo, M. (2005), “Testing for Weak Instruments in Linear IV Regression,” In: Andrews DWK Identification and Inference for Econometric Models. New York: Cambridge University Press, 80-108.

[42] Vissing-Jorgensen, A. (2002), “Limited Asset Market Participation and the Elasticity of Intertemporal Substitution,” *Journal of Political Economy*, 110, 825-853.

[43] Wang, J. and E. Zivot, (1998), “Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments,” *Econometrica*, 66, 1389-1404.

[44] Wu, D. (1973), “Alternative tests of independence between stochastic regressors and disturbances,” *Econometrica* 41, 733-750.

[45] Yogo, M. (2004), “Estimating the elasticity of intertemporal substitution when instrument are weak,” *Review of Economics and Statistics* 86(3), 797-810

Table 3.1. Finite Sample Performance of the Proposed Estimator

	True Value	Case 1	Case 2
β	1	1.002	0.995
		(0.052)	(0.069)
γ	0.9	0.907	-
		(0.165)	-
α_0	0.5	-	0.476
		-	(0.199)
α_1	0.1	-	0.097
		-	(0.082)
β	2	2.002	1.999
		(0.057)	(0.059)
γ	0.5	0.503	-
		(0.173)	-
α_0	0.05	-	0.037
		-	(0.154)
α_1	0.3	-	0.251
		-	(0.097)

Note:

1. Second column shows the results based on DGP 1 and proposed method case 1. Second column shows the results based on DGP 2 and proposed method case 2.

Table 3.2. Heteroskedasticity-Robust Estimate of EIS

Country	Sample Period	CF 1	CF 2	GMM
AUL	1970.3 - 1998.4	0.01	0.03	0.16
CAN	1970.3 - 1999.1	0.15	0.12	0.26
FR	1970.3 - 1998.3	0.05	0.03	0.06
GER	1979.1 - 1998.3	0.09	0.05	0.10
ITA	1971.4 - 1998.1	0.07	0.04	0.08
JAP	1970.3 - 1998.4	0.07	0.08	0.12
NTH	1977.3 - 1998.4	0.11	0.06	0.13
SWD	1970.3 - 1999.2	0.04	0.04	0.04
SWT	1976.2 - 1998.4	0.08	0.06	0.09
UK	1970.3 - 1999.1	0.16	0.08	0.16
USA	1947.3 - 1998.4	0.10	0.09	0.18

Note:

1. This table reports the EIS estimated by the proposed approach and the conventional two step GMM estimator. CF1 and CF2 are estimated by model in case 1 and 2, respectively.

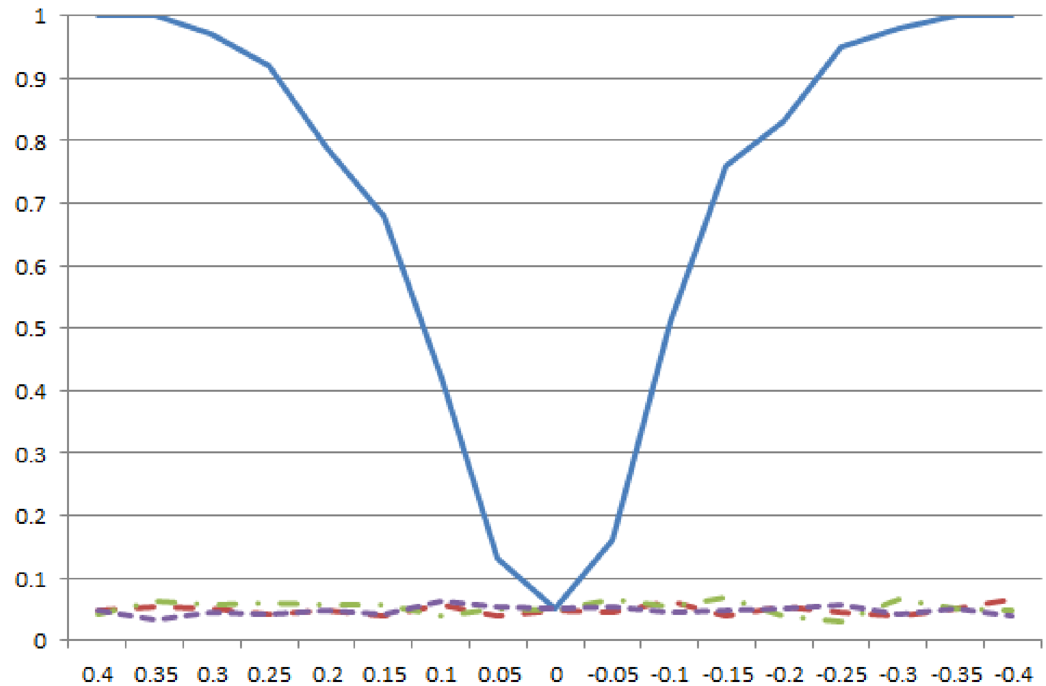
Table 3.3. Heteroskedasticity-Robust Estimate of EIS

Country	CF 1	CF 2	AR	CLR	LM
AUL	[-0.11,0.14]	[-0.13,0.19]	$[-\infty,\infty]$	$[-\infty,\infty]$	$[-\infty,\infty]$
CAN	[0.03,0.28]	[-0.10,0.34]	[0.02,4.03]	[0.05,0.35]	[0.04,0.41]
FR	[-0.03,0.15]	[-0.01,0.08]	[-0.28,0.20]	$[-\infty,\infty]$	[-0.16,0.11]
GER	[-0.01,0.19]	[-0.01,0.11]	$[-\infty,\infty]$	$[-\infty,\infty]$	$[-\infty,\infty]$
ITA	[-0.03,0.17]	[-0.02,0.10]	$[-\infty,\infty]$	$[-\infty,\infty]$	$[-\infty,\infty]$
JAP	[-0.05,0.19]	[-0.04,0.16]	[-0.05,0.32]	[-1.01,0.20]	[-0.02,0.21]
NTH	[0.03,0.19]	[-0.02,0.18]	$[-\infty,\infty]$	$[-\infty,\infty]$	$[-\infty,\infty]$
SWD	[-0.01,0.07]	[-0.02,0.11]	$[-\infty,\infty]$	$[-\infty,\infty]$	$[-\infty,\infty]$
SWT	[-0.04,0.21]	[-0.12,0.24]	$[-\infty,\infty]$	$[-\infty,\infty]$	$[-\infty,\infty]$
UK	[0.02,0.29]	[0.01,0.15]	[-0.51,-0.02]	$[-\infty,\infty]$	$[-\infty,\infty]$
USA	[0.11,0.27]	[0.05,0.13]	[-0.21,-0.02]	$[-\infty,\infty]$	$[-\infty,\infty]$

Note:

