ABSTRACT

MOVCHAN OLEKSANDR VICTOROVICH. Measuring Nonstationary Cycles: a Time-Deformation Approach. (Under the direction of Dr. Atsushi Inoue).

This dissertation consists of three essays on modeling and parameter estimation for covariance non-stationary processes. The first essay considers the non-linear deformation of time scale for $G(\lambda)$ -stationary processes developed by Jiang, Gray and Woodward [2006]. After the appropriate Box-Cox transformation, processes which are nonstationary in the regular time domain, become stationary in the transformed time scale, thus allowing application of traditional econometric tools. We also study the consistency properties of the Q-statistic which is used to estimate parameters of time deformation connecting regular and transformed time scales. As an empirical illustration, the cyclical behavior of the U.S. unemployment series is studied in the context of a structural time series model with explicit trend and cycle modeling. Fitting the model in the deformed time provides different frequency estimates as well as improved inference statistic comparing to the model estimated in the regular time domain.

Second essay investigates the case when the parameters of time deformation for $G(\lambda)$ -stationary processes are time-varying, thereby allowing cyclical behavior to vary over the observed data interval. Forecast of unemployment series performed in the deformed time with varying frequencies provides a better fit to the data over the long-term forecasting horizon. We also estimate the dynamics of parameter λ and show that it can be modeled by first-order Markov chain process. While there are many works considering application of Markov-switching models to macroeconomics series, our approach is different in the sense that we consider regime shifts not in the original data, but in the time scale along which the data is measured.

In the last essay we consider Method of Moments (MM) as an alternative approach for the parameter estimation in State Space models (SSM). Estimation and inference in non-Gaussian or non-linear models is usually carried out using importance sampling or Monte Carlo simulation methods. Our method is different in the way that it allows us to relax the

assumptions about the data distribution or about the potential non-linearity embedded into the model. Thus, it can be used as the general tool for models of unknown or not tractable form. At the same time our approach appears to be more efficient since it does not require analytical or computational solution for the approximating models, and does not rely on simulation or sampling techniques. Using Kalman filter and smoother equations, we derive a set of moment conditions, and investigate performance of MM estimation in the series of Monte Carlo simulations for common types of SSM, as well as for several empirical examples described in literature. Our results indicate that Method of Moments provides adequate results for the regular Gaussian linear SSM. At the same time, for non-Gaussian and non-linear models our approach can compete with Bayesian and importance sampling estimation methods.

Measuring Nonstationary Cycles: a Time-Deformation Approach

by
Oleksandr Victorovich Movchan

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina

2009

APPROVED BY:

_____	_____
Dr. Atsushi Inoue	Dr. Denis Pelletier
Chair of Advisory Committee	Co-chair of the Advising Committee
_____	_____
Dr. David A. Dickey	Dr. John J. Seater

DEDICATION

To my parents Iryna and Victor, and to my wife Yana
for their love, support, understanding, and patience
during all these years.

BIOGRAPHY

Oleksandr Movchan was born in 1977 in Ukraine. He received Bachelor of Science degree in Finance from Cherkasy State Technological University in 1999. In 2003 he earned his Master of Arts degree in Economics from the National University of "Kyiv-Mohyla Academy", Ukraine. In 2004 he joined the Department of Economics at North Carolina State University for pursuing a Doctorate Degree in Economics.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor and dissertation chair Dr. Atsushi Inoue for his guidance, insights, and collaboration, which have been an enormous contribution to this research. I would also like to thank him for all the knowledge I gained from his Econometrics courses which greatly influenced my academic interests.

I would like to deeply thank my advisor and co-chair Dr. Denis Pelletier for his help and advices, as well as for introducing me to the area of State Space modeling, which apparently became the central part of my current research.

I am also indebted to my advisors, Dr. John Seater and Dr. David Dickey for their support and assistance, and for their outstanding example of being a researcher and a teacher.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Nonstationary Cycles: G(λ)-stationary Processes with Constant Time-Deformation Parameters

1.1 Introduction

The evolution path of many pro-cyclical economic time series may change irregularly since they depend on underlying fundamentals striking not systematically. There are several theoretical explanations for this fact, such as models simulating the behavior of risk-averse agents (Chalkley and Lee, [2002]), studies on intertemporal increasing returns (Acemoglua and Scott, [1997]), or adjustment costs models (Caballero and Engel, [1991]). At the same time, empirical research in the given area is more extensive. The change in the volatility and cycle duration of major macro variables was reported by many recent studies. Stock and Watson [2003], Kim, Nelson and Piger [2004] found that cyclical component of real GDP became less volatile in the post-war period. Moreover, decrease in volatility is also present in many production sectors of real GDP as well as in final sales. At the same time, inflation is also marked by changes in persistence and volatility at the corresponding periods of the volatility reduction observed in real GDP. The break in conditional volatility and conditional mean of 214 monthly US variables over 40 years is documented by Sensier and van Dijk [2004]. These findings also suggest that periodic economic data such as GDP, unemployment, inflation series, as well as many others, are characterized by different timing of upturns and downturns. At the same time, most of empirical research is carried out on the data taken at regular intervals of time (days, months, years), which may therefore undermine important relations between variables or provide improper results. Therefore, it is appealing to treat data on a time-scale specific to the process, rather than on the regular time scale. The former may differ significantly from usual time measurement. This paper considers the methodology of transforming the time scale into the one on which data would become covariance stationary and suitable for econometric estimation.

1.2 Literature review

The discussion of nonstationarity of economic variables begins with the pioneering paper by Burns and Mitchell [1946] who suggested that a unit of economic time is defined by business cycle length rather then calendar time. This work initiated numerous empirical studies intended to understand economic fluctuations, which can be divided into a few distinct groups. One group of works deals with modification of the existing econometric models (e.g. ARIMA) to account for different behavior of the data over the cycle. An example would be studies of the asymmetrical behavior of cycles over the periods of expansion and recession. Neftci [1984] considered estimation and testing issues in the model when underlying data are characterized by downturns and expansions of different time length. Introducing an indicator variable for ups and downs into a linear model can potentially improve model fit and forecasting. The major weakness of the studies on asymmetry is considering the whole data series as a stationary process. While downturns and expansions are treated differently, the length of cycles is assumed to remain constant over the time.

Another kind of improvement in this field is provided by Markov-switching common factor models, which allow modeling regime changes in the dynamics of the cycles. Change in the growth rate of output is treated as discrete regime change from high to low state (Hamilton [1989, 1990], Diebold and Rudebusch [1996]). Kim and Nelson [1998] introduced dynamic factor model with regime switching into the state space framework. These authors model the probability of regime change based on the length of recession or expansion phase of a business cycle. In addition to regime switching in output growth, Kim and Nelson [1999] also explored the issue of shifts in the parameters of models based on Markov-switching behavior.

A completely different direction of the research focuses on developing new econometric tools, such as windowed Fourier transformation, band-pass filtering in the frequency domain (Hodrick and Prescott, [1997]), Baxter and King, [1999]), or wavelet

analysis (Yogo, [2003], Raihan, Wen, and Zeng, [2005]) for analyzing data of varying frequency. All these methods concentrate on frequency domain, rather than time domain, to explore dynamic properties of the data. However, these tools suffer from some limitations. Fourier transformation is not applicable for nonstationary signals. Wavelet methods, while nicely capturing data behavior, often lack economic interpretation. Filtering methods usually can not be used for forecasting methods, and as shown by Harvey [1993], may provide spurious cyclical behavior.

Finally, one more approach considers modifying existing data in order to obtain required properties for econometric models. The non-linear transformation of the time scale can change the data from being nonstationary in regular time to being a stationary process in modified scale. Stock [1987] introduced the idea of "economic time", which may be different from regular time. He proposed several non-linear transformations connecting calendar time and economic time for cyclical data. One was the extension of the approach by Burns and Mitchell [1946], based on the phase-averaging procedure. The expansion and contraction of a cycle were split into several phases, and then data were taken as the averages of the observations falling into relevant phase. Other types of deformation were based on the assumption that economic time progresses at different paces over periods of expansion and contraction. To account for this, an indicator variable can be introduced to reflect the corresponding stage of the cycle. This may be considered as a modification of the approach to account for asymmetry introduced by Neftci [1984]. Jorda and Marcellino [2004] raised the issue of matching time scales in data aggregation applications. These authors analyze estimation and inference issues for data which evolves according to its natural time scale (not necessarily linear or regular), but is recorded at the time intervals available to the the observer, such as regular calendar or time scale units. According to their definition of different types of aggregation, our work investigates the Type-II aggregation case when the time series are analyzed based on regularly-spaced calendar time even though it may not represent the original DGP realization time. As shown by the authors, mismatch in timing scales between the

intrinsic and recorded process may pose serious implications for parameter estimation, inference and forecasting when applying regular econometric tools.

Recently, the issue of time-deformation for cyclical data has been widely discussed in the field of signal processing, speech recognition, biology, etc. As a recent development, Gray and Zhang [1988] studied the data obeying multiplicative group composition law. The multiplicative stationary (M-stationary) continuous Euler process, which is characterized by elongating cyclical behavior, was shown to have a stationary dual as a continuous autoregressive process in logarithmically deformed time domain. Later, Vijverberg and Gray [2004] discretized continuous Euler process into stationary discrete ARMA process which enabled application of the model to the discrete empirical data. Gray, Vijverburg, and Woodward [2004] developed forecasting and spectral analysis methodology for discrete M-stationary processes using simulated data and bat echolocation signals. Vijverberg [2006] applied the methodology to study residential investment growth data. The main limitation of M-stationary processes is that they describe data with elongating cycles, which may be not appropriate for many economic variables.

The time-deformation process used in this paper was developed by Jiang, Gray and Woodward [2006][1]. The so-called $G(\lambda)$-stationary process is the generalization of different stationarity concepts, such as additive and multiplicative stationarity. This method may be applied to data with frequency changing systematically in time, either increasing or decreasing. The time scale is transformed using the Box-Cox transformation, where λ is the parameter of transformation. Different values of λ correspond to different cyclical behaviors. Nonstationary data in regular time scale is assumed to be $G(\lambda)$-stationary and can be converted into the dual process, which is stationary in the deformed time. Estimated parameters of the model fitted to such data can be easily transformed back into the regular time.

In the original paper by JGW [2006], time deformation is applied to simulated data

1 Abbreviated as JGW [2006] in the rest of paper for brevity.

and geophysical data with the clear cyclical signal embodied. The novel approach in this paper is application of $G(\lambda)$ transformation to economic data. While logarithmic time scale deformation (which assumes that data follow M-stationary process) was studied in the existing research, fitting $G(\lambda)$ stationary process to the economic series is a completely new way to explore data characteristics. In addition, this paper examines robustness and consistency properties of the Q-test, originally proposed by JGW [2006] to estimate time deformation parameters.

Another new feature in this paper is the introduction of time deformation in the framework of Unobserved Component (UC) time series models. Unlike ARIMA-type models, they provide a convenient tool to model explicitly stochastic trend, seasonal and cyclical components. Traditional Kalman filtering and smoothing algorithms allow us to compute a likelihood function for parameter estimation within the classical or Bayesian framework. Clark [1987], Harvey [1989], and Harvey and Jaeger [1993] developed the general model for structural time series as a linear combination of explicit trend and cycle components and examined model properties in application to several major macroeconomic time series. Later, Tripodis and Penzer [2006] reviewed the composition of the seasonal component in UC models. Koopman, Ooms, and Hindrayanto [2006] discussed issues of estimation, identification and forecasting in the general class of periodic UC models. Different modifications of the original UC model were considered by Perron and Wada [2005] and Morley, Nelson, and Zivot [2002] in application to the U.S. economic activity. Recent studies on UC modeling of business cycles include Koopman and Lee [2005], who used State Space model (SSM) to test the asymmetry in unemployment, private domestic investment, and GDP data. The asymmetry is modeled based on the steepness of the cycle - different frequencies are attached to ascending or descending periods. Varying periodic activity is modeled as the deviation of actual frequency from the underlying basic frequency. Koopman and Wong [2006] employed a time dependent sample spectrum to find time-varying parameters in the UC model of the U.S. GDP and Industrial Product Index.

UC models provide a useful tool to explore structural data, and using this approach along with the time deformation tools may provide new information about behavior of many economic periodic variables.

1.3 Time deformation for nonstationary data

1.3.1 G(λ)-stationary processes

Following Hannan [1965], the most general covariance stationarity property (in its weak form) may be expressed as the following group composition law:

$$E[(X(t)-\mu)(X(f(t,\tau))-\mu)]=C_X(\tau) \tag{1.1}$$

That is, the second moment $C_X(\tau)<\infty$ for covariance-stationary series $\{X(t)\}$ *is* point-pair invariant, and therefore depends only on the distance τ between any two observations, and not on their locations t or $f(t,\tau)$.

Regular (the most common) stationarity assumes additive structure of $f(t,\tau)$: $E[(X(t)-\mu)(X(t+\tau)-\mu)]=C_X(\tau)$. Gray and Zhang [1988] considered the multiplicative composition law: $E[(X(t)-\mu)(X(t*\tau)-\mu)]=R_X(\tau)$. The latter may be applied for data exhibiting only elongating cyclical behavior and is referred to as M-stationary process. Jiang, Gray, and Woodward [2006] develop the $G(\lambda)$-stationary process which obeys the following group composition law: $f(t,\tau)=(t^\lambda+\tau\lambda)^{1/\lambda}$ with $\lambda\in(-\infty,\infty)$. The following definition from JGW [2006] defines the $G(\lambda)$-stationary process.

Definition 1.1. Let $\{X(t)\}$ be a stochastic process defined over $t\in(0,\infty)$ such that if for any $(t^\lambda+\tau\lambda)\in(0,\infty)$, and constant $\lambda\in(-\infty,\infty)$,

i. $E[X(t)]=\mu$,

ii. $var[X(t)]=\sigma^2<\infty$, (1.2)

iii. $E[(X(t)-\mu)(X((t^\lambda+\tau\lambda)^{1/\lambda})-\mu)]=B_X(\tau;\lambda)$

then $X(t)$ will be called a $G(\lambda)$-stationary process (weakly).

$B_X(\tau;\lambda)$ in (1.2) is referred to as the $G(\lambda)$-autocovariance of $\{X(t)\}$ with the

following property: $B_X(-\tau;\lambda)=B_X(\tau;\lambda)\,\forall\,\lambda$.

Definition (1.1) states that a $G(\lambda)$-stationary process $\{X(t)\}$ should have: (i) constant first moment; (ii) finite and constant variance; and (iii) autocovariance which depends only on constant λ and the distance between observations τ, but not on the location of the data points or on the beginning of the series. Also, throughout the paper whenever mentioning $G(\lambda)$ stationarity, we will imply weak stationarity.

It can be shown that $G(\lambda)$-stationarity covers a wide range of different types of stationarity. When $\lambda=1$, $X(t)$ obeys traditional additive stationarity, while $\lambda=0$ leads to *M*-stationarity[2].

Based on (1.2), we may also define *G(λ)*-autocorrelation function of $X(t)$, which depends on the distance between observations, but also on the time-deformation parameter:

$$\rho_X(\tau;\lambda)=\frac{B_X(\tau;\lambda)}{Var(X(t))}=\frac{B_X(\tau;\lambda)}{B_X(0;\lambda)} \tag{1.3}$$

From the Definition 1 it is obvious that for $\lambda\neq 1$ the autocovariance between observations $X(t)$ and $X((t^{\lambda}+\tau\lambda)^{1/\lambda})$ depends only on the distance between them τ, but also on location point t. Thus, a $G(\lambda)$-stationary process appear to be nonstationary on the regular time scale. However, it has its stationary dual process $X(u);u\in(-\infty,\infty)$, which is defined on the time scale over u. In other words, $X(t)=X(u)$ on $t\in(0,\infty)$.

The transformation functions connecting t- and u-based scales are:

$$u=g(t)=\frac{t^{\lambda}-1}{\lambda} \tag{1.4}$$

2 Shown in JGW [2006]

and $t = g^{-1}(u) = (u\lambda + 1)^{1/\lambda}$ (1.5)

When $\lambda \to 0$, then $\lim_{\lambda \to 0} \frac{t^{\lambda} - 1}{\lambda} = \ln(t)$ and (1.4) simplifies to $u = \ln(t)$, which corresponds to the *M*-stationary process.

As shown in JGW [2006], autocovariances of $X(u)$ depend only on lag length τ, and $X(u)$ is stationary in the regular sense over the deformed time $u = \frac{t_1^{\lambda} - 1}{\lambda}, \frac{t_2^{\lambda} - 1}{\lambda}, \frac{t_3^{\lambda} - 1}{\lambda}, \ldots., \frac{t_N^{\lambda} - 1}{\lambda}$.

Another question to be considered is the origin of the observed data. As we just defined, $X(t)$ has it's stationary dual $X(u)$. The origin of the data in t-scale is $t = 1$, which corresponds to $u = 0$ in the u-scale. However, the observed data available to the researcher may not coincide with the origin in the *t*-time. That is, we may observe only $X_1(t)$, which is a subset from $X(t)$ shifted by Λ time intervals from $t = 1$. The corresponding dual for this subset is then $X_1(u)$, which also have the beginning shifted relative to the $u = 0$. For example, if actual data values are timed as $t = 1,2,3,\ldots, N$ and offset is $\Lambda = 100$, they correspond to the $G(\lambda)$-process at time points $101, 102, 103, \ldots., 100 + N$.

In this case (1.4-1.5) should be modified:

$$u = \frac{(t + \Lambda)^{\lambda} - 1}{\lambda} \tag{1.6}$$

$$t = (u\lambda + 1)^{1/\lambda} - \Lambda \tag{1.7}$$

The simplest example of $G(\lambda)$-stationary function would be a trigonometric periodic function with constant amplitude and zero phase:

$$X(t)=cos(\theta\frac{t^{\lambda}-1}{\lambda}) \tag{1.8}$$

The corresponding dual is $X(u)=\cos(\theta u)$, which is stationary on $u\in(-\infty..\infty)$. For the periodic function, parameter λ determines the length of the cycle: $\lambda<1$ generates elongating cycles, while $\lambda>1$ produces fading cycles. The case when $\lambda=1$ corresponds to the cycle which is stationary in the original scale. (see Figure 1.1).

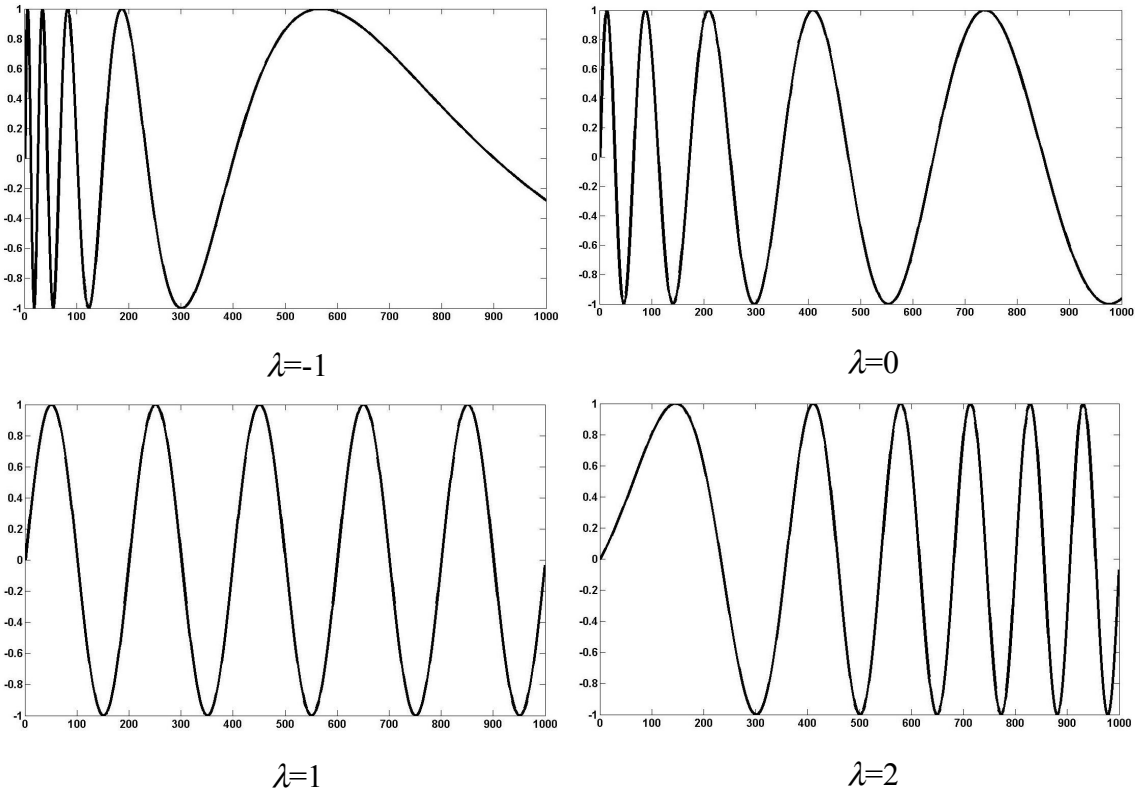


Figure 1.1: Cycle behavior for different values of λ.

The data was generated according to (1.8) with λ =[-1; 0; 1; 2], θ =[5; 2; 0.5; 0.05], t=1..1000.

As it was mentioned, since $X(u)$ is stationary in the domain (u), it may be used for estimation with regular econometric tools. However, several technical issues listed below have to be addressed first.

1.3.2 Interpolating an equally spaced sample from a G(λ) process

The observed discrete data sample $X(t)$ discussed in the previous section will correspond to discrete $X(u)$ process over the *u*-scale. However, the $X(u)$ sample will be spaced over unequal intervals due to non-linear transformation (1.4). This may not be suitable for econometric estimation since many typical models assume that data were taken over equally spaced intervals. One way to obtain equally spaced realizations $\tilde{X}(\tilde{u})$ is via an interpolation technique, for example, linear interpolation between time points t_k and t_{k+1}:

$$\tilde{X}(\tilde{u}, X) = X(u_k) + (\tilde{u} - u_k) * \frac{X(t_{k+1}) - X(t_k)}{u_{k+1} - u_k}, \qquad k = 1..N-1 \tag{1.9}$$

For the convenience of relating two data sets, interpolation may be accomplished for the same number of data points as in the original data. That is, if $t = 1..N$, with $\Delta t = 1$, and $\Lambda = 0$, we have $u = \frac{t_1^{\lambda} - 1}{\lambda}, \ldots, \frac{t_N^{\lambda} - 1}{\lambda}$, and $\Delta u = \frac{t_N^{\lambda} - t_1^{\lambda}}{\lambda(N-1)}$. Figure (1.2) explains how time scales and the data are related.

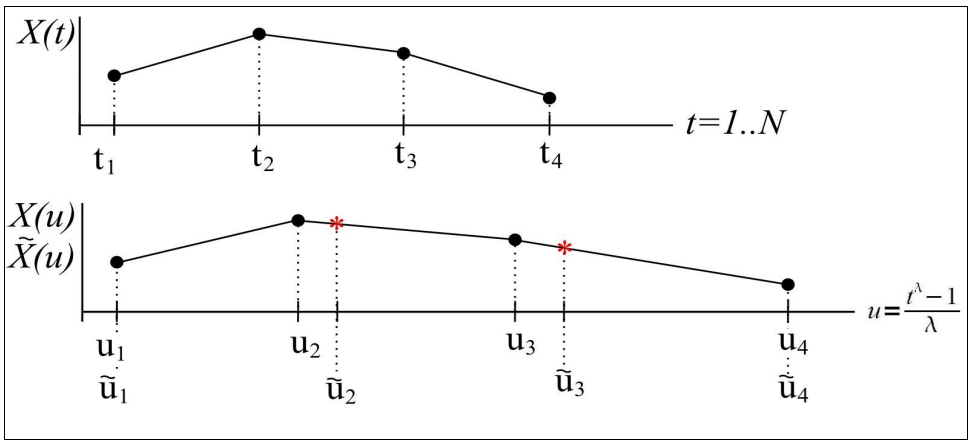


Figure 1.2: Correspondence between regular and transformed time scales

Thus, for all $t=1..N$, we have that $X(t)=X(u)$. However, interpolation over the equally-spaced scale $\tilde{u}$ results in new interpolated data points $\tilde{X}(\tilde{u})$ except for the first and the last observations: $X(t)=X(u)=\tilde{X}(\tilde{u})$ for $t=1,N$. Using (1.9) $\tilde{X}(\tilde{u})$ then can be simply expressed as the linear combination of the original data $X(t)$:

$$\tilde{X}(\tilde{u};X)=X(t_k)+w_k*(X(t_{k+1})-X(t_k)), \qquad k=2,N-1 \tag{1.10}$$

with weights $w_k=\frac{\tilde{u_k}-u_k}{u_{k+1}-u_k}\le 1$, being a differentiable function of λ.

