UNIVERSITY OF CALIFORNIA
RIVERSIDE

Generalized Autocontour: Evaluation of the Density Model in Stable and Unstable Environments

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

Doctor of Philosophy

in

Economics

by

Yingying Sun

June 2013

Dissertation Committee:
Dr. Gloria González-Rivera , Chairperson
Dr. Marcelle Chauvet
Dr. Tae-Hwy Lee
Dr. Aman Ullah

The Dissertation of Yingying Sun is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

Foremost, I would like to express my sincere gratitude to my advisor Prof. Gloria González-Rivera for the continuous support of my Ph.D study and research, for her patience, motivation, enthusiasm, and immense knowledge. Her mentorship helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and guidance for my Ph.D study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Marcelle Chauvet, Prof. Tae-Hwy Lee, and Prof. Aman Ullah, for their encouragement, insightful comments, and hard questions.

My sincere thanks also goes to my best friends: Wei Lin and Mingming Jiang whose patience, carefulness and kindness make the five years in Riverside to be the best days in my life. Also I want to thank all my other friends for their helps in my life and career.

Last but not the least, I would like to thank my family for supporting me spiritually throughout my life.

To my family for all the support.

ABSTRACT OF THE DISSERTATION

Generalized Autocontour: Evaluation of the Density Model
in Stable and Unstable Environments

by

Yingying Sun

Doctor of Philosophy, Graduate Program in Economics
University of California, Riverside, June 2013
Dr. Gloria González-Rivera , Chairperson

In Chapter 1, we propose a generalized version of the autocontour-based methodology for dynamic specification testing (in-sample and out-of-sample) proposed in González-Rivera et al. (2011) (GR2011). The autocontour (ACR) proposed is the basis to construct very powerful tests to detect misspecification in the dynamics of the model and departures from the assumed conditional density but still has some limitations. To overcome these limitations, we propose a generalized autocontour (G-ACR) based on the probability integral transforms (PIT) of the assumed density model. The specification tests will be based now on the G-ACR device. We should also mention that because of the simplicity of G-ACR, the analytical expressions of the asymptotic variance-covariance matrices of the tests have a closed formulation and they depend only the *a priori* probability level associated with the G-ACR. In addition, they will be instances, e.g. multistep predictive densities in nonlinear models, in which the predictive density does not have a closed form solution and we need to resort to simulation or nonparametric methods, but yet we could obtain the PIT process from the simulated density. Once the PITs are in place, our methodology will be able to evaluate the multistep density forecast. In Chapter 2, we extend the G-ACR methodology to the multivariate

case and to random processes that are discrete. Our interest lies on the multivariate process of a vector of counts for which we specify the dynamics of the marginal densities of each process and a copula function that ties up the marginal to produce their multivariate distribution. As an illustration of the G-ACR methodology, we have analyzed a high frequency trivariate system of the number of trades in three US large banking institutions: Bank of America, JP Morgan Chase, and Wells Fargo. In Chapter 3, we propose a robust out-of-sample density forecasting evaluation method in the presence of the instabilities based on Generalized Autocontour . We construct Sup and Avg types of statistics to explore the behavior of the model in unstable environments. We have applied our tests to evaluate the density forecast performance of U.S. inflation produced by linear and Markov-switching Philips Curve.

Contents

List of Figures

List of Tables

Chapter 1

Generalized Autocontour

1.1 Introduction

We propose a generalized version of the autocontour-based methodology for dynamic specification testing (in-sample and out-of-sample) proposed in Gonzlez-Rivera et al. (2011) (GR2011). For a stochastic process $\{Y_t\}$, we would like to test the joint hypothesis of a correct density model, which includes the dynamics of the appropriate conditional moments and the functional form of the assumed conditional density. In GR2011, this hypothesis is equivalent to test the iid-ness of the generalized innovations of the model, say $\{\varepsilon_t\}$ jointly with the functional form of their density. Under this null hypothesis, the autocontour (ACR) is a powerful analytical set of points$(\varepsilon_t, \varepsilon_{t-k})$ that, for a priori probability level α and by projecting the bivariate density of vectors such as $(\varepsilon_t, \varepsilon_{t-k})$ in a two-dimensional space, must contain $\alpha\%$ of the observations. ACR is the basis to construct very powerful tests to detect misspecification in the dynamics of the model and departures from the assumed conditional density. However, the original methodology in GR2011 has some limitations. First, when the density of interest departs from standard densities used in time series econometrics, e.g. Normal, Student-t,

exponential, etc. the analytical expressions of the ACRs for $(\varepsilon_t, \varepsilon_{t-k})$ may be mathematically cumbersome to obtain and we need to resort to numerical methods. Secondly, the asymptotic variance-covariance matrices of the specifications tests in GR2011 do not have all closed-form solutions, some combining parametric with nonparametric expressions. Thirdly, GR2011 considers only continuous random processes with dynamics on the conditional mean and conditional variance and a time-invariant functional form of the density of the standardized innovations $\{\varepsilon_t\}$. To overcome these limitations, we propose a generalized autocontour (G-ACR). Instead of working with the innovations $\{\varepsilon_t\}$, , we work with the probability integral transforms (PIT) of the assumed density model. Under correct specification, the PITs, say $\{u_t\}$ should be an i.i.d. process uniformly distributed on [0,1]. Implementing the original ACR in GR2011 will not be helpful because, for any vector (u_t, u_{t-k}) the ACR obtained by projecting the bivariate uniform density on the plane (u_t, u_{t-k}) will always be a square of area one. We propose a generalized autocontour (G-ACR) based on cutting the domain (instead of the range) of the bivariate uniform density. Under the true density model, the G-ACR is a set of point in the plane (u_t, u_{t-k}) such that any square with $\sqrt{\alpha}$ side must contain $\alpha\%$ of the observations. The specification tests will be based now on the G-ACR device. We should also mention that because of the simplicity of G-ACR, the analytical expressions of the asymptotic variance-covariance matrices of the tests have a closed formulation and they depend only on the parameter -level probability associated with the G-ACR. In addition, they will be instances, e.g. multistep predictive densities in nonlinear models, in which the predictive density does not have a closed form solution and we need to resort to simulation or nonparametric methods, but yet we could obtain the PIT process from the simulated density. Once the PITs are in place, our methodology will be able to evaluate the multistep density forecast.

1.2 Generalized Autocontour and Test Statistics

In this section, we introduce a device -the generalized autocontour- that is the basis to construct statistical tests for the null hypothesis of a well-specified conditional density model. We consider univariate and multivariate models with continuous and discrete stochastic processes.

1.2.1 Generalized Autocontour: G-ACR

Following Diebold, Gunther, and Tay (1998) among others, if the proposed predictive density model for Y_t, i.e. $\{f_t^*(y_t|\Omega_{t-1})\}_{t=1}^T$ coincides with the true conditional density $\{f_t(y_t|\Omega_{t-1})\}_{t=1}^T$, then the sequence of probability integral transforms (PIT) of $\{Y_t\}_{t=1}^T$ w.r.t $\{f_t^*(y_t|\Omega_{t-1})\}_{t=1}^T$ i.e. $\{u_t\}_{t=1}^T$ must be $i.i.d\ U(0,1)$where $u_t = \int_{-\infty}^{y_t} f_t^*(v_t|\Omega_{t-1})dv_t$. Thus, the null hypothesis $H_0 : f_t^*(y_t|\Omega_{t-1}) = f_t(y_t|\Omega_{t-1})$ is equivalent to the null hypothesis $H_0^{'} : \{u_t\}_{t=1}^T$ is $i.i.d\ U(0,1)$.

We construct the G-ACR under i.i.d. uniformity for univariate and multivariate predictive densities. We start with the univariate case. Within the process $\{u_t\}_{t=1}^T$, we choose any vector $(u_{t,}u_{t-k}) \subset R^2$. Under $H_0^{'} : \{u_t\}_{t=1}^T\ i.i.d\ U(0,1)$, the G-ACR$_{\alpha_i,k}$ is defined as the set of points in the plane (u_t, u_{t-k}) such that the square with $\sqrt{\alpha_i}$ side contains $\alpha_i\%$ of observations, i.e.,

$$
\begin{aligned}
& G - ACR_{\alpha_i,k} \\
= \ & \{B(u_t, u_{t-k}) \subset \Re^2 | 0 \leq u_t \leq \sqrt{\alpha_i} \text{and} \, 0 \leq u_{t-k} \leq \sqrt{\alpha_i}, \text{s.t.} \, u_t \times u_{t-k} \leq \alpha_i\}
\end{aligned}
$$

We will call the proposed cutting of the domain 'symmetric' as we impose the same upper bound $\sqrt{\alpha_i}$ for each u in the vector (u_t, u_{t-k}) . However, it is possible to cut the domain in many different ways. The symmetric cutting will be very advantageous on computing the variance-covariance matrix of the test statistics that we present in

the forthcoming sections, and it will also facilitate the construction of the uniform autocontours in the multivariate case, as we explain shortly.

The original autocontour proposed in González-Rivera *et. al.* (2011) was constructed for the process of (standardized) innovations of a dynamic model, say $\{\varepsilon_t\}_{t=1}^T$, with assumed conditional density $f(.)$. Under the null hypothesis of i.i.d. random variables, the bivariate density function of any vector, say $(\varepsilon_{t,}\varepsilon_{t-k}) \subset R^2$ is $f(\varepsilon_{t,}\varepsilon_{t-k})$=$f(\varepsilon_t)f(\varepsilon_{t-k})$, and the autocontour ACR_α was defined as the set of points in the plane $(\varepsilon_{t,}\varepsilon_{t-k}) \subset R^2$ that will contain $\alpha\%$ of observations when we *horizontally* slice the bivariate density function at a fixed value $\overline{f}_\alpha$. Mathematically,

$$ACR_\alpha = \{B(\varepsilon_{t,}\varepsilon_{t-k}) \subset R^2 | \int_l^u \int_{l_k}^{u_k} f(\varepsilon_t)f(\varepsilon_{t-k})d\varepsilon_t d\varepsilon_{t-k} \leq \alpha\}$$

Observe that if were to implement the same approach for the process of PITs $\{u_t\}_{t=1}^T$, under the i.i.d null hypothesis, the joint probability density function of this vector is always a constant, i.e. $f(u_{t,}u_{t-k}) = f(u_t)f(u_{t-k}) = 1 \times 1 = 1$ and, unlike the original ACR, we cannot construct the autocontour by horizontally slicing the joint density. If we horizontally slice $f(u_{t,}u_{t-k})$ at any level, and project down the resulting segment on the plane $(u_{t,}u_{t-k})$ we always obtain a square of area one. For this reason, the proposed G-ACR 'cuts' instead the domain of the bivariate uniform density function that is a square with unit side.

In the multivariate case, our interest is the modeling of an $m \times 1$ random vector $Y_t = (y_{1,t}, y_{2,t,}.....y_{m,t})'$, with a joint probability density $f_t(Y_t|\Omega_{t-1})$. Let us call $\{f_t^*(Y_t|\Omega_{t-1})\}_{t=1}^T$ the sequence of predicted densities. For each t, the joint density can be factorized as the product of the conditional densities and the marginal density, i.e.

$$\begin{aligned} f^*(Y_t|\Omega_{t-1}) &= f^*(y_{1,t}, y_{2,t,}.....y_{m,t}|\Omega_{t-1}) \\ &= f^*(y_{m,t}|y_{1,t}, y_{2,t,}.....y_{m-1,t}) \times .... \times f^*(y_{2,t}|y_{1,t}) \times f^*(y_{1,t}|\Omega_{t-1}) \end{aligned}$$

For each element in the factorization we obtain the sequence of PITs:

$\{u_{1,t}, u_{2|1,t}, ..., u_{m|1,2...m-1,t}\}_{t=1}^{T}$ such that $u_{1,t} = \int_{-\infty}^{y_{1,t}} f_t^*(v_t|\Omega_{t-1})dv_t$, $u_{2|1,t} = \int_{-\infty}^{y_{2,t}} f_{y_2|y_1,t}^*(v_t|\Omega_{t-1})dv_t$, ... , and $u_{m|1,2...m-1,t} = \int_{-\infty}^{y_{m,t}} f_{y_m|y_1,y_2...y_m,t}^*(v_t|\Omega_{t-1})dv_t$.

Under the null hypothesis of a correct density model, the multivariate sequence $\{u_{1,t}, u_{2|1,t}, ..., u_{m|1,2...m-1,t}\}_{t=1}^{T}$ is $i.i.d\ U(0,1)$ (Diebold, Hahn, and Tay, 1999); thus the null hypothesis $H_0 : f_t^*(Y_t|\Omega_{t-1}) = f_t(Y_t|\Omega_{t-1})$ is equivalent to:

$H_0' : \{u_{1,t}, u_{2|1,t}, ..., u_{m|1,2...m-1,t}\}_{t=1}^{T}$ is $i.i.d\ U(0,1)$.

Under this null, we construct the autocontour G-ACR$_{\alpha_i,k}$ by choosing any two m-dimensional vectors in the sequence of multivariate PITs that are k periods apart. Thus, G-ACR$_{\alpha_i,k}$ is now the set of points defined in R^{2m} such that the hyper-cube with $\alpha_i^{1/2m}$ side contains $\alpha_i\%$ of observations:

$$\text{G-ACR}_{\alpha_i,k} = \{B(u_{1,t}, u_{2|1,t}, .., u_{m|1,2...m-1,t}; u_{1,t-k}, u_{2|1,t-k}, ..., u_{m|1,2...m-1,t-k}) \subset \Re^{2m}$$
$$\| \ 0 \leq u_{1,t} \leq \alpha_i^{1/2m}, 0 \leq u_{2|1,t} \leq \alpha_i^{1/2m}, \cdots, 0 \leq u_{m|1,2...m-1,t} \leq \alpha_i^{1/2m}$$
$$0 \leq u_{1,t-k} \leq \alpha_i^{1/2m}, 0 \leq u_{2|1,t-k} \leq \alpha_i^{1/2m}, \cdots, 0 \leq u_{m|1,2...m-1,t-k} \leq \alpha_i^{1/2m}$$
$$s.t. : u_{1,t} \times u_{2|1,t} \times \cdots \times u_{m|1,2...m-1,t} \times u_{1,t-k} \times u_{2|1,t-k} \times \cdots \times u_{m|1,2...m-1,t-k} \leq \alpha_i\}$$

In Figure 1, we present 2-dimensional α_i-autocontours for the univariate case: $B(u_t, u_{t-k}) \subset \Re^2$ (left panel) and 3-dimensional autocontours for a bivariate case (right panel). In the latter, the autocontour is a 4-dimensional hypercube

$$B(u_{1,t}, u_{2|1,t}, u_{1,t-k}, u_{2|1,t-k}) \subset \Re^4,$$

of which we can only plot obviously three dimensions.

1.2.2 Test Statistics for Continuous Distribution Functions

We start by defining an indicator function in the univariate as well as in the multivariate case.

In the univariate case, the indicator will take the value one whenever the observation (u_t, u_{t-k}) falls within the area defined by the $ACR_{\alpha_i,k}$ and zero otherwise, i.e.

$$I_t^{k,\alpha_i} = \mathbf{1}((u_t, u_{t-k}) \subset ACR_{\alpha_i,k}) = \mathbf{1}(0 \leq u_t \leq \sqrt{\alpha_i}, 0 \leq u_{t-k} \leq \sqrt{\alpha_i})$$

In the multivariate case, the indicator will be defined similarly as follows

$$\begin{aligned} &I_t^{k,\alpha_i} = \mathbf{1}((u_{1,t}, u_{2|1,t}, .., u_{m|1,2...m-1,t}, u_{1,t-k}, u_{2|1,t-k}, ..., u_{m|1,2...m-1,t-k}) \subset ACR_{\alpha_i,k}) \\ &= \mathbf{1}(0 \leq u_{1,t} \leq \alpha_i^{1/2m}, 0 \leq u_{2|1,t} \leq \alpha_i^{1/2m}, .., 0 \leq u_{m|1,2...m-1,t} \leq \alpha_i^{1/2m}, \\ &0 \leq u_{1,t-k} \leq \alpha_i^{1/2m}, 0 \leq u_{2|1,t-k} \leq \alpha_i^{1/2m}, ..., 0 \leq u_{m|1,2...m-1,t-k} \leq \alpha_i^{1/2m}) \end{aligned}$$

In both cases, I_t^{k,α_i} are Bernoulli random variables for which $\alpha_i\%$ of the observations will fall inside the $ACR_{\alpha_i,k}$ and $(1-\alpha_i)\%$ observations will fall outside $ACR_{\alpha_i,k}$. Note that the indicator is also autocorrelated; it follows a MA process whose order will depend on k. Consequently, the moments of I_t^{k,α_i} are

$$\begin{aligned} E(I_t^{k,\alpha_i}) &= \alpha_i \\ Var(I_t^{k,\alpha_i}) &= \alpha_i(1-\alpha_i) \\ r_h^{\alpha_i} &= cov(I_t^{k,\alpha_i}, I_{t-h}^{k,\alpha_i}) = \begin{cases} 0 & \text{if } h \neq k \\ \alpha_i^{3/2}(1-\alpha_i^{1/2}) & \text{if } h = k \end{cases} \end{aligned}$$

The indicator I_t^{k,α_i} forms the basis of the following test statistics, which are applicable in the univariate as well as in the multivariate case. The next three propositions follow closely the tests derived in González-Rivera *et. al.* (2011). However, the forthcoming tests offer a great advantage over the previous ones because their asymptotic variance-covariance matrices have closed-form formulations which depend exclusively on the theoretical probability α_i of the autocontour specified under the null. The proofs of the following propositions are all relegated to the Appendix.

Proposition 1

Let $\widehat{\alpha_i} = \frac{\sum_{t=k+1}^{T} I_t^{k,\alpha_i}}{T-k}$ be the sample proportion. Under the null hypothesis of $i.i.d\ U(0,1)$ PITs, the sample proportion $\widehat{\alpha_i}$ is asymptotically normally distributed, i.e.,

$$\frac{\sqrt{T-k}(\widehat{\alpha_i} - \alpha_i)}{\sigma_{k,i}} \to N(0,1)$$

where

$$\sigma^2_{k,\alpha_i} = \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2})$$

In Proposition 1, the lag k and the autocontour α_i are fixed. By letting k and i run through many values, we can construct portmanteau statistics as those in the following propositions.

Proposition 2

For a given contour α_i, let $\ell_{k,\alpha_i} = \sqrt{T-k}(\widehat{\alpha_i} - \alpha_i)$ and stack ℓ_{k,α_i} for $k = 1, ...K$. Let $L_{\alpha_i} = (\ell_{1,\alpha_i}, ...\ell_{K,\alpha_i})'$ be the $K \times 1$ stacked vector. Under the null hypothesis of $i.i.d\ U(0,1)$ PITs, the asymptotic distribution of the vector L_{α_i} is multivariate normal, i.e., $L_{\alpha_i} \to N(0, \Lambda_{\alpha_i})$ and the following quadratic form follows asymptotically a chi-square with K degrees of freedom, i.e.,

$$L'_{\alpha_i}\Lambda^{-1}_{\alpha_i}L_{\alpha_i} \to \chi^2_K$$

where a typical element of the asymptotic covariance matrix Λ_{α_i}, say $\lambda_{j,k}$ is as follows

$$\lambda_{j,k} = \begin{cases} \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2}) & \text{if } j = k \\ 4\alpha_i^{3/2}(1-\alpha_i^{1/2}) & \text{if } j \neq k \end{cases}$$

Proposition 3

For a given lag k, let $c_{k,i} = \sqrt{T-k}(\widehat{\alpha_i} - \alpha_i)$ and stack $c_{k,i}$ for different contours levels $i = 1, 2, ...C$. Let $\mathbf{C}_k = (c_{k,1}, ...c_{k,C})'$ be the $C \times 1$ stacked vector. Under the null hypothesis of $i.i.d\ U(0,1)$ PITs, the asymptotic distribution of the vector $\mathbf{C}_k$ is

multivariate normal, i.e., $\mathbf{C}_k \to N(0, \Omega_k)$ and consequently, the following quadratic form asymptotically follows a chi-square with C degrees of freedom, i.e.,

$$\mathbf{C}_k' \Omega_k^{-1} \mathbf{C}_k \to \chi_C^2$$

where a typical element of the asymptotic covariance matrix Ω_k, say $\omega_{i,j}$ is as follows

$$\omega_{i,j} = \begin{cases} \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2}) & \text{if } i = j \\ \alpha_i(1-\alpha_j) + 2\alpha_i\alpha_j^{1/2}(1-\alpha_j^{1/2}) & \text{if } i < j \\ \alpha_j(1-\alpha_i) + 2\alpha_j\alpha_i^{1/2}(1-\alpha_i^{1/2}) & \text{if } i > j \end{cases}$$

Statistics under parameter uncertainty: Univariate Case

Although we assume that the true parameters for distributions are known in previous sections, in reality, forecasting is done with estimated parameters and we will deal with $u_t(\widehat{\theta})$, which will bring uncertainty into our statistics. This part will discuss the asymptotic distributions of statistics if there exists parameter uncertainty. Rigorous mathematical proofs can show that the asymptotic normality for our statistics still holds under general assumptions.

Now, we are considering about location scale model: $y_t = \mu + \sigma\varepsilon_t$, ε_t is $i.i.d$

Assumption 1:

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) \to N(0, A^{-1}BA^{-1}) \tag{1.1}$$

$$A = -E(H(\theta_0)), H(\theta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_t(\theta_0) \text{ is the Hessian} \tag{1.2}$$

$$B = E(S(\theta_0)S(\theta_0)), S(\theta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_t s_t(\theta_0) \text{ is the score vector} \tag{1.3}$$

Assumption 2:

$$P = \lim_{T\to\infty} E(\frac{\partial\widehat{\alpha}_i(\theta)}{\partial\theta}|_{\theta=\theta_0}) = \lim_{T\to\infty} E(\frac{\partial(\widehat{\alpha}_i(\varepsilon_t(\theta)))}{\partial u_t}.\frac{\partial u_t}{\partial\varepsilon_t}.\frac{\partial\varepsilon_t}{\partial\theta}|_{\theta=\theta_0}) \tag{1.4}$$

P is bounded for all its elements.