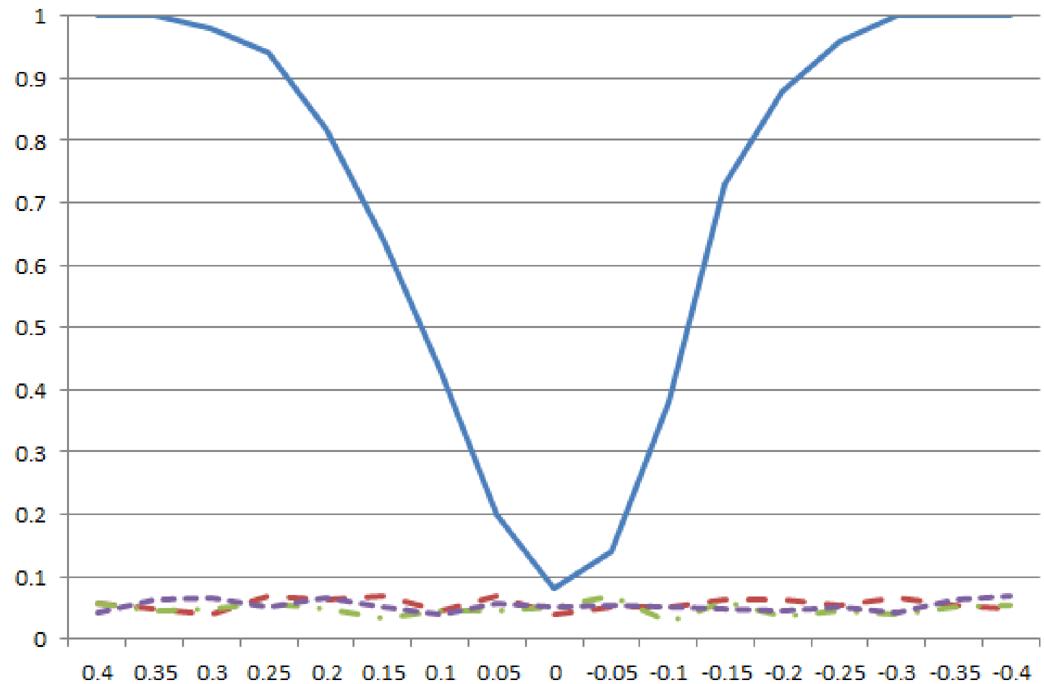
1. The sample period is the same as in Table 3.2.
2. This table reports the CI of EIS estimated by the proposed approach and the weak instrument robust tests. CF1 and CF2 are estimated by model in case 1 and 2, respectively.

Figure 3.1. Power curve under DGP 1.



Note: Power curve under DGP 1 for $H_0 : \beta = 0$ against fixed alternative. X axis denote the value under alternative hypothesis. y axis represent the test power. Solid line is the power curve for the estimation method in case 1. Dotted line is the power curve for CLR test. Dashed-dotted line is the power curve for LM test. Dashed line is the power curve for AR test.

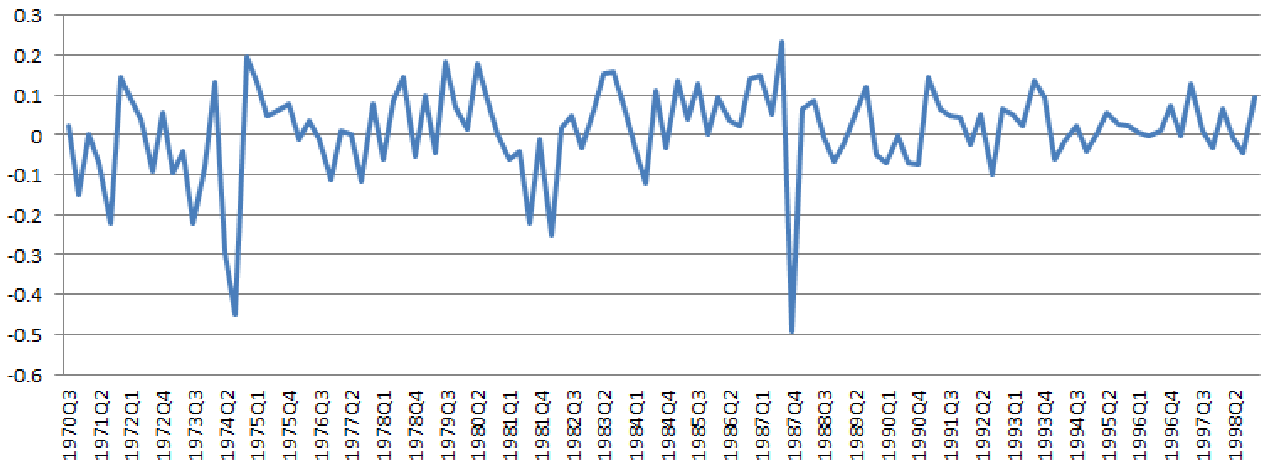
Figure 3.2. Power curve under DGP 2.



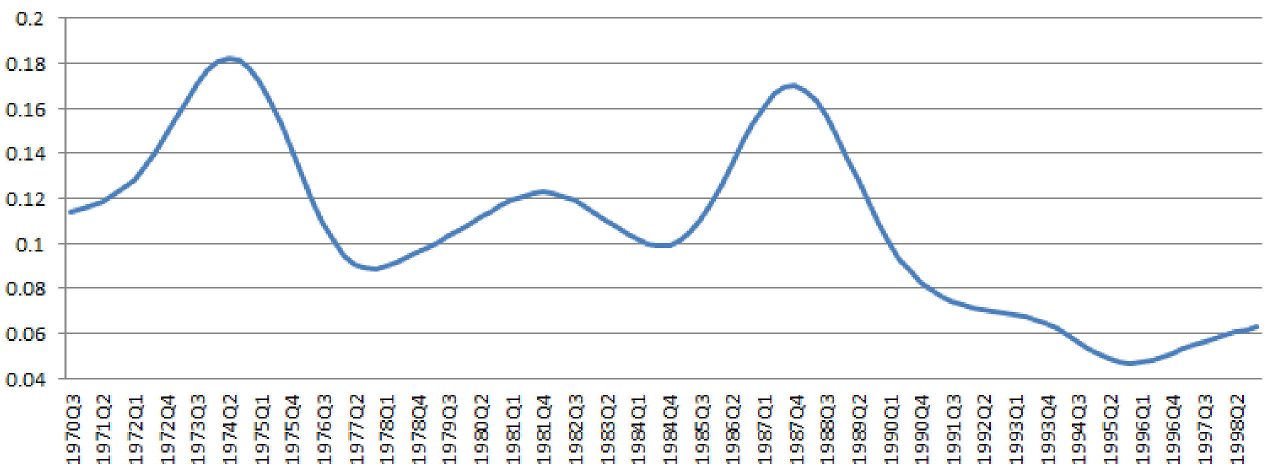
Note: Power curve under DGP 2 for $H_0 : \beta = 0$ against fixed alternative. X axis denote the value under alternative hypothesis. y axis represent the test power. Solid line is the power curve for the estimation method in case 2. Dotted line is the power curve for CLR test. Dashed-dotted line is the power curve for LM test. Dashed line is the power curve for AR test.