The time step Δu in the equally spaced $\tilde{u}$-scale can be normalized to correspond to the time interval Δt in t-scale, so we could plot two series on the same graph. Also, to suppress notation, since $\tilde{u}$ is just a different sampling rate from time scale u, we will use notation $\tilde{X}(u)$ instead of $\tilde{X}(\tilde{u})$ taking the former as interpolated data over the equally spaced scale.

The process of data interpolation using some non-linear weighting scheme in some sense resembles the work by Jorda and Marcellino [2004], who considered constructing aggregate data in the observed time scale by weighting the observations following their intrinsic time.

1.3.3 Estimating λ and offset Λ

Time-deformation literature offers several methods to estimate the parameter λ and offset Λ. To estimate the necessary parameters of time-deformation for M-stationary processes, Gray et al. [2004] fit the ARMA model for the range of offset values, and find the one value to minimize the SSE[3]. Choi et al [2006] also apply ARMA-based dual representation and use Akaike's information criterion to estimate the offset. However,

3 For M-stationary process $\lambda=0$, so the only parameter to be estimated is offset, Λ.

these methods are model-based, since they assume that the original data follow Euler process with the dual having an ARMA form. In contrast, to estimate both parameters - λ and Λ, JGW [2006] offer a model-independent Q-statistic, which is based on the sample autocorrelation functions:

$$Q_n(\lambda,\Lambda)=\sum_{k=0}^{K}[(r_1(k;\lambda,\Lambda)-r_{1.5}(k;\lambda,\Lambda))^2+(r_2(k;\lambda,\Lambda)-r_{1.5}(k;\lambda,\Lambda))^2] \quad (1.11)$$

where $k=1,2,..,K$, and $r_1(k;\lambda,\Lambda)$, $r_{1.5}(k;\lambda,\Lambda)$, and $r_2(k;\lambda,\Lambda)$ are sample $G(\lambda)$-autocorrelations computed from subsamples X_1, $X_{1.5}$, and X_2, respectively, such that

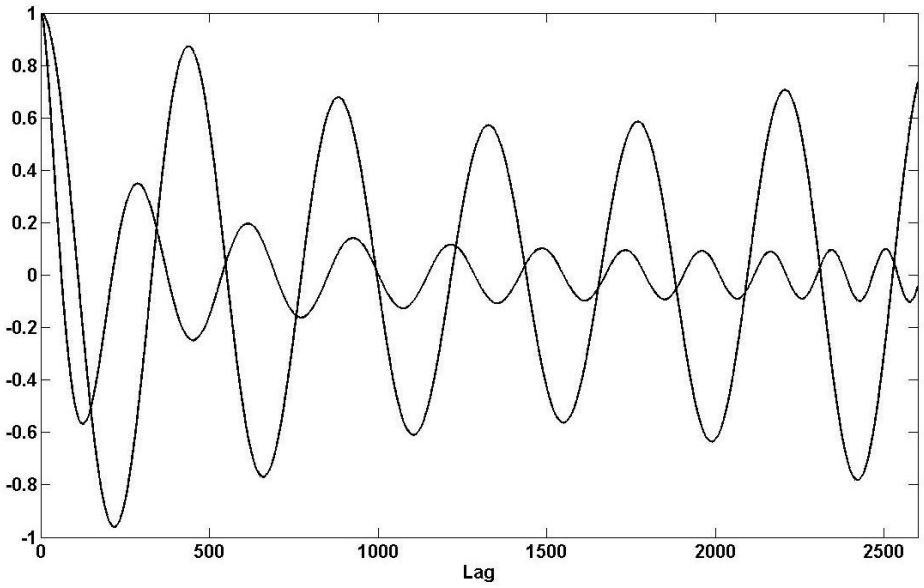
$$r(k)=\frac{\sum_{t=1}^{N-k}(X(t)-E[X])(X(t+k)-E[X])}{\sum_{t=1}^{T}(X(t)-E[X])^2}, \qquad r(-k)=r(k). \quad (1.12)$$

Here X_1 and X_2 represent two equal parts of the data sample. $X_{1.5}$ is a subsample consisting of the second half of X_1 and the first half of X_2. The estimation is carried out by specifying an initial range for λ and Λ, and then computing statistic (1.11) for each pair of values of λ_i and Λ_i in the specified range. The true values are the ones that minimize (1.11). When the data sample is stationary, autocorrelation functions will be similar for each subinterval. Figure 1.3 shows the sample ACF's for the two halves of the data generated as

$$X(t)=cos(\theta\frac{t^{\lambda}-1}{\lambda})+\epsilon_t \quad (1.13)$$

For the original data shown in panel (a), there is a significant difference in ACF's for the first and the second halves of the sample; while for the transformed stationary dual (panel (b)), ACF's are almost identical.

(a)



(b)

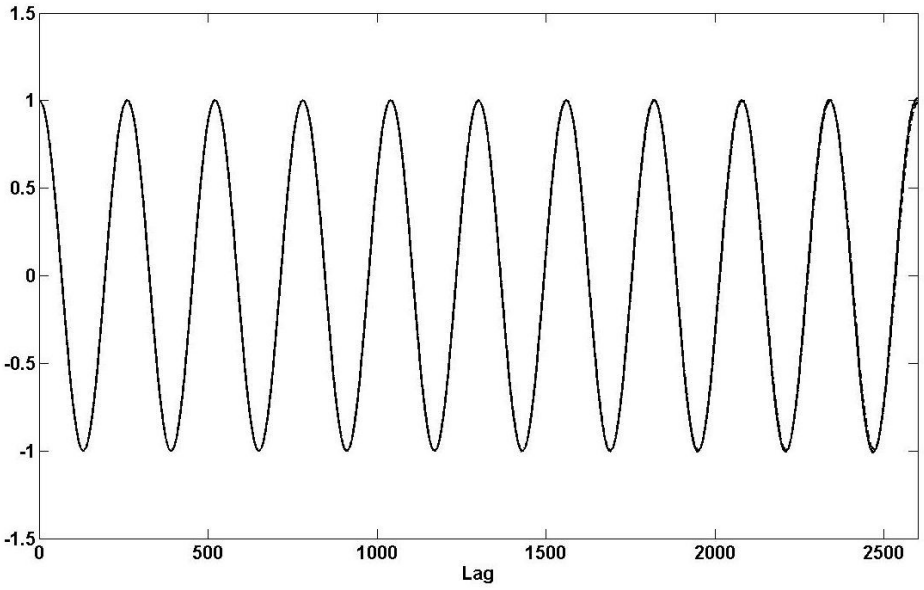


Figure 1.3: Autocorrelation functions for two halves of data over the original and the transformed time scales

The data was generated as in (1.13) with $\theta = 0.1,\ \sigma_\epsilon^2 = 0.04,\ \lambda = 0.5$.

1.4 Consistency of the Qn(λ,Λ) test

Before applying the Q-test in the empirical estimation, we need to explore its asymptotic properties, a natural step before the estimation which, however, was missed by authors in their paper (JGW [2006]).

Let's denote $\hat{\theta} \equiv (\hat{\lambda}, \hat{\Lambda}), \hat{\theta} \in \Theta$, such that $\Theta \subset R^K$ is compact. Then, following Newey and McFadden [1994], $\hat{\theta}$ is a consistent estimator of θ, that is, $\hat{\theta} \to_p \theta$, if there $Q_n(\hat{\theta})$ converges uniformly in probability to $Q(\theta)$, such that $Q(\theta)$ is continuous and has a unique extremum at θ^*.

Establishing convergence in probability of $\hat{\theta}$ requires that sample size $n \to \infty$. However, in (1.11) we have summation of autocorrelations up to the maximum lag K, which also grows with the sample size. Since $r(k)$ may differ from zero for arbitrary large K, its convergence properties may affect the asymptotic properties of $\hat{Q}_n(\theta)$.

Following JGW [2006], a $G(\lambda)$ process, which is a generalization of the Euler (p,q) process has its stationary discrete dual having an $ARMA(p,q)$ form:

$$\phi(L^p)\tilde{X}(u;\lambda^*,\Lambda^*) = \theta(L^q)\epsilon_u \qquad (1.14)$$

where ϕ and θ are p^{th} and q^{th} degree polynomials, L is a backward shift operator, and $\epsilon_t \sim N(0,\sigma^2)$.

Finding correct parameters of time deformation requires computing the Q-test for a range of values of λ and Λ, and finding a pair $\{\lambda_i, \Lambda_j\}$ which would minimize the Q-statistic on Θ. So, each time the original data $X(t)$ are being transformed to $\tilde{X}(u;\lambda,\Lambda)$ based on specific values $\{\lambda_i, \Lambda_j\}$, and only the true values $\{\lambda^*, \Lambda^*\}$ will provide a covariance-stationary ARMA process (1.14).

Definition 1.2[4] An $ARMA(p,q)$ process defined by equations $\phi(L^p)Z_t=\theta(L^q)\epsilon_t$ is said to be *causal* if there exists a sequence of constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty}|\psi_j|<\infty$ and $Z_t=\sum_{j=0}^{\infty}\psi_j\epsilon_{t-j}$, $t=1..N$.

Assumption 1.3 Sequence $\{\tilde{X}\}$ defined in (1.14) follows a causal ARMA(p,q) process. That is, given that the roots of the autoregressive polynomial lie outside the unit circle:

$$\tilde{X}(u;\lambda^*,\Lambda^*)=\sum_{j=0}^{\infty}\psi_j\epsilon_{u-j}, \qquad u=u_1..u_N, \tag{1.15}$$

where $\sum_{j=0}^{\infty}|\psi_j|<\infty$.

Assumption 1.3 expresses $\{\tilde{X}\}$ defined in (1.14) as a moving average of infinite order thereby allowing the application of mixing properties of the process $\{\epsilon_u\}$ to the observed data $\tilde{X}(u;\lambda^*,\Lambda^*)$. However, as we stated before, a covariance-stationary sequence that follows causal ARMA will be obtained only for true parameters $\{\lambda^*,\Lambda^*\}$. For any other pair of values $\{\lambda_i,\Lambda_j\}$, the resulting interpolated data will violate the covariance stationarity property. Due to this fact, summation over ψ in (1.15) may not have a finite limit, which will have serious implications for the convergence of Q-test. Therefore, we need to introduce another concept of limited dependence for a sequence.

4 See Brockwell and Davis [1991] for the discussion of the causality properties of ARMA processes.

Definition 1.4[5] Let $\{V_t\}_{-\infty}^{\infty}$ be a sequence of random variables defined on a common probability space $\{\Omega, \Im, P\}$. For $m \geqslant 0$ let's define the following quantities:

$$\phi(m) = \sup_t \sup_{A \in \Im_{-\infty}^{t},\, B \in \Im_{t+m}^{\infty},\, P(A)>0} |P(B|A) - P(B)|$$

$$\alpha(m) = \sup_t \sup_{A \in \Im_{-\infty}^{t},\, B \in \Im_{t+m}^{\infty}} |P(A \cap B) - P(A)P(B)|$$

Then $\{V_t\}$ is called a *uniform mixing* or ϕ-mixing process if $\lim_{m\to\infty} \phi(m) = 0$, and $\{V_t\}$ is called a *strong mixing* or α-mixing process if $\lim_{m\to\infty} \alpha(m) = 0$. If $\phi(m) = O(m^a)$ for all $a < -\varphi$ then $\phi(m)$ is of size $-\varphi$. Similarly, $\alpha(m)$ is of size $-\varphi$ if $\alpha(m) = O(m^a)$ for $a < -\varphi$.

Proposition 1.5 The sequence $\{\epsilon_u\}_{-\infty}^{+\infty}$ in (1.14) is α-mixing (or strong mixing).

Here we employ only the strong mixing concept omitting the uniform mixing, since as shown by Ibragimov and Linnik [1971], the $ARMA(p, q)$ process is α-mixing of size $-\varphi$ where φ is arbitrary large.

In other words, Proposition (1.5) states that ϵ_t contains no information about ϵ_s if the distance m between t and s is sufficiently large. An attractive feature of defining sequence $\{\epsilon_u\}$ as a mixing process is that a $\Im_{u+m}^{u-m}/B$-measurable function of such process also is a mixing process of the same size[6]. Thus, $\{\tilde{X}_u\}$ in (1.15) is also a mixing process of the size $-\infty$. However, mixing property is not enough for establishing consistency properties, so we need to introduce near-epoch dependence concept for $\{\tilde{X}_u\}$.

5 See Bierens [1996], Gallant and White [1998], Davidson [2004] for the description of mixing processes.

6 See Bierens [1996].

Definition 1.6[7] For a stochastic sequence $\{V_t\}_{-\infty}^{\infty}$ defined on a common probability space $\{\Omega, \Im, P\}$, let $\Im_{t-m}^{t+m} = \sigma(V_{t-m}, \ldots, V_{t+m})$ such that $\{\Im_{t-m}^{t+m}\}_{m=0}^{\infty}$ is an increasing sequence of σ-fields. If, for $p > 0$, a sequence of integrable random variables $\{Z_t\}_{-\infty}^{\infty}$ satisfies

$$\|Z_t - E(Z_t|\Im_{t-m}^{t+m})\|_p \leqslant d_t v_m,$$

where $v_m \to 0$, and $\{d_t\}_{-\infty}^{+\infty}$ is a sequence of positive constants, then Z_t will be said to be *near-epoch dependent* in L_p-norm (L_p-NED) on $\{V_t\}_{-\infty}^{\infty}$.

Proposition 1.7 Stochastic process $\tilde{X}_u$ obtained as a function $\{\tilde{X}_u\} = g(\{\epsilon_u\}_{-\infty}^{+\infty})$ is near-epoch dependent in L_p-norm on on $\{\epsilon_u\}_{-\infty}^{\infty}$. That is, $\left\|\tilde{X}_u - E(\tilde{X}_u|\Im_{u-m}^{u+m})\right\|_p \leq d_u v_m$ where $\{d_u\}_{-\infty}^{+\infty}$ is a sequence of finite scaling constants, $v_m \to 0$, as $m \to \infty$, and $\Im_{u-m}^{u+m} = \sigma(\epsilon_{u-m} \ldots \epsilon_{u+m})$.

Proof[8]**.** Assume, that $\{\epsilon_u\}_{-\infty}^{+\infty}$ is L_p-bounded sequence with $E[\epsilon_u] = 0$, and $\tilde{X}_u$ defined as in (1.15). Then, using Minkowski's[9] inequality:

$$\left\|\tilde{X}_u - E(\tilde{X}_u|\Im_{u-m}^{u+m})\right\|_p = \left\| \sum_{j=m+1}^{+\infty} (\psi_j(\epsilon_{u-j} - E_{u-m}^{u+m}\epsilon_{u-j}) + \psi_{-j}(\epsilon_{u+j} - E_{u-m}^{u+m}\epsilon_{u+j})) \right\|_p \leq d_u v_m$$

where $v_m = \sum_{j=m+1}^{+\infty} (|\psi_j| + |\psi_{-j}|)$. Using the property of absolute summability of $\{\psi_j\}$ defined in (1.15) it follows that $v_m \to 0$. Sequence of constants d_u is defined then as $d_u = 2\sup_s \|\epsilon_s\|_p$ for all u and $s \leq u$.

---

7 See Gallant and White [1998], Davidson [2004] for the description of NED processes.
8 See Davidson [1994] for a similar discussion.
9 See Hardy, Littlewood, and Pólya [1988].

In a less general framework for the economic series when the dependence of the data follows only one way, that is, $\tilde{X}_u = g(\{\epsilon_u\}_{-\infty}^{u})$, as in (1.15), and $\{\epsilon_u\}$ is a L_2-bounded zero-mean sequence, then the dependence of $\tilde{X}_u$ on $\{\epsilon_u\}$ after the distance m vanishes with a certain rate.

Therefore, for the sequence $\tilde{X}_u$ we have:

$$\tilde{X}_u = \sum_{j=0}^{\infty} \psi_j \epsilon_{u-j} = \sum_{j=0}^{m} \psi_j \epsilon_{u-j} + \sum_{j=m+1}^{\infty} \psi_j \epsilon_{u-j} \tag{1.16}$$

Then,

$$\left\| \tilde{X}_u - E(\tilde{X}_u | \Im_{u-m}^{u}) \right\|_2 = \left\| \sum_{j=m+1}^{\infty} \psi_j \epsilon_{u-j} \right\|_2 \le \sum_{j=m+1}^{\infty} \left\| \psi_j \epsilon_{u-j} \right\|_2 \le \sum_{j=m+1}^{\infty} |\psi_j| \left\| \epsilon_{u-j} \right\|_2$$

The sequence $\{|\psi|\}$ decays exponentially fast, so $\{\psi_{m+1}\}$ is arbitrary small and the size of the near-epoch dependence exceeds any arbitrary number φ. Thus, if $\{\tilde{X}_u\}$ follows $ARMA(p, q)$ process (1.14), it is L_2-*NED* of size $-\infty$. Therefore, $\{\tilde{X}_u\}$ is only dependent on the near-epoch of the sequence $\{\epsilon_u\}$, and $(\tilde{X}_u - E(\tilde{X}_u | \Im_{u-m}^{u})) \to 0$ as m is getting large enough.

Important property of *NED* assumption for $\tilde{X}_u$ is that Borel measurable transformation over near-epoch dependent processes, such as expectations, sums and products are also *NED*. This ensures that sequences of autocovariances $\{cov(k)\}$ for the series $\tilde{X}_u$ are also L_2-*NED* on $\{\epsilon_u\}$. Sequences of autocorrelations $\{r(k)\}$ are also then L_2-*NED,* although of a smaller size since $r(k) = \sum cov(k) / \sum cov(0)$. Davidson [2004] shows that this property guarantees that sequences of autocovariances and autocorrelations are summable for any L_2-*NED* process. This fact assures that summation of autocorrelations $\sum_{k=0}^{K} r(k)$ converges to some finite constant even for arbitrary large K.

Having the result of convergence of ARMA autocorrelation functions, we may now establish their consistency properties. Asymptotic properties of sample autocorrelations were extensively studied in the literature[10]. Fore example, Hannan and Heyde [1972] show that for a process $X_t - \mu = \sum_{j=0}^{T} \psi_j \epsilon_{t-j}$, estimate of the autocorrelation function $r(k)$ converges to $\rho(k)$ either in probability, $r(k) \rightarrow_p \rho(k)$, or almost surely $r(k) \rightarrow_{a.s.} \rho(k)$, depending on stationarity properties of $\{X_t\}$. Based on their results, we consider the following proposition:

Proposition 1.8 Let $\{\tilde{X}_u\}$ be a stochastic process such that $\{\tilde{X}_u\} = g(\{\epsilon_u\}_{-\infty}^{u})$, $\sum_0^{\infty} |\psi_j| < \infty$, and $\lim_{n\to\infty} N^{-1} \sum_{u=1}^{n} E(\epsilon_u^2 | \Im_o^{n-1}) = \sigma^2 > 0$ *a.s.* Then $r(k) \rightarrow_p \rho(k)$. Moreover, if $\{\tilde{X}_u\}$ is second order stationary and $E(\epsilon^2 | \Im_o^{n-1}) = \sigma^2 > 0$ *a.s.*, then $r(k) \rightarrow_{a.s.} \rho(k)$.

Immediate implication of Proposition 1.8 is that convergence in probability or almost sure convergence of sample autocorrelations depends only on the properties of disturbances and polynomial coefficients $\{\psi_j\}$.

Corollary 1.9 Let $\{\tilde{X}_u\}$ be a stochastic process as defined in (1.15). Applying Slutsky's corollaries to the result of Proposition 1.4, it follows that $plim(r^2(k)) = \rho^2(k)$, and $plim(r_1^2(k) * r_2^2(k)) = \rho_1^2(k) * \rho_2^2(k)$, where r_1 and r_2 are autocorrelation functions computed for different subsamples of $\{\tilde{X}_u\}$.

Up to this point we considered covariance stationary process $\{\tilde{X}_u\}$ specified in (1.15), which was obtained for the true values of the time transformation $\{\lambda^*, \Lambda^*\}$.

10 See, for example, Hannan and Heyde [1972], Romano and Thombs [1996].

Since any other pair $\{\lambda_i, \Lambda_j\} \neq \{\lambda^*, \Lambda^*\}$ will not guarantee ARMA invertibility requirements for the resulting process, conclusions of the Proposition 1.4 can not be applied to $\tilde{X}(u; \lambda_i, \Lambda_j)$. Nevertheless, in (1.10) we introduced process $\tilde{X}(u; X)$ as an interpolated sequence based on the original data $\{X_t\}$: $\tilde{X}(u; X) = X(t_k) + w_k * (X(t_{k+1}) - X(t_k))$, with weights w_k based on the parameters $\{\lambda_i, \Lambda_j\}$ for any i, j. Since such interpolation works for any time-deformation parameters, choosing appropriate weights we may construct a correspondence between a covariance stationary data $\tilde{X}(u; \lambda^*, \Lambda^*)$ and an arbitrary process $\tilde{X}(u; \lambda_i, \Lambda_j)$. Thus, even if wrong pair $\{\lambda_i, \Lambda_j\}$ is chosen, process $\tilde{X}(u; \lambda_i, \Lambda_j)$ may be represented as a transformation of *NED* sequence $\tilde{X}(u; \lambda^*, \Lambda^*)$, and therefore is itself *NED* of the same order. Then, summability of autocovariances and autocorrelations is guaranteed for interpolated process based on any pair $\{\lambda_i, \Lambda_j\}$. This allows us to conclude that for arbitrary transformed $\tilde{X}(u; \lambda_i, \Lambda_j)$ estimated autocorrelations $r(k)$ may not necessarily converge to $\rho(k)$, however, Q-test (1.11) will converge to some positive constant $\bar{Q}$.

Definition 1.10[11] Let $\{\Omega, \Im, P\}$ be a probability space, and let $\{f_n(\theta), n \in N\}$ be a sequence of stochastic functions $f_n : \Theta \times \Omega \to \mathbb{R}$, $\Im / B$-measurable for $\forall \theta \in \Theta$. The sequence is said to be asymptotically uniformly stochastically equicontinuous if $\limsup_{n \to \infty} P(w(G_n, \delta) \geq \epsilon) < \epsilon$ for all $\epsilon > 0$ $\exists$ $\delta > 0$, and $w(.)$ being referred to as the modulus of continuity such that $w(f_n, \delta) = \sup_{\theta \in \Theta} \sup_{\theta' \in S(\theta, \delta)} |f_n(\theta') - f_n(\theta)|$.

Proposition 1.11 A sequence of functions $\{Q_n(\theta)\}$ is asymptotically uniformly equicontinuous on a bounded parameter space Θ. That is, the modulus of continuity $w(Q_n, \delta) = \sup_{\theta \in \Theta} \sup_{\theta' \in S(\theta, \delta)} |Q_n(\theta') - Q_n(\theta)|$ has the property that $\limsup_n (w(Q_n, \delta)) < \epsilon$ for any ϵ.