Assumption 3:

$$cov(I_t^{k,\alpha_i}, s_t(\theta_0)) < \infty, cov(I_t^{k,\alpha_i}, s_{t-k}(\theta_0)) < \infty \tag{1.5}$$

Proposition 4: under assumption 1 to 3:

$$\sqrt{T}(\widehat{\alpha_i}(\widehat{\theta_T}) - \alpha_i) \to N(0, \xi_{k,i}^2) \tag{1.6}$$

$$\xi_{k,i}^2 = \sigma_{k,i}^2 + PA^{-1}BA^{-1}P + 2E[\sqrt{T}(\widehat{\alpha_i}(\theta_0) - \alpha_i)S^{'}(\theta_0)]A^{-1}P \tag{1.7}$$

The first assumption is a general assumption for consistency of QLME estimators. The second assumption ensures a bound gradient for the statistic, which can be relaxed. Assumption 3 states that the correlation is bounded. The proof of this proposition can follow the similar method used to prove proposition 4 in GSY's paper(2010).

Proposition 5: for Gaussian location –scale model: $y_t = \mu_0 + \sigma_0\varepsilon_t$,where $\varepsilon_t \sim i.i.d\ N(0,1)$

$$P = [\frac{2f(z_{\sqrt{\alpha_i}})\sqrt{\alpha_i}}{\sigma_0}, \frac{f(z_{\sqrt{\alpha_i}})z_{\sqrt{\alpha_i}}\sqrt{\alpha_i}}{\sigma_0^2}]^{'} \tag{1.8}$$

$$E[\sqrt{T}(\widehat{\alpha_i}(\theta_0) - \alpha_i)S^{'}(\theta_0)] = E\begin{bmatrix} \frac{I_t^{k.\alpha_i}\varepsilon_t}{\sigma_0} \\ \frac{I_t^{k.\alpha_i}(\varepsilon_t^2-1)}{\sigma_0^2} \end{bmatrix}^{'} + E\begin{bmatrix} \frac{I_t^{k.\alpha_i}\varepsilon_{t-k}}{\sigma_0} \\ \frac{I_t^{k.\alpha_i}(\varepsilon_{t-k}^2-1)}{\sigma_0^2} \end{bmatrix}^{'} \tag{1.9}$$

$$E[I_t^{k.\alpha_i}\varepsilon_t] = -\sqrt{\alpha_i}f(F^{-1}(\sqrt{\alpha_i})) \tag{1.10}$$

Similarly:

$$E[I_t^{k.\alpha_i}\varepsilon_{t-k}] = -\sqrt{\alpha_i}f(F^{-1}(\sqrt{\alpha_i})) \tag{1.11}$$

$$E[I_t^{k.\alpha_i}\varepsilon_t^2] = \int_{-\infty}^{F^{-1}(\sqrt{\alpha_i})} f(\varepsilon_{t-k})d\varepsilon_{t-k} \int_{-\infty}^{F^{-1}(\sqrt{\alpha_i})} \varepsilon_t^2 f(\varepsilon_t)d\varepsilon_{t.} \tag{1.12}$$

$$E[I_t^{k.\alpha_i}\varepsilon_{t-k}^2]] = \int_{-\infty}^{F^{-1}(\sqrt{\alpha_i})} f(\varepsilon_t)d\varepsilon_t \int_{-\infty}^{F^{-1}(\sqrt{\alpha_i})} \varepsilon_{t-k}^2 f(\varepsilon_{t-k})d\varepsilon_{t-k.} \tag{1.13}$$

Note: Here F denotes the CDF of standard normal distribution and f denotes the PDF of the standard normal distribution.

1.3 Monte Carlo Simulation

1.3.1 Size for Bivariate Normal and Student T Distributions

We evaluate the size properties for the G-ACR statistics under bivariate normal and student-t distributions. The sample size is 200,500,1000. The number of monte carlo simulation is 1000 and of bootstrapping is 500. 13 different contour coverage levels(%) considered are:

$C = [0.01; 0.05; 0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9; 0.95; 0.99]$ and the nominal size is 5

(1) Bivariate Normal: $y_t = \mu + \Sigma^{-1/2}\varepsilon_t$, $\varepsilon_t \sim i.i.d\ N(0,1)$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \ \sigma_1 = \sigma_2 = 1,\ \mu = 0.5,\ \rho = 0.4$$

(2) Bivariate Student-t Distributions $y_t = \mu + \Sigma^{-1/2}\varepsilon_t$, $\varepsilon_t \sim i.i.d\ t(0, I_2, v), v = 5$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$\mu = \begin{bmatrix} \mu_1 \\ \\ \mu_2 \end{bmatrix}$$

$\sigma_1 = \sigma_2 = 1$, $\mu = 0.5$,$\rho = 0.4$

	$t_{1,1}$	$t_{1,2}$	$t_{1,3}$	$t_{1,4}$	$t_{1,5}$	$t_{1,6}$	$t_{1,7}$	$t_{1,8}$	$t_{1,9}$	$t_{1,10}$	$t_{1,11}$	$t_{1,12}$	$t_{1,13}$
200	0.043	0.038	0.042	0.0627	0.063	0.053	0.053	0.077	0.088	0.055	0.06	0.065	0.055
500	0.047	0.053	0.052	0.063	0.072	0.053	0.06	0.067	0.07	0.056	0.062	0.05	0.037
1000	0.038	0.053	0.053	0.057	0.07	0.065	0.063	0.065	0.063	0.063	0.062	0.05	0.041
	$t_{2,7}$	$t_{3,7}$	$t_{4,7}$	$t_{5,7}$	$L_{2,7}$	$L_{3,7}$	$L_{4,7}$	$L_{5,7}$	$C_{1,13}$				
200	0.054	0.068	0.06	0.07	0.056	0.058	0.06	0.05	0.07				
500	0.037	0.065	0.055	0.066	0.065	0.051	0.053	0.048	0.062				
1000	0.051	0.062	0.062	0.063	0.041	0.052	0.058	0.069	0.061				

Notes: $t_{k,7}$ for $k = 1, 2, ...5$, and 7 refers to the 50% autocontour.
$L_{k,7}$ for $k = 2, ...5$ stacking lags up to k and considering the 50% autocontour.
$C_{1,13}$ stacking all 13 autocontours for lag $k = 1$.
1000 Monte Carlo replications and 500 bootstrap samples.

Table 1.1: Size for Bivariate Normal Distribution

	$t_{1,1}$	$t_{1,2}$	$t_{1,3}$	$t_{1,4}$	$t_{1,5}$	$t_{1,6}$	$t_{1,7}$	$t_{1,8}$	$t_{1,9}$	$t_{1,10}$	$t_{1,11}$	$t_{1,12}$	$t_{1,13}$
200	0.062	0.044	0.07	0.048	0.07	0.072	0.062	0.05	0.076	0.082	0.078	0.073	0.074
500	0.058	0.049	0.068	0.048	0.061	0.063	0.058	0.047	0.065	0.069	0.065	0.051	0.064
1000	0.058	0.062	0.062	0.051	0.061	0.068	0.045	0.038	0.067	0.062	0.061	0.055	0.069
	$t_{2,7}$	$t_{3,7}$	$t_{4,7}$	$t_{5,7}$	$L_{2,7}$	$L_{3,7}$	$L_{4,7}$	$L_{5,7}$	$C_{1,13}$				
200	0.065	0.065	0.063	0.068	0.073	0.068	0.063	0.076	0.073				
500	0.063	0.058	0.067	0.062	0.06	0.062	0.045	0.045	0.0583				
1000	0.053	0.052	0.068	0.051	0.068	0.059	0.062	0.058	0.061				

Notes: $t_{k,7}$ for $k = 1, 2, ...5$, and 7 refers to the 50% autocontour.
$L_{k,7}$ for $k = 2, ...5$ stacking lags up to k and considering the 50% autocontour.
$C_{1,13}$ stacking all 13 autocontours for lag $k = 1$.
1000 Monte Carlo replications and 500 bootstrap samples.

Table 1.2: Size for Bivariate Student t Distribution

1.3.2 Power of Statistics

Model under the null:

$y_t = \mu + \Sigma^{-1/2}\varepsilon_t$, $\varepsilon_t \sim i.i.d\ N(0,1)$, $\varepsilon_t \sim i.i.d\ N(0, I_2)$

(1) Power of the misspecification of correlation

The true DGP is the bivariate normal distribution with $\rho = 0.4$

	$t_{1,1}$	$t_{1,2}$	$t_{1,3}$	$t_{1,4}$	$t_{1,5}$	$t_{1,6}$	$t_{1,7}$	$t_{1,8}$	$t_{1,9}$	$t_{1,10}$	$t_{1,11}$	$t_{1,12}$	$t_{1,13}$
200	0.22	0.43	0.455	0.41	0.405	0.33	0.28	0.13	0.135	0.095	0.04	0.02	0.03
500	0.905	0.99	0.995	0.995	0.985	0.98	0.945	0.82	0.725	0.42	0.12	0.06	0.04
1000	0.93	1	1	1	1	0.982	0.96	0.89	0.73	0.43	0.18	0.061	0.053
	$L_{2,7}$	$L_{3,7}$	$L_{4,7}$	$L_{5,7}$	$C_{1,13}$	$C_{2,13}$	$C_{3,13}$	$C_{4,13}$	$C_{5,13}$				
200	0.36	0.33	0.31	0.45	0.43	0.475	0.52	0.465	0.455				
500	0.99	0.99	0.98	0.975	1	1	1	1	1				
1000	1	1	0.99	0.99	1	1	1	1	1				

Notes: $t_{k,7}$ for $k = 1$, and 7 refers to the 50% autocontour.
L_7^k for $k = 2, ...5$ stacking lags up to lag k and considering the 50% autocontour.
C_k^{13} stacking all 13 autocontours for lags $k = 1, ...5$.
1000 Monte Carlo replications and 500 bootstrap samples.

Table 1.3: Power for Misspecification of Correlation

(2) Power of the misspecification of density function

The true DGP is the bivariate student t distribution in which $\varepsilon_t \sim i.i.d\ t(0, I_2, 15)$

	$t_{1,1}$	$t_{1,2}$	$t_{1,3}$	$t_{1,4}$	$t_{1,5}$	$t_{1,6}$	$t_{1,7}$	$t_{1,8}$	$t_{1,9}$	$t_{1,10}$	$t_{1,11}$	$t_{1,12}$	$t_{1,13}$
200	0.01	0.033	0.0533	0.047	0.14	0.35	0.56	0.66	0.67	0.61	0.46	0.197	0.043
500	0.03	0.037	0.09	0.097	0.17	0.7	0.89	0.96	0.98	0.96	0.81	0.39	0.25
1000	0.073	0.103	0.157	0.12	0.25	0.93	1	1	1	0.997	0.99	0.71	0.37
	$L_{2,7}$	$L_{3,7}$	$L_{4,7}$	$L_{5,7}$	$C_{1,13}$	$C_{2,13}$	$C_{3,13}$	$C_{4,13}$	$C_{5,13}$				
200	0.39	0.44	0.4	0.49	0.54	0.55	0.57	0.58	0.54				
500	0.75	0.68	0.75	0.72	0.94	0.89	0.93	0.913	0.92				
1000	0.88	0.89	0.9	0.85	1	1	1	1	1				

Notes: $t_{k,7}$ for $k = 1$, and 7 refers to the 50% autocontour.
L_7^k for $k = 2, ...5$ stacking lags up to lag k and considering the 50% autocontour.
C_k^{13} stacking all 13 autocontours for lags $k = 1, ...5$.
1000 Monte Carlo replications and 500 bootstrap samples.

Table 1.4: Power for Misspecification of Density Functional Form

1.3.3 Evaluation of Multistep Density forecast

The optimal multi-step density forecast will produce serially correlated PITs : $u_t = \int_{-\infty}^{y_t} f(y_t|\Omega_{t-h})dy_t$ with uniform distribution. It is different from the PITs of realizations for the one-step ahead density forecast which is i.i.d $U[0, 1]$. If a h-step-ahead density forecast is correct, then its PITs of realizations at time t are independent of the PITs at time $t \pm (h + i)$, i.e., for 2-step ahead density forecast $\{u_1, u_3, u_5, \cdots\}$ and $\{u_2, u_4, u_6, \cdots\}$ are two random series from i.i.d $U[0, 1]$.

To show the MA$(h-1)$ structure of PITs of the h-step-ahead density forecast: $\forall Y_t$ is a stationary process with mean zero. The Wold representation of Y_t is $Y_t = \sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i}$, ε_t is $WN(0, \sigma^2)$. Under the null hypothesis that $f(y_t|\Omega_{t-h})$ is the correct h-step density forecast for Y_t, where $f(y_t|\Omega_{t-h}) = g(y_t - \sum_{i=h}^{\infty} \theta_i \varepsilon_{t-i}) = g(\xi_t)$, where $\xi_t = y_t - \sum_{i=h}^{\infty} \theta_i \varepsilon_{t-i} = \sum_{i=0}^{h-1} \theta_i \varepsilon_{t-i}$ is the error term. Then,

$$\begin{aligned} u_t &= \int_{-\infty}^{y_t} f(y_t|\Omega_{t-h}) dy_t \\ &= \int_{-\infty}^{y_t} g(y_t - \sum_{i=h}^{\infty} \theta_i \varepsilon_{t-i}) dy_t = \int_{-\infty}^{\xi_t} g(v) dv \\ &= \int_{-\infty}^{\sum_{i=0}^{h-1} \theta_i \varepsilon_{t-i}} g(v) dv \\ &= G(\sum_{i=0}^{h-1} \theta_i \varepsilon_{t-1}). \end{aligned}$$

From the above expression for u_t, we can find that u_t follows the MA$(h-1)$ structure. Therefore, since u_t series is $(h-1)$ dependent, each of the following all h sub-sequences of $\{u_t\}$ will be i.i.d $U[0,1]$ distributed, such as $\{u_1, u_{1+h}, u_{t+2h}, \cdots\}$, $\{u_2, u_{2+h}, u_{2+2h}, \cdots\}$,$\cdots$, $\{u_h, u_{2h}, u_{3h}, \cdots\}$. As the result of that, we can use Bonferroni method to test the null hypothesis if i.i.d. uniformity.

To be specific: suppose that we are interested in the h step ahead density forecast and according to the above arguments, we can create h sub samples from the original sample, each of which exhibits independence. The Bonferroni approach tells us that a test with bound of the size α can be evaluated by performing h tests with size α/h on each of the sub sample of $\{u_t\}$. If the null hypothesis of i.i.d uniformity is rejected by any of the sub sample, we should reject this null that the proposed h step ahead density forecast is correct.

Let E_i represent the event that the i-th sub-series does not reject the null of i.i.d. $U[0,1]$, $i = 1, 2, \cdots, h$. The null hypothesis that the proposed h step head density

forecast is correct is equivalent to the null the any of the sub series does not reject the null of i.i.d uniformity, which is the event $(\bigcap_{i=1}^{h} E_i)$ and thus the rejection of the joint hypothesis is $\overline{\bigcap_{i=1}^{h} E_i}$.

From the Boole's inequality that

$$P(\overline{i = 1^h E_i}) = P(\bigcup_{i=1}^{h} \overline{E_i}) \leq \sum_{i=1}^{h} P(\overline{E_i}) = h\alpha$$

where α represents the size of the test based on the i-th subsample. Then the test with size bounded by α is constructed to test the correctness of the proposed h step ahead density forecast.

We can use the simplest AR(1) model to illustrate the above argument. $y_t = \alpha y_{t-1} + \sigma\varepsilon_t, \varepsilon_t \sim N(0,1), \sigma = 2$

The 2-step ahead density forecast is easily calculated

$$f(y_{t+2}|\mathcal{F}_t) = \frac{1}{\sqrt{2\pi\sigma^2(1+\alpha^2)}} \exp\left(-\frac{(y_{t+2} - \alpha^2 y_t)^2}{2\sigma^2(1+\alpha^2)}\right)$$

and the PIT of the realizations based on this density function is

$$u_{t+2} = \int_{-\infty}^{y_{t+2}} f(y_{t+2}|\mathcal{F}_t) dy_{t+2}$$

. From Diebold, Gunther and Tay (1998), we knwo that the series $\{u_{t+2}\}_{t=1,\cdots,T}$ is uniformly distributed with serial correlation of MA$(h-1)$ structure. However, since $h = 2$, the following two sub-sequence $\{u_1, u_3, u_5, \cdots\}$ and $\{u_2, u_4, u_6, \cdots\}$ and i.i.d. uniformity.

The contour graphs for $\{u_t\}_{t=1}^{T}$ and the two sub-sequences: $\{u_1, u_3, u_5, \cdots\}$ and $\{u_2, u_4, u_6, \cdots\}$ can be used to demonstrate the above argument ($T = 2000$, the size for two sequences is 999).

Model two: $y_t = \alpha y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \exp(\lambda)$, $\lambda = 1$

The 2-step ahead density forecast with exponential error term can be calculated with the convolution method,

$$\begin{aligned} f(y_{t+2}|\mathcal{F}_t) &= \int_{-\infty}^{\infty} f(y_{t+2}|\mathcal{F}_{t+1}) f(y_{t+1}|\mathcal{F}_t) dy_{t+1} \\ &= \int_{\alpha y_t}^{\frac{y_{t+2}}{\alpha}} e^{-(y_{t+2}-\alpha y_{t+1})} e^{-(y_{t+1}-\alpha y_t)} dy_{t+1} \\ &= \frac{1}{1-\alpha}(e^{-(y_{t+2}-\alpha^2 y_t)} - e^{\frac{1}{\alpha}(y_{t+2}-\alpha^2 y_t)}) \end{aligned}$$

The PIT of the realization based on this density function is

$$u_{t+2} = \int_{-\infty}^{y_{t+2}} f(y_{t+2}|\mathcal{F}_t) dy_{t+2}$$

. Similarly, the $\{u_{t+2}\}_{t=1}^{T}$ is uniformly distributed with serial correlation of MA$(h-1)$ structure whereas the two seb-sequences $\{u_1, u_3, u_5, \cdots\}$ and $\{u_2, u_4, u_6, \cdots\}$ are i.i.d. uniformity. ($T = 2000$, the size for two sequences is 999).

Propose steps to evaluate the multi-step density forecast

Step 1): Based on the proposed DGP of Y_t, we can derive the h-step ahead density forecast recursively as follows,

$$\begin{aligned} f(y_{t+h}|\Omega_t) &= \int \cdots \int f(y_{t+h}, y_{t+h-1}, \cdots, y_{t+1}|\Omega_t) dy_{t+h-1} \cdots dy_{t+1} \\ &= \int \cdots \int f(y_{t+h}|\Omega_{t+h-1}) f(y_{t+h-1}|\Omega_{t+h-2}) \cdots f(y_{t+1}|\Omega_t) dy_{t+h-1} \cdots dy_{t+1} \end{aligned}$$

Step 2): Perform PIT of y_{t+h} over $f(y_{t+h}|\Omega_t)$ and obtain $\{u_{t+h}\}_{t=1}^{T}$

Step 3): Select all h sub-sequences of $\{u_{t+h}\}_{t=1}^{T}$:

$\{u_1, u_{1+h}, u_{t+2h}, \cdots\}$, $\{u_2, u_{2+h}, u_{2+2h}, \cdots\}$,$\cdots$, $\{u_h, u_{2h}, u_{3h}, \cdots\}$.

Apply our test statistics to all h subsequences. If all of these sequences are i.i.d. uniformity,the null that the proposed density is correct, and cannot be rejected while if any of these h sequences is not i.i.d. uniformity, the null will be rejected. Since we need to check all h subsequences jointly, we can use Bonferroni method which gives us a test

with the size bounded by α obtained by performing h tests, each of size α/h, on each of the h subsequences of $\{u_t\}$. However, when the analytical expression for the multi period predictive density is not available in the closed form for all distributions, We need to refer to some numerical techniques. We drao the univariate contour for normal and exponential distributions in Figure 1.2 and 1.3, which can graphically illustrate the idea for multistep density forecast with G-ACR methodology.