Figure 3.3. Aggregate Stock Return and Estimated Volatility: Australia

(a)



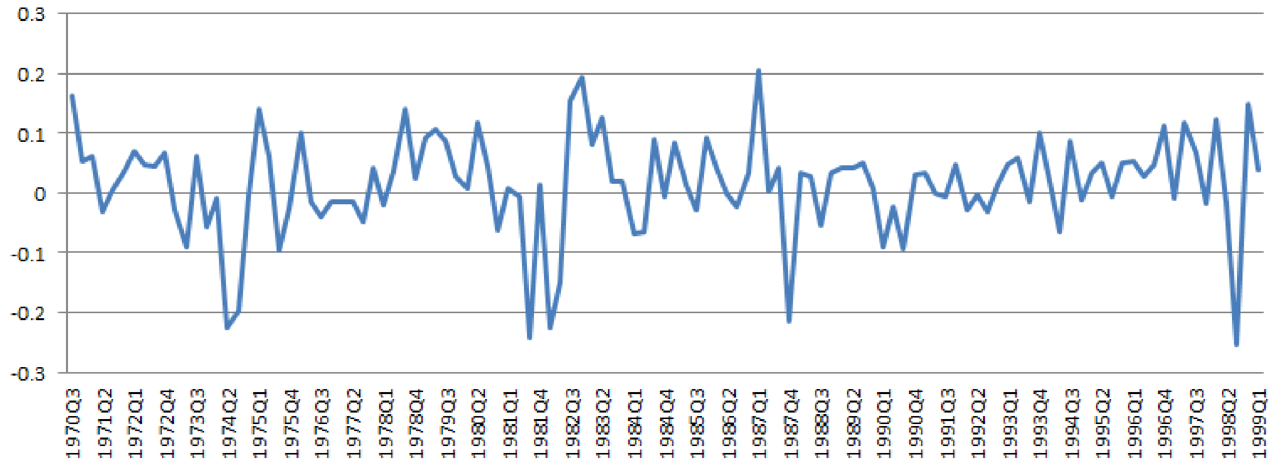
(b)



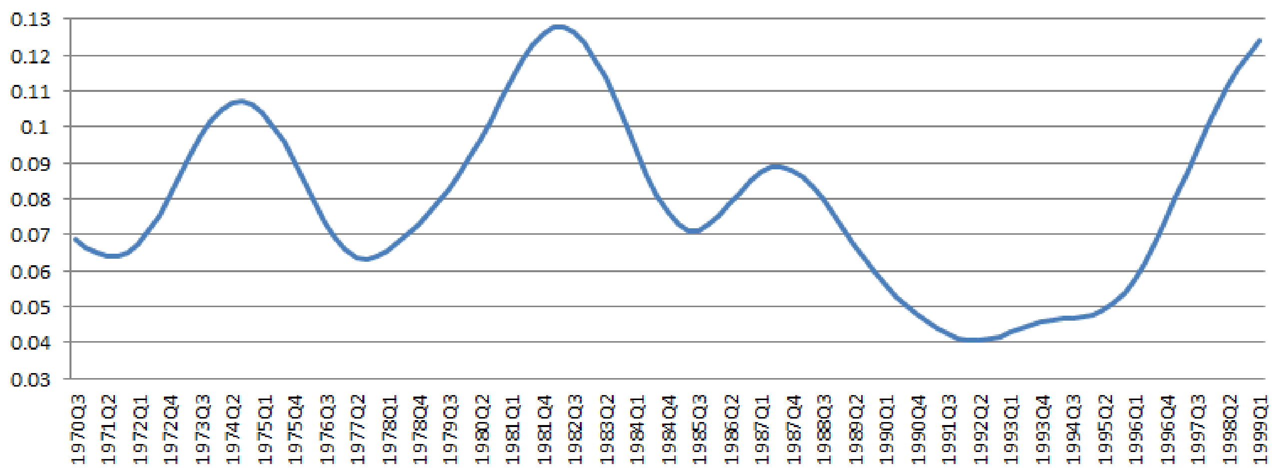
Note: (a) Aggregate Stock Return. (b) Estimated Volatility.

Figure 3.4. Aggregate Stock Return and Estimated Volatility: Canada

(a)



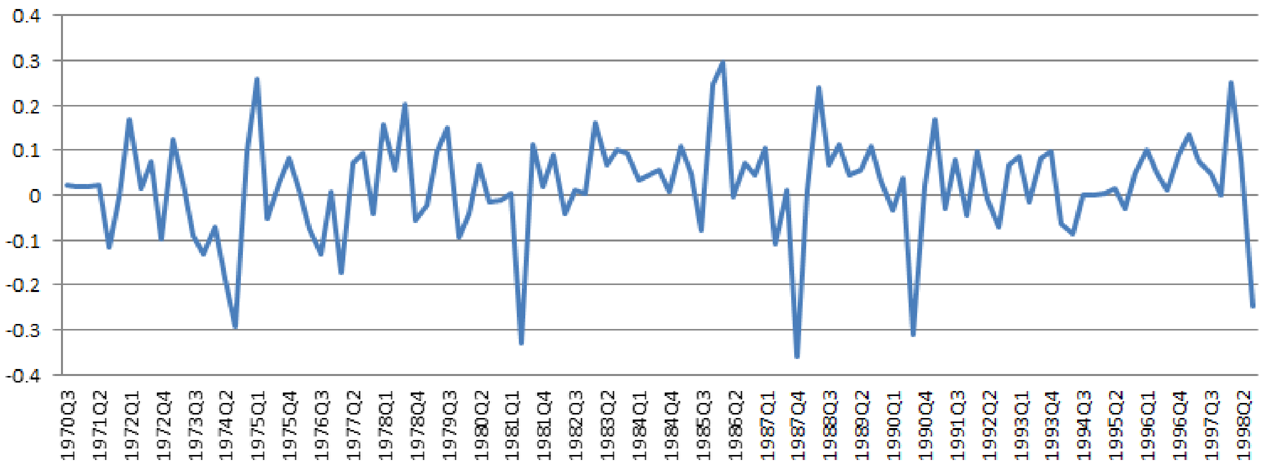
(b)