11 See Davidson [1994]

This proposition ensures that for all Q_n small changes in the parameter values $\{\lambda, \Lambda\} = \theta \in \Theta$ will not result in the breakdown of Q_n, and is a necessary and sufficient condition for the uniform convergence in probability for the sequence Q_n.

Proposition 1.12 Given $\{\Omega, \Im, P\}$ and $\Theta \subset R^K$, a sequence of continuous stochastic functions $Q_n(\hat{\theta})$ converges uniformly in probability to some finite sequence $Q_n(\theta)$ for any $\theta \in \Theta$. That is, there exists $F \in \Im, P(F) = 1$ such that for any $\epsilon > 0$, and for each w in F there exists an integer $N(w, \epsilon) < \infty$ such that for all $n > N(w, \epsilon)$, either $|Q_n(w, \theta) - Q(\theta)| \leq \epsilon$, or $\sup_{\theta \in \Theta} |Q_n(w, \theta) - Q(\theta)| \to 0$ *a.s.* [12]

Proof. For our purposes it is enough to prove only the part about the convergence in probability. By triangle inequality, for any Q_n in the sequence:

$$
\begin{aligned}
&|Q_n(\hat{\theta}) - Q(\theta)| \leq \left|\sum_{k=0}^{K} [(r_1(k;\hat{\theta}) - r_{1.5}(k;\hat{\theta}))^2 + (r_2(k;\hat{\theta}) - r_{1.5}(k;\hat{\theta}))^2]\right| - \\
&- \left|\sum_{k=0}^{K} [(\rho_1(k;\theta) - \rho_{1.5}(k;\theta))^2 + (\rho_2(k;\theta) - \rho_{1.5}(k;\theta))^2]\right| \\
&\leq \left|\sum_{k=0}^{K} [r_1^2(k;\hat{\theta}) - \rho_1^2(k;\theta)]\right| + \left|\sum_{k=0}^{K} [r_2^2(k;\hat{\theta}) - \rho_2^2(k;\theta)]\right| + \\
&+ 2 * \left|\sum_{k=0}^{K} [r_{15}^2(k;\hat{\theta}) - \rho_{15}^2(k;\theta)]\right| + \\
&+ 2 * \left|\sum_{k=0}^{K} [r_1^2(k;\hat{\theta}) * r_{15}^2(k;\hat{\theta}) - \rho_1^2(k;\theta) * \rho_{15}^2(k;\theta)]\right| + \\
&+ 2 * \left|\sum_{k=0}^{K} [r_2^2(k;\hat{\theta}) * r_{15}^2(k;\hat{\theta}) - \rho_2^2(k;\theta) * \rho_{15}^2(k;\theta)]\right|
\end{aligned}
\tag{1.17}
$$

Applying the results of Axioms and Propositions 1.3-1.12 to (1.17) it follows that $|Q_n(\hat{\theta}) - Q(\theta)| \leq \epsilon$, or $\sup_{\theta \in \Theta} |Q_n(\hat{\theta}) - Q(\theta)| \to_p 0$ for any n and for any $Q_{n.}$. Then, using the uniform Law of Large numbers, it follows that $n^{-1} \sum Q_n \to E[Q]$

12 See Gallant and White [1998] for the discussion of uniform convergence.

At this point we established uniform convergence of $Q_n(\hat{\theta})$ to some fixed boundary $Q(\theta)$ To finalize the study of consistency properties of Q-test, we need to make the final assumptions about existence and uniqueness of a minimand for $Q(\theta)$.

Proposition 1.13 Autocorrelation functions computed from the different subsamples of $\tilde{X}_u$ as specified in (1.12) will satisfy equality $\rho_1(k;\theta)=\rho_{1.5}(k;\theta)=\rho_2(k;\theta)$ if and only if $\theta=\theta^*$, $\theta^*\in\Theta$. Thus, $Q(\theta)$ as has a unique minimizer θ^* on a compact set $\Theta\subset R^K$.

Proof. As we established before, $\tilde{X}(u;\theta)$ will follow a causal ARMA process and $r(k;\theta)\rightarrow_{a.s.}\rho(k;\theta)$ only when $\theta=\theta^*$, so $Q(.)\rightarrow 0$. For any other $\theta=\theta'$, $r(k;\theta')\rightarrow\bar{c}$ where $\bar{c}\neq\rho(k;\theta)$.

Corollary 1.14 $\hat{\theta}\equiv(\hat{\lambda},\hat{\Lambda}),\hat{\theta}\in\Theta$, such that $\Theta\subset R^K$, is a consistent estimator of θ_0, or $\hat{\theta}\rightarrow_p\theta_0$. That is, for any fixed non-negative K: $\lim_{n\rightarrow\infty}P(|\hat{\theta}-\theta_0|\geq\epsilon)=0,\forall\epsilon>0$ ◊

So far we proved the consistency of Q-test. The only question remains is what value of K should be chosen to estimate (1.11). Since $r(k,\theta)-\rho(k,\theta)=O_P(N^{-1/2})$, then

$$\sum_{k=1}^{K}(r(k,\theta)-\rho(k,\theta))^2=O_P(K\,N^{-1}), \tag{1.18}$$

and the maximum lag order K should be chosen such that it will guarantee the convergence under the summation. As a possible choice in the simulation studies we use $K=cN^{1/3}$, which corresponds to the convergence rate for the Bartlett kernel (Newey and West [1987]). Other choices are possible as well, for example, $K=c\,N^{2/5}$ which corresponds to the rate of convergence for either Parzen or Quadratic Spectral kernels[13]. Values for the constant c in empirical applications will be determined based on the Monte Carlo simulation in the following section.

13 See Andrews (1991) for discussion and comparison of different kernel functions.

1.5 Small sample performance of the $Q_n(\lambda,\Lambda)$ test

In the original paper, JGW (2006) discuss the estimation of the parameters for several generated data sets. However authors do not study the robustness of the Q-statistic. To evaluate the sensitivity of the proposed Q-test to different data properties in a finite sample, we conduct several Monte Carlo studies.

First, the form of the test (1.11) suggests that the value of the Q-statistic depends on the maximum lag length K chosen to compute ACF's. Since each of three subintervals used in (1.11) has the length of $N/2$, the maximum possible value for K would also be $N/2$. However, longer lags may not necessarily improve the estimation results, since the relation may be weak between observations located too far from each other. To explore how different values of K may affect the time-deformation parameter estimates, Monte-Carlo simulation was used with the maximum lag length being set of order $c*(N/2)^{1/3}$. Table 1.1 summarizes results for the different choices of K.

Table 1.1: Monte Carlo results for different maximum ACF lag orders

$K=c*(N/2)^{1/3}$	**True values: $\lambda=1.5$, $\Lambda=50$**		**True values: $\lambda=0.5$, $\Lambda=50$**	
c, K	λ	Λ	λ	Λ
1, 6	1.483 (0.0019)	41.602 (0.637)	0.462 (0.0022)	58.940 (0.608)
2, 12	1.498 (0.0015)	54.255 (0.635)	0.498 (0.0011)	47.908 (0.357)
5, 30	1.504 (0.0015)	55.070 (0.677)	0.497 (0.0010)	**50.005** (0.322)
10, 60	**1.503 (0.0011)**	55.015 (0.537)	0.498 **(0.0004)**	50.468 **(0.129)**
20, 120	**1.503 (0.0011)**	**53.130 (0.534)**	**0.499 (0.0004)**	50.625 (0.134)
30, 180	1.504 (0.0012)	53.915 (0.571)	0.497 (0.0004)	51.420 (0.162)
40, 240	1.505 (0.0012)	53.885 (0.561)	0.498 (0.0004)	50.930 (0.148)

Reported are mean values of the estimates across 2000 Monte Carlo simulations for the data generated according to (1.13) with sample size N=500, θ=0.1, and σ_ϵ=0.2. Initial specified ranges for λ and Λ are [0..2] and [0..100], respectively. Standard errors of the estimates are given in parentheses. Values with smaller absolute bias or with the smallest standard errors are shown in bold.

Apparently, choosing values for c between 2 and 20 seems to be enough (correspondent values for K are in the range between 12 and 120 for N=500) to estimate both parameters – λ and offset – with sufficient level of precision. Increasing lag horizon beyond that may even reduce the accuracy of estimation in terms of both – bias and the variance of estimates. Based on these results, when estimating Q-statistic in the rest of this paper, the maximum lag length was set as $K = 10 * (N/2)^{1/3}$.

The next test we conduct is intended to examine how sensitive parameter estimates are for different sample sizes. We estimate the values of time-deformation parameters for sample sizes from 100 to 1,000 observations keeping the rest of the parameters fixed. Table (1.2) presents simulation results.

Table 1.2: Monte Carlo results for different sample sizes

	True values: λ=1.5, Λ=50		**True values: λ=0.5, Λ=50**	
Sample size N	λ	Λ	λ	Λ
100	1.919 (0.004)	54.555 (0.681)	0.202 (0.004)	14.212 (0.152)
200	1.660 (0.001)	87.569 (0.316)	0.686 (0.001)	13.910 (0.219)
500	1.503 (0.001)	55.015 (0.537)	0.498 **(0.000)**	50.468 (0.129)
1000	**1.500 (0.000)**	**51.140 (0.048)**	**0.500 (0.000)**	**50.455 (0.020)**

Reported are mean values of the estimates across 2000 Monte Carlo simulations for the data generated according to (1.13) with θ=0.1, lag length $K = 10 * (N/2)^{1/3}$, and $\sigma_\epsilon = 0.2$. Initial specified ranges for λ and Λ are [0..2] and [0..100], respectively. Standard errors of the estimates are given in parentheses. Values with smaller absolute bias or with the smallest standard errors are shown in bold.

Our results suggest that in relatively small samples the Q-statistic performs rather poorly. Only when sample sizes are larger than 200 observations, the test provides adequate estimates. One possible explanation for this phenomenon is that small data range may not contain enough full cycles to capture the underlying cyclical behavior, especially in case of elongating cycles.

Another issue examined is how robust the test is to the presence of noise in the data. For the sample of 500 observations generated according to (1.13), we consider values of the disturbance variance σ_ϵ^2 ranging between 0.02 and 1. Results are given in Table 1.3.

Based on the simulation results, we conclude that Q-test allows us to estimate both parameters of time deformation quite accurately, even in the case of a low signal-to-noise ratio, such as in the case when $\sigma_\epsilon = 1$. Increasing disturbance variance leads to more biased results, however the difference is minor. What creates a larger concern is significantly increased variance of the estimates (especially it is pronounced for the offset values).

Table 1.3: Monte Carlo results for the different noise levels

	True values: λ=1.5, Λ=50		**True values: λ=0.5, Λ=50**	
σ_ϵ^2	λ	Λ	λ	Λ
0.02	**1.500 (0.000)**	**52.335 (0.011)**	**0.500 (0.000)**	**49.860 (0.008)**
0.2	1.503 (0.001)	55.015 (0.537)	0.498 **(0.000)**	50.468 (0.129)
1	1.490 (0.002)	47.228 (0.710)	0.484 (0.002)	58.018 (0.636)

Reported are mean values of the estimates across 2000 Monte Carlo simulations for the data generated according to (1.13) with sample size $N=500$, $\theta=0.1$, maximum lag length $K=10*(N/2)^{1/3}$. Initial specified ranges for λ and Λ are [0..2] and [0..100], respectively. Standard errors of the estimates are given in parentheses. Values with smaller absolute bias or with the smallest standard errors are shown in bold.

Two examples of the simulated series with $\sigma_\epsilon^2 = 1$ and $\sigma_\epsilon^2 = 0.02$ are shown in Figure 1.4.

(a)

(b)

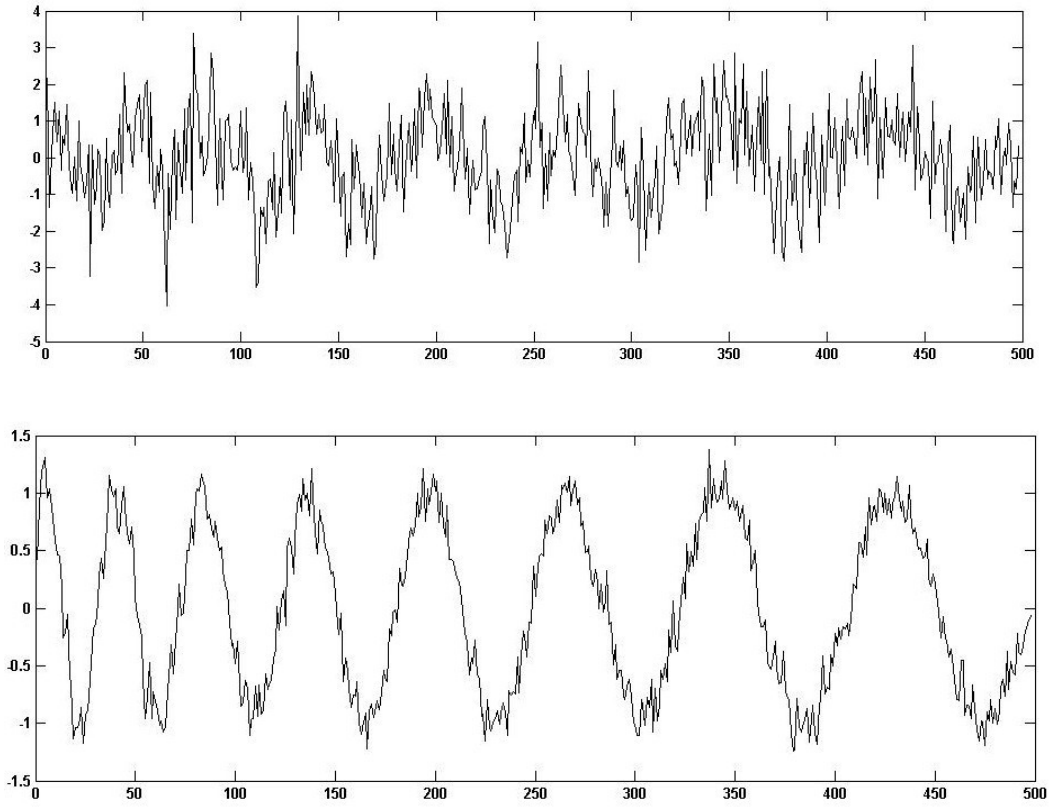


Figure 1.4: Simulated data with different noise levels

The data generated according to (1.13) with (a) $\sigma_\epsilon^2=1$, and (b) $\sigma_\epsilon^2=0.02$.

Up to this point, all the performed simulation studies were based on rather deterministic process (1.13). To assess the adequacy of the Q-test, as well as to test the accuracy of parameter estimation, we carry out the additional set of Monte Carlo experiments for more realistic macroeconomic data.

We model cyclical covariance nonstationary process for the given λ and offset Λ in the form of the structural time series model (Harvey [1993]):

$$X_t = \psi_t + \epsilon_t, \quad t = 1, \ldots, N, \quad \epsilon_t \sim NID(0, \sigma_\epsilon^2). \tag{1.19}$$

where ψ_t is an unobservable stochastic cycle component, specified by the following autoregressive process:

$$\begin{pmatrix} \psi_{(t+1)} \\ \dot{\psi}_{(t+1)} \end{pmatrix} = \rho \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{pmatrix} \psi_t \\ \dot{\psi}_t \end{pmatrix} + \begin{pmatrix} k_t \\ \dot{k}_t \end{pmatrix} \tag{1.20}$$

with $|\rho| < 1$ and $k_t, \dot{k}_t \sim NID(0, \sigma_k^2)$.

Values for the dampening term ρ, frequency θ, as well as cycle and irregular variances are chosen close to the values reported in the literature for macroeconomic time-series (for example, Koopman and Lee [2005]). Estimation of the parameters is carried out via maximizing the likelihood function obtained from the Kalman filter output applied for the original nonstationary data $X(t)$, as well as for the transformed series $\tilde{X}(t)$. Transformation was performed based on the values of λ and Λ estimated using Q-statistic. For the $G(\lambda)$ process we use $\lambda = 0.5$, which will produce elongating cycle. To simplify the exposition, offset value is taken as zero, thus the origin of simulated data will coincide with the origin of $G(\lambda)$ process. Nevertheless, we add offset as an additional parameter to be estimated. Table 1.4 presents the results.

Table 1.4: Monte Carlo results for the structural time series model

	True value	**Original (covariance-nonstationary) data** $X(t)$			**Transformed data** $\tilde{X}(u)$		
		Estimate	**Abs. Bias**	**RMSE**	**Estimate**	**Abs. Bias**	**RMSE**
θ	**.2**	0.166 (0.0007)	0.034	0.047	0.210 (0.0004)	0.010	0.026
ρ	**.9**	0.921 (0.0004)	0.021	0.027	0.930 (0.0002)	0.030	0.041
λ	**.5**				0.462 (0.0021)	0.038	0.152
Λ	**0**				10.26 (0.4423)	10.26	23.99

Reported are mean values of the estimates across 2000 Monte Carlo simulations for the data generated according to (1.19-1.20) with lag length $K=10*(N/2)^{1/3}$. Initial specified ranges for λ and Λ are [0..2] and [0..100], respectively. Standard errors of the estimates are given in parentheses.

Results demonstrate that the time-deformation parameter λ is estimated precisely enough, while offset value is slightly biased. Still, the most important conclusion is that estimation over the transformed data allows us to compute cycle frequency more precisely than for the original data – the estimated value of 0.21 is close enough to the true value. The frequency estimate of 0.166 for non-transformed data considerably underestimates the true frequency of sample cycles. For the rest of the parameters, both data sets provide rather similar estimates. Hence, this may be considered as the evidence that in the case of cyclical data with changing cycle pattern, estimation and forecasting in the deformed time domain may provide superior results.

1.6 Empirical Evidence

In this section, estimation in the deformed time is applied to the historical U.S. unemployment series. The data are monthly percentage unemployment rates reported by BLS, seasonally adjusted at the source, and covering the period between 1970:1 and 2007:12. The sample consists of 456 observations, with at least 3 full cycles present in the data.

1.6.1 Periodic structural time series model

The behavior of the US unemployment data is modeled using structural time series modeling. Unlike commonly used ARMA-type models, a structural model allows us to split the data explicitly into the trend and cycle components. This feature is especially important, because our main interest is in detaching and examining the cycle. Another advantage of the structural model is its robustness to the trend nonstationarity. For the unemployment data used in the following section, Augmented Dickey-Fuller (Dickey and Fuller, [1979]) test reports that the null of nonstationarity can be accepted at 5% level[14]. Thus, ARMA-modeling would require additional nonstationarity fixing procedures.

Following Harvey [1993], the univariate structural time series model can be constructed as follows:

$$y_t = \mu_t + \psi_t + \epsilon_t, \quad t = 1 \ldots N \tag{1.21}$$

where y_t is the observation at time t, μ_t is the trend, ψ_t is the cyclical component, and ϵ_t is an idiosyncratic shock with $\epsilon_t \sim NID(0, \sigma_\epsilon^2)$.

14 Lag length was set at 17 based on Schwarz Info Criterion.

Trend component is built as the local level model:

$$\mu_{t+1}=\mu_t+\beta_t+\xi_t, \quad \xi_t \sim NID(0,\sigma_\xi^2), \tag{1.22}$$

$$\beta_{t+1}=\beta_t+\zeta_t, \quad \zeta_t \sim NID(0,\sigma_\zeta^2), \tag{1.23}$$

where level and slope innovations, ξ_t and ζ_t, respectively, are normally and independently distributed. Such specification allows us to account for different behavior of the trend component μ_t. If $\sigma_\zeta^2=0$, then $\beta_{t+1}=\beta_t=\beta$ and trend follows random walk with drift. When $\sigma_\xi^2=0$, trend becomes an integrated random walk model. In case both innovations, ξ_t and ζ_t, have zero variance, a deterministic trend specification is obtained.

The deterministic cyclical component is specified as follows:

$$\psi_t=a*\cos(\theta_t t-b) \qquad a\neq 0,\theta_t\neq 0 \tag{1.24}$$

with a, θ_t, and b being amplitude, frequency, and phase, respectively.

Frequency, θ_t, is measured in radians with $0<\theta_t<\pi$. The period of the cycle is obtained as $2\pi/\theta_t$. In case frequency is constant, ($\theta_t=\theta$) a stationary cycle is obtained with $\psi_{(t-\tau)}=\psi_{(t+\tau)}$. Time-varying frequency allows us to model elongating and dampening cycles in the time series. Koopman and Lee (2005) modeled asymmetry of the cycle based on its steepness: $\theta_t=\theta+\gamma\dot{\psi}_t$, where $\dot{\psi}_t=\vartheta\psi_t/\vartheta(\theta t)=-a\sin(\theta t-b)$ is the derivative of the cycle component in (1.24), which determines whether the cycle is ascending or descending. Such approach allows us to capture different frequencies over rising or declining periods of a single cycle, yet still assumes that all cycles presented in the data are identical.

Specification (1.24) can be converted into the autoregressive process:

$$\begin{pmatrix}\psi_{(t+1)}\\ \dot{\psi}_{(t+1)}\end{pmatrix}=\begin{bmatrix}\cos\theta_t & \sin\theta_t\\ -\sin\theta_t & \cos\theta_t\end{bmatrix}\begin{pmatrix}\psi_t\\ \dot{\psi}_t\end{pmatrix} \tag{1.25}$$

Adding a damping term ρ and cyclical innovations to (1.25) provides a stochastic version of cycle:

$$\begin{pmatrix}\psi_{(t+1)}\\ \dot{\psi}_{(t+1)}\end{pmatrix}=\rho\begin{bmatrix}\cos\theta_t & \sin\theta_t\\ -\sin\theta_t & \cos\theta_t\end{bmatrix}\begin{pmatrix}\psi_t\\ \dot{\psi}_t\end{pmatrix}+\begin{pmatrix}k_t\\ \dot{k}_t\end{pmatrix} \tag{1.26}$$

with $|\rho|<1$ and $k_t, \dot{k}_t \sim NID(0,\sigma_k^2)$.

All formulas considered above for structural time series can be conveniently put into the linear state-space form with observation and transition equations defined as:

$$y_t=Z_t\alpha_t+\epsilon_t, \qquad \epsilon_t \sim NID(0,H_t) \tag{1.27}$$

$$\alpha_{t+1}=T_t\alpha_t+\eta_t, \quad \eta_t \sim NID(0,Q_t) \tag{1.28}$$

with initial state vector $\alpha_1 \sim NID(\alpha_{1,}P_1)$, and disturbances ϵ_t and η_t being mutually and serially uncorrelated Gaussian processes.

The matrices and vectors of state-space form are given as:

$$\alpha_t=(\mu_t \quad \beta_t \quad \psi_t \quad \dot{\psi}_t)' \tag{1.29}$$

$$\eta_t=(\xi_t \quad \zeta_t \quad k_t \quad \dot{k}_t)' \tag{1.30}$$

$$Z_t=[1 \quad 0 \quad 1 \quad 0] \tag{1.31}$$

$$T_t=\begin{bmatrix}1 & 1 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & \rho\cos\theta_t & \rho\sin\theta_t\\ 0 & 0 & -\rho\sin\theta_t & \rho\cos\theta_t\end{bmatrix} \tag{1.32}$$

$$Q_t = \begin{bmatrix} \sigma_\xi^2 & 0 & 0 & 0 \\ 0 & \sigma_\zeta^2 & 0 & 0 \\ 0 & 0 & \sigma_k^2 & 0 \\ 0 & 0 & 0 & \sigma_k^2 \end{bmatrix} \qquad (1.33)^{15}$$

Estimation of the parameters of the model is implemented using Kalman filtering algorithm.