1.4 Conclusion

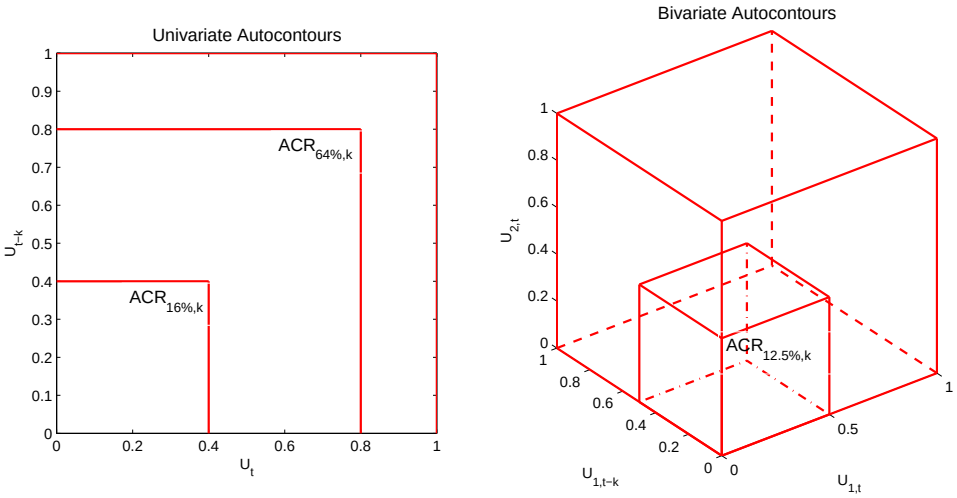
The Generalized Autocontour (G-ACR) is a generalized version of the *auto-contour* methodology proposed by González-Rivera *et al.* (2011) (GR2011) to detect misspecification in the dynamics of a time series model and departures from the assumed conditional density model. The G-ACR will overcome some important limitations of the original methodology in GR2011. First, when the conditional density of interest departs from standard densities in financial econometrics, e.g. Normal, Student-t, Exponential, etc., the analytical expressions of the autocontours may be mathematically cumbersome to obtain and we need to resort to numerical methods to compute their density mass. The difficulty is compounded when the system is multivariate. In contrast, the G-ACR is very easy to obtain for any density because it is based on the probability integral transforms (PIT) instead of standardized innovations, which are the basis of the original ACR. Second, GR2011 considers only continuous stochastic processes with dynamics restricted to the conditional mean and conditional variance, and a time-invariant functional form of the density of the standardized innovations of the model. The advantage of G-ACR is that it is applicable to continuous or discrete random processes, either uni-

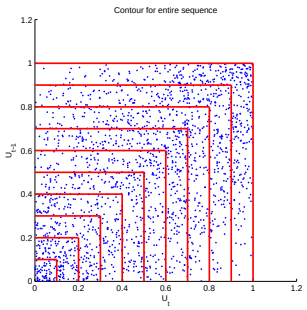
variate or multivariate. In a multivariate framework, the dimensionality of the system is not a constraint because the information contained in the vector of PITs is condensed into an indicator, which constitutes the basis of the proposed tests. Furthermore, the components of the multivariate system may have different marginal densities, which could be individually tested, but more importantly, the multivariate density, obtained as a copula function linking the marginals, can also be jointly tested. As a result, out statistics based on G-ACR are also useful diagnostics for correct copula specification. G-ACR does not restrict the dynamics of the model to any particular moment(s) and it is also applicable to cases when the predictive density does not have a closed form solution, e.g. a multistep predictive densities in nonlinear models, and we have to resort to simulation or nonparametric methods, but yet we could obtain the PIT process from the simulated density. Third, the tests proposed in GR2011 have asymptotic variance-covariance matrices that do not all enjoy closed-form solutions, some combining parametric and nonparametric expressions. In contrast and because of the simplicity of G-ACR, the asymptotic variances of the tests have all closed formulations that depend on only one parameter, the *a priori* specified probability level associated with the G-ACR.

As a brief introduction to G-ACR, which will be explained in detail in the forthcoming sections, suffices to say that the basis of our testing techniques is the construction of hyper-cubes of different sizes within the maximum hyper-cube formed by a multidimensional uniform density $[0,1]^n$. We assess the location of the empirical PITs (duplex, triplex, n-plex of observations) within the corresponding population hyper-cubes. If the multivariate model is correct, the volumes of the population hyper-cubes must be the same as those in their empirical counterparts. Our tests will evaluate these differences statistically to either reject or fail to reject the proposed density model. This

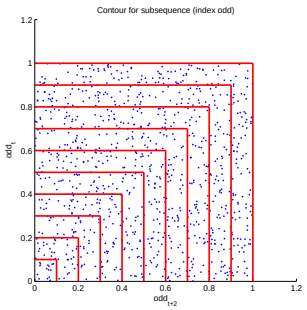
approach will also permit to focus on different areas of the conditional density model to assess those regions of interest. There is also a graphical visualization aspect of our approach that will be very helpful on guiding the modeling process.

Figure 1.1: Autocontours for the Uniform Density

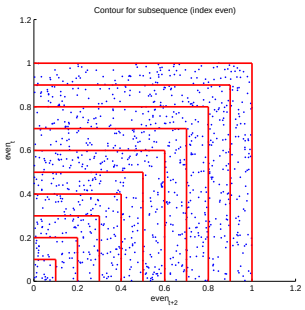




(a) Whole sample

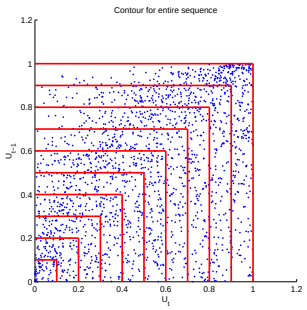


(b) Odd Subsample

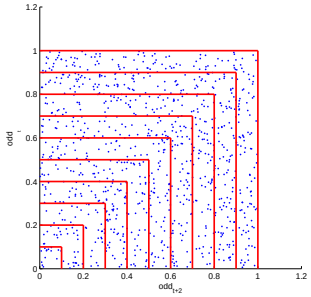


(c) Even Subsample

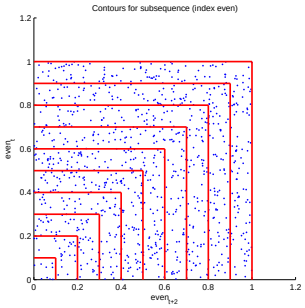
Figure 1.2: Univariate Contour Plots for Normal Distribution:Multistep



(a) Whole sample



(b) Odd Subsample



(c) Even Subsample

Figure 1.3: Univariate Contour Plots for Exponential Distribution:Multistep

Bibliography

[1] Bauwens, L., and Hautsch, N. (2006), "Stochastic Conditional Intensity Process", *Journal of Financial Econometrics*, vol. 4, 3, pp. 450-493.

[2] Cameron, C., and Trivedi, K. (1998). *Regression Analysis of Count Data*. Cambridge University Press, Cambridge.

[3] Cameron, C., and Trivedi, K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press, Cambridge.

[4] Diebold, F., Gunther, A., and Tay, A. (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management", *International Economic Review*, vol. 39, 4, pp. 863-883.

[5] Diebold, F., Hahn, J. and Tay, A. (1999), "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange", *The Review of Economics and Statistics*. vol. 81, 4, pp. 661-673.

[6] Denuit, M., and Lambert, P. (2005), "Constraints on Concordance Measures in Bivariate Discrete Data", *Journal of Multivariate Analysis*, vol. 93, pp. 40-57.

[7] Epps, T. (1979), "Comovements in Stock Prices in the Very Short Run ", *Journal of American Statistical Association*, vol. 74, 336, pp. 291-298.

[8] Easley, D., and O'Hara, M. (1992), "Time and the Process of Security Price Adjustment", *Journal of Finance*, vol. 27, 2, pp. 577-605.

[9] Engle, R., and Russell, R. (2007), "Autoregressive Conditional Duration: A New Model For Irregularly Spaced Transaction Data", *Econometrica*, vol. 66, 5, pp. 1127-1162.

[10] González-Rivera, G., Senyuz,Z., and Yoldas, E. (2011), "Autocontours: Dynamic Specification Testing", *Journal of Business and Economic Statistics*, vol. 29, 1, pp. 186-200.

[11] Hautsch, N. (2004), *Modelling Irregularly Spaced Financial Data*, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin.

[12] Heinen, A., and Rengifo, E. (2007), "Multivariate Autoregressive Modeling of Time Series Count Data Using Copulas", *Journal of Empirical Finance*, Vol. 14, pp. 564-583.

[13] Jung, R., Lisenfeld, R. and Richard, J. (2011), "Dynamic Factor Models for Multivariate Count data: An Application to Stock-Market Trading Activity", *Journal of Business and Economic Statistics*, vol. 29, 1, pp. 72-85.

[14] Marshall, A. (1996), "Copulas, Marginals, and Joint Distributions", *Lecture Notes-Monograph Series*. Vol. 28, 1, pp. 213-222.

[15] Madhavan, A. (2000), "Market Microstructure: A Survey", *Journal of Financial Markets*, vol. 3, pp. 205-258.

[16] Nelsen, R. (2005). *An Introduction to Copulas*. Second Edition, Springer.

[17] O'Hara, M. (1995). *Market Microstructure Theory*. Blackswell.

[18] Pacurar, M. (2008), "Autoregressive Conditional Duration Models in Finance: A Survey of The Theoretical and Empirical Literature", *Journal of Economic Surveys*, vol. 22, 4, pp. 711-751.

[19] Patton, A.J. (2006), "Estimation of Multivariate Models for Time Series of Possibly Different Lengths", *Journal of Applied Econometrics*, vol. 21, 4, pp. 147-173.

[20] Quoreshi, A.M.M.S. (2008), "A Vector Integer-Valued Moving Average Model for High Frequency Financial Count Data", *Economics Letters*, vol. 101, 3, pp. 258-261.

Chapter 2

Evaluation of Multivariate Counts Models. An Application to Trading Activity in U.S. Large Banks

2.1 Introduction

A point process $\{t_i\}_{i=1,2..}$ is a sequence of non-negative random variables associated with random arrival times $0 \leq t_i \leq t_{i+1}$. Intensity models, duration models, count models are alternative approaches to model a point process. From an information perspective, the three approaches are equivalent as a process with high intensity will have short durations and a high number of counts, and vice versa a low number of counts is associated with long durations and low intensity. In the financial econometrics literature, duration models have been very popular, and since the introduction of the

autoregressive conditional duration (ACD) model by Engle and Russell (1998), we have witnessed an explosion of alternative specifications to accommodate the characteristics of duration data. Extensive surveys are provided by Pacurar (2008) and Hautsch (2004). While univariate duration models have attracted much interest, research on multivariate models is much thinner. The difficulty arises from the nature of duration data. Given that durations are observations irregularly spaced in time, the components of a vector of durations will not be synchronized and it will be difficult to model their dependence as well as to incorporate information that arrives during the duration spells. For these reasons, multivariate modeling has either shifted to a continuous-time framework focusing on the dynamics of intensity functions, which are defined at any point in time (Bauwens and Hautsch, 2006) or, on keeping the discrete-time framework, the analysis has moved from durations to counts to facilitate the synchronization of multiple series, see Heinen and Rengifo (2007), Quoreshi (2008), and Jung, Lisenfeld and Richard (2011).

If dynamic multivariate specifications are few, specification testing and forecast evaluation techniques, either in-sample or out-of-sample, are fewer. In general, testing is limited to partial aspects of the model and there is not a global assessment of the multivariate system. Motivated by this state of affairs, we propose a new tool, the Generalized Autocontour (G-ACR), as the basis of in-sample or out-of-sample specification testing techniques, which will allow for the evaluation of the adequacy of the full multivariate system regardless of its dimension. Specifically, we will focus on multivariate counts models to avoid the asynchronization of information. The problem that we are facing is the evaluation of a vector of discrete stochastic processes, whose individual components could follow well-known discrete marginal density functions, such as Poisson, Negative Binomial, etc. but when analyzed jointly, we confront an open question regarding the specification of their multivariate density function. As an illustration of

the proposed G-ACR methodology, we will specify a multivariate counts model for the number of stock trades of three large U.S. banking institutions: Bank of America, JP Morgan Chase, and Wells Fargo. Dynamic trading is important because reflects arrival of news and it is intimately related to issues of liquidity risk and market microstructure, see O'Hara (1995) and Madhavan (2000) among others. We proceed by specifying an autoregressive system for the number of trades of each bank. We will entertain different distributional assumptions for the marginal densities of each component of the system but, most importantly, we are interested in the modeling of contemporaneous correlations of the trades as those may have implications for the risk that these large institutions pose to the banking system and beyond. We use a copula function to understand the contemporaneous correlation among the three banks. Heinen and Rengifo (2007) also implemented a copula approach but they restrict themselves to a normal copula where the dependence is contained in a correlation coefficient. As the recent financial crisis has shown, the correlation among institutions varies during episodes of low or high liquidity. We explore the possibility of asymmetric contemporaneous correlation such that the correlation may be different when the number of trades is large (the market is very active) or when the number is small (the market is slow). This modeling exercise will allow us to showcase the proposed specification testing methodology and, in particular, the use of visualization techniques to drive the specification process.

2.2 Extension of G-ACR to Discrete Stochastic Processes

2.2.1 Test Statistics for Discrete Multivariate Distribution Functions

In this section, we extend our methodology to assess the multivariate modeling of a vector of discrete random variables. For instance, in our empirical illustration we

are interested in modeling a vector of counts, in particular the number of trades in a day for several stocks. For each component of the vector, e.g. daily trades for each stock, we will assume a marginal density function, for instance, a Poisson distribution. In order to assess the multivariate model with the proposed generalized autocontour-based tests, we need the multivariate density of the vector process when its components are contemporaneously correlated. We will choose a copula function to tie up the assumed marginal densities of each component of the vector so that the multivariate distribution will be fully specified. However, on implementing this approach we encounter two problems. The first deals with the uniqueness of the copula function for discrete marginal densities (Marshall, 1996). The copula function is unique only for continuous densities. The second problem is that for the implementation of the autocontour-based tests we need uniformly distributed PITs associated with the conditional and marginal densities, but when the data is discrete the uniformity of the PITs does not hold. The solution to both problems is to find a continuous extension of the discrete random variables, which is achieved by adding an independent perturbation taking values in [0,1] to the discrete random variables (Denuit and Lambert, 2005).

Suppose that Y is a discrete random variable with PDF $f_y = \Pr(Y = y)$. We associate Y with a continuous random variable Y^*, such that $Y^* = Y + (U - 1)$, where U is a continuous random variable in (0,1), independent of Y, with strictly increasing CDF and no sharing any parameter with the CDF of Y. Assuming a uniformly distributed random variable U will satisfy all these conditions. Then, it is said that Y is *continued* by U, if for any $y \in N$, N being the set of non-negative integers, $F(y) = F^*(y)$ and $f_y = \int_{y-1}^{y} f^*(s)ds$. To see that this is the case, let $[s]$ be the integer part of $s \in \Re$. For U uniformly distributed and independent of Y, we can write $F^*(s) = \Pr(Y^* \leq$

$s) = \Pr(Y + (U-1) \leq s) = \sum_{y \leq [s]} \Pr(Y = y) + \Pr(Y = [s]+1) \times \Pr(U \leq s - [s]) =$
$= F([s]) + \Pr(Y = [s]+1) \times (s - [s])$, so that for any $y \in N$, it follows that $F(y) = F^*(y)$.

As we mention above, we specify the multivariate model of a vector of counts by choosing a copula function that ties up the marginal distribution of each component in the vector. Next, we explain how to implement the proposed generalized autocontour-based tests under Gaussian and Clayton copulas. In the empirical sections, we also entertain the Gumbel copula.

2.2.1.1 Multivariate Distribution of a Vector of Counts with Gaussian Copula

Let $(N_1, ...N_n)$ be a discrete random vector, with marginal distributions $F_i(N_i)$. We specify the joint distribution $H(N_{1,t}, ...N_{n,t})$ of the vector by choosing a copula function C such that $H(N_{1,t}, ...N_{n,t}) = C(F_1(N_{1,t}), ..., F_n(N_{n,t}))$. For a Gaussian copula, the multivariate distribution is

$$H(N_{1,t}, ...N_{n,t}) = \Phi^n(\Phi^{-1}(F_1(N_{1,t})), ..., \Phi^{-1}(F_n(N_{n,t})); \Sigma)$$

with Φ^n as the n-dimensional multivariate standard normal with correlation matrix Σ.

Let us call $q_{i,t} = \Phi^{-1}(F_i(N_{i,t}))$, where Φ is the univariate standard normal distribution. Then $H(N_{1,t}, ...N_{n,t}) = \Phi^n(q_{1,t}, ..., q_{n,t}; \Sigma)$ with corresponding density $f(q_{1,t}, ..., q_{n,t}) = \phi^n(q_{1,t}, ..., q_{n,t}; \Sigma)$, which is a multivariate standard normal density with correlation matrix Σ.

Suppose that we wish to evaluate a proposed predictive density $h_t^*(N_t|\Omega_{t-1})\}_{t=1}^T$, generated by a Gaussian copula, for the discrete random vector $(N_{1,t}, ...N_{n,t})$. To test the null hypothesis that $h_t^*(N_t|\Omega_{t-1})\}_{t=1}^T$ is the correct density model for $(N_{1,t}, ...N_{n,t})$

is equivalent to test that the correct density for $(q_{1,t}, ..., q_{n,t})$ is multivariate normal, i.e.,

$$H_0 : \{f_t^*(q_{1,t}, ..., q_{n,t}|\Omega_{t-1}) = \phi^n(q_{1,t}, ..., q_{n,t}|\Omega_{t-1}; \Sigma)\}_{t=1}^T$$

Therefore, we will proceed as follows,

1. Obtain the *continued* variable for each element of the random vector $(N_{1,t}, ...N_{n,t})$, i.e., $N_{i,t}^* = N_{i,t} + (U_i - 1)$, where U_i is a draw from the uniform U[0,1] density.

2. Obtain the *continued* PIT for each element of the random vector $(N_{1,t}, ...N_{n,t})$, i.e., $u_{1,t}^* = F_1^*(N_{1,t}^*), ..., u_{n,t}^* = F_n^*(N_{n,t}^*)$, where F_i^* is the marginal CDF of the *continued* random variable associated with $N_{i,t}$.

3. Obtain $q_{i,t}$ for each $N_{i,t}$ by transforming the *continued* PITs as $q_{i,t} = \Phi^{-1}(F_i^*(N_{i,t}^*))$ where Φ is the univariate standard normal distribution.

4. Factorize the multivariate density of the random vector $Q_t = (q_{1,t}, ..., q_{n,t})$ as the product of conditional and marginal densities as

$$f^*(q_{1,t}, q_{2,t,}.....q_{n,t}|\Omega_{t-1}) = f^*(q_{n,t}|q_{1,t}, q_{2,t,}.....q_{n-1,t}) \times .... \times f^*(q_{2,t}|q_{1,t}) \times f^*(q_{1,t}|\Omega_{t-1})$$

For a Gaussian copula, $f_t^*(q_{1,t}, ..., q_{n,t}|\Omega_{t-1}) = \phi^n(q_{1,t}, ..., q_{n,t}|\Omega_{t-1}; \Sigma)$, the marginal and conditional densities are also normal. Thus, we can now easily obtain the corresponding PITs, i.e., $u_{1,t} = \int_{-\infty}^{q_{1,t}} f_{q_{1,t}}^*(v_t|\Omega_{t-1})dv_t$, ..., $u_{m|1,2...m-1,t} = \int_{-\infty}^{q_{m,t}} f_{q_m|q_1,...q_{m-1},t}^*(v_t|\Omega_{t-1})dv_t$, ..., which under the null hypothesis of a correctly specified density model, must be i.i.d. U[0,1] random variables.

2.2.1.2 Multivariate Distribution of a Vector of Counts with Clayton Copula

For the Clayton copula, the multivariate distribution of the vector $(N_1, ...N_n)$ is given by

$$\begin{aligned} H(N_{1,t}, ...N_{n,t}) &= C(F_1(N_{1,t}), ..., F_n(N_{n,t})) = (1 - n + \sum_{i=1}^{n} (F_i(N_{i,t}))^{-\theta})^{-1/\theta} \\ &= (1 - n + \sum_{i=1}^{n} (q_{i,t})^{-\theta})^{-1/\theta} \end{aligned}$$

where $q_{i,t} = F_i(N_{i,t})$, and the corresponding density is given by

$$c(q_{1,t}, ...q_{n,t}) = (1 - n + \sum_{i=1}^{n} q_{i,t}^{-\theta})^{-n-\frac{1}{\theta}} \prod_{i=1}^{n} (q_{i,t}^{-\theta-1}\{(i-1)\theta + 1\})$$

To obtain the PITs under the null hypothesis of a correct model under Clayton, proceed as follows

1. Obtain the *continued* variable for each element of the random vector $(N_{1,t}, ...N_{n,t})$, i.e., $N^*_{i,t} = N_{i,t} + (U_i - 1)$, where U_i is a draw from the uniform U[0,1] density.

2. Obtain the *continued* PIT for each element of the random vector $(N_{1,t}, ...N_{n,t})$, i.e., $q_{1,t} = F^*_1(N^*_{1,t}), ..., q_{n,t} = F^*_n(N^*_{n,t})$, where F^*_i is the marginal CDF of the *continued* random variable associated with $N_{i,t}$.

3. Factorize the multivariate density of the random vector $Q_t = (q_{1,t}, ..., q_{n,t})$ as the product of conditional and marginal densities as

$$c^*(q_{1,t}, q_{2,t,}.....q_{n,t}|\Omega_{t-1}) = c^*(q_{n,t}|q_{1,t}, q_{2,t,}.....q_{n-1,t}) \times .... \times c^*(q_{2,t}|q_{1,t}) \times c^*(q_{1,t}|\Omega_{t-1})$$

In order to calculate the conditional densities, we exploit the property that all

n-marginal distributions of a Clayton copula are identical, i.e.,

$$C(F_1(N_{1.t}), ..., F_n(N_{n-1,t}), 1) = (1 - (n-1) + \sum_{i=1}^{n-1}(q_{i,t})^{-\theta})^{-1/\theta}.$$

With the marginal and conditional distributions in place, we obtain the PITs as previously described, i.e.,

$$u_{1,t} = \int_{-\infty}^{q_{1,t}} c^*_{q_{1,t}}(v_t|\Omega_{t-1})dv_t \cdots du_{m|1,2...m-1,t} = \int_{-\infty}^{q_{m,t}} c^*_{q_m|q_1 \cdots q_{m-1},t}(v_t|\Omega_{t-1})dv_t,$$

and proceed with the implementation of the autocontour-based tests.

2.3 Monte Carlo Simulations

In this section we perform extensive Monte Carlo simulations to assess the finite sample properties of the tests proposed in Propositions 1 to 3 when the data is discrete and our interest is to model a vector of counts $N_t = (N_{1,t}, N_{2,t})'$ under different multivariate distributions functions.