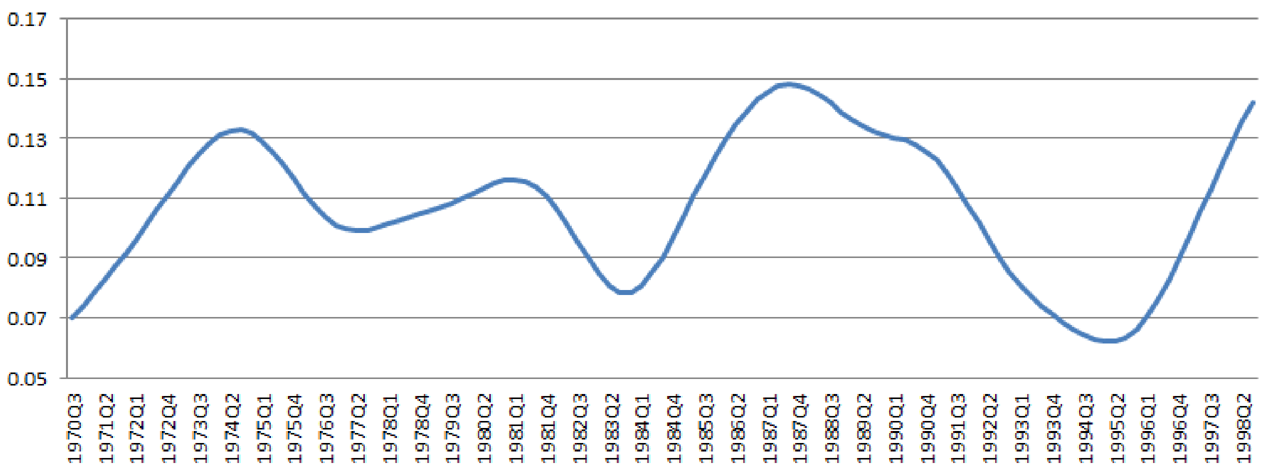
Note: (a) Aggregate Stock Return. (b) Estimated Volatility.

Figure 3.5. Aggregate Stock Return and Estimated Volatility: France

(a)



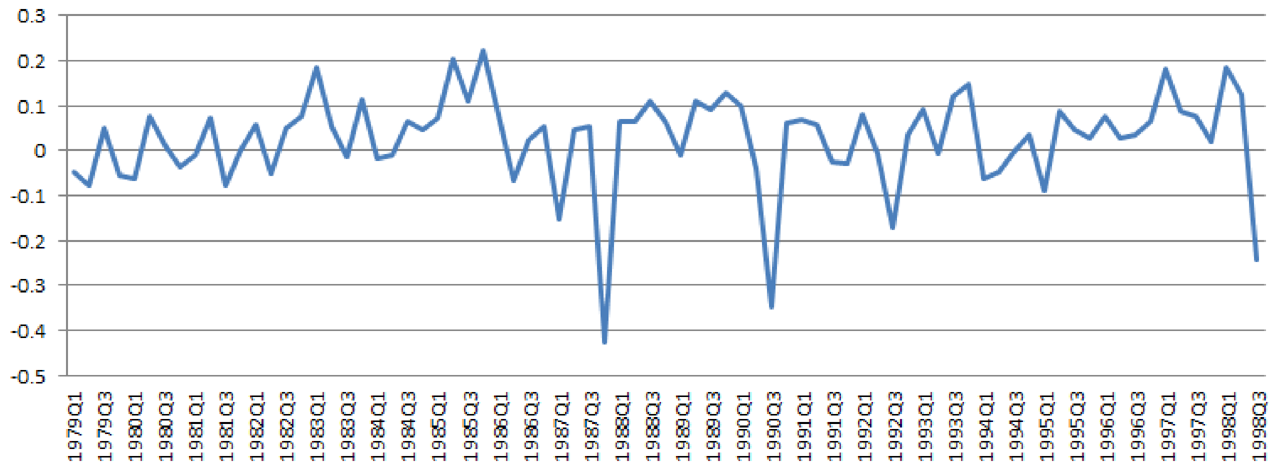
(b)



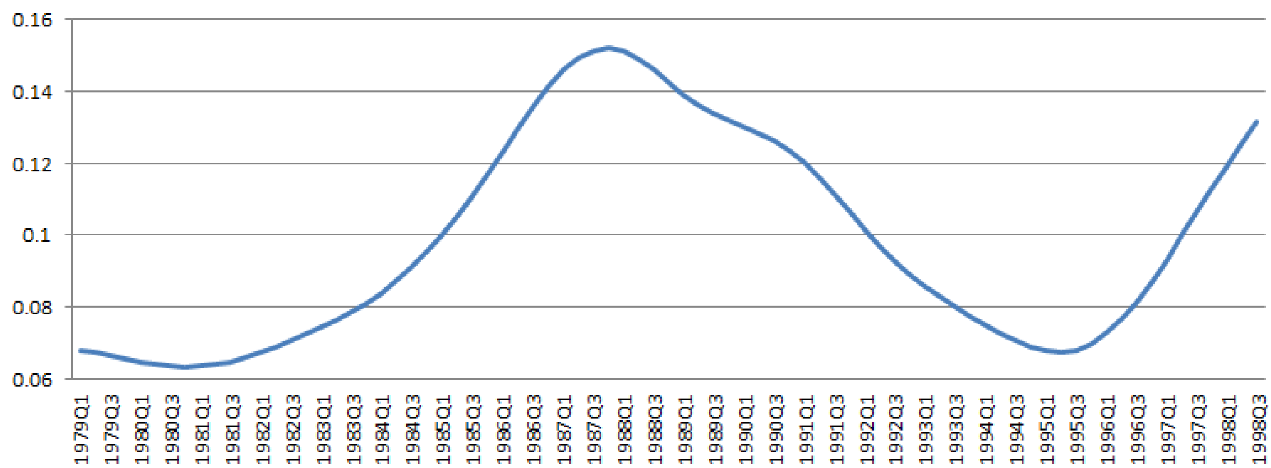
Note: (a) Aggregate Stock Return. (b) Estimated Volatility.

Figure 3.6. Aggregate Stock Return and Estimated Volatility: Germany

(a)



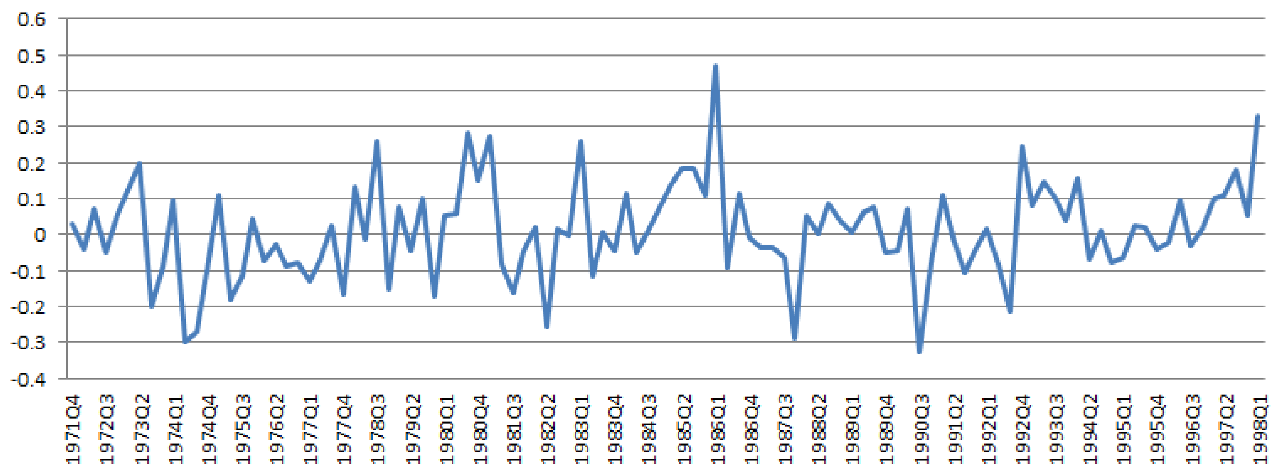
(b)