1.6.2 Estimation of the time-scale parameters

For the unemployment data described before, the estimated values of the time-deformation parameter λ and offset are found to be equal to 0.8 and 18, respectively. The value of λ is smaller than 1, indicating the elongating cyclical nature of the data, with the cyclical behavior being in between M-stationary process and a regular stationary process. Also, the sample is shifted by 18 time intervals relative to the original $G(\lambda = 0.8)$ process. Elongating pattern can also be supported by visual inspection of the original data (Figure 1.5). Original data series were transformed according to the estimated parameters and then interpolated on the equally spaced time scale. To match the transformed data to the original one, time interval in the equally-spaced deformed *u*-scale was normalized to be equal to the regular step in the t-scale. A plot of the original series, along with the transformed data, is given in Figure 1.5. While values of unemployment rates remain the same in both series, the timing is different. At the beginning of the time scale the data are prolonged, while near the end the data points are compressed.

15 Here the time subscript for matrix Q can actually be omitted, since its components are not time varying.

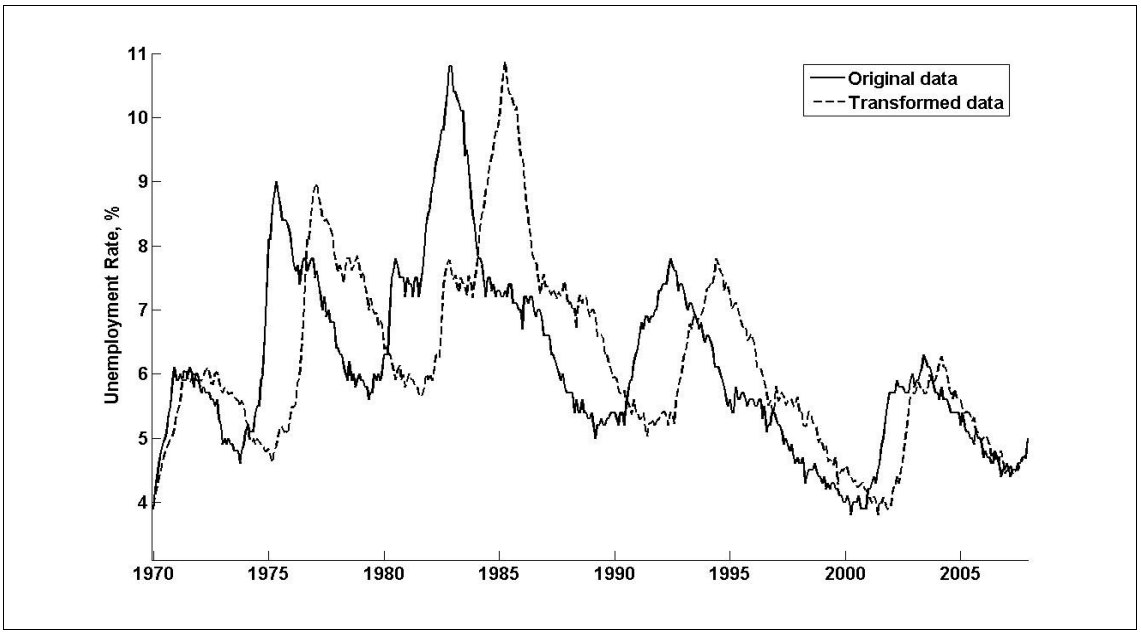


Figure 1.5: Original and transformed series for the U.S. unemployment rate

1.6.3 Fitting the SSM model to the data

After finding time-deformation parameters, unobserved component model in the form of (1.21-1.26) was estimated for the original and transformed data[16]. Results are reported in Table 1.5.

16 For the unemployment data innovations variance for the level of trend σ_{ξ}^{2} was found to be insignificant in this as well as in many previous studies, thereby is excluded from the UC specification (2.2).

Table 1.5: UC model estimation results for the U.S. unemployment

	Original series	**Transformed series**
σ_{ϵ}^2	0	0
σ_{ς}^2	7.87e-6 (0.0436)	6.55e-6 (0.0588)
σ_{k}^2	0.025 (0.0031)	0.014 (0.0032)
Frequency, θ	0.073 (0.0046)	0.077 (0.0034)
Period (months)	86.1	81.5
Dampening term, ρ	0.986 (0.0187)	0.992 (0.0227)
Log-likelihood	138.7	257.9
Jarque-Bera normality test	8.3 (0.02)	5.5 (0.06)
Ljung-Box (20) test	193.9 (0.00)	56.3 (0.00)

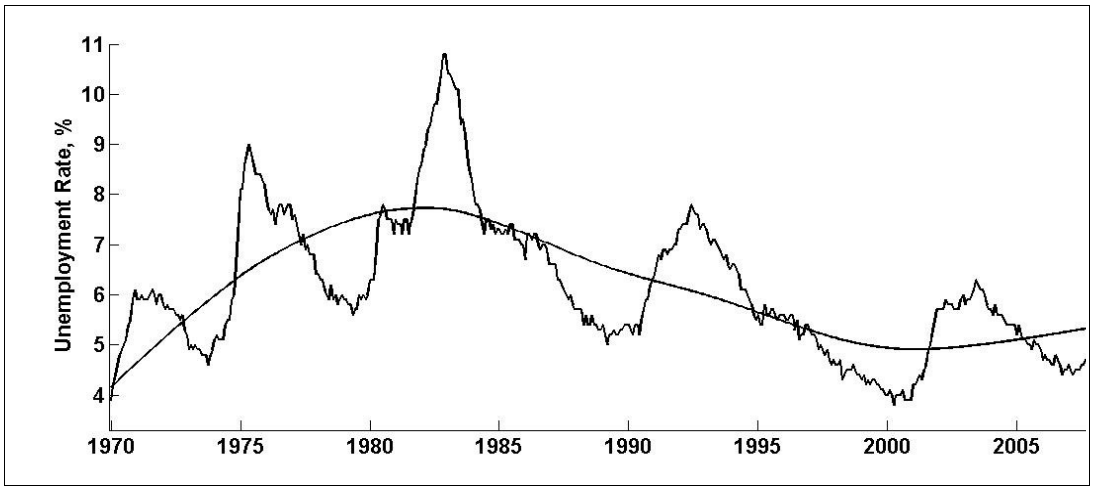
Reported are MLE parameter estimates for the UC model specified in (1.21-1.26) fitted to the U.S. unemployment rate series covering the period from 1970:1 to 2007:12. Disturbance variances were restricted to prevent non-positive values. Damping term was restricted as $0<\rho<1$. Standard errors of the estimates are given in parentheses.

Obtained results are consistent with the existing literature. Irregular variance and trend variance appear to be insignificantly different from zero, which is a common feature of many macroeconomic time series. Estimated variance of the cycle component is found to be significantly smaller for the transformed data. This result can be explained by the fact that in the deformed time we created a stationary and therefore more homogenous cyclical component. At the same time, for the original series, higher variance reflects different cyclical structure across our data sample.

Average cycle length is also found to be different – for the transformed data, average cycle length is shorter by 4.5 months on average. The transformed data set is also marked with an increase in the persistence coefficient, although the difference is not substantial. Test results report a significant increase in the likelihood values for the estimation over the transformed series. Jarque-Bera normality test suggests that normality assumption can be accepted at 2% and 6% levels of significance for the original and transformed

series, respectively. Ljung-Box serial correlation test up to twenty lags suggests that both specifications reject the Null about the absence of serial correlation. However, results for the transformed data demonstrate a considerable improvement in reducing the correlation. Plots of the decomposed smoothed trend and cycle components are shown in Figures 1.6-1.7. As illustrated in the plots, cyclical behavior of the original data is subject to periods of increasing length, while deformed data is characterized by approximately equal periodicity.

(a)



(b)

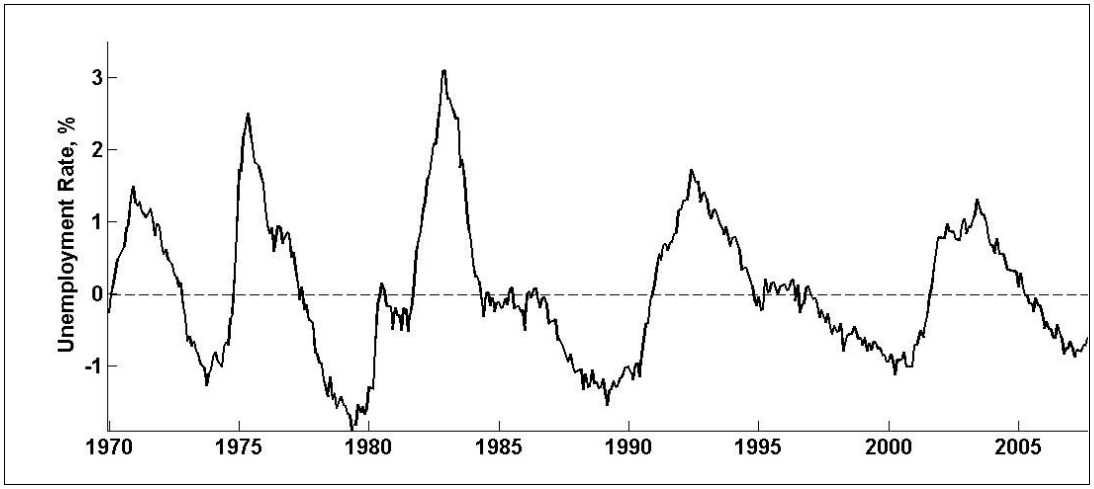


Figure 1.6: Trend-cycle decomposition of the original unemployment data

(a) The data and smoothed trend. (b) Smoothed cycle.

(a)

(b)

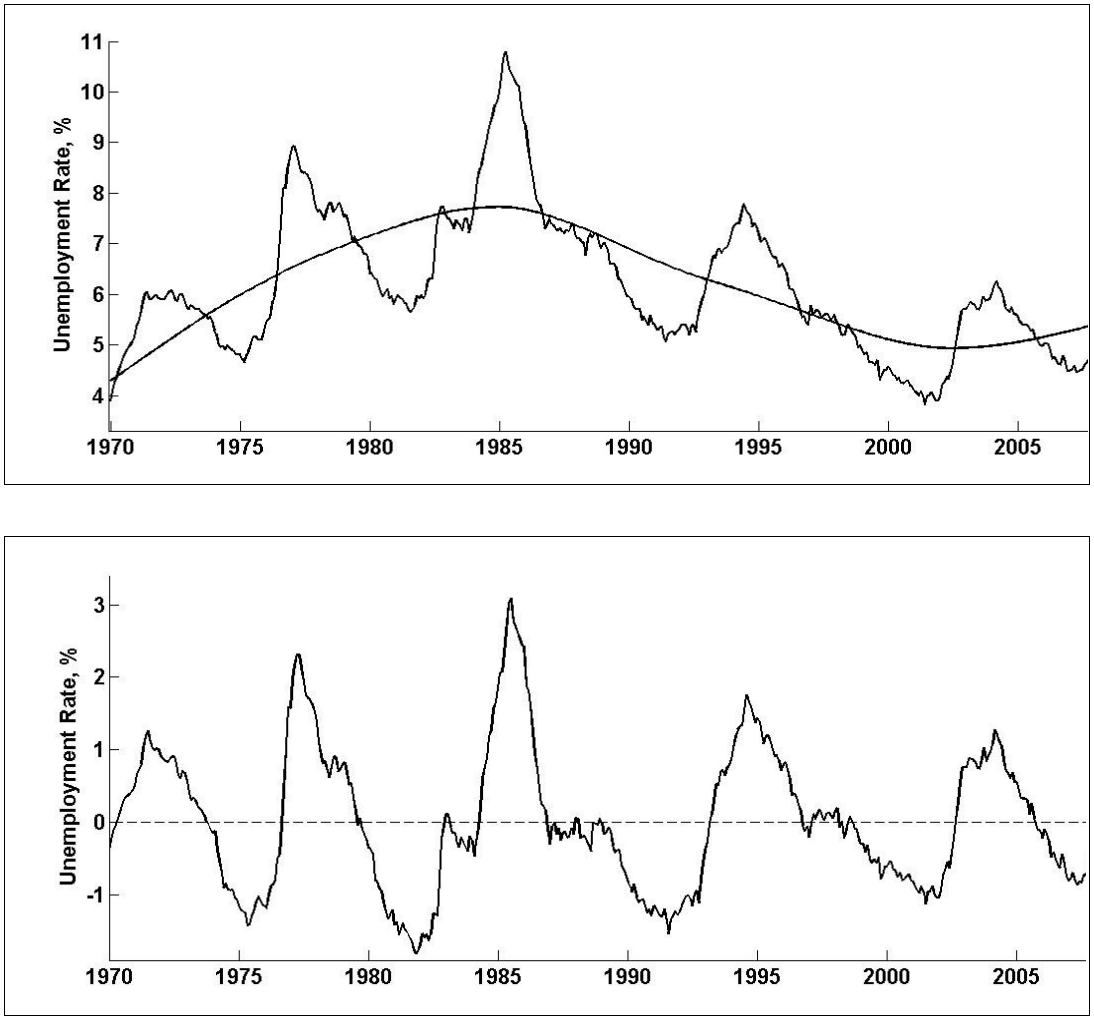


Figure 1.7: Trend-cycle decomposition of the transformed unemployment data

(a) The data and smoothed trend. (b) Smoothed cycle.

1.7 Conclusions

The purpose of this paper was to discuss the issues of exploring covariance nonstationary data. As the novel feature in the given area, we suggests reconsidering the original (calendar) timing of occurrence for the economic data. This is implemented by transforming the original time scale into a new scale, in which nonstationarity of the data is removed.

We proved the consistency properties of the Q-test introduced by Jiang, Gray, and Woodward [2006] which can be used to estimate parameters relating regular and transformed time scales. In several sets of simulations we also investigated robustness of the test with respect to the data features.

As the empirical estimation illustration, time-deformation methods were used to study the behavior of the U.S. unemployment data. We found that estimated cyclical component of unemployment clearly demonstrated elongating behavior for the whole sample. This finding suggests that regular estimation techniques may provide inadequate results in the economic studies, since many econometric models are built on the assumption of stationarity of the underlying processes. To examine this statement, we compared estimation results for unemployment series in the framework of the structural time series models. In addition to the significant improvement in estimation statistics we found that our approach reports a shorter cycle length and smaller variability of irregular components comparing to the traditional model.

2 Time Deformation with Varying Cyclical Behavior

2.1 Introduction

In the previous section we introduced $G(\lambda)$-stationary process which allows us to model the stationary data with frequencies varying over time. The value for the time-deformation parameter for the unemployment data was estimated to be 0.8, which means that unemployment rates in general exhibited elongating pattern throughout the whole range (1970:1 – 2007:12). However, it may not necessarily be the case when considering smaller intervals inside the whole series – as it was mentioned before, the cyclical speed of major macroeconomic variables may change at different times or phases of the cycle. Therefore, time-deformation process outlined in Section 1 may be revised when applying to the shorter intervals. To access the plausibility of this question, we estimate time-deformation parameters for shorter periods, and evaluate the results in terms of forecasting performance. As will be shown later, the development path of the time-deformation parameter λ possesses its own dynamics, and may be modeled as an independent process with underlying parameters changing values at different periods.

The use of Markov-switching (MS) models for the study of macroeconomic time series began with the seminal work of Hamilton [1989] in which he introduced a non-linear filtering procedure and used it to estimate autoregressive MS model for US GNP growth data. Since then, Markov-type switching models have been widely used in macroeconomics and finance – some examples are the Garcia and Perron [1996] study of ex-post real interest rates, Kim and Nelson [1998] model of stock returns with regime-changing, and many more[17].

The distinguishing feature of this research is the analysis of regime-switching features not in the original data, but in the process determining the timing of the observations.

17 Kaufman and Scheicher [1996] present a selective overview of works in this field.

2.2 Forecasting with varying frequencies

To evaluate the adequacy of the estimation in the deformed time, forecasting performance in both time scales was compared based on the unemployment data discussed in the previous chapter. Forecasting was performed using a rolling window of 360 observations (30 years) for the parameter estimation, and computing the forecast for the following 12 months. Then the rolling window was moved one step forward, providing a total of 96 forecasts for each lead time. In the case when time transformation was used, parameters relating the two time scales - λ and Λ - were estimated for the corresponding window, and the data series were transformed based on those values.

Plot of obtained lambdas from each of 96 windowed intervals is provided in Figure 2.1. Time axis shows months and year corresponding to the last observations of each 360-months interval for the corresponding λ estimate.

Figure 2.1: Estimated values of λ.

The results presented in Figure (2.1) suggest that at the beginning of the considered sample and during the period of 2004:7-2005:9, the estimated $\lambda's$ were found to be smaller than 1, indicating elongating cycles. However, in the period 2004:02-2004:06 and by the end of the sample, computed values of lambda were larger than 1, meaning that cycles were exhibiting increasing frequency at those periods. This feature of changing cyclical behavior is also supported by the visual inspection of the data – at the beginning of the considered period, average cycle length was smaller than in the middle of the sample. However, decline in the unemployment rates in 2007 was relatively short, compared to the previous cycles. This suggests that using a rolling window allows us to capture the varying cyclical behavior when the data is subject to a changing frequency.

One potential problem related to forecasting in the deformed time is that the number of discrete time intervals does not coincide between the original and the deformed time scales. For example, if in the original scale forecasting period is $t^f = 12$, then for a specific case with $\lambda = 0.5, \Lambda = 20$, the forecast length in u-scale will correspond to $u^f = 8$ time intervals. At the same time, for $\lambda > 1$, forecast in the deformed time will require more values to be forecasted than those in the original scale. Additional source of reduced forecast accuracy for the transformed series comes from the necessity to interpolate values forecasted in the u-scale back to the regular t-scale.

Nevertheless, we found that using transformed stationary data provides adequate results. Based on the estimated time-deformation parameters, we transformed time scale at each rolling window and fitted structural time series model (1.21-1.26) to the unemployment series. Based on the estimated parameters, 1 to 12-month unemployment forecasts were created. Forecasted values of the unemployment rates obtained in the deformed time were interpolated back into the original time scale and compared to the true data.

Table 2.1 compares forecast results from the UC model of the form (1.21-1.26) for the unemployment data for the original and deformed time scales.

Table 2.1: UC model forecast results in the original and deformed time scale

	Lead Time (months)											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
PMSE: Original time	**0.015**	**0.027**	**0.045**	**0.068**	**0.091**	**0.125**	**0.161**	**0.202**	0.254	0.310	0.377	0.454
PMSE: Transformed time	0.021	0.036	0.056	0.078	0.101	0.132	0.166	0.207	**0.240**	**0.287**	**0.335**	**0.407**
DM test	**-2.28 (0.02)**	-1.31 (0.19)	-0.18 (0.86)	-0.16 (0.87)	0.10 (0.92)	0.39 (0.69)	0.64 (0.52)	1.35 (0.17)	**1.71 (0.08)**	**2.28 (0.02)**	**3.64 (0.00)**	**3.68 (0.00)**

Values in the table are Prediction Mean Squared Errors based on the out-of-sample forecast using a rolling window of 360 observations. For PMSE's smaller numbers are given in bold. For the Diebold-Mariano test p-values are given in parentheses. Diebold-Mariaono test is based on squared error loss function.

In terms of prediction mean squared errors, for up to 8-month ahead, forecast over the regular series appears to provide superior results. Then, for the longer leads, forecast in the deformed time provides a better fit to the data. However, as Diebold-Mariano test reports, the null of equal predictive accuracy can not be rejected for all forecast leads from the second to the ninth. The forecast in the original scale is significantly different only for the first month. At the same time, for longer leads – 9 months and above – forecast with the deformed data gives significantly better results.

Such a difference in forecast performance over different horizons may be explained by the fact that for the short periods PMSE for the deformed data may result from the interpolation errors, while the changing nature of the cycle may not be so pronounced. However, as forecast horizon increases, the gain from considering the changing frequency becomes more important and time-varying model provides better forecast as the cycle changes the speed over time.

Similar difference in the forecasting performance for shorter and longer leads was also confirmed when using a rolling window of 300 and 400 observations.

Figure 2.2 presents a forecast for the last 12-month interval (2007:1 – 2007:12) based on 360-month estimation window.

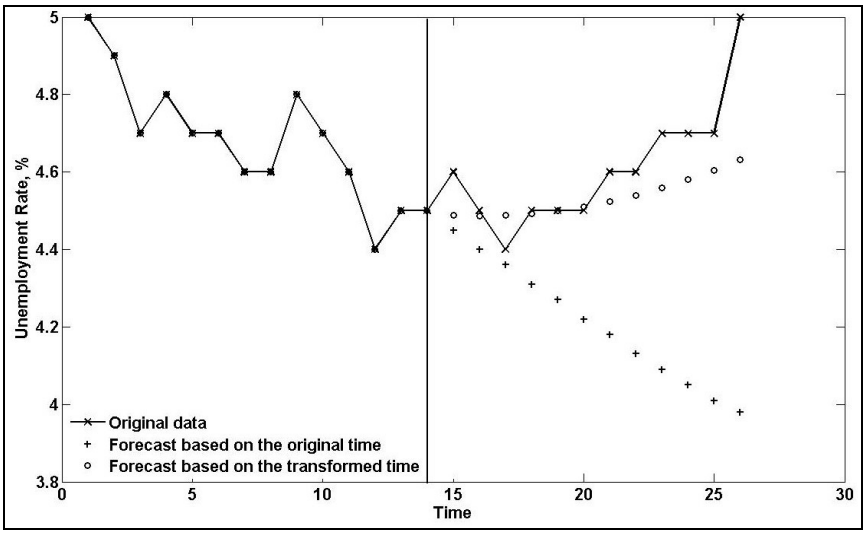


Figure 2.2: Unemployment rate forecast for 2007:1 – 2007:12

For this window, estimated λ was found to be equal to 1.2, suggesting that over that interval unemployment was exhibiting a shortening cycle. Forecast in the transformed time allowed us to take into account faster cycle climbing up beginning in 2007, while the forecast based on the original series failed to capture this feature.

For the additional comparison, we also estimate a simple $AR(p)$ model and perform a 12-month forecast for the original and the transformed data. Similar to the unobserved component model estimation, the forecast was obtained based on the estimated AR model parameters using a rolling window of 360 observations. In case of the transformed data, parameters of time-deformation were estimated for each window, and the data was transformed correspondingly. After this, $AR(p)$ model was fitted to the data with AR order being set in the range $p=1...12$, and the resulting choice of p was

based on the Akaike Information criterion. Table 2.2 summarizes forecast results.

Table 2.2: AR(p) model forecast results in the original and deformed time scales

	Lead Time (months)											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
PMSE: Original time	**0.024**	**0.061**	**0.110**	0.223	0.363	0.552	0.798	1.098	1.455	1.873	2.335	2.878
PMSE: Transformed time	0.030	0.064	0.124	**0.169**	**0.248**	**0.357**	**0.490**	**0.658**	**0.858**	**1.086**	**1.348**	**1.645**
DM test	-1.43 (0.15)	-0.32 (0.75)	0.90 (0.37)	**2.59 (0.01)**	**3.62 (0.00)**	**3.64 (0.00)**	**3.48 (0.00)**	**3.36 (0.00)**	**3.25 (0.00)**	**3.20 (0.00)**	**3.13 (0.00)**	**3.09 (0.00)**

Values in the table are Prediction Mean Squared Errors based on the out-of-sample forecast using a rolling window of 360 observations. For PMSE's smaller numbers are given in bold. For the Diebold-Mariano test p-values are given in parentheses. Diebold-Mariaono test is based on squared error loss function.