2.3.1 Size of the tests

We consider five models for which the conditional mean of counts, $\mu_{i,t} \equiv E[N_{i,t}|\Omega_{t-1}]$ for $i = 1, 2$, obeys dynamics of order one, i.e.,

$$E(N_t|\Omega_{t-1}) = W + A \times N_{t-1} + B \times \mu_{t-1} \tag{2.1}$$

where Ω_{t-1} is the information set, and W is a vector and A and B are matrices of parameters with the following values: $W = \begin{pmatrix} 5 \\ \\ 5 \end{pmatrix}$, $A = \begin{pmatrix} 0.4 & 0 \\ \\ 0.15 & 0.45 \end{pmatrix}$, $B = \begin{pmatrix} 0.5 & 0 \\ \\ 0 & 0.45 \end{pmatrix}$

The difference among the following five models lies on the assumed marginal densities and whether the elements of the vector are contemporaneously correlated or not. We entertain the following density specifications:

Model S1: Bivariate Conditional Autoregressive Poisson (P) Model

$$\begin{aligned} N_{i,t}|\Omega_{t-1} &\sim P(\mu_{i,t}), \text{ for } i = 1, 2 \\ f(n_{t,i}|\Omega_{t-1}) &= \frac{e^{-\mu_{i,t}}\mu_{i,t}^{n_{t,i}}}{n_{t,i}!} \\ E[N_{i,t}|\Omega_{t-1}] &= \mu_{i,t} \\ Var[N_{i,t}|\Omega_{t-1}] &= \mu_{i,t} \end{aligned}$$

In empirical financial applications, we observe overdispersion in the data very frequently. A Poisson model does not allow for overdispersion because the mean and the variance are the same. For this reason, we consider next Negative Binomial marginal densities.

Model S2: Bivariate Conditional Autoregressive Negative Binomial (NB) Model ($\kappa_i = 4$)

$$\begin{aligned} N_{i,t}|\Omega_{t-1} &\sim NB(\mu_{i,t}), \text{ for } i = 1, 2 \\ f(n_{t,i}|\Omega_{t-1}) &= \frac{\Gamma(n_{t,i} + \kappa_i)}{\Gamma(\kappa_i)\Gamma(n_{t,i} + 1)} \left(\frac{\kappa_i}{\kappa_i + \mu_{i,t}}\right)^{\kappa_i} \left(\frac{\mu_{i,t}}{\mu_{i,t} + \kappa_i}\right)^{n_{t,i}} \\ E[N_{i,t}|\Omega_{t-1}] &= \mu_{i,t} \\ Var[N_{i,t}|\Omega_{t-1}] &= \mu_{i,t}(1 + \frac{\mu_{i,t}}{\kappa_i}) \end{aligned}$$

Because $\kappa_i > 0$ and $\mu_{i,t} > 0$, the conditional variance is greater than the conditional mean, thus the model generates overdispersion in the data (Cameron and Trivedi, 1986). When $1/\kappa_i \to 0$, the Negative Binomial converges to the Poisson density.

The three following models allow for contemporaneous dependence among the elements of the vector. The dependence is modeled by a copula function. We consider three one-parameter Archimedean copulas (Nelsen, 2005): a Normal copula for which the dependence is summarized by the correlation coefficient, and the Clayton and Gumbel copulas that allow for asymmetric dependence among the elements of the vector.

Model S3: Bivariate Conditional Autoregressive Negative Binomial with Gaussian Copula Model (correlation coefficient ρ=0.8).

This model has the same marginal distributions as in Model S2 but the contemporaneous cross-correlation between the vector components is generated by using the Gaussian copula, which functional form is described in section 2.3.1.

Model S4: Bivariate Conditional Autoregressive Negative Binomial with Clayton Copula Model (copula parameter $\theta = 2$).

This model has the same marginal distributions as in Model S2 but the contemporaneous cross-correlation between the vector components is generated by using the Clayton copula, which allows for stronger correlation at low values than at high values of the data. The functional form of the Clayton copula is described in section 2.3.2.

Model S5: Bivariate Conditional Autoregressive Negative Binomial with Gumbel Copula Model (copula parameter $\theta = 2$).

This model has the same marginal distributions as in Model S2 but the contemporaneous cross-correlation between the vector components is generated by the Gumbel copula, which allows for stronger correlation at high values than at low values of the data. Gumbel functional form is

$$C(u_1, u_2, ..., u_n) = \exp(-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta + ... + (-\ln u_n)^\theta\right]^{1/\theta}), \theta > 0$$

We have estimated the five models S1 to S5 by maximum likelihood. The log-likelihood function is constructed according to the distributional assumptions specified in each model. For models with a copula function we follow a two-stage estimation procedure as in Patton (2006) by first estimating the parameters in the marginal model, and secondly using these estimates to estimate the copula parameter. Since parameter

uncertainty is more important in-sample testing, we conduct the experiments in sample. We obtain the PITs associated with the one-step-ahead conditional expectation of the count vector following the procedures explained in section 2.3, and proceed to implement the tests described in Propositions 1 to 3. We have implemented a parametric bootstrap procedure to approximate the asymptotic variance of the tests. [1] We obtain bootstrap samples for each model by replacing the true value θ_0 with the the estimate $\hat{\theta}_T$, computing the conditional mean $\mu_{i,t}$, and making draws $N_{i,t}$ from the specified parametric distributions. This is a standard procedure to overcome the difficulties of estimating asymptotic variances when parameter uncertainty is relevant. The following experiments will show that bootstrapping the variance of the tests and using standard asymptotic critical values provides statistics with the right size.

In Tables 1 and 2, we show the size of the tests for different sample sizes $T = 250, 500$ and 1000 observations. The overall performance of the tests is very good. Across models and across sample sizes, the average empirical size is 5%. We do not observe any instance in which the tests are grossly over- or undersized. In Table 1, we also include the size of the test for model S1 without bootstrapping the variance of the test. As we expected, the size is very distorted and the tests are all uniformly undersized so that the tests do not reject the null hypothesis as much as they should. These results support the practice of implementing a bootstrap procedure when parameter uncertainty is a concern.

[1] When testing in-sample specification, ignoring parameter uncertainty may cause severe distortions in the size of the tests. When testing out-of-sample specification, the importance of parameter uncertainty will depend on the forecasting scheme and on the size of the estimation sample (R) relative to the prediction sample (P). Under the assumption of $\sqrt{R}$-consistent estimators, if $R \to \infty$, $P \to \infty$, and $P/R \to 0$ as $T \to \infty$, parameter uncertainty is asymptotically negligible and no adjustment is needed in the tests provided in Propositions 1 to 3.

Table 2.1: **Size of t_{k,α_i}-statistics for 13 autocontours (Nominal size 5% and $k = 1$)**

	$t_{1,1}$	$t_{1,2}$	$t_{1,3}$	$t_{1,4}$	$t_{1,5}$	$t_{1,6}$	$t_{1,7}$	$t_{1,8}$	$t_{1,9}$	$t_{1,10}$	$t_{1,11}$	$t_{1,12}$	$t_{1,13}$
T	Model S1												
250	0.051	0.044	0.066	0.054	0.061	0.062	0.046	0.044	0.045	0.045	0.046	0.056	0.050
500	0.058	0.042	0.064	0.058	0.052	0.057	0.052	0.048	0.042	0.048	0.045	0.050	0.046
1000	0.045	0.053	0.047	0.044	0.054	0.052	0.054	0.056	0.057	0.056	0.054	0.050	0.045
T	Model S1 (no bootstrap variance)												
250	0.037	0.015	0.016	0.005	0.002	0.004	0.008	0.014	0.012	0.013	0.021	0.030	0.028
500	0.040	0.024	0.018	0.008	0.004	0.008	0.009	0.016	0.013	0.015	0.030	0.033	0.034
1000	0.032	0.031	0.017	0.011	0.006	0.006	0.007	0.012	0.017	0.018	0.027	0.041	0.034
T	Model S2												
250	0.045	0.038	0.039	0.042	0.042	0.046	0.043	0.04	0.052	0.048	0.052	0.047	0.053
500	0.055	0.039	0.051	0.052	0.054	0.055	0.051	0.057	0.05	0.045	0.043	0.052	0.047
1000	0.062	0.051	0.048	0.054	0.045	0.044	0.042	0.04	0.058	0.049	0.048	0.056	0.052
T	Model S3												
250	0.053	0.043	0.047	0.051	0.050	0.049	0.060	0.046	0.046	0.056	0.052	0.058	0.043
500	0.048	0.041	0.048	0.062	0.065	0.059	0.053	0.065	0.056	0.043	0.043	0.043	0.041
1000	0.05	0.051	0.042	0.047	0.055	0.054	0.050	0.048	0.062	0.067	0.053	0.048	0.048
T	Model S4												
250	0.034	0.047	0.047	0.049	0.050	0.050	0.060	0.064	0.057	0.049	0.049	0.038	0.060
500	0.036	0.049	0.05	0.047	0.055	0.055	0.063	0.057	0.058	0.048	0.052	0.044	0.049
1000	0.047	0.056	0.046	0.047	0.054	0.052	0.053	0.05	0.048	0.055	0.045	0.046	0.042
T	Model S5												
250	0.049	0.043	0.042	0.047	0.041	0.042	0.043	0.039	0.051	0.043	0.036	0.056	0.043
500	0.067	0.033	0.051	0.053	0.058	0.053	0.047	0.049	0.044	0.039	0.045	0.043	0.053
1000	0.052	0.062	0.049	0.062	0.060	0.056	0.052	0.048	0.053	0.051	0.041	0.050	0.046

Notes: The 13 autocontours are $C = [0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8.0.9, 0.95, 0.99]$.
1000 Monte Carlo replications and 500 bootstrap samples.

Table 2.2: **Size of t_{k,α_i} , L_{α_i}, C_k statistics (Nominal size 5%)**

	$t_{1,7}$	$t_{2,7}$	$t_{3,7}$	$t_{4,7}$	$t_{5,7}$	$L_{2,7}$	$L_{3,7}$	$L_{4,7}$	$L_{5,7}$	$C_{1,13}$
T	Model S1									
250	0.046	0.049	0.046	0.043	0.051	0.053	0.047	0.062	0.061	0.050
500	0.052	0.046	0.048	0.053	0.056	0.056	0.055	0.055	0.052	0.045
1000	0.054	0.05	0.048	0.055	0.055	0.047	0.052	0.054	0.056	0.058
T	Model S2									
250	0.043	0.036	0.045	0.04	0.039	0.037	0.044	0.048	0.052	0.047
500	0.051	0.057	0.064	0.058	0.044	0.057	0.055	0.043	0.053	0.054
1000	0.042	0.047	0.047	0.04	0.045	0.043	0.051	0.053	0.052	0.049
T	Model S3									
250	0.060	0.046	0.043	0.049	0.045	0.055	0.051	0.057	0.055	0.055
500	0.053	0.055	0.054	0.061	0.056	0.058	0.065	0.06	0.058	0.040
1000	0.05	0.052	0.048	0.053	0.054	0.044	0.048	0.058	0.053	0.054
T	Model S4									
250	0.06	0.06	0.059	0.063	0.066	0.069	0.057	0.065	0.059	0.064
500	0.063	0.053	0.057	0.062	0.065	0.067	0.047	0.060	0.052	0.063
1000	0.053	0.045	0.049	0.043	0.041	0.053	0.052	0.056	0.054	0.053
T	Model S5									
250	0.043	0.039	0.04	0.037	0.04	0.04	0.06	0.053	0.049	0.066
500	0.047	0.048	0.051	0.047	0.042	0.041	0.045	0.049	0.052	0.042
1000	0.052	0.046	0.047	0.058	0.049	0.055	0.052	0.048	0.052	0.056

Notes: $t_{k,7}$ for $k = 1, 2, ...5$, and 7 refers to the 50% autocontour.
$L_{k,7}$ for $k = 2, ...5$ stacking lags up to k and considering the 50% autocontour.
$C_{1,13}$ stacking all 13 autocontours for lag $k = 1$.
1000 Monte Carlo replications and 500 bootstrap samples.

2.3.2 Power of the tests

To study the power properties of the tests we consider as a null hypothesis Model S3: a bivariate autoregressive model of order one as in (2.1) for the vector of conditional means, with negative binomial marginal densities ($\kappa_i = 4$) and a normal copula function with contemporaneous correlation $\rho = 0.8$. We consider four data generating mechanisms:

Model P1: Conditional Autoregressive Poisson Model with Normal Copula ($\rho = 0.8$). In this case, we maintain the same dynamic structure and the same copula characteristics and we study departures from the hypothesized marginal densities.

Model P2: Conditional Autoregressive Negative Binomial Model with a Clayton copula ($\theta = 2$). We maintain the same dynamics and marginal densities and study departures from the hypothesized copula, and in particular, detecting asymmetric contemporaneous dependence.

Model P3: Conditional Autoregressive Negative Binomial Model with a Gumbel copula ($\theta = 2$). This case is similar to the previous one but the contemporaneous asymmetric dependence in Gumbel runs in opposite direction to that of Clayton.

Model P4: Conditional Autoregressive Negative Model with Normal Copula ($\rho = 0.8$) with high order dynamics in the conditional means, i.e.,

$$E(N_t|\Omega_{t-1}) = D + A_1 \times N_{t-1} + A_2 \times N_{t-2} + A_3 \times N_{t-3} + A_4 \times N_{t-4} + B \times \mu_{t-1}$$

where $D = \begin{bmatrix} 0.8 \\ 0.7 \end{bmatrix}$, $A_1 = \begin{bmatrix} 0.3 & 0 \\ 0.25 & 0.28 \end{bmatrix}$, $A_2 = \begin{bmatrix} 0.09 & 0 \\ 0.05 & 0.06 \end{bmatrix}$, $A_3 = \begin{bmatrix} 0.005 & 0 \\ 0.003 & 0.008 \end{bmatrix}$, $A_4 = \begin{bmatrix} 0.003 & 0 \\ 0.001 & 0.002 \end{bmatrix}$, and $B = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.35 \end{bmatrix}$. We study departures from the hypothesized dynamics and maintain the distributional assumptions on the marginal densities

and the copula.

We present the power results in Tables 3 and 4. In Table 3, we analyze the behavior of the t-statistics, t_{k,α_i} for a fixed lag $k = 1$ and 13 autocontours that span the entire uniform density. The Poisson case, model P1, is very easy to detect because of the property of overdispersion generated by the Negative Binomial densities. The power is one regardless of sample size not only for the t-statistics but also for the portmanteau tests L_{α_i} and C_k reported in Table 4. In general, the tests are more powerful at detecting departures from distributional assumptions than at detecting misspecified dynamics. In models P2 and P3, we find very high power even with small sample sizes. In model P4 we need a large sample (above 500 observations) to find power above 50%. At the lowest autocontours, i.e. 1, 5, or 10 %, and the highest autocontour 99%, and mainly for small samples ($T = 250$), the power is lower because there are only a few (or a lot) observations within each autocontour and, consequently there is not much variability in the indicator function, which is at the core of the tests. The power for model P2 (Clayton) is lower than that for model P3 (Gumbel) for the intermediate autocontours around the 10- 40% level, which is expected as the correlation of the observations in this range for a Clayton copula is closer to that of the hypothesized Normal than the correlation of those observations generated by Gumbel copula. The opposite happens when we examine the high autocontours 50-99% precisely because of the opposite reason, the Gumbel correlation for observations in the upper contours is closer to the Normal correlation than that generated by the Clayton.

In Table 4, we present the power of the t-statistics, $t_{k,7}$, and of the portmanteau statistics L_7^k and C_k^{13} for several values of k. In general, the power is excellent across models and sample sizes. The overall findings are similar to those from Table

3. Examining the t-statistics for different lags helps to detect dynamic misspecification. In model P4 the highest power happens for $t_{1,7}$ and for $k > 1$, the power stabilizes indicating that lag 1 is more problematic than the rest. The tests C_k^{13} are obviously the most powerful as they collect information for all the autocontours. However, if the researcher is interested in partial aspects of the densities, such as a quantile or a collection of quantiles, it would be more informative examining the $L_{\alpha_i}^k$ statistics, where we fix the level of the quantile α_i, and the individual t-statistics that provide information about the quantile desired.

2.4 Empirical Illustration: Trading Activity in Large Banks

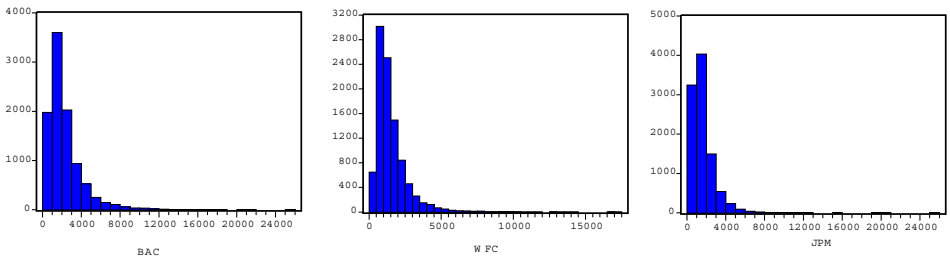
In this section, we will iilustrate how to use the proposed testing methodology and, in particular, the use of visualization techniques to drive the specification process. We estimate a multivariate counts model for the trades of three large U.S. banking institutions: Bank of America, JP Morgan Chase, and Wells Fargo. We are interested in exploring models that produce asymmetric contemporaneous correlation among institutions in times of intense versus low trading activity.

2.4.1 Data Description

We collect transaction data from the TAQ database for three U.S. large commercial banks: Bank of America (BOA), JP Morgan Chase (JPM), and Wells Fargo (WF), trading in the New York Stock Exchange, from January 3rd to June 30th, 2011 for a total of 125 trading days. We record the number of trades at the 5-minute frequency and remove any trades before 9:30 am (opening time) and after 4:00 pm (closing time) for a total of 9,750 observations per bank. In Figure 2, we plot the histograms

of the number of trades; they show evidence for overdispersion as there is a large tail to the right with most of the observations concentrated around the mean. The range in the number of trades is very wide, from minima in the 100s trades to maxima in the 20,000s.

Figure 2.1: Histograms of the data



It is very common to find intra-day seasonality in high frequency data. Trading activity is intense at the the beginning and towards the end of the trading day and substantially lighter during the mid-day hours, which gives rise to a U-shaped curve of diurnal effects. We take care of these effects by estimating a set of half-hour dummies.

2.4.2 Dynamic Models and Evaluation

We estimate several dynamic models for the conditional mean $\mu_{i,t} \equiv E[N_{i,t}|\Omega_{t-1}]$ for $i = 1(BAC),\ 2(WFC), 3(JPM)$ under different distributional assumptions. To illustrate how our methodology works, we will present our results in a sequential fashion. First, we will show the results of the Poisson model and the Negative Binomial Model ignoring contemporaneous correlation. Though *a priori* we know that the Poisson model will not fit our data because of overdispersion, it is interesting to contrast the estimation results with those of a Negative Binomial model and observe the behavior of the corresponding PITs and the generalized autocontour-based specification tests. Secondly, we

will introduce contemporaneous correlation by estimating the dependence parameters under Gaussian, Clayton, and Gumbel copulas, and we will show the specification improvements by analyzing the autocontour-based specification tests and the responses of the corresponding PITs.

2.4.2.1 Models with no contemporaneous correlation

In Table 5, we present maximum likelihood estimates for the parameters of the best dynamic model for the conditional mean of trade counts with Poisson and Negative Binomial marginal densities and without contemporaneous correlation. We have experimented with different lag structures and, through standard specification tests on the Pearson residuals, we finally settled in a low order model such as $\mu_t \equiv E(N_t|\Omega_{t-1}) = w + A \times N_{t-1} + B \times \mu_{t-1}$ where N_t is a 3×1 vector of trade counts, A is a 3×3 matrix, not necessarily symmetric, with typical element $\{a_{i,j}\}$, and B is 3×3 diagonal matrix with typical element $\{b_{i,i}\}$.

For each equation in the system, the trading dynamics for each bank are mainly driven by the most recent activity of the bank itself as we can see in the magnitude of the estimates of $a_{i,i}$ and $b_{i,i}$. The effect of trading in the other institutions, i.e. estimates of $a_{i,j}$, are much smaller in magnitude although statistically significant at the conventional levels. The overall system is stationary as the eigenvalues of the matrix $A+B$ are inside the unit circle. The overdispersion parameter $1/\kappa_i$ is evidently different from zero, as we expected. The last panel of Table 6 provides some descriptive statistics of the Pearson residuals, i.e. $(N_{i,t} - \mu_{i,t})/\sigma_{i,t}$. If the model is well specified, the residuals should have mean zero and variance one. This is not the case for the Poisson model because of the overdispersion in the data. The Q-statistics are two orders of magnitude smaller than the raw Q-statistics, indicating that the estimated dynamics are adequate. Some

Q-statistics are statistically significant but the residual autocorrelation that they are picking up is extremely low, for instance correlations as low as 0.007, 0.02, 0.05, etc., so that such magnitude is not economically meaningful. As expected, we find enough evidence to reject the Poisson model in favor of the Negative Binomial.

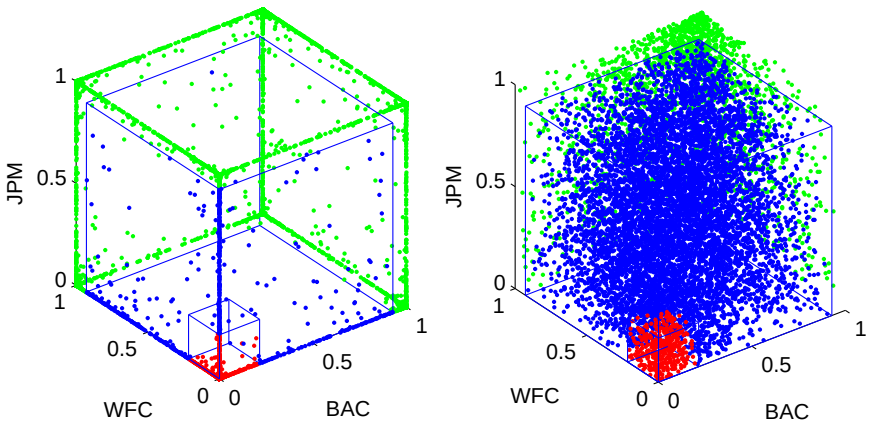
We proceed to implement the generalized autocontour-based specification tests proposed in section 2 for the Negative Binomial model with no contemporaneous correlation. In Table 6, we present the values of the t-tests for 5 lags and 13 autocontours, the L_{α_i} tests stacking up to 2, 3, 4, and 5 lags for a given autocontour α_i, and the C_k tests stacking the 13 autocontours for a given lag $k = 1, \cdots, 5$. The t-tests reject very strongly the model at the lower and middle (1 to 60%) and upper (95 and 99%) autocontour levels. At the lower levels, the test indicate that the number of observations within the autocontours is much larger than what we should expect given the assumed dynamics and contemporaneously independent Negative Binomial marginals. At the upper levels, the opposite happens, there are fewer observations than expected. However, since the t-tests show similar values across lags, the rejection of the model is not due to misspecified dynamics but to the assumed distributional assumptions. The umbrella tests L_{α_i} and C_k, which aggregate over lags or over contours, strongly confirm the message delivered by the t-tests.