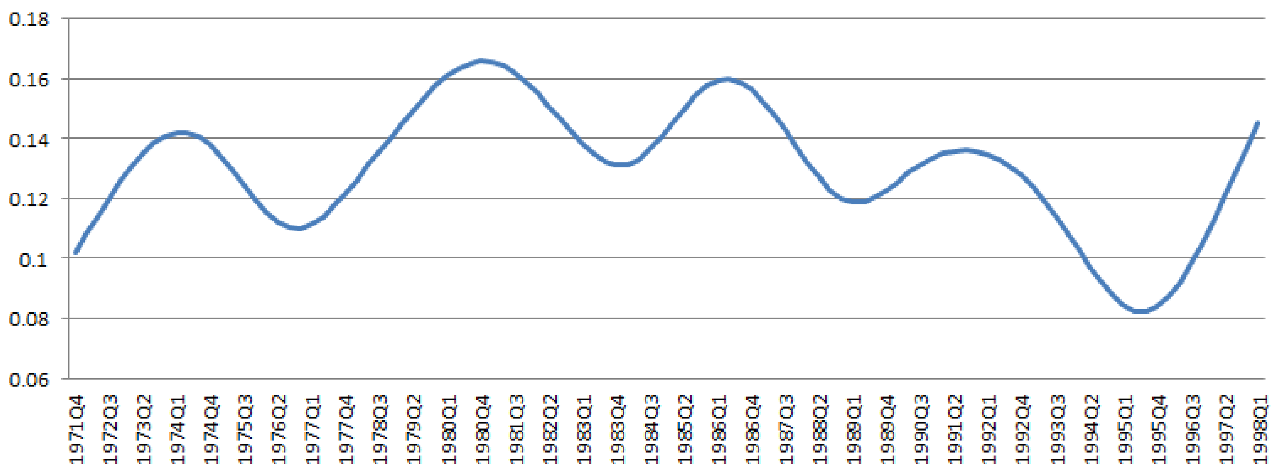
Note: (a) Aggregate Stock Return. (b) Estimated Volatility.

Figure 3.7. Aggregate Stock Return and Estimated Volatility: Italy

(a)



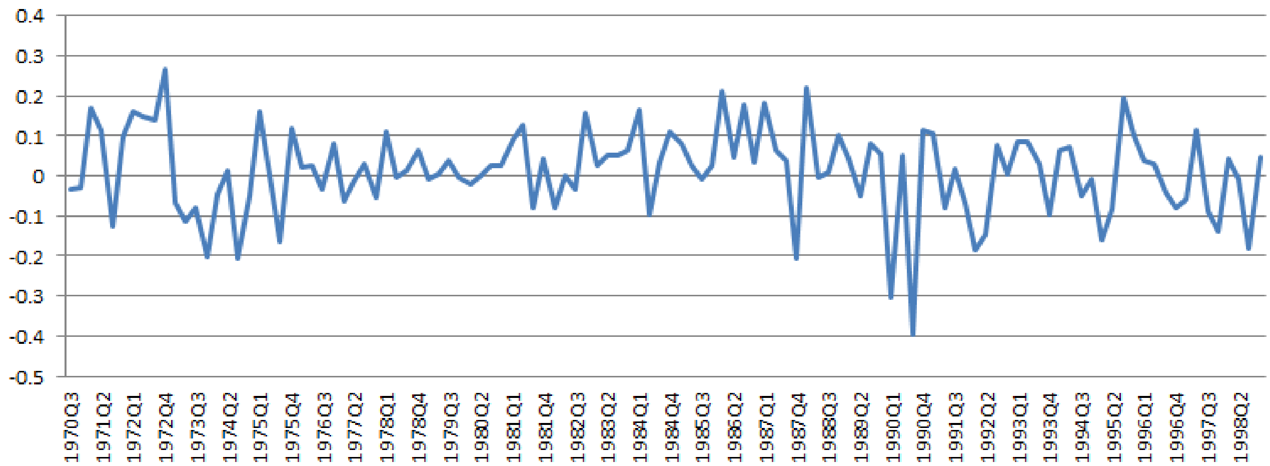
(b)



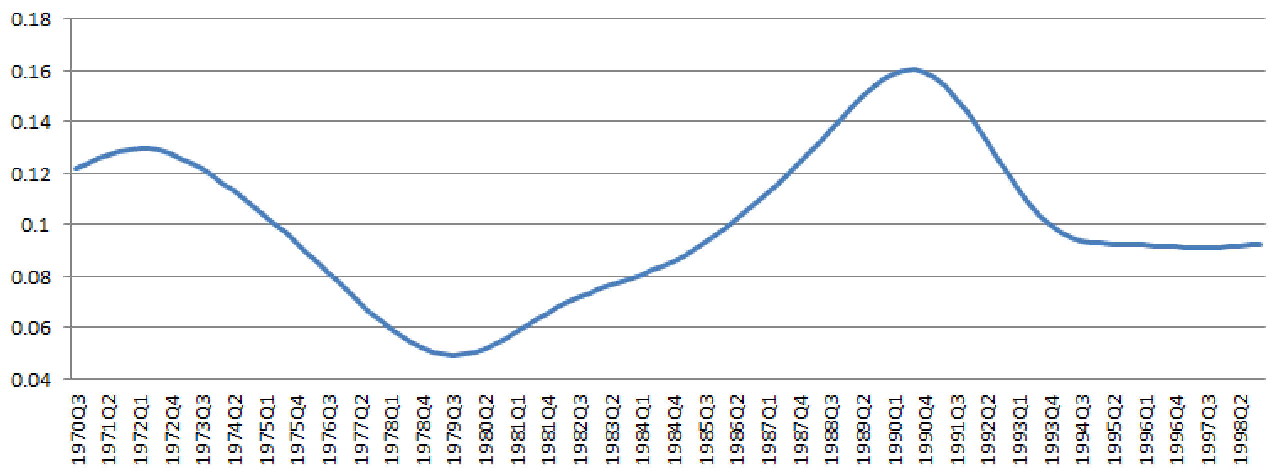
Note: (a) Aggregate Stock Return. (b) Estimated Volatility.