In general, forecast results for the AR specification appear to be much worse than for the UC model (shown in Table 2.2) as measured by prediction MSE's for either original or transformed series. Nevertheless, the predictive performance in different time scales has the same pattern. For $AR(p)$ model, forecast in the original time scale provides a smaller MSE for the short leads – up to 3 months. However, as reported by the Diebold-Mariano test, predictive accuracy is insignificantly different between two series. However, for longer forecast horizons – from 4 months and up – forecast over the deformed time scale provides significantly better fit to the data. The results are similar to those obtained for the UC model, although for the AR model the gain in terms of the forecast precision is obtained also for shorter forecast horizons.

2.3 Determining regime changes

In Section 2.2 we estimated values for the time-deformation parameter λ for the monthly data using the period from 1970:1 to 2007:12 using rolling window of 360 observations. As Figure (2.3)[18] suggests, values of λ gradually change over the long range with several significant breaks occurring for the windows ending at 10/2000, 02/2004, 07/2004, 08/2005. Much smaller change in the dynamics of λ was captured on 05/2002 as well as an outlier (or possible regime switch) on 10/2007.

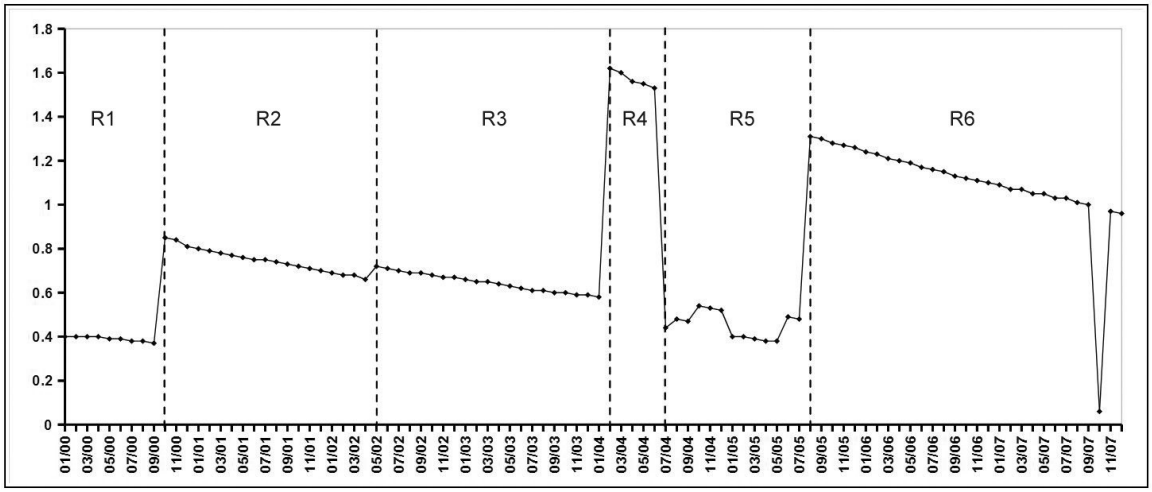


Figure 2.3: Estimated values of λ.

Such abrupt changes in the values of lambda are due to the turning points of the unemployment cycles. $Q(\lambda, \Lambda)$ -test used for the estimation, is based on the sample autocorrelations, which change when a cycle passes a turning point. Since the sign of the autocorrelations changes as well as their magnitude – since a cycle has different speeds for ups and downs – this changes the nature of the time-deformation pattern. In Figure (2.4) we plot the original unemployment rate data along with the coverage for each

18 For the ease of exposition, Figure (2.1) is replicated with added regimes.

regime based on the rolling window size and the length of each regime. From the figure it is obvious that either the beginning or the end of each regime corresponds to the peak or to the trough in the unemployment cycles. Thus, estimated lambdas experience a change when a rolling window passes the peak or trough.

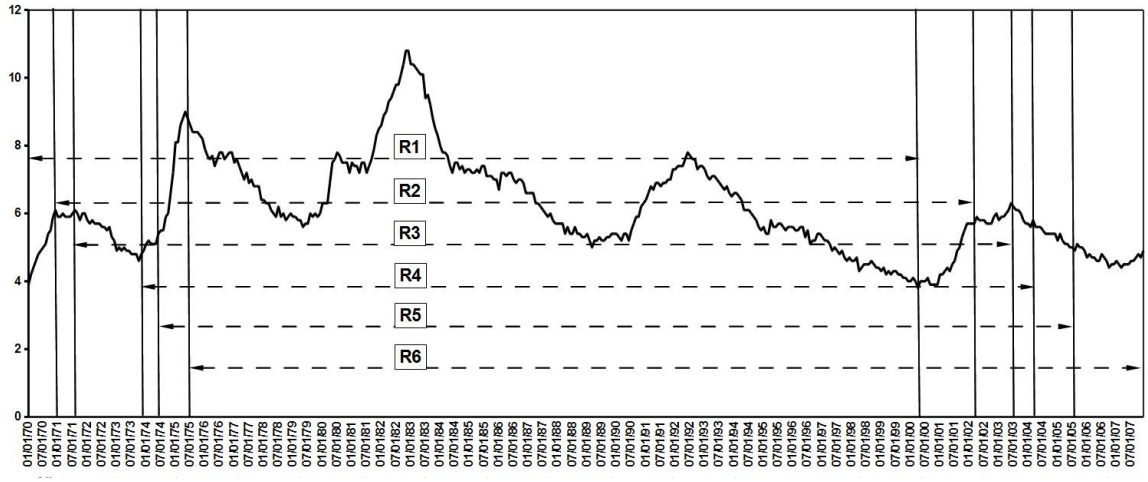


Figure 2.4: Regime zones for the unemployment rate data

The length of rolling window was selected of 360 observations in order to capture at least two complete cycles of unemployment fluctuations. As our data suggests, the longest full cycle (from trough to peak) run from 05/1979 through 06/1992 covering 157 months. Estimating the parameter of time-deformation over shorter windows – 240, 300 observations – generates a similar dynamics for lambda, although with more volatile estimates. Longer windows of 400 or more observations provide more averaged value for λ, which is close to the average value of 0.8, estimated in section (1.6.2).

As was mentioned before, the nature of the correspondence between time scales is governed by two parameters – λ and offset Λ. However, offset plays an important role in locating our sample relative to the origin of the series, while λ itself is used to find a specific process corresponding to our data. The latter seems to be more important than just finding the exact location, so in this chapter we concentrate only on exploring the dynamics of a single lambda. Also, estimated values of the offset were found to be

fluctuating in the range between 1 and 20 without any specific pattern, which, as suggested by the results of section (1.5), may be due to the estimation problems in a small sample.

To test if the breaks in the parameters governing the dynamics of lambda are significant, we assume that values of s_t are known a-priori, and run OLS regression with lambda being a regressand and a set dummy variables representing discrete regimes outlined previously. As an additional variable, a linear time trend was included into the model:

$$\lambda_t = \mu_t + \beta_1 R_1 + \beta_2 R_2 + .. + \beta_6 R_6 + \beta_7 time + \epsilon_t \tag{2.1}$$

Estimation results are summarized in Table (2.3).

Table 2.3: Results of OLS regression

μ	**R2**	**R3**	**R4**	**R5**	**R6**	**Time**
0.456 (0.034)	0.544 (0.046)	0.710 (0.069)	1.810 (0.096)	0.812 (0.102)	1.736 (0.133)	-0.013 (0.0016)

Results for the OLS regression (2.1). Standard errors are given in parentheses.
Regression $R^2 = 0.92$, Durbin-Watson statistics = 1.97, F-statistic=162.52 (p-val=0.00)

All the parameters appear to be highly significant indicating that regression intercept does change for different regimes. To evaluate the significance of these parameters breaks, Chow Breakpoint test was employed. As shown in Table (2.4), the null about parameter constancy can be clearly rejected for all the periods.

Table 2.4: Results of Chow Breakpoint test

Regimes	**R1-R2**	**R2-R3**	**R3-R4**	**R4-R5**	**R5-R6**
Chow test statistic (p-value)	6159.54 (0.00)	181.03 (0.00)	52299.15 (0.00)	308.57 (0.00)	76.72 (0.00)

Chow test of the Null hypothesis that model parameters remain constant between the consecutive regimes R_i and $R_{i=1}$. Test p-values are given in parentheses.

2.3.1 Markov-Switching Regime Model

Abrupt and significant breaks in the path of lambda observed previously give the potential to endogenize its behavior assuming that dynamics of lambda is driven by an unobserved Markov process. Following Hamilton [1994], the general form of a process with parameters changing at different regimes may be written as follows:

$$y_t = \beta_{1,S_t} x_{1,t} + \beta_{2,S_t} x_{2,t} + \ldots + \beta_{p,S_t} x_{p,t} + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma_{S_t}) \tag{2.2}$$

where s_t represent the state at time t, and takes integer values $s_t = 1, \ldots, K$.

The value of s_t is unobserved and it is assumed that it depends only on its recent value following a first-order Markov chain with probability:

$$p_{ij} = p(s_t = j | s_{t-1} = i, s_{t-2} = k, \ldots) = p(s_t = j | s_{t-1} = i) \tag{2.3}$$

where p_{ij} is the transition probability that state i will be followed by state j.

For *K* regimes, probabilities of switching between regimes can be cast into the transition matrix *P*:

$$P = \begin{bmatrix} p_{11} & p_{21} & .. & p_{K1} \\ p_{12} & p_{22} & .. & p_{K2} \\ .. & .. & .. & .. \\ p_{1K} & p_{2K} & .. & p_{KK} \end{bmatrix} \tag{2.4}$$

with all individual probabilities bounded by the unit interval: $0 \le p_{ij} \le 1, \quad \forall i, j = 1, .., K$, and all columns entries summing up to unity: $\sum_{j=1}^{K} p_{ij} = 1$.

As suggested by OLS results, in each phase the developing path of λ contains the linear time trend. Therefore the dynamics of time switching parameter λ, was estimated in the following form:

$$\lambda_t = \mu_{S_t} + \beta_{S_t} * time + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma_{S_t}) \tag{2.5}$$

with all three parameters – intercept term, slope, and irregular variance allowed to change their value according to the state.

Essential questions arises how many regime breaks should we specify when estimating model (2.5). Markov-switching literature suggests several tests to determine how many phases are present in the data[19]. Hansen [1992] and Garcia [1998] examined the asymptotic properties of likelihood ratio tests. Carrasco, Hu, and Ploberger [2004, 2009] developed the test based on the Information matrix, which requires estimating the model only under the null. For the purpose of given research, we employ LR-test. However, as mentioned by several authors (Hamilton [1989, 2005], Carrasco et al [2004]), the LR test is not identified under the null hypothesis of no regime switching and does not have asymptotic χ^2 distribution. To solve this problem, Di Sanzo [2007] suggests bootstrapping with resampling to approximate the distribution of the LR-test. Monte Carlo studies show that in small samples this approach outperforms similar tests in many cases.

The estimated log-likelihood under the null of linearity (no switching in regimes) is equal to 18.9. Log-likelihood for the Markov-switching model with two regimes was found to be 112.48, thereby giving LR test statistic of 187.16. Following the procedure described by Di Sanzo [2007], the model (2.5) was estimated under the null hypothesis of linearity using maximum likelihood estimation. The obtained residuals were saved for

19 Di Sanzo (2007), Carrasco (2004) provide a brief summary of works in the given area.

the later bootstrap procedure. Then the model was estimated under the alternative hypothesis of two distinct regimes, and the corresponding LR statistic was computed. Then saved residuals were used to generate the bootstrap samples and to obtain the values of the LR test. This procedure was performed 1,000 times, and the distribution of the test under the null is shown in Figure (2.5). Clearly, the null hypothesis should be rejected at all the conventional significance levels.

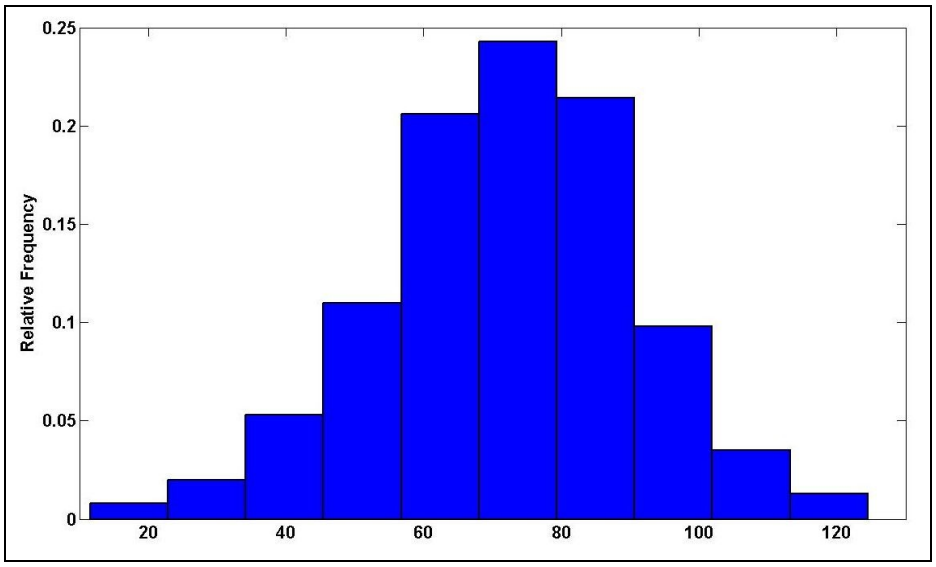


Figure 2.5: Bootstrap distribution of the LR test under the Null of no regime switching.

To assess if there are more than just two regimes, the model was estimated for three possible regimes. The value of the log-likelihood for the latter case is 113.62, which gives the LR statistics (with respect to the Null of two regimes) of 2.28 (p-value=0.89). Therefore, the dynamics of lambda was constructed allowing for only two different regimes.

There exists a variety of different approaches to carry out estimation in Markov-type models using either Classical or Bayesian econometric tools[20]. In this paper estimation

20 See Kim and Nelson [1999], Hamilton [2005] for the discussion of different approaches.

was performed by using a Hamilton filter (Hamilton [1994]). A Hamilton filter is a forward-backward non-linear filtering-smoothing algorithm, which allows us to compute filtered estimates (based on the information set up to time $1,..,t$) and smoothed estimates (based on all data $1,..,n$) probabilities of observing a specific state at each time step t. It also provides the likelihood of the observed variables, which we can maximize to get the estimates of unobserved parameters of the model. Table (2.5) summarizes results[21].

Table 2.5: Estimated parameters for the MS-model with two regimes

Parameter	**Regime 1**	**Regime 2**
μ	0.983 (0.121)	1.951 (0.192)
β	-0.019 (0.003)	
σ	0.059 (0.009)	
Estimation results for the model (2.5) with $s_t=1,2$. Estimated log-likelihood=112.48. Standard errors of the estimates are given in parentheses. Standard errors of the parameters were computed using White's heteroscedasticity robust variance-covariance matrix.		

The mean value of lambda in the first regime is 0.983, for the second regime it is 1.951 – therefore two regimes represent elongating and compacting behavior of the unemployment cycles. Unlike the OLS estimation (Table (2.3), which considered six significant breaks in the value of the model's intercept, the Markov-switching model distinguishes only two breaks in the parameter values which characterize different cycle speeds, and follow each other.

The estimated transition probabilities matrix is (standard errors are given in brackets):

$$P=\begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix}=\begin{bmatrix} 0.93\,(0.14) & 0.34\,(0.07) \\ 0.07\,(0.02) & 0.66\,(0.08) \end{bmatrix} \tag{2.6}$$

21 Estimation is carried out using Matlab toolbox developed by Marcelo Perlin, [2009]

Estimated transition probabilities indicate very high persistency of the first regime – probability of s_t to change from regime 1 to regime 2 is very low (p_{12}=0.07). At the same time, the opposite switching of regimes has much higher chances: p_{21}=0.34.

Knowing the estimated off-diagonal probabilities in (2.6), we may also compute the expected duration time of each regime[22]. Expected duration of Regime 1 is 1/(1-0.93) = 12.5 months, while for Regime 2 it is 1/(1-0.66)=2.94 periods. These findings support previous empirical results (Neftci [1984], Stock [1987], Diebold and Rudebusch [1996]) that unemployment rate has more prolonged periods of decline (cycle expansion phase when $\lambda < 1$) than periods when unemployment rates grow (cycle contraction, $\lambda > 1$) .

Figure (2.6) plots the smoothed estimated probabilities for each regime along with the original series for lambda.

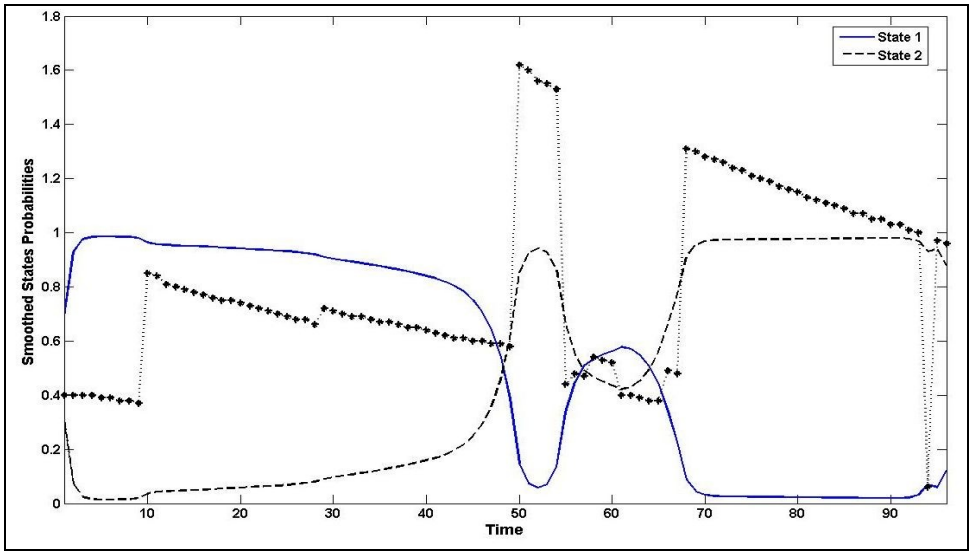


Figure 2.6: Smoothed probabilities for each regime

Obtained results allow us to predict the developing path of lambda. Figure (2.7) shows the out-of-sample forecast for the last 12 months of data along with the original series. The value of λ was computed conditionally on smoothed probabilities for each

22 See Kim and Nelson (1999) for the detailed derivations.

state at the previous step weighted by the probability of remaining in a particular regime or switching to another one:

$$\hat{\lambda}_{t+1} = \hat{\mu}_{S_t} * P * \hat{p}_t^{sm} + \hat{\beta} * time \tag{2.7}$$

where P is the estimated transition probabilities matrix (2.6), and $\hat{p}_t^{sm}$ is the vector of the estimated smoothed probabilities corresponding to each regime at time t obtained from Hamilton smoother recursion.

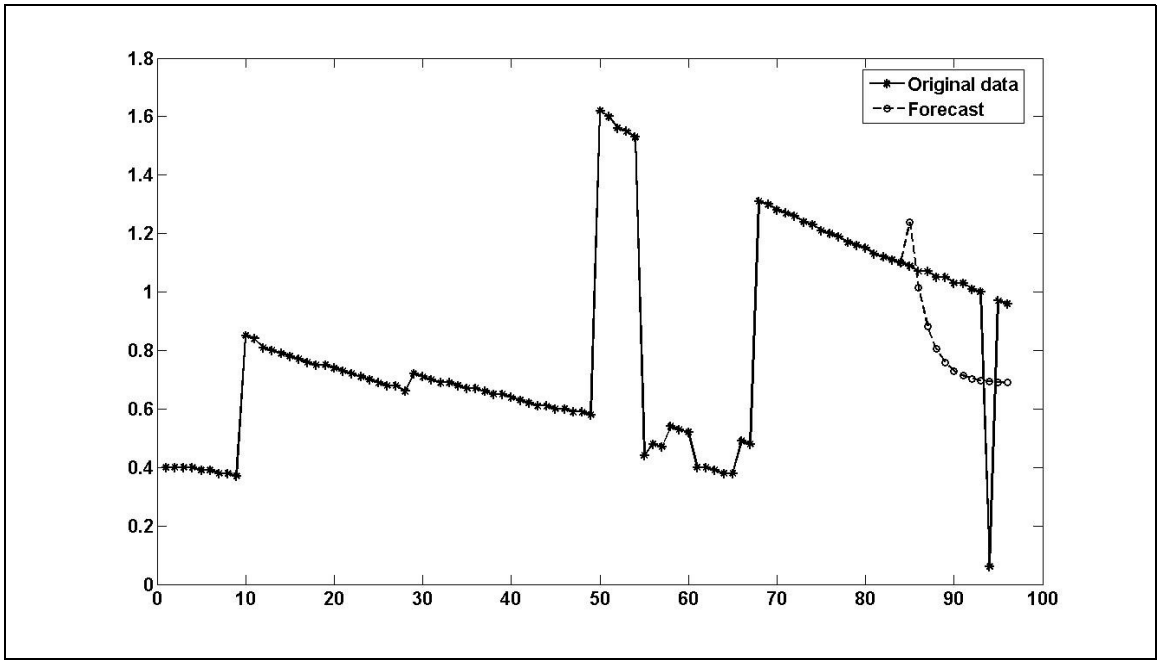


Figure 2.7: Twelve-month forecast for λ

Based on the predicted values of the time-deformation parameter, we also constructed the forecast for corresponding unemployment series. Twelve one-month unemployment forecasts were created for the original time scale and also for the data transformed based on the predicted value of λ. Forecast was computed based on the parameters obtained from the unobserved component model fitted to the data, and the model was re-estimated each time after the new forecast value was added to the existing

sample. Figure (2.8) presents 12-months forecast for the unemployment rates based on both time scales.

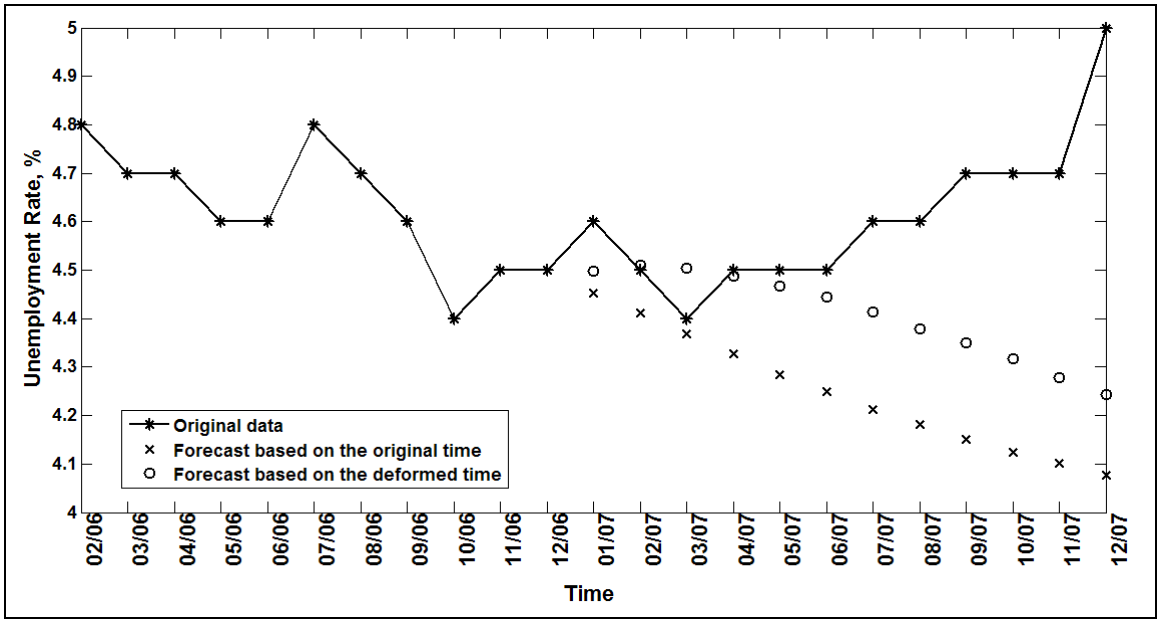


Figure 2.8: Unemployment rate forecast based on the predicted λ's

The first thing to note is that the forecast performance in the deformed time was found to be worse when it was based on the predicted values of λ rather than on the estimated values[23]. Nevertheless, the forecast performed in the deformed time is still more accurate than in the original time. Prediction MSE's for the original and deformed series are 0.195 and 0.094, respectively. Moreover, Diebold-Mariano test value is 2.56 with p-value of 0.011, thus the Null of equal predictive accuracy can be rejected even at 1% level of significance. These results can serve as an evidence that despite the fact, that forecasting in the deformed time scale is subject to additional λ -forecast errors and to inevitable interpolation errors, it still provides superior results as compared to the traditional techniques.