The graphs in Figure 3 reveal why the rejection of the Negative Binomial model is so strong. We plot the 3-dimensional vector of contemporaneous continued PITs $\{u_{t,i}\}$ for the Poisson model (left panel) and for the Negative Binomial model (right panel) for three autocontours, the 10% level (red dots), the 10-80% levels (blue dots), and the 80-100% levels (green dots). The rejection of Poisson is very evident because most of the points are distributed along the boundaries of the contour levels, whose location is

very far from where we expect the i.i.d. uniform observations to fall. The unmodeled overdispersion drives the location of the observations towards the boundaries of the cube. The Negative Binomial model corrects this problem as the observations are more evenly distributed within the cubes associated with the levels considered. Nevertheless, we observe an elliptic concentration of the observations towards the diagonals of the cubes, which is the reason why the autocontour-based tests reject this model and, more importantly, it points out the need to model the contemporaneous dependence among the components of the vector.

Figure 2.2: Contemporaneous PITs:

Poisson model (left), Negative Binomial with No Contemporaneous Correlation (right)



2.4.2.2 Models with contemporaneous correlation

We model the contemporaneous dependence of trade counts by using copula functions, which offer great flexibility to account for the dependence structure among

several series regardless of the specification of the their marginal distributions. We have selected three copula functions: Gaussian, Gumbel, and Clayton. These functions summarize the dependence of the data with one parameter. The Gaussian copula characterizes dependence through the correlation coefficients and, as such, they provide a global measure of dependence for the entire collection of observations. On the contrary, the dependence parameters in the Gumbel and Clayton copulas allow for stronger or weaker dependence in different regions of the observations. Clayton (Gumbel) produces stronger (weaker) dependence for low values and weaker (stronger) dependence for high values of the observations. We also consider the inverse Gumbel copula, which by inverting its shape, delivers an asymmetric dependence similar to that from Clayton. [2]

The maximum likelihood estimates of the correlation matrix for the Gaussian copula are presented in Table 7. The pairwise contemporaneous correlation is significant and it runs between 0.40 and 0.50. The maximum likelihood estimates of the dependence parameter for the inverse Gumbel copula is $\theta = 1.37$ with a t-ratio of 163.70, and for the Clayton copula $\theta = 0.65$ with a t-ratio of 47.18. These values suggest very strongly that the dependence is asymmetric.

In Table 8, we present the results of the autocontour-based t-tests for the Negative Binomial model with contemporaneous correlation modeled by Gaussian, inverse Gumbel, Clayton, and a mixture of Gaussian/Clayton copulas. By comparing these results with those in Table 6, we observe the substantial overall reduction in the values of the t-tests, which indicates the need to model the contemporaneous dependence among the series. Focusing on the results from the Gaussian copula, the tests fail to reject the model in the lowest (1% and 5%) and central autocontours (40% to 70% levels). At

[2] We are grateful to Andrew Patton for suggesting the inverse Gumbel copula to us. If $\{u_t\}$ are the PITs from Gumbel, $\{1 - u_t\}$ will be the corresponding PITs in the inverse Gumbel copula.

the upper autocontours 80% to 99%, however, there is a strong rejection as the model imposes more dependence between large counts than what is granted in the data. For the upper autocontours, the Clayton copula is more responsive to the needs of the data and the values of the tests are significantly lower than those in the Gaussian case. However, the Clayton copula imposes too much dependence in the observations in the lower autocontours so that the model is also rejected. In between the Gaussian and Clayton results, we have those from an inverse Gumbel copula model. In this case, the values of the tests at the low and central autocontours are smaller than those from Clayton and larger than those from Gaussian; at the upper autocontours, the opposite happens, the tests are smaller than those from Gaussian and slightly larger than those from Clayton. The inverse Gumbel copula straddles between the dependence provided by the Gaussian copula and that provided by the Clayton copula.

In Figure 4, we plot 2-dimensional contours of several copula functions with Negative Binomial marginal densities with $1/\kappa_i = 0.2$ and dependence parameters similar to the estimates obtained for this data set. Observe the characteristics that prompt the rejection that we have discussed above. The contours of the Gaussian copula exhibit the most elliptical shapes compared to those from the remaining copulas. The rejection of this copula is mainly due to too much correlation imposed by the model on the observations in the upper north-east corner; the inverse Gumbel and Clayton relax the dependence in this area at the expense of imposing stronger dependence in the lower south-west region.

Given these results, it seems sensible to think about a model that combines the properties of different copulas. The most natural mixtures are Gaussian with Clayton copulas or inverse Gumbel with Clayton. In the last three columns of Table 8, we report

the results for the t-tests for the Gaussian/Clayton mixture. We estimated by maximum likelihood the weights of the mixture and the estimate is 0.534 on the Gaussian and 0.466 on Clayton. In addition, the estimates of the copula-based correlations are larger (from 0.5 to 0.7) than those from the pure Gaussian model, and the estimate of the Clayton dependence parameter θ is 0.45, which is smaller than that from the pure Clayton model. This mixture provides a remarkable improvement in the t-tests for all autocontours from 1% to 70% with values that fail to reject the model at conventional significance levels, while for the high autocontours 80% to 99%, the t-tests are substantially lower than those from Gaussian but still larger than those from Clayton. We should also note that the mixture inverse Gumbel/Clayton produces improvements in these high autocontours but at the expense of some rejections in the central autocontours. We report the t-tests of this mixture in the Appendix. In Figure 4, we can see why the mixture model performs better than the other models for this data set. The contours of the mixture Gaussian/Clayton show that the correlation in the observations in the south-west corner is higher than in the Gaussian case but lower than in Clayton or inverse Gumbel cases, and though this behavior is favored by this data set, we still need further independence among the observations in the north-east corner.

In summary, the combination of statistical tests and graphical devices offers plenty of evidence on the highly asymmetric dependence structure among trades of these three banking institutions. When the number of trades is a few, the dependence is high, and when the number of trades is very large, the dependence is low or non-existent. This finding is analogous to the empirical Epps effect (Epps, 1979) in high frequency returns, which shows that the sample correlation among stock returns goes to zero as the sampling interval shrinks. If we translate trade counts into durations, our findings can be interpreted as an Epps-like effect: when the durations become shorter (number

of trade counts is large), the correlation among trades goes to zero, and vice versa, for long durations (number of trade counts is low), the correlation is strong. This empirical effect is relevant for measurement of liquidity risk because when liquidity drains (lack of trading) , we should expect a concurrent effect among similar institutions.

In Figure 5, we plot the 3-dimensional vector of contemporaneous PITs for the Negative Binomial model with Gaussian, inverse Gumbel, Clayton, and mixture Gaussian/Clayton. It is evident that the four models have captured the contemporaneous dependence among the three series observed in Figure 3 (right panel). Now the observations are more evenly distributed within the cubes considered, as it is expected from i.i.d. data. With the Gaussian copula, observe the uneven location of the observations in the contours 80%-99% (green dots), which is the reason why the tests strongly reject this model for the higher contours. Inverse Gumbel and Clayton models correct substantially the distribution of the observations in the upper contours but at the expense of misplacing observations in the 10%-80% levels (blue dots). The mixture Gaussian/Clayton model offers the best fit as it is able to produce the right distribution of observations in the 10%-80% levels with no so large distortions in the very upper 80%-99% levels.

2.5 Conclusion

We have proposed a new tool, the Generalized Autocontour (G-ACR) as the basis for a battery of dynamic specification tests. G-ACR overcomes some important limitations of the autocontour methodology proposed in González-Rivera *et. al.* (2011), and in doing so, G-ACR tests are useful diagnostics to assess the conditional density model of either discrete or continuous random processes in a univariate or multivariate model, either in-sample or out-of-sample. The tests enjoy standard distributions

and have outstanding finite sample properties, they are correctly sized and are very powerful to detect departures from the assumed conditional density. In addition, the G-ACR methodology brings a visual aspect to the modeling exercise that is helpful on driving the specification process. To illustrate the usefulness of our approach we have focused on the modeling of a trivariate system of counts. The literature on evaluation of multivariate counts model is rather thin, so that our approach will also fill a void in this area of statistics and econometrics. We have analyzed a high frequency trivariate system of the number of trades in three US large banking institutions. We have shown the need to specify not only the correct dynamics and the individual marginal densities of trade counts but also the contemporaneous dependence of the three banks. We have modeled such dependence with copula functions and we have found that there is a highly asymmetric response depending on whether trading activity is dense or thin. When the number of trades is low, the contemporaneous dependence is stronger than when the number is high, so that when liquidity dries out in one institution, we should expect similar behavior in the rest.

Table 2.3: **Power of t_{k,α_i}-statistics for 13 autocontours (Nominal size 5% and $k = 1$)**

	$t_{1,1}$	$t_{1,2}$	$t_{1,3}$	$t_{1,4}$	$t_{1,5}$	$t_{1,6}$	$t_{1,7}$	$t_{1,8}$	$t_{1,9}$	$t_{1,10}$	$t_{1,11}$	$t_{1,12}$	$t_{1,13}$
T	Model P1 (Poisson/Normal)												
250	0.498	1.0	0.26	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
500	0.88	1.0	0.37	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1000	1.0	1.0	0.517	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
T	Model P2 (Negative Bin/Clayton)												
250	0.055	0.065	0.055	0.145	0.195	0.290	0.48	0.59	0.77	0.81	0.825	0.89	0.87
500	0.07	0.08	0.07	0.165	0.235	0.52	0.78	0.89	0.915	0.99	0.985	0.985	0.98
1000	0.182	0.384	0.240	0.427	0.510	0.68	0.94	1.0	1.0	1.00	1.00	1.00	1.00
T	Model P3 (Negative Bin/Gumbel)												
250	0.005	0.14	0.195	0.255	0.255	0.37	0.435	0.47	0.47	0.495	0.405	0.445	0.34
500	0.01	0.185	0.35	0.43	0.505	0.60	0.58	0.64	0.67	0.645	0.70	0.675	0.60
1000	0.05	0.26	0.48	0.59	0.68	0.81	0.72	0.85	0.842	0.862	0.95	0.90	0.83
T	Model P4 (Dynamic Misspecification)												
250	0.035	0.055	0.07	0.12	0.165	0.215	0.30	0.305	0.30	0.335	0.335	0.26	0.19
500	0.05	0.07	0.115	0.27	0.32	0.375	0.435	0.49	0.475	0.51	0.445	0.44	0.32
1000	0.075	0.115	0.255	0.415	0.49	0.60	0.665	0.675	0.70	0.745	0.69	0.60	0.505

Notes: The 13 autocontours are $C = [0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8.0.9, 0.95, 0.99]$.
1000 Monte Carlo replications and 500 bootstrap samples.
The null hypothesis is a bivariate conditional autoregressive model with Negative Binomial marginal densities and a Normal copula function.

Table 2.4: **Power of t_{k,α_i}, L_{α_i}, C_k statistics (Nominal size 5%)**

	$t_{2,7}$	$t_{3,7}$	$t_{4,7}$	$t_{5,7}$	L_7^2	L_7^3	L_7^4	L_7^5	C_1^{13}	C_2^{13}	C_3^{13}	C_4^{13}	C_5^{13}
T	Model P1 (Poisson/Normal)												
250	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
500	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1000	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
T	Model P2 (Negative Bin/Clayton)												
250	0.465	0.44	0.485	0.495	0.445	0.42	0.395	0.385	0.965	0.975	0.98	0.985	0.97
500	0.755	0.765	0.77	0.71	0.705	0.665	0.65	0.625	1.0	1.0	1.0	0.995	0.99
1000	0.921	0.930	0.935	0.932	0.927	0.873	0.854	0.839	1.0	1.0	1.0	1.0	1.0
T	Model P3 (Negative Bin/Gumbel)												
250	0.415	0.455	0.45	0.42	0.39	0.387	0.385	0.38	0.825	0.785	0.84	0.805	0.8
500	0.58	0.56	0.61	0.555	0.53	0.505	0.505	0.505	0.97	0.965	0.955	0.955	0.96
1000	0.723	0.719	0.723	0.708	0.705	0.701	0.704	0.702	1.0	1.0	0.994	0.992	1.0
T	Model P4 (Dynamic Misspecification)												
250	0.235	0.27	0.28	0.26	0.245	0.235	0.24	0.25	0.52	0.515	0.455	0.525	0.48
500	0.415	0.41	0.425	0.47	0.42	0.395	0.36	0.335	0.775	0.72	0.745	0.73	0.775
1000	0.585	0.605	0.59	0.60	0.585	0.57	0.56	0.52	0.935	0.875	0.91	0.93	0.95

Notes: $t_{k,7}$ for $k = 2, ...5$, and 7 refers to the 50% autocontour.
L_7^k for $k = 2, ...5$ stacking lags up to lag k and considering the 50% autocontour.
C_k^{13} stacking all 13 autocontours for lags $k = 1, ...5$.
1000 Monte Carlo replications and 500 bootstrap samples.
The null hypothesis is a bivariate conditional autoregressive model with Negative Binomial marginal densities and a Normal copula function.

Table 2.5: Estimation Results. Maximum Likelihood Estimates

	Poisson Model			Negative Binomial Model		
Parameter	BAC	WFC	JPM	BAC	WFC	JPM
w	157.08	139.48	185.72	170.68	156.55	106.32
	(12.42)	(13.44)	(12.22)	(29.63)	(22.32)	(15.48)
$a_{1,1}$	0.418			0.466		
	(41.51)			(43.07)		
$a_{1,2}$	0.027			0.0279		
	(2.47)			(3.72)		
$a_{1,3}$	0.0743			0.129		
	(5.88)			(3.29)		
$a_{2,1}$		0.0063			0.0287	
		(2.09)			(3.93)	
$a_{2,2}$		0.461			0.4557	
		(26.56)			(30.13)	
$a_{2,3}$		0.0296			0.029	
		(3.41)			(4.11)	
$a_{3,1}$			0.032			0.0258
			(6.36)			(6.53)
$a_{3,2}$			0.040			0.038
			(3.24)			(5.10)
$a_{3,3}$			0.482			0.433
			(23.98)			(24.72)
$b_{1,1}$	0.431			0.321		
	(32.26)			(31.13)		
$b_{2,2}$		0.394			0.310	
		(19.13)			(24.73)	
$b_{3,3}$			0.310			0.406
			(10.12)			(18.76)
Dispersion						
$1/\kappa_1$	0			0.196		
				(55.45)		
$1/\kappa_2$		0			0.186	
					(53.06)	
$1/\kappa_3$			0			0.141
						(39.03)
Pearson Residuals						
mean	-0.15	-0.63	1.09	0.01	0.09	-0.01
variance	436.28	256.58	313.91	1.23	1.10	1.36
Q(10)	15.19	76.89	147.97	84.47	90.00	73.81
Q(20)	60.79	112.61	212.73	154.64	219.25	126.13

Note: t-statistics in parenthesis. The models also include five dummies to take care of diurnal effects. The Pearson residual is defined as $(N_{i,t} - \mu_{i,t})/\sigma_{i,t}$.

Table 2.6: Autocontour-based Tests:

Negative Binomial Model with no Contemporaneous Correlation

	t-tests					L_{α_i}			
α_i	lag 1	lag 2	lag 3	lag 4	lag 5	lag 2	lag 3	lag 4	lag 5
1%	28.57	28.17	29.15	29.83	26.70	1235.36	1527.06	1673.36	1748.38
5%	28.81	27.78	30.08	29.19	27.58	1072.97	1259.76	1357.17	1382.56
10%	25.80	23.43	24.72	25.72	24.16	751.97	839.72	886.40	894.86
20%	20.42	19.54	20.61	20.27	19.69	452.32	496.74	504.11	505.65
30%	17.66	17.30	17.59	17.74	17.26	338.05	348.40	354.56	354.93
40%	13.80	13.60	13.30	13.74	13.25	200.80	201.69	204.83	204.83
50%	10.08	10.14	9.63	10.44	10.04	107.15	107.17	112.07	112.48
60%	7.61	7.28	7.47	8.36	7.71	58.06	58.78	72.31	72.61
70%	4.15	3.84	3.88	4.21	4.38	17.41	17.42	18.98	21.74
80%	0.15	-0.19	-0.13	-0.15	0.17	2.36	2.51	2.99	4.52
90%	-5.38	-5.72	-5.59	-5.76	-5.42	35.37	35.73	38.66	40.50
95%	-9.36	-9.70	-9.34	-9.56	-9.46	99.84	102.90	104.30	104.32
99%	-16.70	-16.95	-16.93	-16.81	-16.94	304.00	309.90	311.60	315.97
	C_k								
	1688.21	1612.38	1758.39	1772.77	1546.43				

Notes: t-tests critical values: ± 1.96 (5% level), ± 2.58 (1%), ± 3.5 (0.05%).
L_{α_i} stacks lags up to 2, 3,4,5 for α_i autocontour. Critical values at 5% level:
5.99 (2 lags), 7.81 (3 lags), 9.49 (4 lags) and 11.1 (5 lags).
C_k stacks all 13 autocontours for lag k. Critical values: 22.4 (5% level), 27.7 (1%), 34.5 (0.1%).

Table 2.7: Copula-based Correlation Matrix

	BAC	WFC	JPM
BAC	1.00	0.38	0.39
WFC	0.38	1.00	0.48
JPM	0.39	0.48	1.00

Table 2.8: Autocontour-based t-tests:

Negative Binomial Model with Contemporaneous Correlation

	Gaussian Copula			Inv. Gumbel Copula			Clayton Copula			Gaussian/Clayton		
α_i	lag 1	lag 2	lag 3	lag 1	lag 2	lag 3	lag 1	lag 2	lag 3	lag 1	lag 2	lag 3
1%	-0.72	0.33	0.52	-1.00	-0.94	-1.28	1.47	3.46	0.71	-0.77	-0.82	-0.60
5%	1.20	1.86	3.27	3.38	2.54	2.38	9.12	6.71	6.59	0.62	0.21	0.50
10%	3.43	3.52	4.10	3.25	2.44	3.02	10.89	8.58	9.98	1.75	1.74	1.41
20%	3.95	4.00	5.23	4.43	3.54	4.48	10.59	9.02	9.71	1.81	1.93	1.83
30%	3.70	3.56	4.73	4.53	3.29	4.13	10.16	8.78	9.19	2.82	2.63	3.31
40%	2.05	1.89	2.75	3.96	2.73	3.17	9.06	8.29	8.35	1.51	1.43	1.92
50%	0.78	0.57	1.51	3.42	2.61	2.86	7.61	6.80	7.11	0.22	0.16	0.32
60%	-0.74	-0.36	-0.23	1.39	1.27	1.60	5.97	5.59	6.06	-0.52	-0.23	-0.44
70%	-2.58	-2.54	-2.42	0.22	-0.14	0.28	3.78	3.54	3.71	-2.16	-2.22	-2.46
80%	-5.45	-5.41	-5.29	-3.19	-3.53	-3.40	-0.12	-0.13	-0.10	-4.19	-4.31	-4.21
90%	-9.70	-10.31	-10.24	-6.33	-6.68	-6.52	-4.88	-5.19	-5.04	-7.70	-8.26	-8.04
95%	-12.58	-13.05	-12.86	-10.41	-10.98	-10.69	-9.24	-9.61	-9.29	-10.51	-10.87	-10.54
99%	-20.60	-20.76	-20.87	-16.07	-16.25	-16.14	-15.06	-15.33	-15.29	-18.16	-18.25	-18.15

Notes: t-tests critical values: ± 1.96 (5% level), ± 2.58 (1%), ± 3.5 (0.05%).

Figure 2.3: Contours of Copulas with Negative Binomial Marginal Densities

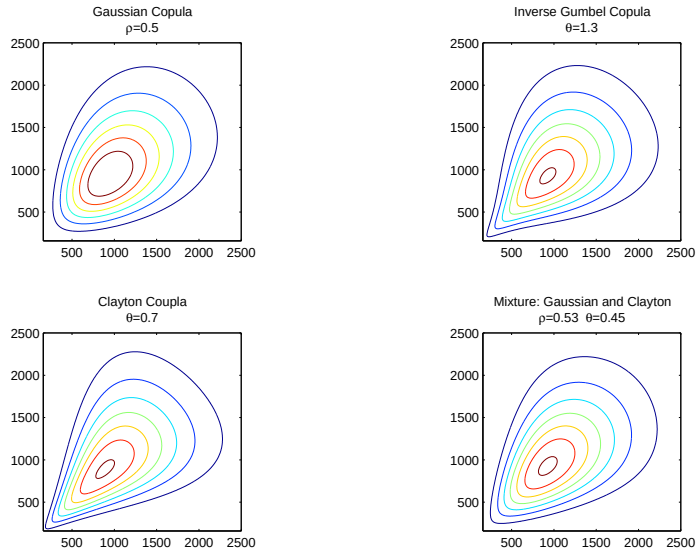
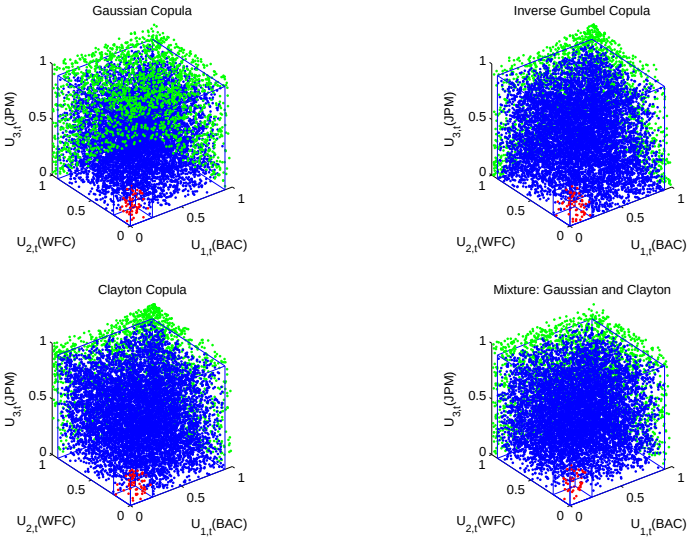


Figure 2.4: Contemporaneous PITs: Negative Binomial Model with Copula Dependence



Bibliography

[1] Bauwens, L., and Hautsch, N. (2006), "Stochastic Conditional Intensity Process", *Journal of Financial Econometrics*, vol. 4, 3, pp. 450-493.