Figure 3.8. Aggregate Stock Return and Estimated Volatility: Japan

(a)



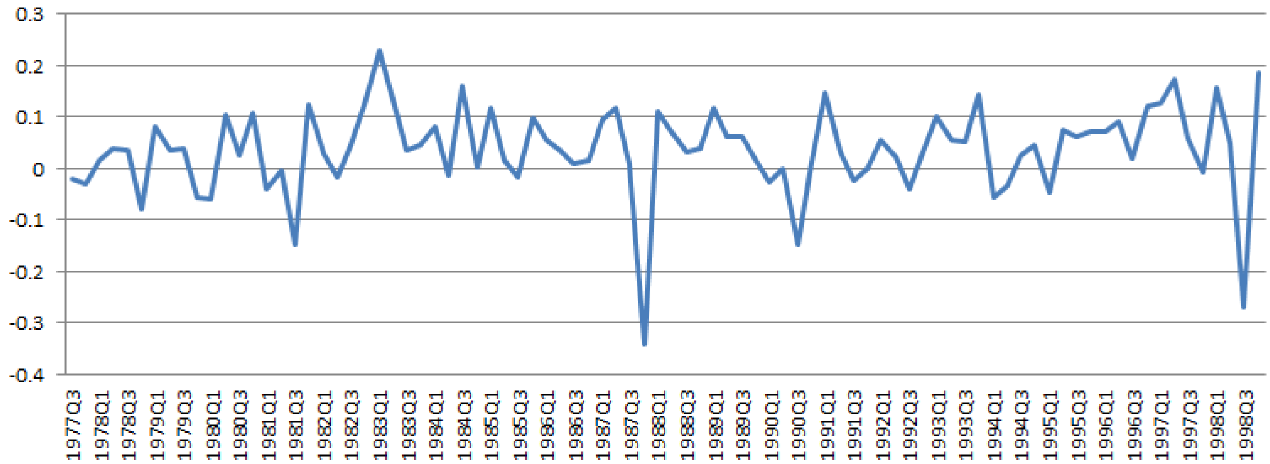
(b)



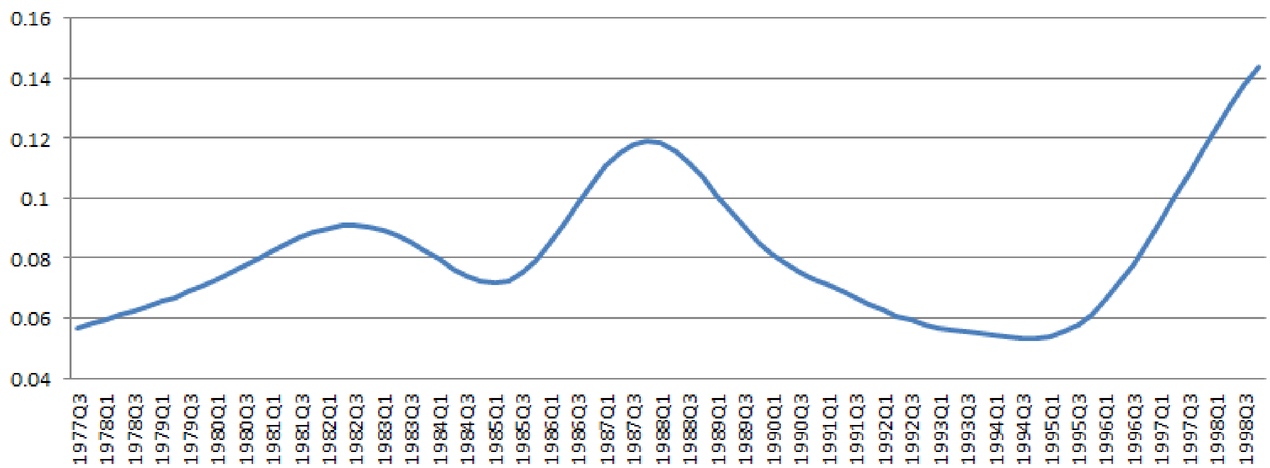
Note: (a) Aggregate Stock Return. (b) Estimated Volatility.

Figure 3.9. Aggregate Stock Return and Estimated Volatility: Netherland

(a)



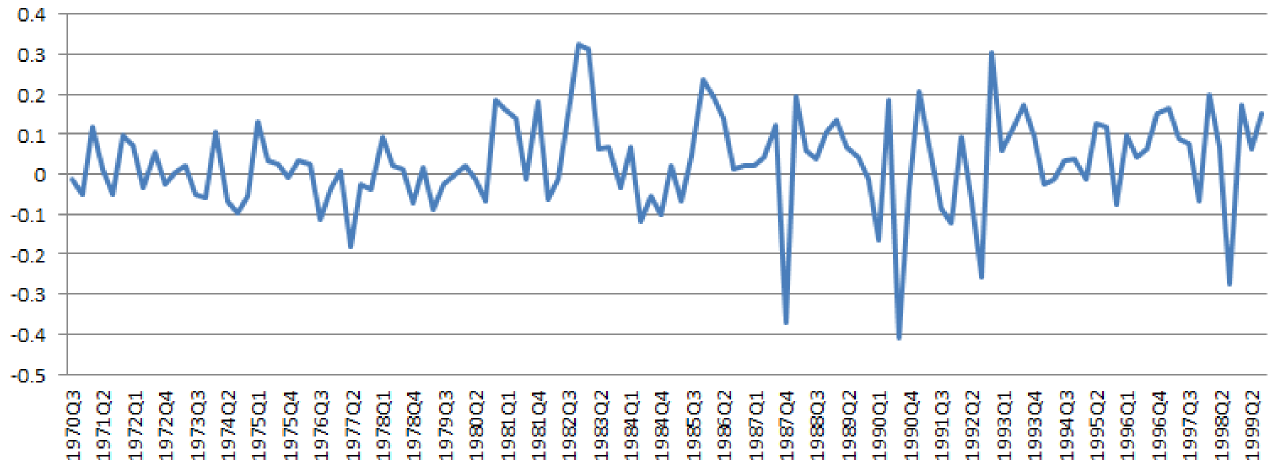
(b)