23 Shown in Figure (2.2)

2.4 Conclusions

This paper considers modeling and parameter estimation for covariance non-stationary processes with time-varying frequencies. We used the $G(\lambda)$ -transformation considered in the previous chapter to deform the time scale of the underlying process and obtain a process which would be stationary in the deformed time. However, in this work we assume that frequency of cyclical data, and, correspondingly, the parameter of time deformation may experience abrupt changes.

We assess the validity of our proposed model by comparing forecast performance for the structural time series and autoregressive model fitted to the unemployment series. Obtained results indicate that our model provides a better fit to the data over the long-run horizon than regular models, and allows us to capture potentially changing behavior of the cyclical processes.

The estimated parameter λ , which governs the dynamics of cyclical data for $G(\lambda)$ -stationary processes, was found to follow a first-order Markov switching process with two distinct regimes – one corresponding to the elongating phase of the cycle, and another responsible for the period of shortening cycle. These findings support the evidence that macroeconomic cycles are characterized by the different length of upturns and downturns. However, unlike previous research, we show that varying cyclical behavior may result from the changes in the process governing the timing of realized observations.

Furthermore, we show that process of time deformation can be predicted based on its Markov-switching behavior and be used to forecast the underlying macroeconomic data. Despite the fact that such forecast will be affected by more sources of errors due to the double forecasting process and due to the interpolation methods, it provides superior results. Our finding implies that failure to account for the changing cyclical behavior can significantly deteriorate estimation and inference in the empirical research.

3 Method of Moments Estimation for Non-Gaussian and Non-Linear State Space Models

3.1 Introduction

Unobservable component (UC) models provide an attractive framework for modeling economic series which may be driven by some unobserved process. Many of these models can be easily cast into the State Space form for the estimation and forecasting purposes using techniques well developed in the literature[24]. Kalman filtering and smoothing (Kalman [1960], Harvey [1989], Harvey and Shephard [1993]) iterative procedure can be applied to the linear State Space models (SSM) with Gaussian disturbances to compute filtered and smoothed state estimates and their disturbances, and also provide likelihood function for the observed data. For the case where the theoretical model suggests non-linear dependence between the observed process and unobserved state variables, the Extended Kalman Filter (EKF) can be applied. The linearization of the system dynamic is implemented by approximating state matrices with their partial derivatives, and then applying filtering and smoothing algorithms in a similar fashion to the regular Kalman filter. Being widely used in practice, this method is known to have many drawbacks – it appears to be not the optimal filter, since it is based on an approximation. In addition, it is difficult to tune and the filter may diverge when the system is highly non-linear. Recently, Julier and Uhlmann [1997, 2004] developed Unscented Kalman filter (UKF) which instead of linearizing the system, uses weighted points to parametrize the means and covariances of the data distribution.

While classical econometric tools are the most common method for dealing with SSM, Bayesian methods can be also used for the problems involving either linear or non-linear SSM. The model parameters, as well as unobserved states are treated as unknown random variables and their posterior means and variances can be estimated via Gibbs-

24 See Kim and Nelson [1999] for the overview of applied research involving UC models.

sampling techniques (Carter and Kohn [1994], Carlin, Polson, and Stoffer [1992]). Bayesian inference for non-Gaussian SSM[25] turns out to be more complicated task and requires using Monte Carlo Markov chain analysis (Shephard and Pitt [1997]). The main disadvantage of Bayesian methods is that they are rather computationally involved. Also, as noted by Durbin and Koopman [1997, 2000], they involve approximation errors of unknown size which may undermine estimation results.

Durbin and Koopman [1997, 2000, 2001] developed another approach when non-linear or non-Gaussian models are approximated by their linear Gaussian counterparts. They developed simulation methods to approximate the log-likelihood in Monte Carlo simulations using importance sampling techniques and antithetic variables. Drawing observations from the simulated approximating importance density can be applied for parameter estimation under either classical or Bayesian perspectives.

However, despite the fact that importance sampling method is more accurate and efficient than competing approaches, it still remains computationally burdensome, because of relying on simulations. In addition, model linearization requires the correct information about distributional properties of the data and about the correct underlying model, which often may be not available to the researcher. For example, Durbin and Koopman develop approximating model details for several specific cases covering Poisson, binary, binomial, and exponential family distributions. However, this approach becomes not applicable in case when distribution has unknown or analytically not tractable form.

In this paper we offer Method of Moments as an alternative approach for estimation in State Space models framework. While our intention is to introduce an easier way of dealing with non-linear or non-Gaussian cases, it can be equivalently used for the "classical" linear Gaussian models. Using the properties of prediction errors and axillary residuals obtained from Kalman filter and smoother, we construct several moment conditions for the Method of Moments or Generalized Method of Moments. Parameter

25 See Durbin and Koopman [1997, 2000] for the detailed overview of works in the given area.

estimation is then carried out by minimizing objective function. In case of non-linear models, Extended Kalman filter output can be used. Also, we additionally find that the issue of non-convergence for EKF is greatly reduced comparing to the traditional MLE tools.

The main advantage of our approach is that it does not require knowledge about the data distribution. This feature allows us to investigate more general model specifications than covered in the literature, and is of a special importance when the distributional properties of the data are not known. At the same time it may help to avoid possible mistakes when the incorrect model specification is being assumed by the researcher.

Another attractive feature of the given method is easiness of implementation since it does not involve analytical or computational derivations for linearizing each specific problem and can be applied as the general tool. This, in turn, again reduces the source of potential errors.

Finally, given that our method does not rely on sampling or simulation procedures, it provides significant gain in computational efficiency. In several comparative experiments described in the following sections, Method of Moments estimation required much less time comparing to the importance sampling techniques.

3.2 Overview of State Space modeling

3.2.1 Linear Gaussian SSM

The linear Gaussian SSM usually is represented in the following form (Harvey [1993], Durbin and Koopman [2001]):

$$y_t = Z_t \alpha_t + \epsilon_t, \qquad \epsilon_t \sim N(0, H_t) \tag{3.1}$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \quad \eta_t \sim N(0, Q_t) \tag{3.2}$$

Where (3.1) is the observation equation with $Z_t \alpha_t$ referred to as a signal, and (3.2) is the state equation for the unobserved state variable α_t, and $t = 1..N$. For convenience, the dimensions of each component are summarized in the following table:

Table 3.1: Dimension for the SSM vectors and matrices

Vector	Dimension	Matrix	Dimension
y_t	$p \times 1$	Z_t	$p \times m$
α_t	$m \times 1$	T_t	$m \times m$
ϵ_t	$p \times 1$	H_t	$p \times p$
η_t	$r \times 1$	R_t	$m \times r$
		Q_t	$r \times r$
		P_t	$m \times m$

Disturbances ϵ_t and η_t are mutually and serially uncorrelated Gaussian processes, however, this assumption may be easily relaxed. Observation and state matrices Z_t and T_t are known and may depend on y_t or α_t. However, for most of applications they are time invariant, in which case the time subscript may be omitted. R_t is commonly modeled as an identity matrix. In this case it is called a selection matrix, which

determines which state components will have non-zero disturbances.

The SSM framework provides a powerful tool for modeling economic data. For example, the $MA(1)$ model

$$y_t = \epsilon_t + \theta \epsilon_{t-1} \tag{3.3}$$

can be easily cast into the state space form:

$$y_t = [1 \quad 0]\alpha_t, \tag{3.4}$$

$$\alpha_t = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 \\ \theta \end{bmatrix} \epsilon_t \text{ , where } \alpha_t = \begin{bmatrix} y_t \\ \theta \epsilon_t \end{bmatrix} \tag{3.5}$$

An alternative set up is also possible as:

$$y_t = [1 \quad \theta]\alpha_t, \tag{3.6}$$

$$\alpha_t = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \epsilon_t \text{ , where } \alpha_t = \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix} \tag{3.7}$$

Estimating models that include unobservable state variables is usually carried out with a Kalman filtering and smoothing. A Kalman Filter is a recursive algorithm which allows us to estimate conditional mean and variance of the state of the process α_t based on the past observations $Y_{t-1} = \{y_{1,\cdots}, y_{t-1}\}$ by minimizing the mean squared error. When disturbances ϵ_t and η_t are Gaussian, the Kalman filter is the optimal estimator which minimizes the mean square error. When the model is not normal, this filter remains the optimal linear estimator[26]. In addition to the estimated states, filter output enables to construct a likelihood function used for the parameter estimation. There are several representations of the Kalman filter equations including fixed interval KF, contemporaneous KF, etc. In this paper we follow the representstion by DeJong [1993] and Koopman [1993, 1994].

26 See Harvey [1993].

Assuming that the initial state vector α_1 is $\sim N(a_{1,}P_1)$, the following set of equations summarizes the Kalman Filter iterative procedure:

$$v_t = y_t - Z_t a_t \tag{3.8}$$

$$F_t = Z_t P_t Z_t' + H_t \tag{3.9}$$

$$K_t = T_t P_t Z_t' F_t^{-1} \tag{3.10}$$

$$L_t = T_t - K_t Z_t \tag{3.11}$$

$$a_{t+1} = T_t a_t + K_t v_t \tag{3.12}$$

$$P_{t+1} = T_t P_t L_t' + R_t Q_t R_t' \tag{3.13}$$

where v_t is being referred to as a prediction error with variance F_t, K_t is a Kalman gain, a_t is the predicted value of the state vector at time t, given the information set up to the time $t-1$, that is: $a_t = E(\alpha_t | y_1 \ldots y_{t-1})$, with variance $P_t = Var(\alpha_t | y_1 \ldots y_{t-1})$.

In addition to estimating the state vector, one may be interested in estimating additional parameters, such as variances of disturbances or coefficients on additional explanatory variables, which may be embedded into the model (3.1-3.2). Let θ denote a stacked vector of the parameters of interest. Then, if ϵ_t *and* η_t are Gaussian, the log-likelihood is:

$$\log L(\theta | y) = \sum_{t=1}^{n} \log p(y_t | Y_{t-1}; \theta) \tag{3.14}$$

Since $E(y_t | Y_{t-1}) = Z_t a_t$, and $var(y_t | Y_{t-1}) = var(y_t - Z_t a_t) = var(v_t) = F_t$, the output of the Kalman filter may be used to form the likelihood function as

$$\log L(\theta | y) = \frac{-np}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{n} (\log |F_t| + v_t' F_t^{-1} v_t) \tag{3.15}$$

Computation of the parameter estimate $\hat{\theta}$ is then implemented via maximization of log-likelihood (3.15). Although MLE estimation is a rather popular approach to the parameter estimation in SSM, Bayesian analysis for those purposes is also well developed in the literature – see for example, DeJong and Shephard [1995], Koop [2003].

A Kalman smoother represents a backward recursion used to estimate smoothed state vector $\hat{\alpha}_t$ based on the whole set of observations $Y_n = \{y_{1,..}, y_n\}$ using the following set of equations. There exists several versions of the Kalman Smoother. In this paper we use so-called *fixed-interval* smoother (Anderson and Moore [1979]):

$$L_t = T_t - K_t Z_t \tag{3.16}$$

$$r_{t-1} = Z_t' F_t^{(-1)} v_t + L_t' r_t \tag{3.17}$$

$$N_{t-1} = Z_t' F_t^{(-1)} Z_t + L_t' N_t L_t \tag{3.18}$$

$$\hat{\alpha}_t = a_t + P_t r_{t-1} \tag{3.19}$$

$$V_t = P_t - P_t N_{t-1} P_t \tag{3.20}$$

with smoothed state vector $\hat{\alpha}_t = E(\alpha_t | Y_n)$ having the smallest variance $V_t = Var(\alpha_t | Y_n)$. Smoothing recursion starts from time $t = n$, and runs backward setting initial values for the dummy vector r_N to 0 and its variance to $N_N = 0$. Values of the prediction error v_t, its variance F_t and Kalman gain K_t are obtained from the Kalman filter stage, which is implemented before the smoothing stage.

In addition to estimating the smoothed state vector, a researcher may be interested in estimating smoothed disturbances $\hat{\epsilon}_t$ and $\hat{\mu}_t$. Koopman [1998] discusses how they may be used for the parameter estimation and diagnostic tests.

A disturbance smoother (de Jong [1988], Kohn and Ansley, [1989]) is implemented as follows:

$$\hat{\epsilon}_t = E(\epsilon_t | Y_n) = H_t(F_t^{-1} v_t - K_t' r_t) \tag{3.21}$$

$$\hat{v}_t = E(\eta_t | Y_n) = Q_t R_t' r_t \tag{3.22}$$

In addition to estimating only smoothed disturbances, disturbance smoothing can be used for the so-called fast smoother (Koopman [1998]) which computes the values of $\hat{\alpha}_t = E(\alpha_t | Y_n)$ with smaller computation burden than (3.16-3.20). In case we are not interested in quantities N_t and V_t, (3.19) can be replaced by the following equation:

$$\hat{\alpha}_t = T\hat{a}_t + R_t \eta_t \tag{3.23}$$

which is implemented using disturbance smoother results.

3.2.2 Non-Gaussian SSM

Non-Gaussian state space models have a form similar to (3.1-3.2). However, in this case either observation or state disturbances are not normally distributed, although they still are assumed to be mutually and serially uncorrelated. Following Durbin and Koopman [2004], we may consider the two following cases:

1. Observation disturbances in (3.1) come from a non-Gaussian distribution:

$$y_t = Z_t \alpha_t + \epsilon_t, \qquad \epsilon_t \sim p(\epsilon_t). \tag{3.24}$$

For example ϵ_t may come from fat-tailed distribution, such as Student's t distribution.

2. Data come from the exponential family of distributions and can be cast into the following form (Shephard [1997]):

$$p(y_t|Z_t\alpha_t)=\exp[y_t'Z_t\alpha_t-b_t(Z_t\alpha_t)+c_t(y_t)],\qquad -\infty<Z_t\alpha_t<\infty \tag{3.25}$$

where $b_t(.)$ and $c_t(.)$ are known functions.

For example, for the Poisson distribution we have:

$$\log p(y_t|\mu_t)=y_t\log\mu_t-\mu_t-\log(y_t!), \tag{3.26}$$

where $\mu=\exp(Z_t\alpha_t)$ is the parameter of the Poisson distribution. Similar decomposition may be obtained for the Binary or Binomial densities[27].

Estimating the model parameters θ as well as unobserved state series for non-Gaussian SSM in either Classical or Bayesian econometrics is usually carried out by applying importance sampling from an approximating Gaussian importance density $g(\alpha_t|y_t;\theta_t)$ which should resemble a non-Gaussian density $p(\alpha_t|y_t;\theta_t)$. To implement this, we need first to find an approximating Gaussian model for each specific case in (3.24-3.25). Second, the obtained density has to be linearized to be used in the Kalman filter equations. Sampling from the importance density (possibly using antithetic variables) then allows construction of a likelihood function and application of traditional maximization techniques for the parameter estimation, or the obtaining posterior parameters in a Bayesian framework. Durbin and Koopman [2004] provide a detailed description of linearizing and sampling for several specific models.

27 see Durbin and Koopman (2004) for more details.

3.2.3 Nonlinear SSM

In nonlinear SSM either observations y_t or states α_t depend nonlinearly on the state vector:

$$y_t = Z_t(\alpha_t) + \epsilon_t, \qquad \epsilon_t \sim N(0, H_t) \tag{3.27}$$

$$\alpha_{t+1} = T_t(\alpha_t) + R_t \eta_t, \quad \eta_t \sim N(0, Q_t) \tag{3.28}$$

with $Z_t(.)$ and $T_t(.)$ being some differentiable functions. There exist several approaches for parameter estimation in nonlinear SSM. Commonly used in empirical research, the Extended Kalman Filter (EKF) is based on linearization of system dynamics:

$$y_t = \dot{Z}_t(\alpha_t)\alpha_t + \epsilon_t, \tag{3.29}$$

$$\alpha_{t+1} = \dot{T}_t(\alpha_t)\alpha_t + R_t \eta_t, \tag{3.30}$$

where $\dot{Z}_t$ and $\dot{T}_t$ are Jacobian matrices of partial derivatives of $Z_t(.)$ and $T_t(.)$ with respect to α_t.

Filtering and smoothing steps are then implemented similar to the regular Kalman filter (3.8-3.13):

$$v_t = y_t - Z_t(a_t), \quad F_t = \dot{Z}_t P_t \dot{Z}_t' + H_t \tag{3.31}$$

$$K_t = T_t P_t \dot{Z}_t' F_t^{-1}, \quad L_t = \dot{T}_t - K_t \dot{Z}_t \tag{3.32}$$

$$a_{t+1} = T_t(a_t) + K_t v_t, \quad P_{t+1} = \dot{T}_t P_t L_t' + R_t Q_t R_t' \tag{3.33}$$

$$r_{t-1} = \dot{Z}_t' F_t^{(-1)} v_t + L_t' r_t, \quad N_{t-1} = \dot{Z}_t' F_t^{(-1)} \dot{Z}_t + L_t' N_t L_t \tag{3.34}$$

$$\hat{\alpha}_t = a_t + P_t r_{t-1}, \quad V_t = P_t - P_t N_{t-1} P_t \tag{3.35}$$

While being computationally easy to implement, the EKF has well-known drawbacks. It is not an optimal filter since it is based on a set of approximations. It may also diverge if the linearization does not provide a good approximation for the underlying nonlinear dynamics. However, the most important problem arises from the fact that linearization may lead to non-Gaussianity of the model. Since the Method of Moments does not require normality of the data, moment conditions derived in the next section can be easily applied to the EKF filter in case other model linearization methods are not applicable.

3.3 Method of Moments estimation details

In this section we consider Method of Moments estimation as an alternative to MLE and Bayesian techniques for estimating SSM parameters in a situation when we may have "classical" linear, or either non-linear or non-Gaussian SSM.

Generalized Method of Moments estimation (Hansen [1982]) is carried out in the traditional way by finding a minimand for the criterion function

$$Q_T(\theta) = f_n(\theta)' W_n f_n(\theta) \ , \tag{3.36}$$

where $f_N(\theta) = n^{-1} \sum_{t=1}^{n} f(x_t, \theta)$ (3.37)

is used to approximate sample moment conditions $E[f(.,.)]$ satisfying the condition $E[f(.,.)] = 0$, and W_n is a weighting matrix. GMM literature discusses different choices for W_n, as well as minimization procedures for obtaining the GMM minimand[28].

In our work we use a two-step estimation procedure with moment conditions constructed as the difference between sample moments and their population counterparts

In a case of exact identification, when the number of moment conditions is equal to the number of estimated parameters, the Method of Moments can be used instead of GMM. It minimizes the the following objective function:

$$Q_n(\theta) = f_n(\theta)' f_n(\theta) \tag{3.38}$$

28 See Hall [2005] for the detailed exposition.

3.3.1 Basic moment conditions

Kalman filter equations (3.8-3.13) allow us to establish the following properties of the prediction error $v_t = y_t - E(y_t|Y_{t-1})$, namely:

$$E[v_t|Y_{t-1}] = E[y_t - E(y_t|Y_{t-1})] = E[Z\alpha_t + \epsilon - Za_t|Y_{t-1}] = \\ = Z\,E[\alpha_t - a_t|Y_{t-1}] = 0 \tag{3.39}$$

The conditional variance of v_t is given by:

$$E[v_t v_t'|Y_{t-1}] = var[Z_t\alpha_t + \epsilon_t - Z_t a_t|Y_{t-1}] = Z_t\,var(\alpha_t|Y_{t-1})\,Z_t' + var(\epsilon_t) = \\ = Z_t P_t Z_t' + H_t = F_t \tag{3.40}$$

where conditional variance-covariance matrix of the state vector is defined as:

$$P_t = var(\alpha_t|Y_{t-1}) = E[(\alpha_t - E(\alpha_t|Y_{t-1}))(\alpha_t - E(\alpha_t|Y_{t-1}))'] \tag{3.41}$$

It is also known that prediction errors are uncorrelated:

$$E[v_t, v_s'|Y_{t-1}] = 0 \text{ for } \forall t \neq s, \quad t, s = 1,..n \tag{3.42}$$

Given these properties, we may establish the following moment conditions for the MM estimation:

M1. $E[v_t v_t' - F_t] = E[E(v_t v_t' - F_t|Y_{t-1})] = 0$ (3.43)

M2. $E[v_t v_{t+l}] = E[E(v_t v_{t+l}|Y_{t-1})] = 0$ for $\forall l \geq 1$ (3.44)

One issue in the empirical estimation refers to the diffuse initialization of the Kalman filter. Since the filtering recursion starts with $P_1 = c * I_k$ where c is taken

arbitrary large number[29], the first few filtered variances of the prediction error may be inappropriately large, and potentially deteriorate moment condition M1. To resolve this issue, it is plausible to drop the first few observations when computing sample moment function $f_N(\theta)=N^{-1}\sum_{t=1}^{N} f(x_t,\theta)$. In case of a large data set this should not pose any problems. Koopman [2000] showed that usually it takes 3-4 first iterations to get filter initialized.

The two moment conditions laid out previously can be rewritten in an alternative way using the properties of the prediction error and Kalman Filter relations. These additional relations may be used instead of M1-M2 for the parameter estimation.

First, consider the cross-covariance between the prediction error and the filtered state estimate which is formulated as:

$$E[v_t a_{t+1}'|Y_{t-1}]=E[v_t(T_t a_t+K_t v_t)'|Y_{t-1}]=F_t K_t'$$

$$E[v_t a_{t+2}'|Y_{t-1}]=E[v_t(T_t a_{t+1}+K_{t+1} v_{t+1})'|Y_{t-1}]=F_t K_t' T_t'$$

and therefore can be written as the following moment conditions:

$$\begin{aligned} &E[v_t a_{t+j}'-F_t K_t'(T_t')^j]=E[E(v_t a_{t+j}'-F_t K_t'(T_t')^j|Y_{t-1})]=0 \\ &\text{for } \forall\, j>1 \text{, and} \\ &E[v_t a_{t-j}']=E[E(v_t(T_t a_{t-j-1}+K_{t-j-1} v_{t-j-1})|Y_{t-1})]=0 \\ &\text{for } \forall\, j=0..t \end{aligned} \tag{3.45}$$

29 For example, we used $c=10^7$ as the starting value for filter initialization in experiments discussed further.

Similarly, the cross-covariance between the prediction error and the observation vector:

$$E[v_t y_{t-j}'] = E[E(v_t(Z a_{t-j} + v_{t-j})'|Y_{t-1})] = 0 \text{ for } \forall\, j \geq 1$$

$$E[v_t y_t' - F_t] = E[E(v_t(Z a_t + v_t)' - F_t|Y_{t-1})] = 0$$

and (3.46)

$$E[v_t y_{t+j}' - F_t K_t'(T_t')^j Z'] = E[E(v_t y_{t+j}' - F_t K_t'(T_t')^j Z'|Y_{t-1})] = 0$$

for $\forall\, j \geq 1$

These relations provide the same information for the parameter estimation as conditions M1-M2, and can be used in a similar fashion.