[2] Cameron, C., and Trivedi, K. (1998). *Regression Analysis of Count Data*. Cambridge University Press, Cambridge.

[3] Cameron, C., and Trivedi, K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press, Cambridge.

[4] Diebold, F., Gunther, A., and Tay, A. (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management", *International Economic Review*, vol. 39, 4, pp. 863-883.

[5] Diebold, F., Hahn, J. and Tay, A. (1999), "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange", *The Review of Economics and Statistics*. vol. 81, 4, pp. 661-673.

[6] Denuit, M., and Lambert, P. (2005), "Constraints on Concordance Measures in Bivariate Discrete Data", *Journal of Multivariate Analysis*, vol. 93, pp. 40-57.

[7] Epps, T. (1979), "Comovements in Stock Prices in the Very Short Run ", *Journal of American Statistical Association*, vol. 74, 336, pp. 291-298.

[8] Easley, D., and O'Hara, M. (1992), "Time and the Process of Security Price Adjustment", *Journal of Finance*, vol. 27, 2, pp. 577-605.

[9] Engle, R., and Russell, R. (2007), "Autoregressive Conditional Duration: A New Model For Irregularly Spaced Transaction Data", *Econometrica*, vol. 66, 5, pp. 1127-1162.

[10] González-Rivera, G., Senyuz,Z., and Yoldas, E. (2011), "Autocontours: Dynamic Specification Testing", *Journal of Business and Economic Statistics*, vol. 29, 1, pp. 186-200.

[11] Hautsch, N. (2004), *Modelling Irregularly Spaced Financial Data*, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin.

[12] Heinen, A., and Rengifo, E. (2007), "Multivariate Autoregressive Modeling of Time Series Count Data Using Copulas", *Journal of Empirical Finance*, Vol. 14, pp. 564-583.

[13] Jung, R., Lisenfeld, R. and Richard, J. (2011), "Dynamic Factor Models for Multivariate Count data: An Application to Stock-Market Trading Activity", *Journal of Business and Economic Statistics*, vol. 29, 1, pp. 72-85.

[14] Marshall, A. (1996), "Copulas, Marginals, and Joint Distributions", *Lecture Notes-Monograph Series*. Vol. 28, 1, pp. 213-222.

[15] Madhavan, A. (2000), "Market Microstructure: A Survey", *Journal of Financial Markets*, vol. 3, pp. 205-258.

[16] Nelsen, R. (2005). *An Introduction to Copulas*. Second Edition, Springer.

[17] O'Hara, M. (1995). *Market Microstructure Theory*. Blackswell.

[18] Pacurar, M. (2008), "Autoregressive Conditional Duration Models in Finance: A Survey of The Theoretical and Empirical Literature", *Journal of Economic Surveys*, vol. 22, 4, pp. 711-751.

[19] Patton, A.J. (2006), "Estimation of Multivariate Models for Time Series of Possibly Different Lengths", *Journal of Applied Econometrics*, vol. 21, 4, pp. 147-173.

[20] Quoreshi, A.M.M.S. (2008), "A Vector Integer-Valued Moving Average Model for High Frequency Financial Count Data", *Economics Letters*, vol. 101, 3, pp. 258-261.

Chapter 3

Density Forecast Evaluation in Unstable Environment

3.1 Introduction

This paper proposes new methodologies for evaluating the out-of-sample density forecast performance of models in the presence of instabilities. The definition of instabilities may vary from different papers. In most of the literature, people specifically refer to the conditional mean parameter instabilities. It can be illustrated with the following simple model, where the data are generated by: $y_{t+h} = \beta_t x_t + \varepsilon_{t+h}$, where $\varepsilon_{t+h} \sim i.i.d.(0, \sigma^2)$. The parameter instability means that β_t is time-varying, which can have one break, multiple breaks in the history or even be different at each time point. However, in addition to the parameter instabilities, in some paper, (Giacomini, R. and B. Rossi (2010b)), the instabilities also contain the structural break to the model's variance. Therefore, in this paper, the instabilities are defined as time variation in the density function of a stochastic process. These variations include changes of mean, variance and/or the functional form of the underlying density function. The definition is

general enough to contain most of the important types of instabilities discussed widely in current literature.

There are a lot of works about detecting the parameter instability such as Chow (1960), Andrews (1993), Andrews and Ploberger (1994),Elliott and Mller (2003) etc. Beside those tests, instabilities are also a significant and practical concern for forecasters interested in evaluating predictive ability. Since the traditional forecast evaluation methods are invalid in the presence of instabilities, there are several robust evaluation methods proposed by researchers. In fact, a predictive content is reliable in one period is no necessary useful in the subsequent periods. The existence of unstable time series predictors over time have been documented in massive empirical studies in macroeconomics, finance and international finance. In the presence of instabilities, Rossi (2005) shows that the traditional Granger-causality test of a potential predictor on an economic variable of interests is invalid. Therefore, she constructs a robust Granger-causality test, which has been used by Rossi(2006) to show the existence of time-varying relationship between exchange rate and the fundamentals, used by Giacomini and Rossi (2006) to demonstrate that the term structure Granger-causes future output growth, used by Rapach and Wohar (2006) to provide empirical evidence on predictive ability of asset returns, and used by Chen, Rogoff and Rossi (2010) to provide empirical evidence that exchange rates Granger-causes commodity prices.

The estimation of models in the presence of instabilities is another major concern for researchers. There are several ways to estimate model with structural breaks. Andrew's (1993) and Andrew and Ploberger's (1994) tests are typically used to estimate one time discrete breaks. Pesaran and Timmermann (2002) proposed a Reversed Ordered Cusum (ROC) test to estimate the latest break for forecasting. There are also several works about estimating models with multiple discrete breaks (Pesaran, Pet-

tenuzzo and Timmermann, 2006, and Koop and Potter, 2007). Stock and Watson (2007) proposed a Unobserved Components Stochastic Volatility model, which allows the parameters to vary at each time point. They estimate this model with Markov Chain Monte Carlo and show that this model can successfully forecast inflation in U.S. In addtion, intercept corrections is another alternative approach to estimate models with parameter instabilities proposed by Clemens and Hendry (1996). With instabilities, to perform inference robust to the choice of the estimation window size to is another important topic. Pesaran and Timmermann (2002, 2005 and 2007) compared the performance between rolling and recursive estimation schemes in the presence of structural break in the parameters and discussed how to chose the optimal window size. Inoue and Rossi (2010) prosed a new method to evaluate out-of-sample forecasting performance that are robust to the choice of the estimation window size. Furthermore, in order to improve forecasting, researchers have suggested to use forecast combination with equal, time-varying weights (Stock and Watson (2003, 2004), Aiol and Timmermann(2006), Pesaran and Timmermann (2007), Clark and McCarken (2008), Timmermann (2009), etc.) or Bayesian model averaging method (Kozicki and Tinskey (2001), Cogley and Sargent (2005), Wright (2008), Clark and McCracken (2010), Inoue and Rossi (2011), etc.). More details about those methods can be found in Rossi (2011), which provide a thorough review about current researches about forecasting under instabilities.

When there exist instabilities of the predictive relationship, practitioners will concern more about the evaluation of the out-of-sample forecasts in the presence of instabilities. The impact of structural instability for forecast evaluation has not been formally investigated until very recently. The traditional forecast evaluation based on the assessment of the expectation of the difference of the loss from some proposed forecast models to the that of a benchmark model. It assumes the existence of a forecast

whose performance is superior to its competitors in every period. Typically, the researchers are looking by the average out-of-sample forecast error loss difference to find the best model forecast such as the works by Diebold and Mariano (1995) and West (1996). However, the averaging loss method has been in applied literature (e.g. Stock and Watson (2003)) which found that the models' performance varies across subsample. The tests of out-of-sample forecasting comparisons and evaluation have no power in the presence of instabilities. There are several important tools to compare the forecast model with instable predictive relationships. Rossi (2005) proposed a optimal test for nested model selection with parameter instabilities, which allows the parameters to be time varying or is unknown vector. The optimal tests jointly test for both parameter instability and a null hypothesis on a subset of parameters. This method can be applied to test whether the variables of interests in some proposed economic model are statistically signicant and to test whether this relationship between variables is stable over time. Along this paper, she (2006) used this optimal tests to test whether exchange rates are random walk and provided empirical evidence on parameter instability of the out-sample forecast performance in models used to determine nominal exchange rate. Giacomini and Rossi (2009, 2010a and 2010b) have recently introduced formal methods for forecast evaluation in the presence of instability. Giacomini and Rossi (2009) focused on "absolute measures" of accuracy, wheareas Giacomini and Rossi (2010a and 2010b) considered"relative measures".

In Giacomini and Rossi (2009), they proposed a method for detecting and predicting forecast breakdowns, which is defined as a situation in which the out-of-sample performance of a forecast model is significantly worse than its in-sample performance. They defined a surprise loss at time $t+\tau$ as the difference between the out-of-sample loss at time $t+\tau$ and the average in-sample loss. They then considered the out-of-sample

mean of the surprise losses and proposed a test based on the idea that, if a forecast is reliable, this mean should be close to zero. This test focused on absolute measure of forecast accuracy. They analyzed the possible causes of forecast break down and thus relate their test to structural break tests and showed that Forecast Breakdowns may happen because of instabilities and over-fitting. However, while their test detects both instabilities and over-fitting, in practice it cannot identify the exact source of the breakdown. Rossi and Sekhposyan (2011a) follow the same decomposition framework of Giacomini and Rossi (2010a) to measure time variation in models relative forecasting performance by averaging relative predictive ability over rolling windows. They decompose the out-of-sample loss function differences calculated in rolling windows of size m into following three components: the difference relative to the average loss, an average forecast error loss expected on the basis of the in-sample performance,and an average unexpected forecast error loss. This decomposition is helpful to identify three possible sources of models forecasting performance: time-varying forecasting ability due to parameter instabilities or unmodeled changes of the stochastic process, predictive content and over-fitting. They implemented the test in fixed rolling window as well as recursive window environments to study the performance of models of exchange rate determination in out-of-sample forecast environment.

Giacomini and Rossi (2010a) investigated the related problem of estimating and testing the time variation in the out-of-sample relative performance of models. They said that in the presence of structural change, the relative performance of two models may be time-varying and thus chose model which merely perform best on average will result in a loss of information. Similarly, they proposed the Fluctuation test and the One-Time Reversal test in the out-of-sample context to analyze and then compare the out-of-sample forecast performance of two competing models to understand which model

performs better at each point in time. The researchers have divided the sample of size T into an in-sample portion of size R and an out-of-sample portion of size P, and obtained two competing sequences of h-step ahead out-of-sample forecasts. For a general loss function L, there is a sequence of P out-of-sample forecast loss differences.

$$\{\triangle L_t(\hat{\theta}_{t-h,R}, \hat{\gamma}_{t-h,R})\}_{R+h}^{T} \equiv \{L^1(y_t, \hat{\theta}_{t-h,R}) - L^2(y_t, \hat{\gamma}_{t-h,R})\}_{t=R+h}^{T}$$

which depend on the realizations of the variables and on the in-sample parameter estimates for each model,$\hat{\theta}_{t-h,R}$ and $\hat{\gamma}_{t-h,R}$. These parameters are estimated under fixed and rolling scheme. They define the local relative loss for the two models as the sequence of out-of-sample loss differences computed over centered rolling windows of size m (over the evaluation sample P) and thus the fluctuation test is constructed as follows:

$$F_{t,m}^{OOS} = \hat{\sigma}^{-1} m^{-1} \sum_{j=t-\frac{m}{2}}^{t+\frac{m}{2}-1} \triangle L_t(\hat{\theta}_{j-h,R}, \hat{\gamma}_{j-h,R}), \tau \in [0,1], t = R+h+\frac{m}{2}, ..., T-\frac{m}{2}+1$$

where $\hat{\sigma}_2$ is a HAC estimator of σ^2

$$H_0 : E(\triangle L_t(\hat{\theta}_{j-h,R}, \hat{\gamma}_{j-h,R})) = 0$$

for all $t = R+h, ..., T$. The null is rejected when $max_t|F_{t,m}^{OOS}| > k_\alpha$, where k_α is the significance level. The asymptotic distribution of $F_{t,m}^{OOS}$ will be a functional of Brownian motion and the $F_{t,m}^{OOS}$ statistic is equivalent to Diebold and Mariano's (1995) and Giacomini and White's (2006) unconditional predictive test statistic, which is computed over rolling out-of-sample window of size m. (In Giacomini and White's (2006), they developed a framework for out-of-sample conditional predictive ability testing and forecast selection when there are model misspecification due to inadequately modeled dynamics, inadequately modeled heterogeneity, incorrect functional form, or any combination of these). The Fluctuation test can reveal the existence of break but the Fluctuation test may sacrifice its power to be a robust test. The One-Time Reversal test can be used to

estimate the timing of the break at which the reversal of relative models' performance may be observed. However, it is impossible for this test to detect the unique source of rejection if the test are rejected for both of its two components: the instabilities in the relative performance and a model being constantly better than its competitor. Their tests are only about the point forecast comparison between two competitive models, which can not be extended to the density forecast evaluation area.

In their companion paper, Giacomini, R. and B. Rossi (2010b) analyzed the relative in-sample performance of two competing, misspecified non-nested models in unstable environment. The unstable environment in their paper can be illustrated with the following design: $y_t = \beta_t x_t + \gamma_t z_t + \varepsilon_t, \varepsilon_t \sim i.i.d.N(0,1)$, where $x_t \sim N(0, \sigma^2_{x,t}), \quad z_t \sim N(0, \sigma^2_{z,t})$. In addition to allow the conditional mean parameters (β_t or γ_t) to change over time, they showed that the relative model's performance will change when one of the variances is time-varying ($\sigma^2_{x,t}$), even if the parameters in the conditional mean (β_t or γ_t) are constant. Therefore, in this paper, the unstable environment is not necessarily related to the instability of model's parameters but also related to the structural break to the model's variance. To detect the change of competing models' relative performance in the presence of structural change, they proposed a nonparametirc local Kullback-Leibler information criterion (KILC), based on which they constructed two in-sample tests (the Fluctuation test and the One-Time Reversal test). Similarly, the in-sample Fluctuation test and the One-Time Reversal test will suffer the same problems as their out-of-sample version mentioned in Giacomini and Rossi (2009). They claimed that in the presence of structural change, the model's performance may itself be changing over time, which is necessarily related to the presence of instability in the models parameters. That is to say that when the parameters of the competing models are stable, the model's relative performance may vary over time. However, it is also possible that the relative

performance of the models is constant with the time-varying parameters. Therefore, in this paper, the unstable environment is not defined merely by parameter instability but is more closely related to structural break to the distribution of error term. This paper focused on in-sample point forecast and their method is not applicable to nested competing model which is the common limitation to KILC method.

Rossi and Sekhposyan (2011b) proposed forecast rationality and optimality tests that are robust to the presence of instabilities. They proposed a "fluctuation optimality" test which estimates and tests for forecast optimality in rolling windows over the out-of-sample forecast portion of the data. This test is applied to the absolute predictive ability and forecast optimality. They tested whether the h-step ahead out-of-sample direct forecasts for the variable are optimal in the presence of instabilities by performing the forecast unbiasedness, forecast efficiency, forecast encompassing and serial uncorrelated tests. The parameter estimates computed over rolling window size m. They consider estimating regression using data of the evaluation sample with rolling regression from $t-m+1$ up to t, for $t=m,...,P$, to avoid averaging out instabilities. Their tests are based on the framework developed by West and McCracken ("Regression based tests of predictive ability" IER 1998): v_{t+h} is the forecast error associated to the h-steps ahead forecast made at time t and in the case of a simple linear regression model with h-period lagged regressors x_t, the forecast error is $v_{t+h} = y_{t+h} - \widehat{\gamma}_{t,R} x_t$. $\widehat{\gamma}_{t,R}$ is the in-sample estimates, which can be obtained under fixed scheme, rolling scheme and recursive scheme.

Consider the general regression:

$$v_{t+h} = g_t^{'}\theta + \eta_{t+h}, \quad t = R, ..., T \tag{3.1}$$

where v_{t+h} is the forecast error associated to the h-steps ahead forecast made at time t,

θ is a parameter vector and g_t is a vector of variable known at time t such that $E(g_t g_t^{'})$ has full rank.

West and McCracken (1998) consider the following leading cases:

- forecast unbiasedness test, where $g_t = 1$
- forecast efficiency, where $g_t = y_{t+h|t}$
- forecast encompassing tests, where g_t is the forecast of the encompassed model
- serial uncorrelation tests, where $g_t = v_t$

West and McCracken's forecast rationality tests focus on testing the null hypothesis:

$$H_0 : \theta = \theta_0, H_A : \theta \neq \theta_0$$

Wald test:

$$W_p = (\hat{\theta}_P - \theta_0)' \widehat{V}_{\theta,P}^{-1} (\hat{\theta}_P - \theta_0)$$

$\hat{\theta}_P$ denote the estimate of θ in the general regression 3.1, $\widehat{V}_{\theta,P}$ is a consistent estimate of the long run variance of $\hat{\theta}_P$.

The main interest in this paper is testing forecast optimality in the presence of instabilities and they considered the rolling regression in 3.1 to avoid averaging out instabilities. Let $\hat{\theta}_{t,R}$ be the parameter estimate in regression 3.1 computed over rolling window of size m. They consider estimating regression 3.1 using data (the evaluation sample) from $t-m+1$ up to t, for $t = m, ..., P$, then the Wald test in the corresponding regressions is defined as:

$$W_{t,m} = (\hat{\theta}_t - \theta_0)' \widehat{V}_{\theta,t}^{-1} (\hat{\theta}_t - \theta_0)$$

for $t = m, ..., P$, where $W_{t,m}$ is referred as the Fluctuation Optimality test. Under the assumptions given in this paper, $W_{t,m}$ will obey the Functional Central Limit Theorem

and its limiting distribution can be expressed as the function of standard *Brownian* motion. In order to construct the statistics robust to the instabilities, that is, to reject the null hypothesis:

$$H_0 : E(\hat{\theta}_t) = \theta_0$$

for all $t = m, ..., P$, they construct the statistic of the maximum of $W_{t,m}$ i.e $max_t W_{t,m}$. If $max_t W_{t,m} > \kappa_{\alpha,q}$, where $\kappa_{\alpha,q}$ are the critical values at the $100\alpha\%$ significance level. They showed that while the traditional tests are more powerful when the data are stationary, the Fluctuation-type tests are more powerful in presence of instabilities. The above paper is closely related to Giacomini and Rossi (2010), in which a similar Fluctuation test is proposed to compare model's relative forecasting performance. However, the above robust procedures are all about in-sample or out-sample point forecasting evaluation or model selection. Existing econometric techniques are inadequate for conducting density forecast evaluation in an unstable environment. In this paper, we try to fill the gap by proposing the robust density forecasting evaluation method in the presence of instabilities.