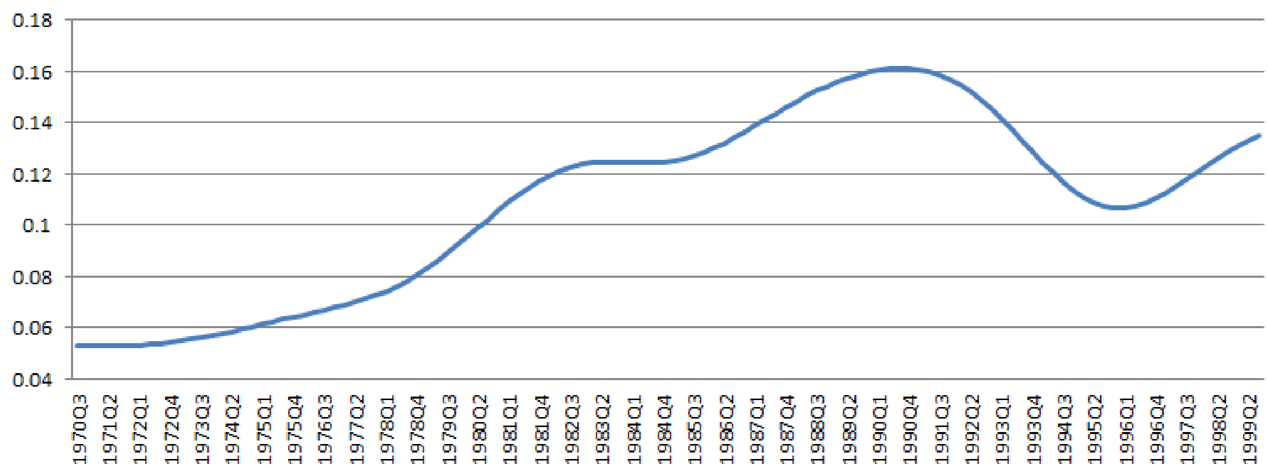
Note: (a) Aggregate Stock Return. (b) Estimated Volatility.

Figure 3.10. Aggregate Stock Return and Estimated Volatility: Sweden

(a)



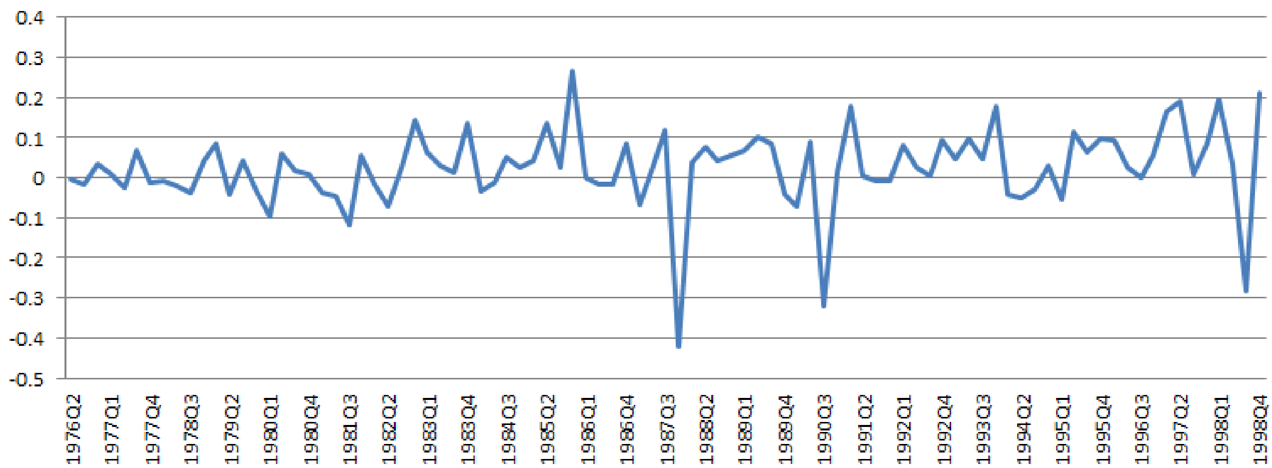
(b)



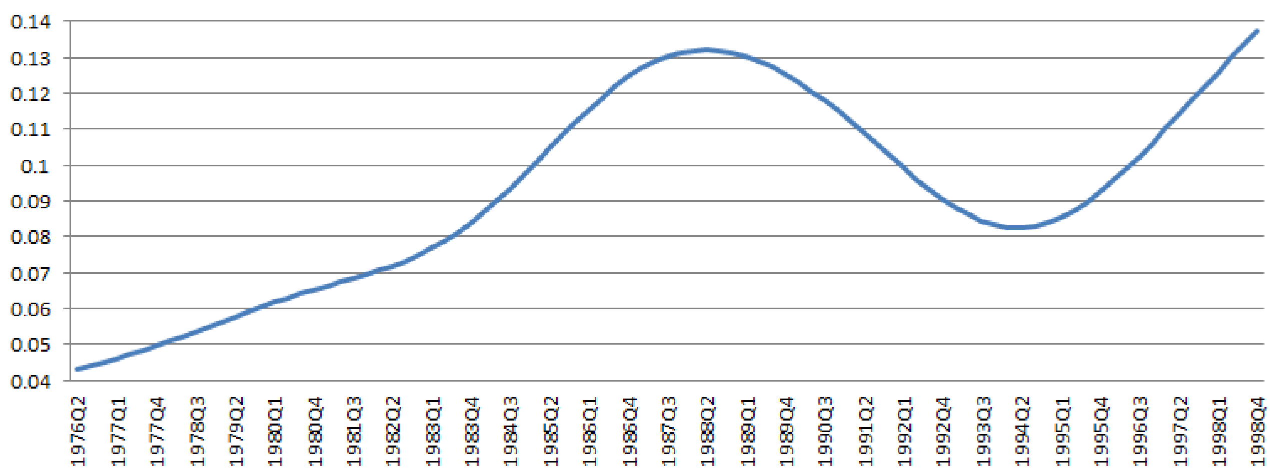
Note: (a) Aggregate Stock Return. (b) Estimated Volatility.

Figure 3.11. Aggregate Stock Return and Estimated Volatility: Switzerland

(a)



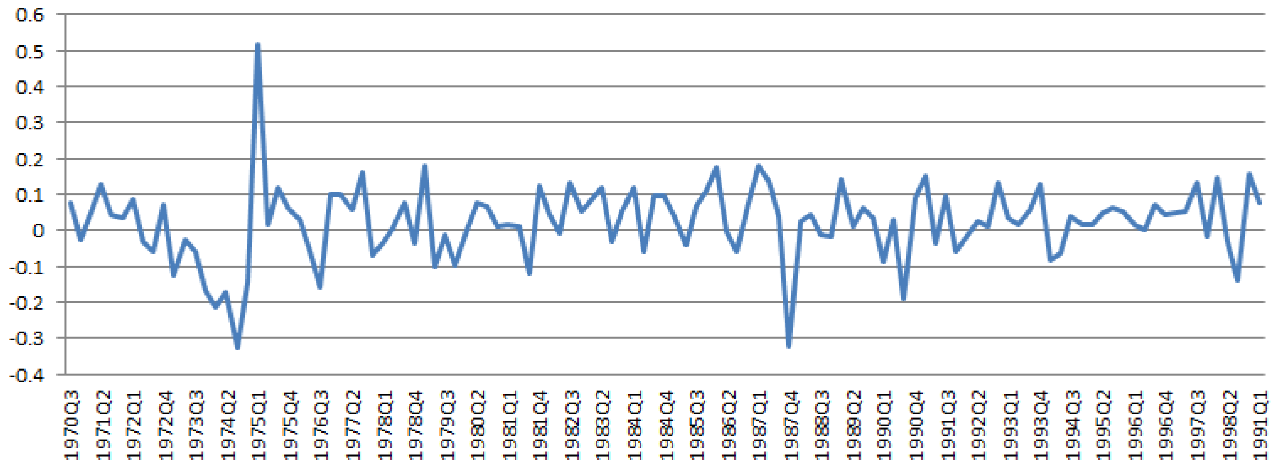
(b)