3.3.2 Additional moment conditions.

When we need to estimate more than two parameters, we may employ the output of Kalman smoother (3.16-3.20). As shown in Koopman [1993], unlike the true disturbances ϵ_t and ν_t, their smoothed estimates (so-called auxiliary residuals) $\hat{\epsilon}_t$ and $\hat{\nu}_t$ are correlated between different lags and between themselves. We will use these relations to construct the desired moment conditions.

The covariances matrices between smoothed state estimators were first derived by de Jong and MacKinnon [1988], who used geometric proof in their computations. These results then were re-derived in an easier way by Harvey and Proietti [2005]. Koopman [1993] showed how to construct contemporaneous and lagged mean square error matrices for the smoothed disturbances and used them for the diagnostic testing. Later,

Durbin and Koopman [2001] developed formulas for the unconditional covariances of the auxiliary residuals in a less sophisticated manner. Similar results were also stated by Harvey and Proietti [2005]. In this section we re-derive the main relations outlined in these works in order for them to be used as unconditional moment conditions for the MM estimation.

First, given that:

$$\hat{\epsilon}_t = E(\epsilon_t | Y_n) = H_t(F_t^{-1} v_t - K_t' r_t) \tag{3.47}$$

$$\hat{v}_t = E(\eta_t | Y_n) = Q_t R_t' r_t \tag{3.48}$$

We may establish the following unconditional moment conditions:

$$\textbf{M3.} \quad E[\hat{\epsilon}\hat{\epsilon}' - H_t(F_t^{-1} + K_t' N_t K_t) H_t'] = \\ = E[E(\hat{\epsilon}\hat{\epsilon}' - H_t(F_t^{-1} + K_t' N_t K_t) H_t' | Y_n)] = 0 \tag{3.49}$$

$$\textbf{M4.} \quad E[\hat{\eta}_t \hat{\eta}_t' - Q_t R_t' N_t R_t Q_t'] = E[E(\hat{\eta}_t \hat{\eta}_t' - Q_t R_t' N_t R_t Q_t' | Y_n)] = 0 \tag{3.50}$$

When the disturbance vectors consist of more than one observation or state process, then the off-diagonal elements of (3.49-3.50) can be used for setting up moments as well.

Next, we compute the covariances between smoothed disturbances for different lags.

The first-order autocovariance for $\hat{\epsilon}$ is:

$$E[\hat{\epsilon}_t, \hat{\epsilon}'_{t+1} | Y_n] = E[H_t(F_t^{-1} v_t - K'_t r_t)(F_{t+1}^{-1} v_{t+1} - K'_{t+1} r'_{t+1}) H'_{t+1} | Y_n]$$

Since $E(v_t v_t' | Y_{t-1}) = F_t$, $E[v_t v'_{t+l} | Y_{t-1}] = 0$ for $l \geq 1$, $E[v_t r_t' | Y_{t-1}] = 0$,

and $r_t = Z'_{t+1} F_{t+1}^{(-1)} v_{t+1} + L'_{t+1} r_{t+1}$, we have that

$$E(\hat{\epsilon}_t, \hat{\epsilon}_{t+1} | Y_n) = E[H_t(-K_t' Z_t' F_{t+1}^{-1} v_{t+1} v'_{t+1} (F_{t+1}^{-1})' + K_t' L_{t+1}' r_{t+1} r'_{t+1} K_{t+1}) H'_{t+1} | Y_n] =$$

$$= H_t K_t' (-Z'_{t+1} F_{t+1}^{-1} + L'_{t+1} N_{t+1} K_{t+1}) H'_{t+1}$$

Similar, the second order autocovariance is:

$$E[\hat{\epsilon}_t, \hat{\epsilon}_{t+2}|Y_n] = E[H_t(F_t^{-1}v_t - K_t'r_t)(F_{t+2}^{-1}v_{t+2} - K_{t+2}'r_{t+2})'H_{t+2}'|Y_n]$$

Making the appropriate substitutions for r_t and r_{t+1}, it follows that

$$E[\hat{\epsilon}_t, \hat{\epsilon}_{t+2}|Y_n] =$$

$$= E[H_t(-K_t'L'_{t+1}Z'_{t+2}F_{t+2}^{-1}v_{t+2}v'_{t+2}(F_{t+2}^{-1})' + K_t'L'_{t+1}L'_{t+2}r_{t+2}r'_{t+2}K_{t+2})H'_{t+2}|Y_n] =$$

$$= H_tK_t'L_{t+1}'(-Z_{t+2}'F_{t+2}^{-1} + L_{t+2}'N_{t+2}K_{t+2})H_{t+1}'$$

Having these results, we may establish the following relation to be used as the moment condition based on the autocovariance function for smoothed observation residuals:

M5. $$E[\hat{\epsilon}_t\hat{\epsilon}_j' - H_tK_t'G_{t+1,j-1}(-Z_j'F_j^{-1} + L_j'N_jK_j)H_j'] =$$
$$= E[E(\hat{\epsilon}_t\hat{\epsilon}_j' - H_tK_t'G_{t+1,j-1}(-Z_j'F_j^{-1} + L_j'N_jK_j)H_j'|Y_n)] = 0 \quad (3.51)$$

where

$$G_{t+1,j-1} = I_m \text{ for } j = t+1,$$
$$G_{t+1,j-1} = L'_{t+1} \text{ for } j = t+2, \quad (3.52)$$
$$G_{t+1,j-1} = L'_{t+1}L'_{t+2}\ldots L'_{j-2}L'_{j-1} \text{ for } j > t+2.$$

Analogously, using the properties of v_t and r_t, autocovariances for $\hat{\eta}_t$ may be derived in the following manner:

$$E[\hat{\eta}_t, \hat{\eta}_{t+1}|Y_n] = E[Q_tR_t'r_tr_{t+1}'R_{t+1}Q_{t+1}'|Y_n] =$$

$$= E[Q_tR_t'(Z'_{t+1}F_{t+1}^{-1}v_{t+1} + L'_{t+1}r_{t+1})r'_{t+1}R_{t+1}Q'_{t+1}|Y_n] = Q_tR_t'L'_{t+1}N_{t+1}R_{t+1}Q'_{t+1}$$

$$E[\hat{\eta}_t, \hat{\eta}_{t+2}|Y_n] = E[Q_t R_t' r_t r'_{t+2} R_{t+2} Q'_{t+2}|Y_n] =$$

$$= E[Q_t R_t' L'_{t+1}(Z'_{t+2} F_{t+2}^{-1} v_{t+2} + L'_{t+2} r_{t+2}) r'_{t+2} R_{t+2} Q'_{t+2}|Y_n] =$$

$$= Q_t R_t' L_{t+1}' L'_{t+2} N_{t+2} R_{t+2} Q'_{t+2}$$

which leads to the following moment condition:

$$\textbf{M6.}\quad E[\hat{\eta}_t \hat{\eta}_j' - Q_t R_t' G_{t+1,j} N_j R_j Q_j'] = \\ = E[E(\hat{\eta}_t \hat{\eta}_j' - Q_t R_t' G_{t+1,j} N_j R_j Q_j'|Y_n)] = 0 \tag{3.53}$$

where

$$G_{t+1,j} = L_j' \text{ for } j = t+1, \\ G_{t+1,j} = L'_{t+1} L'_{t+2} \ldots L'_{j-2} L_j' \text{ for } j > t+1 \tag{3.54}$$

Next, we derive cross-covariances of the smoothed disturbances:

$$E[\hat{\eta}_t, \hat{\epsilon}_t|Y_n] = E[Q_t R_t' r_t (F_t^{-1} v_t - K_t' r_t)' H_t'|Y_n] = Q_t R_t' N_t K_t H_t'$$

which gives us the following moment condition:

$$\textbf{M7.}\quad E[\hat{\eta}_t \hat{\epsilon}_t' - Q_t R_t' N_t K_t H_t'] = E[E(\hat{\eta}_t \hat{\epsilon}_t' - Q_t R_t' N_t K_t H_t'|Y_n)] = 0 \tag{3.55}$$

Cross-covariances of smoothed state disturbances with future smoothed observation disturbances are:

$$E[\hat{\eta}_t, \hat{\epsilon}_{t+1}|Y_n] = E[Q_t R_t'(Z'_{t+1} F_{t+1}^{-1} v_{t+1} + L'_{t+1} r_{t+1})(F_{t+1}^{-1} v_{t+1} - K'_{t+1} r_{t+1})' H'_{t+1}|Y_n] =$$

$$= Q_t R_t'(Z'_{t+1} F_{t+1}^{-1} + L'_{t+1} N_{t+1} K_{t+1}) H'_{t+1}$$

$$E[\hat{\eta}_t, \hat{\epsilon}_{t+2}|Y_n] = E[Q_t R_t' L'_{t+1}(Z'_{t+2} F_{t+2}^{-1} v_{t+2} + L'_{t+2} r_{t+2})(F_{t+2}^{-1} v_{t+2} - K'_{t+2} r_{t+2})' H'_{t+2}|Y_n] =$$

$$= Q_t R_t' L_{t+1}'(Z'_{t+2} F_{t+2}^{-1} + L'_{t+2} N_{t+2} K_{t+2}) H'_{t+2}$$

With the following general pattern used to state the following condition:

$$\textbf{M8.}\quad E[\hat{\eta}_t\hat{\epsilon}_j' - Q_t R_t' G_{t+1,j-1}(Z_j' F_j^{-1} + L_j' N_j K_j) H_j'] =$$

$$= E[E(\hat{\eta}_t\hat{\epsilon}_j' - Q_t R_t' G_{t+1,j-1}(Z_j' F_j^{-1} + L_j' N_j K_j) H_j' | Y_n)] = 0 \qquad (3.56)$$

for $j > t$

where $G_{t+1,j-1}$ is defined as in (3.20).

Finally, covariances of observation disturbances with future state disturbances are:

$$E(\hat{\epsilon}_t, \hat{\eta}_{t+1} | Y_n) = E[H_t(F_t^{-1} v_t - K_t' r_t) r'_{t+1} R_{t+1} Q'_{t+1} | Y_n] =$$

$$= E[-H_t K_t (Z'_{t+1} F_{t+1}^{-1} v_{t+1} + L'_{t+1} r_{t+1}) r'_{t+1} R_{t+1} Q'_{t+1} | Y_n] =$$

$$= -H_t K_t' L'_{t+1} N_{t+1} R_{t+1} Q'_{t+1}$$

$$E[\hat{\epsilon}_t \hat{\eta}_{t+2} | Y_n] = E[-H_t K_t' L'_{t+1} (Z'_{t+2} F_{t+2}^{-1} v_{t+2} + L'_{t+2} r_{t+2}) r'_{t+2} R_{t+2} Q'_{t+2} | Y_n] =$$

$$= -H_t K_t' L'_{t+1} L'_{t+2} N_{t+2} R_{t+2} Q'_{t+2}$$

With the following general form, used as the moment condition:

$$\textbf{M9.}\quad E[\hat{\epsilon}_t \hat{\eta}_j' + H_t K_t' G_{t+1,j} N_j R_j Q'_j] =$$

$$E[E(\hat{\epsilon}_t \hat{\eta}_j' + H_t K_t' G_{t+1,j} N_j R_j Q'_j | Y_n)] = 0 \qquad (3.57)$$

where $G_{t+1,j}$ was defined in (3.54).

Conditions M3-M9 are based on the covariances between axillary residuals for different leads and lags, and therefore provide a plenty of moment conditions for parameters estimation. When using exact identification and method of moments, the number of conditions should be chosen equal to the number of estimated parameters. Also it is possible to use Generalized Method of Moments constructing extra moment conditions by selecting different combinations of M3-M10 with different lags.

3.4 Monte-Carlo simulation.

3.4.1 Linear SSM.

First, we compare MM and MLE performance for the "classical" local level with trend model with Gaussian disturbances:

$$y_t = \mu_t + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{3.58}$$

where the trend component is built as the local level model with slope:

$$\mu_{t+1} = \mu_t + \beta_t + \xi_t, \qquad \xi_t \sim NID(0, \sigma_\xi^2), \tag{3.59}$$

$$\beta_{t+1} = \beta_t + \zeta_t, \qquad \zeta_t \sim NID(0, \sigma_\zeta^2), \tag{3.60}$$

Level and slope innovations, ξ_t and ζ_t respectively, are normally and independently distributed. Such specification allows us to account for different behavior of the trend component μ_t. If $\sigma_\zeta^2 = 0$, then $\beta_{t+1} = \beta_t = \beta$, trend follows random walk with drift. If $\sigma_\xi^2 = 0$, then trend resembles an integrated random walk model. When both innovations, ξ_t and ζ_t, have zero variance, a deterministic trend specification is obtained.

All formulas considered above for structural time series can be conveniently put into the linear state-space form with observation and transition equations defined as:

$$y_t = Z_t \alpha_t + \epsilon_t \quad \epsilon_t \sim NID(0, H_t) \tag{3.61}$$

$$\alpha_{t+1} = T_t \alpha_t + \eta_t \quad \eta_t \sim NID(0, Q_t) \tag{3.62}$$

with initial state vector assumed to be $\alpha_1 \sim NID(a_1, P_1)$, and disturbances ϵ_t and η_t being mutually and serially uncorrelated Gaussian processes.

The matrices of state-space form for the local level with trend model are:

$$\alpha_t=(\mu_t \quad \beta_t)', \quad \eta_t=(\xi_t \quad \zeta_t)', \quad Z_t=[1 \quad 0]$$

$$T_t=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}, \quad Q_t=\begin{bmatrix}\sigma_\xi^2 & 0\\ 0 & \sigma_\zeta^2\end{bmatrix}$$

Parameters to be estimated are $\theta=(\sigma_\epsilon^2, \sigma_\xi^2, \sigma_\zeta^2)'$. The Kalman filtering process is initialized with $\alpha_1=(y_1 \quad 0)'$, and $P_1=I_m*1000000$. It is also possible to use exact filter initialization, however, as shown by empirical studies the gain in the number of iterations between exact and diffuse initialization is negligible.

In table 3.2 we compare Monte Carlo simulation results for MLE and MM with moment conditions based on the Kalman filter output for a random walk with drift (Model 1) and integrated random walk (Model 2) specifications of SSM.

Table 3.2: Monte Carlo results for linear Gaussian SSM with basic moments

Model	**Parameter**	**True value**	**MLE**				**Method of Moments** **moments used: M1, M2 (lag=1)**			
			Mean estimate	**s.e.**	**Median estimate**	**RMSE**	**Mean estimate**	**s.e.**	**Median estimate**	**RMSE**
1	σ_ϵ^2	5e-1	**5.03e-1**	1.23e-3	5.02e-1	5.6e-2	4.97e-1	1.57e-3	4.94e-1	7.0e-2
	σ_ξ^2	5e-2	**4.82e-2**	4.25e-4	4.60e-2	1.9e-2	5.55e-2	6.48e-4	4.64e-2	3.0e-2
2	σ_ϵ^2	5e-1	**5.01e-1**	1.19e-3	5.01e-1	5.4e-2	4.99e-1	1.39e-3	4.94e-1	6.2e-2
	σ_ζ^2	5e-3	**5.00e-3**	4.02e-5	4.92e-3	1.8e-3	5.80e-3	8.05e-5	4.89e-3	3.7e-3

Reported are mean values across 2,000 Monte Carlo simulations for data generated according to (3.58-3.60) with sample size *N=200.* In addition, for each estimation method we report median values for the MC estimates (third column). Starting values for MLE and GMM optimization are $\theta_{SV}=(0.2; 0.2; 0.2)'$. To preserve non-negativity of the variance estimates, all the parameters were log-transformed. Estimates with the smallest Root Mean Square Error are shown in bold.

In general, MM provides results similar to those of MLE in terms of absolute bias. However, MM estimates appear to be less efficient then MLE results, although, the difference is minor. The possible explanation for this fact is that MLE utilizes all the possible information about distributional properties of the parameters, while the Method of Moments does not.

To evaluate performance of moment condition derived from the Kalman smoother, we consider Monte Carlo estimation for 3 parameters with three random moment conditions selected. Results are summarized in Table 3.3.

Table 3.3: MC results for linear Gaussian SSM with additional moments

Para-meter	**True value**	**MLE**	**MM**		
			M1, M2, M4	**M1, M4**	**M7, M8**
σ_ϵ^2	5e-1	**5.003e-1** (3.169e-5)	4.977e-1 (3.721e-5)	5.013e-1 (3.191e-5)	5.024e-1 (5.230e-5)
σ_ξ^2	5e-2	4.926e-2 (1.545e-5)	4.733e-2 (1.900e-5)	**4.898e-2** (1.397e-5)	4.168e-2 (1.800e-5)
σ_ζ^2	5e-4	4.887e-4 (1.997e-7)	**5.459e-4** (1.877e-7)	4.667e-4 (4.456e-7)	5.597e-4 (2.139e-7)

Reported are mean values across 2,000 Monte Carlo simulations for data generated according to (3.58-3.60) with sample size *N=200*. Starting values for MLE and GMM optimization are $\theta_{SV}=(0.2;\ 0.2;\ 0.2)'$. To preserve non-negativity of the variance estimates, all the parameters were log-transformed. Standard errors of the estimates are given in parentheses. Estimates with the smallest standard errors are shown in bold.

Similar to the previous case, the Method of Moments provides consistent parameter estimates similar to MLE. However, unlike the previous case, for the state vector

parameters, MM outperforms the MLE in terms of the efficiency of the estimates.

To access if efficiency of the estimates may increase with adding additional moment conditions, model (3.58-3.60) with just two parameters was estimated with three and more moments. Overidentified estimation was carried via two-step GMM procedure. Table (3.4) summarizes results.

Table 3.4: GMM results for linear Gaussian SSM with overidentified parameters

		M1,M2		**M1,M2,M4**		**M1,M2,M4,M7**		**M1,M2,M4,M5,M7**	
Para meter	**True value**	**Mean (s.e.)**	**Abs.Bias (RMSE)**	**Mean (s.e.)**	**Abs.Bias (RMSE)**	**Mean (s.e.)**	**Abs.Bias (RMSE)**	**Mean (s.e.)**	**Abs.Bias (RMSE)**
σ_ϵ^2	5e-1	4.969e-1 (1.57e-3)	0.0031 (0.0702)	4.957e-1 (1.52e-3)	0.0043 (0.0684)	4.951e-1 (1.53e-3)	0.0049 (0.0688)	4.896e-1 **(1.49e-3)**	0.0104 (0.0674)
σ_ξ^2	5e-2	5.46e-2 (6.51e-3)	0.0046 (0.0295)	5.31e-2 (5.05e-4)	0.0031 (0.0228)	5.07e-2 **(4.49e-4)**	0.0007 (0.0201)	5.39e-2 (5.59e-4)	0.0039 (0.0253)

Results are based on 2,000 Monte Carlo simulations for data generated according to (3.58-3.60) with sample size *N=200*. Starting values for MLE and GMM optimization are θ_{SV}*=(0.2; 0.2)'*. To preserve non-negativity of the variance estimates, all the parameters were log-transformed. Standard errors of the estimates are given in parentheses. Estimates with smaller standard errors are shown in bold.

Apparently, adding more moment conditions improves the efficiency of the estimates, although MLE still outperforms GMM for all moment choices (see Table 3.2 for comparison). At the same time, adding more moment conditions while increasing efficiency, may lead to more biased results – in Table (3.4) the observation variance estimate has the smallest standard error when five moment conditions were used, however, this estimate appears to be the most biased one.

In the above illustrations we used several randomly selected moment conditions. However, in an empirical application a large number of moment conditions may pose a question exactly which moments should be chosen. Although all conditions M1-M9 are

based on different relations, and should not be correlated, they may contain the same information about the estimated parameters which may provide biased results. A potential solution to this problem is to evaluate performance of different combinations of moment condition in the simulated studies before applying the model to the real data.

To examine this question, we compare the performance of different sets of moment conditions for a random walk with drift with just two parameters – observation variance and variance of the state process (with true values are 0.5 and 0.05 respectively). Table 3.5 summarizes the results. These results can be also compared with the MLE performance in Table 3.2.

Based on the results of Table 3.5 we may draw several conclusions. Overall, different combinations of moment conditions provide adequate estimation for both parameters – for the observation and from the state equation. In many cases estimated parameters are comparable in terms of efficiency and bias to ML estimates. Shown in bold are the estimates which are more efficient than MLE estimates from Table (3.2). However, when including moments based only on the observation residuals (pair M3-M5), the model fails to properly estimate the state disturbances variance. Similarly, the combination M1-M5 performs poorly, when neither of the moment conditions provide information about parameters from the state vector. Also, the estimation failed to converge twice when M8 (which is based on the covariance with future values of the observation auxiliary residuals) was used together with M3 and M7.

Table 3.5: MC results for SSM for different combinations of moments.

	M2	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**
M1	0.497 (1.57e-3) 0.055 (6.48e-4)	0.507 (1.34e-3) **0.055** (4.02e-4)	0.497 (1.39e-3) 0.051 (4.47e-4)	0.202 (5.05e-3) 0.631 (1.10e-2)	0.498 (1.36e-3) **0.048** (3.80e-4)	0.497 (1.43e-3) 0.051 (4.92e-4)	0.500 (1.43e-3) **0.049** (4.25e-4)	0.514 (4.83e-3) 0.059 (1.21e-3)
M2		0.498 (1.32e-3) 0.052 (6.48e-4)	0.5225 (2.95e-3) 0.054 (5.37e-4)	0.5108 (3.49e-3) 0.053 (5.14e-4)	0.515 (3.44e-3) 0.052 (4.47e-4)	0.495 (1.50e-3) 0.054 (6.93e-4)	0.493 (1.43e-3) 0.057 (6.71e-4)	0.498 (1.57e-3) 0.056 (6.71e-4)
M3			0.447 (3.31e-3) **0.050** (4.25e-4)	0.070 (3.18e-3) 145.10 (447.11)	0.498 (1.32e-3) 0.049 (4.47e-4)	0.492 (1.34e-3) 0.051 (4.70e-4)	0.343 (8.27e-4) 0.376 (1.65e-3)	0.480 (1.77e-3) 0.059 (1.03e-3)
M4					0.502 (1.27e-3) 0.044 (4.92e-4)	0.500 (1.48e-3) **0.051** (4.25e-4)	0.499 (1.39e-3) **0.050** (4.02e-4)	0.478 (1.99e-3) 0.052 (6.48e-4)
M5					0.514 (3.87e-3) 0.053 (4.47e-4)	0.480 (1.72e-3) 0.066 (1.03e-3)	0.480 (1.81e-3) 0.065 (1.01e-3)	0.453 (1.50e-3) 0.057 (6.48e-4)
M6						0.500 (1.54e-3) 0.049 (4.70e-4)	0.501 (1.50e-3) 0.048 (4.25e-4)	0.514 (4.43e-3) **0.052** (4.02e-4)
M7							0.413 (1.34e-3) 0.199 (2.68e-4)	0.485 (1.83e-3) 0.058 (1.01e-3)
M8								0.494 (2.62e-3) 0.059 (1.05e-3)

Results are based on 1,000 Monte Carlo simulations for data generated according to ((3.58-3.60) with sample size $N=200$. True parameters values are values are $\theta_0=(0.5;\ 0.05)'$. Starting values for GMM optimization are $\theta_{SV}=(0.2;\ 0.2)'$. For each pair of moment conditions table reports mean value across all MC estimates and standard error in parentheses. To preserve non-negativity of the variance estimates, all the parameters were log-transformed. For all the lag-based moments the first lag was used. Estimates with smaller RMSE as compared to MLE estimates are shown in bold.

Finally, we estimated State Space model for the unemployment data in the form (1.27-1.33) for the original and deformed time scales using Method of Moments with several sets of different moment conditions. The results are similar to those discussed in section (1.6.3) for both - deformed and original - series when MLE approach was used.