3.2 Statistics and asymptotic distribution

3.2.1 Definition of Statistics

Our main interest is the evaluation of density forecast in the presence of instabilities. We propose a robust out-of-sample density forecasting evaluation method in the presence of the instabilities based on G-ACR. The tests are constructed with the following rolling evaluation scheme. Let Y_t denote the random process of interest with conditional density function $f(y_t|\Omega_{t-1})$, where Ω_{t-1} is the information set available at time t-1. if the proposed predictive density model for Y_t, i.e. $\{f_t^*(y_t|\Omega_{t-1})\}_{t=1}^T$ coincides with the

true conditional density $\{f_t(y_t|\Omega_{t-1})\}_{t=1}^T$, then the sequence of probability integral transforms (PIT) of $\{Y_t\}_{t=1}^T$ w.r.t $\{f_t^*(y_t|\Omega_{t-1})\}_{t=1}^T$ i.e. $\{u_t\}_{t=1}^T$ must be $i.i.d\ U(0,1)$where $u_t = \int_{-\infty}^{y_t} f_t^*(v_t|\Omega_{t-1})dv_t$. Thus, the null hypothesis $H_0 : f_t^*(y_t|\Omega_{t-1}) = f_t(y_t|\Omega_{t-1})$ is equivalent to the null hypothesis $H_0^{'} : \{u_t\}_{t=1}^T$ is $i.i.d\ U(0,1)$. In this section we construct the generalized autocontours under i.i.d. uniformity for predictive densities. Under $H_0^{'} : \{u_t\}_{t=1}^T\ i.i.d\ U(0,1)$, the generalized autocontour $GACR_{\alpha_i,k}$ is defined as the set of points in the plane (u_t, u_{t-k}) such that the square with $\sqrt{\alpha_i}$ side contains $\alpha_i\%$ of observations, i.e.,

$$\begin{aligned} & \text{G-ACR}_{\alpha_i,k} \\ = \ & \{B(u_t, u_{t-k}) \subset \Re^2 \| \ 0 \leq u_t \leq \sqrt{\alpha_i} \text{ and } 0 \leq u_{t-k} \leq \sqrt{\alpha_i},\ s.t. : u_t \times u_{t-k} \leq \alpha_i\} \end{aligned} \tag{3.2}$$

The indicator series constructed under the G-ACR is as follows

$$I_t^{k,\alpha_i} = \mathbf{1}((u_t, u_{t-k}) \subset G - ACR_{\alpha_i,k}) = \mathbf{1}(0 \leq u_t \leq \sqrt{\alpha_i}, 0 \leq u_{t-k} \leq \sqrt{\alpha_i})$$

The total sample size T is divided into in-sample portion (R) and out-of-sample portion (P). First, we evaluate one subsample (with size rol) of the evaluation sample, using data from $t-rol+1$ up to t, where $t = R+rol, \cdots, T$, to evaluate the assumed predictive density. According to G-ACR, for one subsample, we can obtain three different types of statistics: t, C and L. Second, we roll one observation ahead, with fixed subsample size rol, until the last observation in the evaluation sample. For $t = R + rol, \cdots, T$, we obtain a collection of $T - rol - R + 1$ statistics of for t, C and L statistics, respectively, i.e., $\{t_i\}_{i=1}^{T-s-R+1}$, $\{C_i\}_{i=1}^{T-s-R+1}$ and $\{L_i\}_{i=1}^{T-s-R+1}$. Third, to take care of the instabilities, we construct the Sup-type statistic by taking supreme over the absolute value of $\{|t_i|\}_{i=1}^{T-s-R+1}$, $\{C_i\}_{i=1}^{T-s-R+1}$ and$\{L_i\}_{i=1}^{T-s-R+1}$ respectively to obtain supreme

of these statistics respectively: S_t, S_C and S_L; the Avg-type statistic is constructed by taking average over collections of statistics, i.e.,A_t, A_C and A_L, which explore the average behavior of the environmental instabilities.

3.2.2 Asymptotic Properties of statistics

- Assumption 1: $T \to \infty$, $R \to \infty$, $P \to \infty$, $\lim_{T\to\infty} \frac{P}{R} = 0$, $\lim_{T\to\infty} \frac{rol}{P} = m$,as $rol, P \to \infty$. Where, rol is the size of rolling window in the evaluation sample, $m \in (0, \infty)$, k is the lag.

- Assumption 2: $\{y_t, x_t\}$ are mixing with either ϕ of size $-r/2(r-1)$ or α of size $-r/2(r-1)$

Proposition 1

Let $\widehat{\alpha}_i(J) = \frac{\sum_{t=R+1+J-rol+k}^{R+J} I_t^{k,\alpha_i}}{rol-k}$, where $J = [Ps]$, $s \in [m, 1]$.

$$
\begin{aligned}
S_t &= \sup_J \left| \frac{\sqrt{rol-k}(\widehat{\alpha}_i(J) - \alpha_i)}{\sigma_{k,\alpha_i}} \right| \\
A_t &= \frac{1}{P - rol + 1} \sum_{\underline{J}}^{\bar{J}} \left| \frac{\sqrt{rol-k}(\widehat{\alpha}_i(J) - \alpha_i)}{\sigma_{k,\alpha_i}} \right|
\end{aligned}
$$

Where $\sigma^2_{k,\alpha_i} = \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2})$

Under the null hypothesis of $i.i.d\ U(0,1)$ PITs, given assumption 1,2 and 5:

$$S_t \xrightarrow[P\to\infty]{} \sup_{s\in[m,1]} \frac{|W(s) - W(s-m)|}{\sqrt{m}} \tag{3.3}$$

$$A_t \xrightarrow[P\to\infty]{} \int_{\underline{s}}^{\bar{s}} \frac{|W(s) - W(s-m)|}{\sqrt{m}} ds \tag{3.4}$$

Where $J = [Ps]$, $rol - k = [Pm]$ and $W(.)$ is a standard univariate Brownian motion.

Proposition 2 For a given lag k, let $c_{k,i}(J) = \sqrt{rol-k}(\widehat{\alpha}_i - \alpha_i)$ and stack $c_{k,i}(J)$ for different contours levels $i = 1, 2, ...C$. Let $\mathbf{C}_k(J) = (c_{k,1}(J), ...c_{k,C}(J))'$ be the $C \times 1$ stacked vector.

$$\begin{aligned} S_C &= \sup_J |\mathbf{C}_k(J)'\Omega_k^{-1}\mathbf{C}_k(J)| \\ A_C &= \frac{1}{P-rol+1}\sum_{\underline{J}}^{\bar{J}} |\mathbf{C}_k(J)'\Omega_k^{-1}\mathbf{C}_k(J)| \end{aligned}$$

Where Ω_k is the asymptotic variance and covariance matrix for the random vector $\mathbf{C}_k(J)$

Under the null hypothesis of $i.i.d\ U(0,1)$ PITs, given the assumption 1,3 and 5:

$$\begin{aligned} S_C &\xrightarrow[P\to\infty]{} \sup_{s\in[m,1]} \frac{(\mathbf{W}(s)-\mathbf{W}(s-m))'(\mathbf{W}(s)-\mathbf{W}(s-m))}{m} \\ A_C &\xrightarrow[P\to\infty]{} \int_{\underline{s}}^{\bar{s}} \frac{(\mathbf{W}(s)-\mathbf{W}(s-m))'(\mathbf{W}(s)-\mathbf{W}(s-m))}{m} ds \end{aligned} \tag{3.5}$$

Where $J = [Ps]$, $rol - k = [Pm]$ and $\mathbf{W}(.)$ is a standard C-variate Brownian motion.

Proposition 3

For a given contour α_i, let $\ell_{k,\alpha_i}(J) = \sqrt{rol-k}(\widehat{\alpha}_i(J) - \alpha_i)$ and stack ℓ_{k,α_i} for $k = 1, ...K$. Let $L_{\alpha_i} = (\ell_{1,\alpha_i}(J), ...\ell_{K,\alpha_i}(J))'$ be the $K \times 1$ stacked vector.

$$\begin{aligned} S_L &= \sup_J |L_{\alpha_i}(J)'\Lambda_{\alpha_i}^{-1}L_{\alpha_i}(J)| \\ A_L &= \frac{1}{P-rol+1}\sum_{\underline{J}}^{\bar{J}} |L_{\alpha_i}(J)'\Lambda_{\alpha_i}^{-1}L_{\alpha_i}(J)| \end{aligned}$$

Where Λ_{α_i} is the asymptotic variance and covariance matrix for the random vector L_{α_i}.

Under the null hypothesis of $i.i.d\ U(0,1)$ PITs, given the assumption 1, 4 and 5:

$$
\begin{aligned}
S_L &\xrightarrow[P\to\infty]{} \sup_{s\in[m,1]} \frac{(\mathbf{W}(s)-\mathbf{W}(s-m))'(\mathbf{W}(s)-\mathbf{W}(s-m))}{m} \\
A_L &\xrightarrow[P\to\infty]{} \int_{\underline{s}}^{\bar{s}} \frac{(\mathbf{W}(s)-\mathbf{W}(s-m))'(\mathbf{W}(s)-\mathbf{W}(s-m))}{m} ds
\end{aligned} \tag{3.6}
$$

Where $J=[Ps]$, $rol-k=[Pm]$ and $\mathbf{W}(.)$ is a standard L-variate Brownian motion.

Proof of the Proposition 1:

The indicator function I_t^{k,α_i} is a Bernoulli random variable with the following moments: $E(I_t^{k,\alpha_i})=\alpha_i$, $Var(I_t^{k,\alpha_i})=\alpha_i(1-\alpha_i)$ and covariance

$$
r_h^{\alpha_i} \equiv cov(I_t^{k,\alpha_i}, I_{t-h}^{k,\alpha_i}) = \begin{cases} 0 & \text{if } h\neq k \\ \alpha_i^{3/2}(1-\alpha_i^{1/2}) & \text{if } h=k \end{cases}
$$

Given the assumption 1, 2 and 5, according to **Theorem 7.16** in ("Asymptotic theory for Econometricians :Halber White (Revised Edition)"), we need to prove that :$Z_t^{k,i}$ is globally stationary. That is to show:

$$
\sigma^2_{k,\alpha_i} \equiv \lim_{rol\to\infty} var((rol-k)^{-\frac{1}{2}} \sum_{t=R+1+J-rol+k}^{R+J} Z_t^{k,i}) > 0
$$

$$
\begin{aligned}
\sigma^2_{k,\alpha_i} &= \lim_{rol\to\infty} var((rol-k)^{-\frac{1}{2}} \sum_{t=R+1+J-rol+k}^{R+J} (I_t^{k,\alpha_i}-\alpha_i)) \\
&= \lim_{rol\to\infty} var((rol-k)^{-\frac{1}{2}} \sum_{t=R+1+J-rol+k}^{R+J} Z_t^{k,i}) \\
&= \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2})
\end{aligned}
$$

It shows that $Z_t^{k,i}$ is globally covariance stationary process.

$$
\begin{aligned}
\widehat{\alpha}_i(J) &= \frac{\sum_{t=R+1+J-rol+k}^{R+J} I_t^{k,\alpha_i}}{rol-k} = \frac{\sum_{t=R+[P(s-m)]+1}^{R+Ps} I_t^{k,\alpha_i}}{[Pm]} \\
W_P(s) &= \frac{\sqrt{[Pm]}(\widehat{\alpha}_i(J)-\alpha_i)}{\sigma_{k,i}} = \frac{\sum_{t=R+[P(s-m)]+1}^{R+Ps}(I_t^{k,\alpha_i}-\alpha_i)}{\sqrt{[Pm]}\sigma_{k,i}}
\end{aligned}
$$

Where $J=[Ps]$, $\underline{J}=[P\underline{s}]$, $\bar{J}=[P\bar{s}]$, $s\in[m^*,1]$, $rol-k=[Pm]$

$$\begin{aligned}
W_P(s) &= \frac{\sqrt{[Pm]}(\widehat{\alpha}_i-\alpha_i)}{\sigma_{k,i}}=\frac{\sum_{t=R+[P(s-m)]+1}^{R+Ps}(I_t^{k,\alpha_i}-\alpha_i)}{\sqrt{[Pm]}\sigma_{k,i}}\\
&= \frac{\sum_{t=R+1}^{R+Ps}(I_t^{k,\alpha_i}-\alpha_i)}{\sqrt{[Pm]}\sigma_{k,i}}-\frac{\sum_{t=R+1}^{R+[P(s-m)]}(I_t^{k,\alpha_i}-\alpha_i)}{\sqrt{[Pm]}\sigma_{k,i}}\\
&= \frac{\sqrt{P}}{\sqrt{[Pm]}}\frac{\sum_{t=R+1}^{R+Ps}(I_t^{k,\alpha_i}-\alpha_i)}{\sqrt{P}\sigma_{k,i}}-\frac{\sqrt{P}}{\sqrt{[Pm]}}\frac{\sum_{t=R+1}^{R+[P(s-m)]}(I_t^{k,\alpha_i}-\alpha_i)}{\sqrt{P}\sigma_{k,i}}\\
&= \frac{\sqrt{P}}{\sqrt{[Pm]}}\frac{\sum_{t=R+1}^{R+Ps}Z_t^{k,i}}{\sqrt{P}\sigma_{k,i}}-\frac{\sqrt{P}}{\sqrt{[Pm]}}\frac{\sum_{t=R+1}^{R+[P(s-m)]}Z_t^{k,i}}{\sqrt{P}\sigma_{k,i}}
\end{aligned}$$

Since $Z_t^{k,i}$ is globally stationary and $E(Z_t^{k,i})=E(I_t^{k,\alpha_i}-\alpha_i)=0$, according to **Theorem 7.17** in ("Asymptotic theory for Econometricians :Halber White (Revised Edition)"):

$$\begin{aligned}
W_P^1(s) &= \frac{\sqrt{P}}{\sqrt{[Pm]}}\frac{\sum_{t=R+1}^{R+Ps}Z_t^{k,i}}{\sqrt{P}\sigma_{k,i}}\xrightarrow[P\to\infty]{}\frac{1}{\sqrt{m}}W(s)\\
W_P^2(s) &= \frac{\sqrt{P}}{\sqrt{[Pm]}}\frac{\sum_{t=R+1}^{R+[P(s-m)]}Z_t^{k,i}}{\sqrt{P}\sigma_{k,i}}\xrightarrow[P\to\infty]{}\frac{1}{\sqrt{m}}W(s-m)
\end{aligned}$$

Where $W(.)$ is the standard Brownian motion.

$$\begin{aligned}
W_P(s) &= W_P^1(s)-W_P^2(s)\\
&\xrightarrow[P\to\infty]{} \frac{1}{\sqrt{m}}(W(s)-W(s-m)) \qquad (3.7)
\end{aligned}$$

By continuous mapping theorem and (3.7) we can find that:

$$\begin{aligned}
S_t &= \sup_J|\frac{\sqrt{rol-k}\widehat{\alpha}_i(J)}{\sigma_{k,\alpha_i}}|\xrightarrow[P\to\infty]{}\sup\frac{|W(s)-W(s-m)|}{\sqrt{m}}\\
A_t &= \frac{1}{P-rol+1}\sum_{\underline{J}}^{\bar{J}}|\frac{\sqrt{rol-k}\widehat{\alpha}_i(J)}{\sigma_{k,\alpha_i}}|\xrightarrow[P\to\infty]{}\int_{\underline{s}}^{\bar{s}}\frac{|W(s)-W(s-m)|}{\sqrt{m}}
\end{aligned}$$

Proof of Proposition 2

Given assumption 1, 3 and 5, according to **Theorem 7.30** in ("Asymptotic theory for Econometricians :Halber White (Revised Edition)"), we need to prove random vector $\mathbf{Z}_t^C$ is global covariance stationary, where $\mathbf{Z}_t^C = (Z_t^{k,1}, ..., Z_t^{k,C})$ and $Z_t^{k,i} = I_t^{k,\alpha_i} - \alpha_i$.

First, we show $\mathbf{Z}_t^C$ is globally covariance stationery, that is to show

$$\Omega_k = \lim_{rol\to\infty} var((rol-k)^{-1/2}\Sigma_{t=k+1}^{rol}\mathbf{Z}_t^C)$$

, where Ω_k is nonsingular global covariance matrix of $\{\mathbf{Z}_t^C\}$

The elements $\omega_{i,j}$ of the asymptotic variance-covariance Ω_k are obtained as follows. We need to calculate

$$cov(c_{k,i}, c_{k,j}) = cov(I_t^{k,\alpha_i}, I_t^{k,\alpha_j}) + cov(I_t^{k,\alpha_i}, I_{t-k}^{k,\alpha_j}) + cov(I_{t-k}^{k,\alpha_i}, I_t^{k,\alpha_j}) + o(1)$$

If $i = j$, by Proposition 1, $\omega_{i,i} = var(\sqrt{T-k}(\widehat{\alpha}_i - \alpha_i)) = \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2})$. If $i < j$ (and similarly for $i > j$), $\alpha_i \subset \alpha_j$. Then, we have that

$$cov(I_t^{k,\alpha_i}, I_t^{k,\alpha_j}) = E(I_t^{k,\alpha_i} \times I_t^{k,\alpha_j}) - \alpha_i \times \alpha_j = \alpha_i(1-\alpha_j)$$

$$cov(I_t^{k,\alpha_i}, I_{t-k}^{k,\alpha_j}) = E(I_t^{k,\alpha_i} \times I_{t-k}^{k,\alpha_j}) - \alpha_i \times \alpha_j = \alpha_i \times \alpha_j^{1/2} - \alpha_i \times \alpha_j$$

$$cov(I_{t-k}^{k,\alpha_i}, I_t^{k,\alpha_j}) = E(I_{t-k}^{k,\alpha_i} \times I_t^{k,\alpha_j}) - \alpha_i \times \alpha_j = \alpha_i \times \alpha_j^{1/2} - \alpha_i \times \alpha_j$$

Therefore, Ω_k is the nonsingular global covariance matrix of $\mathbf{Z}_t^C$ and thus $\{\mathbf{Z}_t^C\}$is global covariance stationary.

$$\begin{aligned}\mathbf{W}_P(s) &\equiv \Omega_k^{-1/2}[Pm]^{-1/2} \sum_{t=R+[P(s-m)]+1}^{R+Ps} \mathbf{Z}_t^C \\ &\equiv \Omega_k^{-1/2}[Pm]^{-1/2} \sum_{t=R+1}^{R+Ps} \mathbf{Z}_t^C - \Omega_k^{-1/2}[Pm]^{-1/2} \sum_{t=R+1}^{R+P(s-m)} \mathbf{Z}_t^C\end{aligned}$$

Since $\mathbf{Z}_t^C$is global covariance stationary, $E(Z_t^{k,i}) = E(I_t^{k,\alpha_i} - \alpha_i) = 0$, according to **Theorem 7.30** in ("Asymptotic theory for Econometricians :Halber White (Revised Edition)"), we can show that:

$$\begin{aligned}
\mathbf{W}_P^1(s) &= \Omega_k^{-1/2}[Pm]^{-1/2} \sum_{t=R+1}^{R+Ps} \mathbf{Z}_t^C \xrightarrow[P\to\infty]{} \frac{1}{\sqrt{m}}\mathbf{W}(s) \\
\mathbf{W}_P^2(s) &= \Omega_k^{-1/2}[Pm]^{-1/2} \sum_{t=R+1}^{R+P(s-m)} \mathbf{Z}_t^C \xrightarrow[P\to\infty]{} \frac{1}{\sqrt{m}}\mathbf{W}(s-m)
\end{aligned}$$

Where $\mathbf{W}$(s) is a C-variate Brownian process.

$$\mathbf{W}_P(s) \equiv \mathbf{W}_P^1(s) - \mathbf{W}_P^2(s) \xrightarrow[P\to\infty]{} \frac{1}{\sqrt{m}}(\mathbf{W}(s) - \mathbf{W}(s-m)) \tag{3.8}$$

From continuous mapping theorem and 3.8

$$\begin{aligned}
S_C &= \sup_J |R_k(J)'\Omega_k^{-1}R_k(J)| \\
&\xrightarrow[P\to\infty]{} \sup_{s\in[m^*,1]} \frac{(\mathbf{W}(s) - \mathbf{W}(s-m))'(\mathbf{W}(s) - \mathbf{W}(s-m))}{m}
\end{aligned}$$

$$\begin{aligned}
A_C &= \frac{1}{P - rol + 1} \sum_{\underline{J}}^{\bar{J}} |R_k(J)'\Omega_k^{-1}R_k(J)| \\
&\xrightarrow[P\to\infty]{} \int_{\underline{s}}^{\bar{s}} \frac{(\mathbf{W}(r) - \mathbf{W}(r-m))'(\mathbf{W}(r) - \mathbf{W}(r-m))}{m}
\end{aligned}$$

Proof of Proposition 3:

Given assumption 1, 4 and 5, according to **Theorem 7.30** in ("Asymptotic theory for Econometricians :Halber White (Revised Edition)"), we need to prove random vector $\mathbf{Z}_t^L$ is global covariance stationary. where $\mathbf{Z}_t^L = (Z_t^{1,c}, ..., Z_t^{L,c})$ and $Z_t^{k,i} = I_t^{k,\alpha_i} - \alpha_i$.