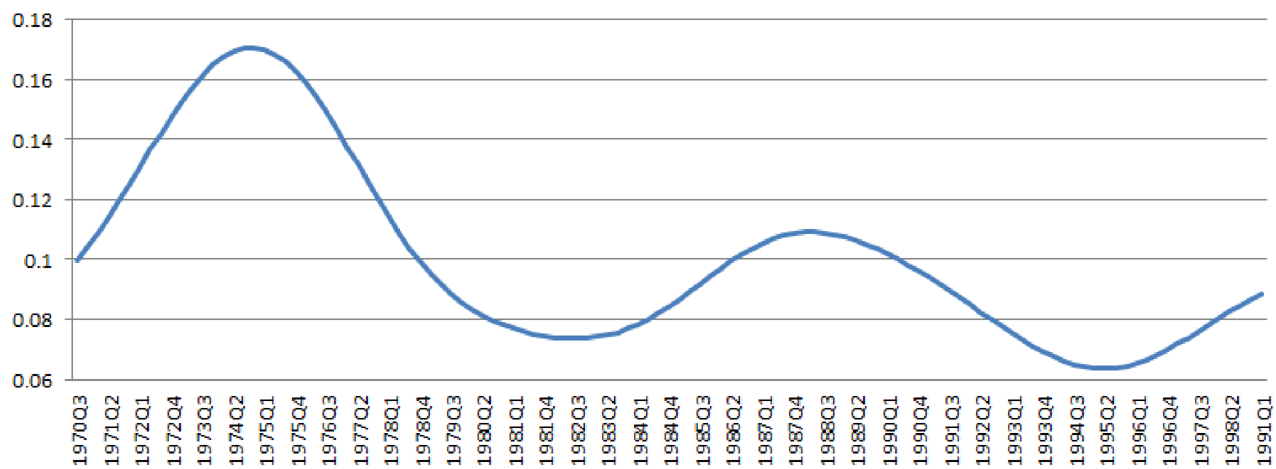
Note: (a) Aggregate Stock Return. (b) Estimated Volatility.

Figure 3.12. Aggregate Stock Return and Estimated Volatility: United Kingdom

(a)



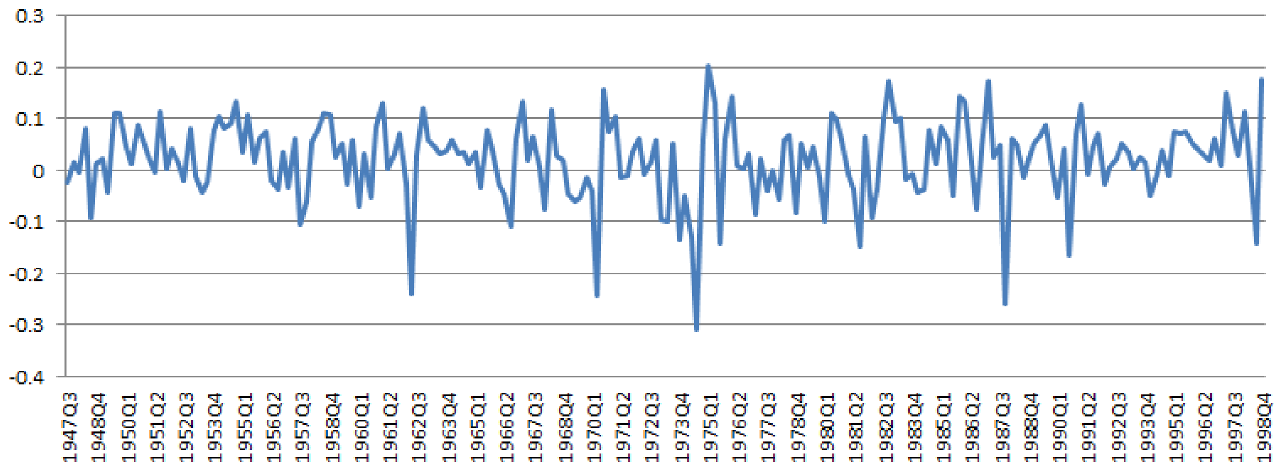
(b)



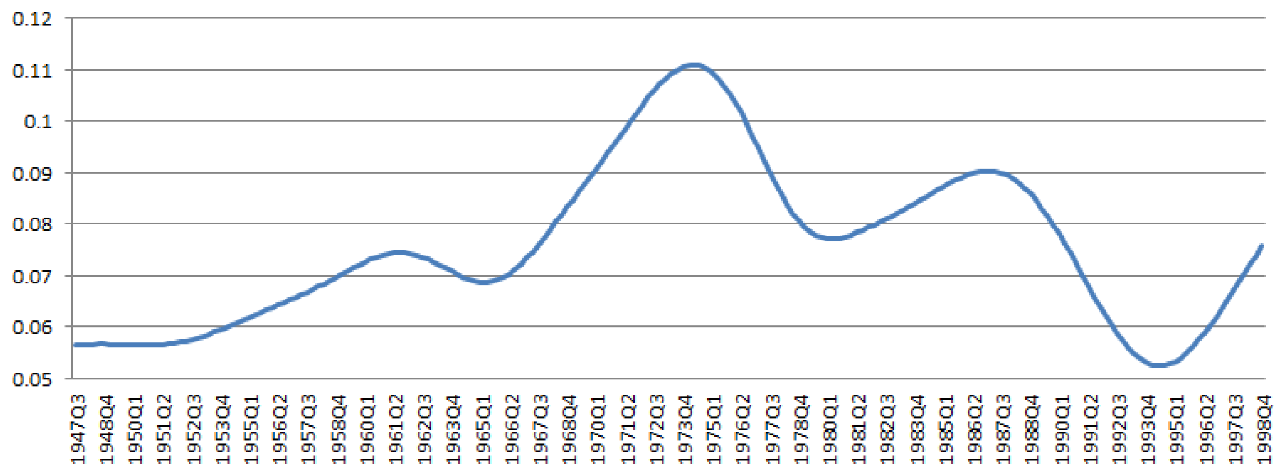
Note: (a) Aggregate Stock Return. (b) Estimated Volatility.

Figure 3.13. Aggregate Stock Return and Estimated Volatility: United States

(a)



(b)



Note: (a) Aggregate Stock Return. (b) Estimated Volatility.