3.4.2 Non-Gaussian UC Models

Although MM works quite well for the regular Gaussian SSM, the main purpose of our research is to carry out the estimation, when regular MLE would be inappropriate. In this section we compare estimation results in the State Space model of the form (3.58-3.60) when the observation disturbance comes from Student's distribution as specified in (3.24). Table 3.6 summarizes estimation results for the regular MLE, which incorrectly assumes that $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, MLE based on importance sampling (IS-MLE), and MM with several different sets of moment conditions.

Table 3.6: Estimation comparison for fat-tailed distribution

True value σ_ϵ^2 σ_η^2	**MLE based on Gaussianity assumption**	**MLE based on importance sampling**[30]	**MM**			
			M1, M2	**M1, M4**	**M6, M7**	**M8, M9**
5e-1	**0.507** (3.108e-3) RMSE=0.140	0.473 (6.462e-3) RMSE=0.290	0.493 (3.533e-3) RMSE=0.156	0.498 (3.309e-3) RMSE=0.147	0.495 (3.265e-3) RMSE=0.146	0.488 (3.667e-3) RMSE=0.164
5e-2	0.0499 (4.249e-4) RMSE=0.018	**0.0372** (8.944e-5) RMSE=0.013	0.0536 (5.814e-4) RMSE=0.026	0.0497 (4.249e-4) RMSE=0.019	0.051 (4.472e-4) RMSE=0.020	0.042 (7.826e-4) RMSE=0.036

Results are based on 2,000 Monte Carlo simulations for data generated according to (3.58-3.60) with sample size *N=200*. Observation disturbances come from Student's t-distribution with 2 degrees of freedom. Starting values for MLE and GMM optimization are $\theta_{SV}=(0.2;\ 0.2)'$. Standard errors of the estimates are given in parentheses. Also reported root mean square error. Estimates with smaller RMSE are shown in bold.

30 Estimation is carried out using SSM toolbox developed by Jyh-Ying Peng and John A. D. Aston, [2007]

Surprisingly, the least biased estimates are provided by the regular MLE. Method of Moments provides satisfactory and consistent results. MLE based on importance sampling is significantly more biased, although for the second parameter it has the smallest standard error and RMSE. Nevertheless, despite the fact that regular MLE provides the most adequate results, it may fail to capture data outliers. As an empirical illustration proving this statement, we consider an example with gas consumption in UKfrom 1980 to 1986 (Harvey [1994], Koopman [1999, 2003])[31].

The quarterly demand for gas is modeled as

$$y_t = \mu_t + \gamma_t + \epsilon_t \tag{3.63}$$

where the trend follows a local level linear model with trend as in (3.58-3.60), and the seasonal component is based on Harvey's [1994] specification:

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t+1-j} + \omega_t, \qquad \omega_t \ \ N(0,\sigma_\omega^2) \tag{3.64}$$

We estimate four parameters of the model using all three techniques. Results are summarized in Table 3.7.

Table 3.7: Gas consumption in the UK

	Gaussian MLE	**IS-MLE[32]**	**MM (M1,M2,M4,M8)**
σ_ϵ^2	1.822e-3	2.665e-3	2.353e-3
σ_μ^2	7.689e-10	7.690e-10	8.051e-10
σ_β^2	7.875e-6	7.768e-6	7.301e-6
σ_ω^2	3.308e-3	1.483e-3	1.858e-3

31 Data is taken from Harvey [1993].
32 Estimation is carried out using SSM toolbox developed by Jyh-Ying Peng and John A. D. Aston, [2007]

MLE based on importance sampling and MM estimation show similar estimates. At the same time, as pointed by Koopman (1999), the main outcome of the given application is to demonstrate the inability of Gaussian MLE to capture the unexpected change in the gas consumption which occurred in 1970-1971. These outliers are successfully identified by IS-MLE and GMM models, as shown in the following figure.

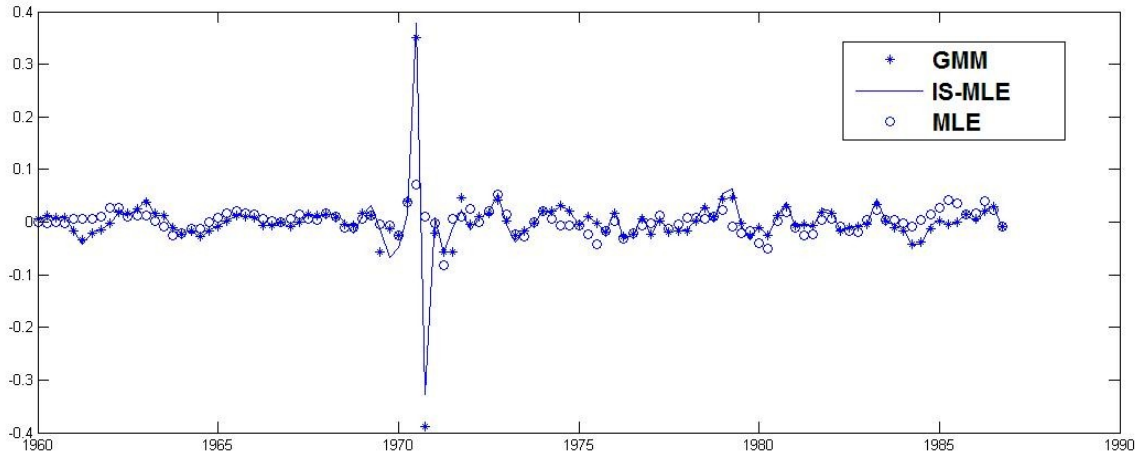


Figure 3.1: **Smoothed disturbance residuals for the gas consumption model**

3.4.3 Non-linear UC Models

First, we consider the case when non-linearity in the model is represented only by the relation between the observed value and the unobserved state:

$$y_t = Z_t(\alpha_t) + \epsilon_t, \qquad \epsilon_t \sim N(0, H_t) \tag{3.65}$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \quad \eta_t \sim N(0, Q_t) \tag{3.66}$$

Linearizing $Z_t(\alpha_t)$ by taking partial derivatives with respect to the state variable allows applying the Extended Kalman Filter to the data. Since the disturbance error

enters (3.65) linearly, and is normally distributed, maximum likelihood estimation remains an appropriate procedure for the parameter estimation. Moment conditions for the Method of Moments are constructed as described previously by replacing state matrices with their Jacobians. Table 3.8 summarizes results for the quadratic signal.

Table 3.8: Monte Carlo results for Extended Kalman Filter (quadratic signal)

	Model: $y_t = \alpha_t^2 + \epsilon_t, \qquad \epsilon_t \sim N(0, H_t)$					
Parameter	**True value**		**MLE**	**M1,M2**	**M3,M6**	**M4, M7**
σ_ϵ^2	5e-2	Not-corr. (%) Corr. (s.e) Abs.Bias/RMSE	7.061e-2 (14.42%) **5.963e-2** (4.772e-4) 9.636e-3/2.340e-2	9.665e-2 (6.2%) 4.28e-2 (7.070e-4) 7.182e-3/3.239e-2	5.396e-2 **(0%)** 5.396e-2 (6.456e-4) 3.963e-3/2.896e-2	**1.157e-2** (5.7%) 5.309e-2 (5.532e-4) 3.093e-3/2.491e-2
σ_μ^2	5e-2	Not-corr. (%) Corr. (s.e) Abs.Bias/RMSE	4.679e-2 (0.7%) **4.372e-2** (2.176e-4) 6.27e-3/1.062e-2	4.595e-2 (0.6%) 4.514e-2 (2.28e-4) 4.862e-3/1.129e-2	7.196e-2 (0.6%) 5.959e-2 (1.042e-3) 9.598e-3/4.756e-2	4.447e-2 **(0%)** 4.447e-2 (6.800e-4) 5.532e-3/3.089e-2

Results are based on 2,000 Monte Carlo simulations for data generated according to (3.65-3.66) with sample size *N=200*. Starting values for MLE and GMM optimization are θ_{SV}*=(0.2; 0.2)'*. To preserve non-negativity of the variance estimates, all the parameters were transformed as *log(θ)*. Standard errors of the estimates are given in parentheses.

As mentioned before, the Extended Kalman Filter may sometimes diverge thus providing biased estimates. In Table (3.8) first row (Not-corr) reports the mean estimate value across 2,000 Monte Carlo runs including cases when either MLE or GMM estimation did not converge – parameter estimates either remained at their starting values or diverged to infinity. Those non-convergent values were removed. The percentage in parentheses shows the proportion of estimates which were dropped, and second row (Corr) represents the mean value across cleared estimates. Bias and RMSE are reported for corrected data. The first result we can observe, is that there are significantly fewer cases when estimation failed to converge for MM than for MLE. Second, MLE estimates again appear to be more efficient and they have smaller RMSE. Despite that, MM estimates almost in all cases are less biased as reported by absolute bias numbers.

As a more complicated empirical illustration for the MM estimation in non-linear UC models we consider Stochastic Volatility example discussed in Harvey [1994], Koopman [1993, 2000][33]. This case differs from the previous one by introducing a multiplicative disturbance into the observation equation.

The log difference of daily pound/dollar exchange rates is modeled as

$$y_t = \sigma \exp(\frac{1}{2}\theta_t) u_t \quad u_t \sim N(0,1) \tag{3.67}$$

where a signal θ_t follows the AR(1) process:

$$\theta_{t+1} = \phi \theta_t + \eta_t \qquad \eta_t \sim N(\sigma_\eta^2) \tag{3.68}$$

The model is log-linearized as $\log(y_t^2)$ and estimated using regular MLE, importance sampling MLE and MM techniques. Since the linearization process produces observation disturbances that are approximately distributed as *log(χ^2)*, this poses a problem for the traditional MLE. Linearization of the model is necessary for IS-MLE as discussed in detail in Koopman [2004]. For the regular MLE and MM the model is estimated as a simple SSM with AR(1)-driven state realizations. Monte Carlo results for the simulated SV data are presented in Table 3.9.

33 Data is provided in Harvey (1994)

Table 3.9: Monte Carlo results for Stochastic Volatility model

	True value	**Regular MLE**	**IS-MLE**[5]	**MM (M2, M6, M7)**
σ	0.6	0.568 (1.051e-3) RMSE=0.056	**0.585** (4.472e-5) RMSE=0.0156	0.573 (2.245e-3) RMSE=0.103
ϕ	0.95	0.863 (4.562e-3) RMSE=0.221	**0.964** (2.236e-5) RMSE=0.014	0.970 (7.714e-4) RMSE=0.040
σ_η^2	0.1	0.104 (3.309e-3) RMSE=0.1483	0.116 (1.118e-4) RMSE=0.017	**0.101** (2.124e-3) RMSE=0.0951

Results are based on 2,000 Monte Carlo simulations for data generated according to (3.67-3.68) with sample size $N=400$. Starting values for MLE and GMM optimization are $\theta_{SV}=(0.2;\ 0.8;\ 0.2)'$. Standard errors of the estimates are given in parentheses.

Importance sampling MLE gives the least biased estimates for σ and AR-parameter, MM provides second-best results. At the same time MM estimates trend variance with the smallest bias. Regular MLE shows the worst performance among all techniques considered.

This example gives evidence that Method of Moments estimation can compete with more complicated IS-MLE technique. The IS-MLE technique requires linearizing the model individually for each empirical application, which assumes knowing the exact model and data generating process, while the Method of Moments can be used as a general tool which does not depend on the model specifics or distributional properties of the data. Another gain from using MM is the significantly smaller amount of time required for the estimation, as compared to the IS-MLE.

To conclude, we consider an illustration of applying MM estimation for a Stochastic Volatility model based on real data. Table (3.10) reports the estimates for the SV model (3.67-3.68) fitted to the daily pound/dollar exchange rates data covering the period from

1/10/1981 to 28/6/1985, and examined in Harvey [1993], and Koopman [2000]

Table 3.10: Estimation results for pound/dollar exchange rates

Parameter	**MLE**	**Importance sampling MLE**[34]	**MM (M2, M6, M7)**
σ	0.534	0.639	0.611
ϕ	0.991	0.977	0.931
σ_η^2	0.009	0.023	0.017

The parameter estimates for IS-MLE and MM are much closer to each other than to the regular MLE, although MM reports a smaller state variance as well as a smaller value for the autocorrelation parameter ϕ.

34 Estimation is carried out using SSM toolbox developed by Jyh-Ying Peng and John A. D. Aston, 2007

3.5 Conclusions

In this paper we introduce the Method of Moments as an alternative estimation technique for the parameters in State Space models framework. So far, likelihood-maximization approaches remain the most common tool for SSM investigation, however, they become inadequate or too complicated when the focus shifts to non-linear or non-Gaussian models. To overcome this difficulty, using Kalman filter and smoother equations we derived a set of relations, which can be used to construct many moment conditions for either MM or GMM estimation.

Simulation results and empirical applications provided evidence that our approach provides the results similar to the MLE method for the regular linear Gaussian SSM, while for non-linear or non-Gaussian models, it outperforms regular MLE and can compete with estimation procedures based on sampling techniques. At the same time, it is much easier to implement, since constructing moment conditions does not require knowledge about the distributional properties or model's non-linearity for each individual case. Another gain from using MM is significantly smaller computational burden involved as compared to the sampling and simulation tools.

BIBLIOGRAPHY

Anderson B.D.O., and Moore, J.B. “Optimal Filtering”, Prentice Hall, 1979.

Andrews, D.W. “Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation”, Econometrica, vol.59 (3), 1991.

Baxter, M., and King, R.G. "Measuring business cycles: Approximate band-pass filters for economic time series". The Review of Economics and Statistics, vol 81 (4), 1999.

Bierens H.J. “Topics in Advanced Econometrics: Estimation, Testing, and Specification of Cross-Section and Time Series Models”, Cambridge University Press, 1996.

Brockwell, P.J, and Davis R.A. “Time series: theory and methods”, 2e, Springer, 1991.

Burns, A.F. and Mitchell, W.C. “Measuring Business Cycles”. New York: NBER, 1946.

Carrasco, M., Hu, L., and Ploberger, W. “Optimal Test for Markov Switching”, Econometric Society North American Summer Meetings, N396, 2004.

Carrasco, M., Hu, L., and Ploberger, W. “Optimal Test for Markov Switching Parameters”, 2009.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. “'A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling”, Journal of the American Statistical Association, vol. 87 (418), 1992.

Carter, C. K., and Kohn, P. ”On Gibbs Sampling for State Space Models”, Biometrica, vol. 81, 1994.

Choi, E., Woodward, W. A., and Gray, H. L. “Euler (p,q) Processes and their Application to Nonstationary Time Series with Time-Varying Frequencies”, Communications in Statistics - Theory and Methods, vol. 35, 2006.

Clark, P. “The cyclical component of U.S. Economic Activity”. The Quaterly Journal of Economics, vol.102 (4), 1987.

Commandeur, J. J. F., and Koopman, S.J. “An Introduction to State Space Time Series Analysis”, Oxford University Press, 2007.

Davidson, J. “Stochastic limit theory: an introduction for econometricians”, Oxford University Press, 1994.

Davidson, J. “Econometric theory”, Wiley-Blackwell, 2000.

De Jong, P., and M. Mackinnon. “Covariances for Smoothed Estimates in State Space Models”, Biometrika, vol.75 (3), 1988

DeJong, P., and N. Shephard. “The Simulation Smoother for Time Series Models”, Biometrika, vol. 82 (2), 1995.

Dickey, D. A., and Fuller, W. A. . "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", Journal of the American Statistical Association, vol. 74, 1979.

Diebold, F.X. and Rudebusch, G.D. "Measuring Business Cycles: A Modern Perspective". The Review of Economics and Statistics, vol. 78 (1), 1996.

Diebold, F. X. and Mariano, R. S. “Comparing Predictive Accuracy”, Journal of Business & Economic Statistics, vol. 13 No. 3, 1995.

Durbin, J., and Koopman, S.J. "Time Series Analysis by State Space Methods". Oxford: Oxford University Press, 2001.

Durbin, J., and Koopman, S.J. "A simple and efficient simulation smoother for state space time series analysis" Biometrika vol. 89, issue 3, 2002.

Durbin, J., and Koopman, S.J. “Time Series Analysis of Non-Gaussian Observations Based on State Space Models from both Classical and Bayesian Perspectives”, Journal of Royal Statistical Society, 62, part 1, 2000.

Durbin, J., and Koopman, S.J. “Monte Carlo maximum likelihood estimation for non-Gaussian state space models”, Biometrika, 84(3), 1997

Gallant, R., White, H “A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models”, Basil Blackwell Ltd, 1988.

Gray, H.L., Vijverburg, C.P., and Woodward, W.A. "Nonstationary Data Analysis by Time Deformation". Communication in Statistics, A34 (2005), 163-192.

Gray, H.L. and Zhang, N. "On a class of nonstationary processes". Journal of Time Series Analysis, vol. 9, issue 2, 1988.

Hall, A.R. “Generalized Method of Moments”, Oxford University Press, 2005.

Hansen, L. “Large Sample Properties of Generalized Method of Moments Estimators”, Econometrica, vol. 50 (3), 1982.

Hamilton, J.D. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle". Econometrica, vol. 57 Issue 2, 1989.

Hamilton, J.D. "Analysis of time series subject to changes in regime". Journal of Econometrics, vol. 45, 1990.

Hamilton, J. D. “Time Series Analysis”, Princeton University Press, 1994.

Hannan, E. J., and Heyde, C.C. “On limit theorems for quadratic functions of discrete time series”. Annals of Mathematical Statistics, vol. 43, 1972.

Hannan, E. J. “Group Representation and Applied Probability”, Journal of Applied Probability, vol. 2(1), 1965.

Hardy, G. H., Littlewood, J. E., and Pólya, G. “Inequalities”, 2e, Cambridge

University Press, 1988.

Harvey, A.C. "Forecasting, Structural Time Series Models and the Kalman Filter". Cambridge: Cambridge University Press, 1989.

Harvey, A.C., and Jaeger, A. “Detrending, stylised facts and the business cycle.” Journal of Applied Econometrics (8), 1993.

Harvey, A.C., and Shephard, N. “Structural Time Series Models”, Handbook of Statistics, vol. 11, Elsevier Science Publishers, 1993.

Harvey, A., and T. Proietti. “Readings in unobserved components models”, Oxford University Press, 2005.

Hodrick, R.J., and Prescott, E.C. “Post-War U.S. Business Cycles: An Empirical Investigation”. Journal of Money, Credit and Banking 29 (1), 1997.

Ibragimov, I.A., and Linnik, Yu.V., “Independent and Stationary Sequences of Random Variables”, Groningen: Wolters-Noordhoff, 1971.

Jiang H., Gray, H.L. and Woodward, W.A. "Time-frequency analysis - G(λ)-stationary processes". Computational Statistics & Data Analysis, vol. 51 (3), 2006.

Jorda O., and Marcellino, M. “Time-Scale Transformations of Discrete Time Processes”. Journal of Time Series Analysis, vol. 25 (6), 2004.

Julier, S. J., and Uhlmann, J. K. “Unscented filtering and nonlinear estimation”, Proceedings of the IEEE, vol. 92 (3), 2004.

Kalman, R. E. "A New Approach to Linear Filtering and Prediction Problems", Transactions ASME Journal of Basic Engineering, 82 (series D), 1960.

Kaufman S., and Scheicher, M. “Markov-regime Switching in Economic Variables”, Institute for Advanced Studies, Vienna, Economic Series N38, 1996.

Kim, C.J., Nelson, C.R., and Piger, J. "The Less-Volatile U.S. Economy: A Bayesian Investigation of Timing, Breadth, and Potential Explanations". Journal of Business and

Economic Statistics, vol. 22, Issue 1, 2004.

Kim, C.J. and Nelson, C.R. "Has the U.S. economy become more stable? A Bayesian approach based on a Markov-Switching model of the business cycle". The Review of Economics and Statistics, vol. 81, No. 4, 1999.

Kim, C.J. and Nelson, C.R. " Business Cycle Turning Points, A New Coincident Index, and Test of Duration Dependence Based on a Dynamic Factor Model with Regime Switching".The Review of Economics and Statistics, 80, 1998.

Kim, C.J. and Nelson, C.R. “State-space Models With Regime Switching : Classical and Gibbs-sampling Approaches With Applications”, MIT Press, 1999.

Koop, G. “Bayesian Econometrics”, Wiley-Interscience, 2003.

Koopman, S.J. “Disturbance Smoother for State Space Models”, Biometrika, vol. 80 (1), 1993.

Koopman, S.J. and Wong, S.Y. "Extracting Business Cycles using Semi-parametric Time-varying Spectra with Applications to US Macroeconomic Time Series". Tinbergen Institute Discussion Paper No.06-105/4, 2006.

Koopman, S.J., Ooms, M., and Hindrayanto, I. "Periodic Unobserved Cycles in Seasonal Time Series with an Application to US Unemployment". Tinbergen Institute Discussion Paper No.06-101/4, 2006.

Koopman, S.J. and Lee, K.M. "Measuring Asymmetric Stochastic Cycle Components". Tinbergen Institute Discussion Paper No.05-081/4, 2005.

Morley, J.C., Nelson, C.R., and Zivot, E. “Why Are the Beveridge-Nelson and Unobserved-Components Decompositions of GDP So Different?”. Review of Economics and Statistics, vol.85, Issue 2.

Neftci, S.N. "Are Economic Time-Series Asymmetric Over The Business-Cycle". Journal of Political Economy 92 (2), 1984.

Nelson C. and Murray, C. "The Uncertain Trend in U.S. GDP". Journal of Monetary Economics, vol. 46, Number 1, 2000.

Newey, W. K., and K. D. West. “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix”, Econometrica, vol. 55 (3), 1987.

Newey, W. K., and McFadden, D. “Large sample estimation and hypothesis testing”, Handbook of Econometrics, vol. 4, 1994.

Perron, P., and Wada, T. “Let’s Take a Break: Trends and Cycles in US Real GDP?”. Boston University Working Papers Series #WP2005-031, 2005.

Penzer, J. and Tripodis, Y. "Single season heteroscedasticity in time series.' Journal of Forecasting, 2006.

Perlin, M. “Estimation, Simulation and Forecasting of a Markov Regime Switching Regression in Matlab”, MATLAB Central file exchange, 2009

Raihan S.M., Wen, Y., and Zeng B. "Wavelet: a new tool for business cycle analysis". Federal Reserve Bank of St. Louis Working Papers Num. 2005-050A, 2005.

Sensier, M. and van Dijk, D. "Testing for volatility changes in U.S. macroeconomic time series'. Review of Economics and Statistics 86, vol.3, 2004.

Romano, J.L., and Thombs L. A.. “Inference for autocorrelations under weak Assumptions”. Journal of American Statistical Association, 91, 1996

Shephard, N., and M. K. Pitt. “Likelihood Analysis of Non-Gaussian Measurement Time Series”. Biometrika, vol.84 (3), 1997.

Stock, J.H. “Measuring Business cycle time”. Journal of Political Economy, vol. 95, no. 6, 1987.

Stock, J.H. and Watson, M.W. "Has the business cycle changed? Evidence and

explanations". Federal Reserve Bank of Kansas City, journal "Proceedings", 2003.

Vijverberg, C.P. "Time Deformation, Continuous Euler Processes and Forecasting". Journal of Time Series Analysis, vol. 27, No. 6, 2006

Vijverberg, C.P. and Gray, H.L. "Discrete Euler Processes and Their Applications". Southern Methodist University Technical Report No.314, 2004.

Yaffee, R.A., and McGee, M. “Introduction to time series analysis and forecasting: with applications of SAS and SPSS”, Academic Press, 2000.

Yogo, M. "Measuring Business Cycles: Wavelet Analysis of Economic Time Series". Department of Economics, Harvard University, Cambridge, MA, 2003.

Zarchan, P. “Fundamentals of Kalman Filtering - A Practical Approach”, American Institute of Aeronautics and Astronautics, 2005.