That is to show:

$$\Lambda_k \quad = \quad \lim_{rol\to\infty} var((rol-k)^{-1/2}\Sigma_{t=k+1}^{rol}\mathbf{Z}_t^L)$$

, where Λ_k is nonsingular global covariance matrix of $\{\mathbf{Z}_t^L\}$

$$\lambda_{j,k} = cov(\ell_{j,\alpha_i}, \ell_{k,\alpha_i}) = var(\sqrt{T-k}(\widehat{\alpha}_i - \alpha_i)) = \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2})$$

The elements $\lambda_{j,k}$ of the asymptotic variance-covariance Λ_{α_i} are obtained as follows. When $j = k$, we have

$$\lambda_{j,k} = cov(\ell_{j,\alpha_i}, \ell_{k,\alpha_i}) = var(\sqrt{T-k}(\widehat{\alpha}_i - \alpha_i)) = \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2})$$

When $j > k$ (and similarly for $j < k$),

$$\begin{aligned} & cov(\ell_{k,\alpha_i}, \ell_{j,\alpha_i}) \\ = \quad & cov(I_t^{k,\alpha_i}, I_t^{j,\alpha_i}) + cov(I_t^{k,\alpha_i}, I_{t-k}^{j,\alpha_i}) + cov(I_{t-j}^{k,\alpha_i}, I_t^{j,\alpha_i}) + cov(I_t^{j,\alpha_i}, I_{t-j+k}^{k,\alpha_i}) + o(1) \end{aligned}$$

from which each covariance term is

$$cov(I_t^{k,\alpha_i}, I_t^{j,\alpha_i}) = E(I_t^{k,\alpha_i} \times I_t^{j,\alpha_i}) - \alpha_i^2 = \alpha_i^{3/2} - \alpha_i^2 = \alpha_i^{3/2}(1-\alpha_i^{1/2})$$

and taking into account that $cov(I_t^{k,\alpha_i}, I_t^{j,\alpha_i}) = cov(I_t^{k,\alpha_i}, I_{t-k}^{j,\alpha_i}) = cov(I_{t-j}^{k,\alpha_i}, I_t^{j,\alpha_i}) = cov(I_{t-j+k}^{k,\alpha_i}, I_t^{j,\alpha_i})$, Therefore, Λ_k is the nonsingular global covariance matrix of $\{\mathbf{Z}_t^L\}$ and thus $\{\mathbf{Z}_t^L\}$is global covariance stationary.

$$\begin{aligned} \mathbf{W}_P(s) \quad & \equiv \quad \Omega_k^{-1/2}[Pm]^{-1/2} \sum_{t=R+[P(s-m)]+1}^{R+Ps} \mathbf{Z}_t^C \\ & \equiv \quad \Omega_k^{-1/2}[Pm]^{-1/2} \sum_{t=R+1}^{R+Ps} \mathbf{Z}_t^C - \Omega_k^{-1/2}[Pm]^{-1/2} \sum_{t=R+1}^{R+P(s-m)} \mathbf{Z}_t^C \end{aligned}$$

Since $\mathbf{Z}_t^L$is global covariance stationary, $E(Z_t^{k,i}) = E(I_t^{k,\alpha_i} - \alpha_i) = 0$, according to **Theorem 7.30** in ("Asymptotic theory for Econometricians :Halber White (Revised Edition)"), we can show that:

$$\mathbf{W}_P^1(s) = \Omega_k^{-1/2}[Pm]^{-1/2}\sum_{t=R+1}^{R+Ps}\mathbf{Z}_t^C \xrightarrow[P\to\infty]{} \frac{1}{\sqrt{m}}\mathbf{W}(s)$$

$$\mathbf{W}_P^2(s) = \Omega_k^{-1/2}[Pm]^{-1/2}\sum_{t=R+1}^{R+P(s-m)}\mathbf{Z}_t^C \xrightarrow[P\to\infty]{} \frac{1}{\sqrt{m}}\mathbf{W}(s-m)$$

Where $\mathbf{W}(s)$ is a L-variate Brownian process.

$$\mathbf{W}_P(s) \equiv \mathbf{W}_P^1(s) - \mathbf{W}_P^2(s) \xrightarrow[P\to\infty]{} \frac{1}{\sqrt{m}}(\mathbf{W}(s) - \mathbf{W}(s-m)) \tag{3.9}$$

From continuous mapping theorem and 3.9

$$\begin{aligned} S_L &= \sup_J |R_l(J)'\Lambda_l^{-1}R_l(J)| \\ &\xrightarrow[P\to\infty]{} \sup_{s\in[m^*,1]} \frac{(\mathbf{W}(s)-\mathbf{W}(s-m))'(\mathbf{W}(s)-\mathbf{W}(s-m))}{m} \end{aligned}$$

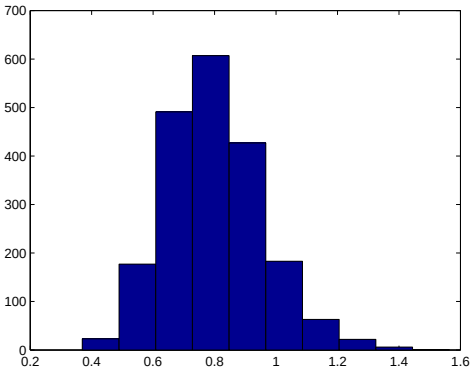
$$\begin{aligned} A_L &= \frac{1}{P-rol+1}\sum_{\underline{J}}^{\bar{J}} |R_l(J)'\Lambda_l^{-1}R_l(J)| \\ &\xrightarrow[P\to\infty]{} \int_{\underline{s}}^{\bar{s}} \frac{(\mathbf{W}(s)-\mathbf{W}(s-m))'(\mathbf{W}(s)-\mathbf{W}(s-m))}{m} \end{aligned}$$

3.3 Monte Carlo Simulation

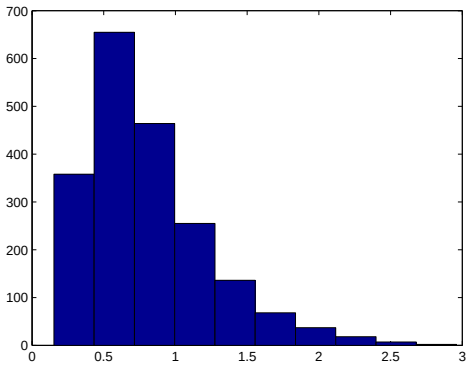
3.3.1 Critical Values for Asymptotic Distributions

In this session, we calculated the theoretical critical values for the asymptotic distributions of our statistics. The number of Monte Carlo replication is 2000. Since the asymptotic distributions for the statistics depend on m, which is proportion of the rolling window size to the whole evaluation sample, we consider all the critical values for the asymptotic distribution with $m \in (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$. Table

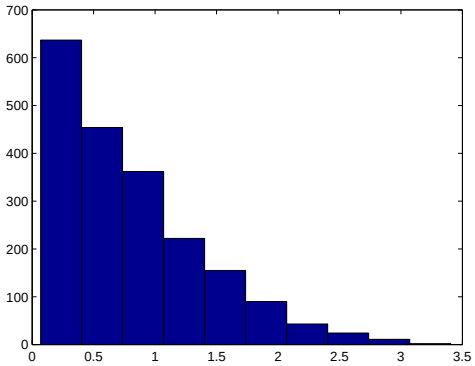
3.1 and Table 3.2 report the critical values for S_t and A_t statistics. Table 3.3 and Table 3.4 report the critical values for S_C and A_C statistics of 13 contour levels. Table 3.5 and Table 3.6 report the critical values for S_L and A_L of 5 lags. With $m = 0.1, 0.5$ and 0.9, the histograms for S_t are plotted in Figure 3.1, for A_t in Figure 3.2, for S_C in Figure 3.3, for A_C in Figure 3.4, for S_L in Figure 3.5, for A_L in Figure 3.6.



(a) Histogram of A_t with $m = 0.1$



(b) Histogram of A_t with $m = 0.5$

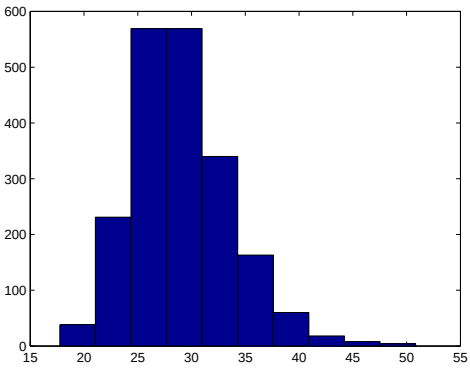


(c) Histogram of A_t with $m = 0.9$

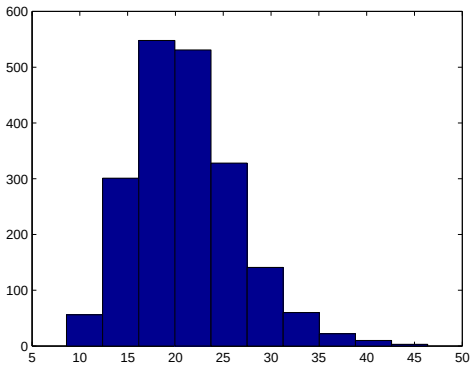
Figure 3.2: Histogram of A_t with $m = 0.1, 0.5$ and 0.9

	m								
Percentile	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
99%	3.950	3.713	3.510	3.396	3.428	3.213	3.089	3.112	2.951
95%	3.502	3.267	3.066	2.926	2.866	2.679	2.537	2.439	2.406
90%	3.292	2.987	2.845	2.642	2.571	2.361	2.240	2.121	2.015
80%	3.020	2.684	2.522	2.372	2.217	2.053	1.935	1.849	1.655
70%	2.843	2.489	2.315	2.156	1.999	1.834	1.697	1.582	1.392
60%	2.702	2.343	2.145	1.983	1.819	1.653	1.529	1.383	1.197
50%	2.569	2.203	2.015	1.830	1.664	1.515	1.365	1.221	1.021
40%	2.457	2.090	1.884	1.684	1.515	1.373	1.212	1.066	0.867
30%	2.339	1.975	1.747	1.540	1.379	1.228	1.071	0.915	0.730
20%	2.201	1.822	1.586	1.384	1.226	1.074	0.923	0.791	0.606
10%	2.033	1.629	1.420	1.199	1.055	0.901	0.761	0.631	0.479
5%	1.922	1.492	1.290	1.065	0.943	0.813	0.663	0.535	0.394
1%	1.730	1.265	1.072	0.829	0.739	0.631	0.498	0.409	0.287

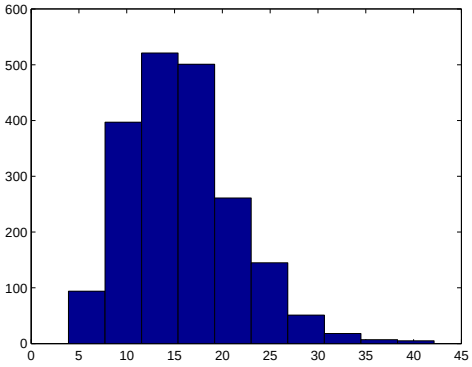
Table 3.1: Critical Values for S_t Statistics



(a) Histogram of S_C with $m = 0.1$



(b) Histogram of S_C with $m = 0.5$

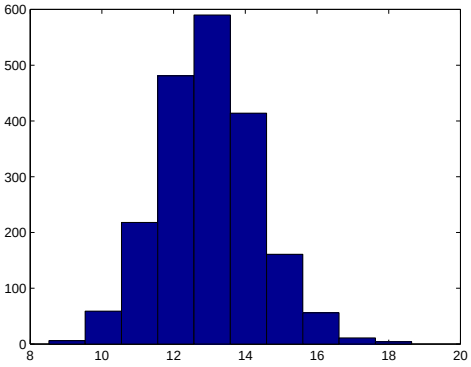


(c) Histogram of S_C with $m = 0.9$

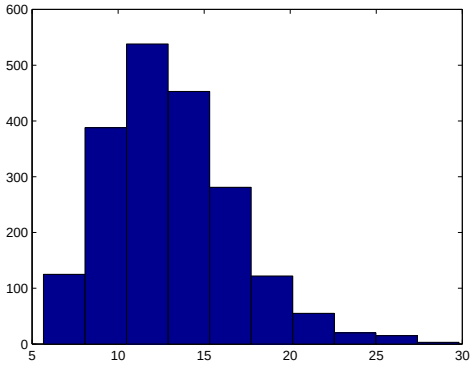
Figure 3.3: Histogram of S_C with $m = 0.1, 0.5$ and 0.9

	m								
Percentile	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
99%	1.241	1.435	1.656	1.824	2.204	2.181	2.360	2.542	2.583
95%	1.078	1.199	1.355	1.487	1.697	1.694	1.774	1.854	1.998
90%	1.004	1.088	1.206	1.300	1.418	1.406	1.490	1.586	1.666
80%	0.918	0.970	1.035	1.101	1.115	1.128	1.183	1.264	1.256
70%	0.870	0.891	0.922	0.948	0.939	0.928	0.966	1.017	1.019
60%	0.825	0.825	0.827	0.837	0.818	0.796	0.807	0.824	0.825
50%	0.789	0.760	0.760	0.741	0.710	0.685	0.682	0.675	0.661
40%	0.752	0.706	0.690	0.653	0.615	0.581	0.551	0.534	0.507
30%	0.711	0.649	0.617	0.568	0.524	0.487	0.453	0.417	0.378
20%	0.673	0.589	0.544	0.497	0.451	0.409	0.358	0.320	0.274
10%	0.608	0.517	0.465	0.409	0.359	0.319	0.269	0.234	0.186
5%	0.567	0.458	0.413	0.342	0.310	0.260	0.219	0.188	0.140
1%	0.478	0.381	0.325	0.251	0.235	0.200	0.152	0.135	0.096

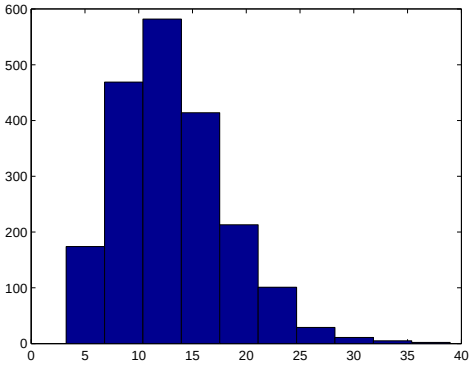
Table 3.2: Critical Values for A_t Statistics



(a) Histogram of A_C with $m = 0.1$



(b) Histogram of A_C with $m = 0.5$

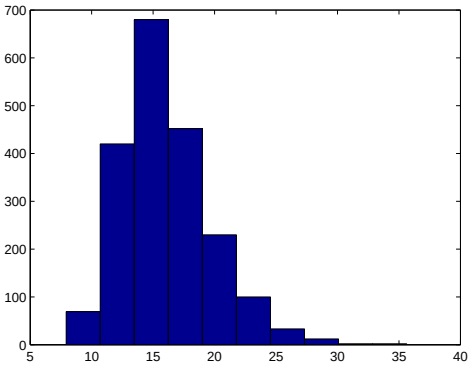


(c) Histogram of A_C with $m = 0.9$

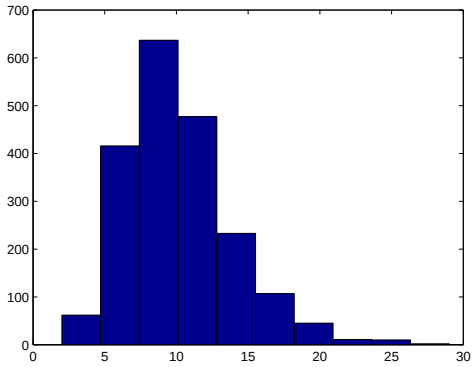
Figure 3.4: Histogram of A_C with $m = 0.1, 0.5$ and 0.9

	m								
Percentile	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
99%	42.488	39.592	38.534	37.694	37.361	35.925	36.051	34.591	31.804
95%	37.159	34.956	32.983	32.277	30.902	29.597	29.008	27.773	25.962
90%	35.155	32.562	30.722	29.378	28.069	27.176	26.181	24.685	23.356
80%	32.624	29.964	28.071	26.673	25.329	24.182	22.935	21.621	20.175
70%	30.932	28.330	26.359	24.906	23.421	22.132	20.932	19.423	18.125
60%	29.689	26.991	24.908	23.416	21.858	20.579	19.179	17.901	16.550
50%	28.540	25.798	23.648	22.111	20.626	19.136	17.708	16.501	15.241
40%	27.482	24.522	22.468	20.900	19.254	17.672	16.440	15.111	13.829
30%	26.508	23.385	21.197	19.510	17.873	16.413	15.058	13.685	12.477
20%	25.351	22.174	19.836	18.193	16.514	14.927	13.582	12.250	10.898
10%	23.731	20.400	18.067	16.333	14.568	13.061	11.741	10.480	8.987
5%	22.376	18.902	16.821	14.879	13.338	11.876	10.383	9.037	7.852
1%	20.394	16.833	14.737	12.807	11.208	9.884	8.275	6.929	5.905

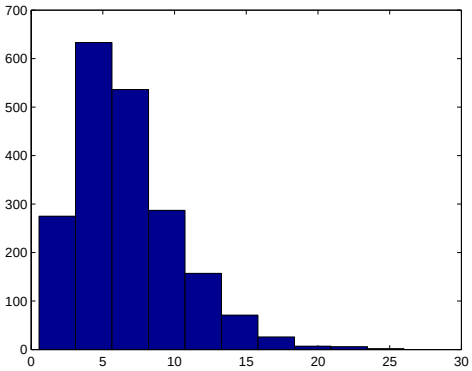
Table 3.3: Critical Values for S_C Statistics



(a) Histogram of S_L with $m = 0.1$



(b) Histogram of S_L with $m = 0.5$

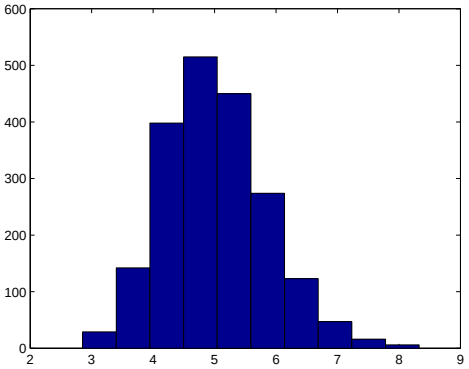


(c) Histogram of S_L with $m = 0.9$

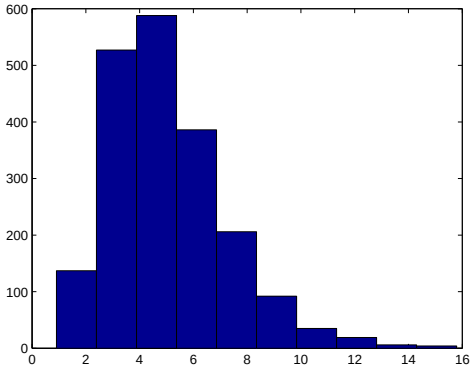
Figure 3.5: Histogram of S_L with $m = 0.1, 0.5$ and 0.9

	m								
Percentile	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
99%	16.342	18.610	20.265	22.301	24.658	26.534	27.422	28.202	27.987
95%	15.298	16.576	17.794	18.927	19.924	20.838	21.866	22.084	22.533
90%	14.762	15.734	16.456	17.287	18.022	18.629	18.996	19.612	19.965
80%	14.099	14.648	15.143	15.477	16.014	16.449	16.761	16.975	17.096
70%	13.672	14.029	14.209	14.439	14.705	14.937	15.107	15.111	15.290
60%	13.295	13.393	13.478	13.566	13.597	13.552	13.629	13.790	13.772
50%	12.955	12.873	12.847	12.820	12.666	12.591	12.534	12.514	12.547
40%	12.623	12.388	12.178	12.014	11.722	11.471	11.290	11.362	11.311
30%	12.262	11.851	11.535	11.227	10.909	10.576	10.364	10.166	10.073
20%	11.828	11.264	10.837	10.423	9.991	9.555	9.288	8.965	8.786
10%	11.298	10.483	9.871	9.330	8.746	8.297	7.860	7.558	7.143
5%	10.840	9.895	9.120	8.436	7.849	7.358	6.750	6.380	5.967
1%	10.030	8.791	7.993	7.279	6.363	5.659	5.069	4.629	4.382

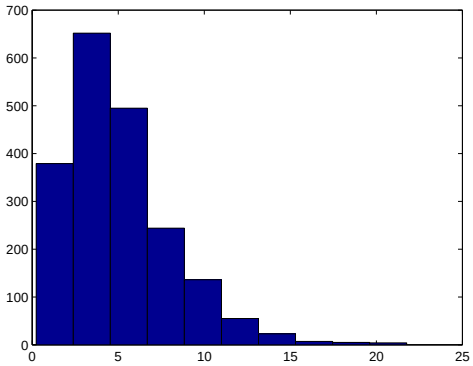
Table 3.4: Critical Values for A_C Statistics



(a) Histogram of A_L with $m = 0.1$



(b) Histogram of A_L with $m = 0.5$

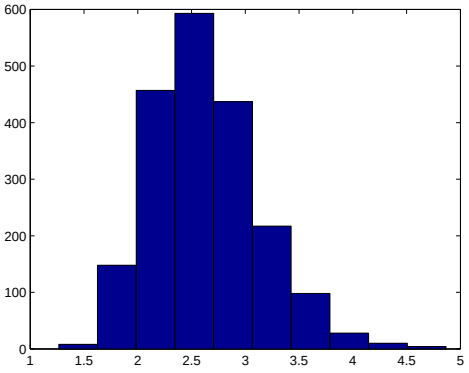


(c) Histogram of A_L with $m = 0.9$

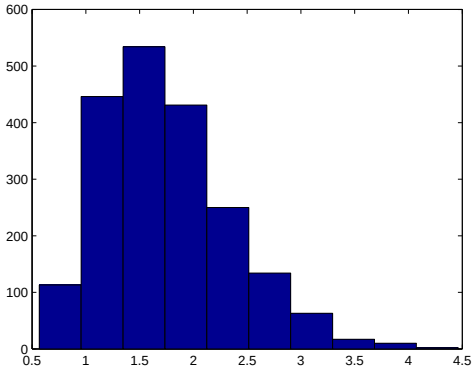
Figure 3.6: Histogram of A_L with $m = 0.1, 0.5$ and 0.9

	m								
percentile	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
99%	26.825	25.069	23.492	23.010	22.057	20.621	19.801	19.370	17.756
95%	22.719	21.121	19.941	18.372	17.199	16.042	15.645	14.470	13.631
90%	20.867	19.109	17.489	16.318	15.203	14.241	13.301	12.552	11.636
80%	18.859	16.979	15.273	14.045	12.859	11.973	11.213	10.267	9.325
70%	17.452	15.505	13.917	12.671	11.665	10.585	9.708	8.788	7.949
60%	16.389	14.279	12.920	11.587	10.492	9.537	8.610	7.765	6.927
50%	14.810	12.425	11.052	9.796	8.827	7.916	6.830	6.071	5.192
40%	15.534	13.300	11.910	10.616	9.576	8.678	7.698	6.917	6.068
30%	13.970	11.675	10.164	8.891	7.999	6.948	6.048	5.213	4.423
20%	13.066	10.697	9.249	8.023	7.039	6.054	5.113	4.388	3.646
10%	11.896	9.535	8.158	7.011	5.863	4.966	4.213	3.484	2.743
5%	11.145	8.544	7.070	6.093	5.099	4.286	3.492	2.786	2.187
1%	9.583	7.267	5.857	4.821	3.950	3.173	2.585	1.907	1.510

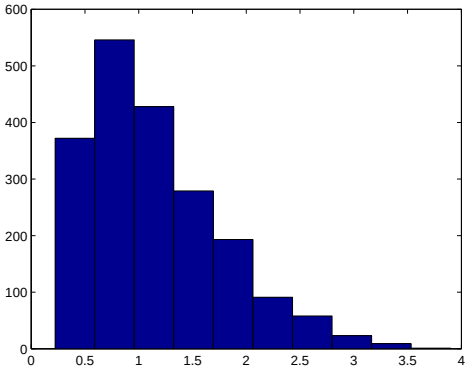
Table 3.5: Critical Values for S_L Statistics



(a) Histogram of S_t with $m = 0.1$



(b) Histogram of S_t with $m = 0.5$



(c) Histogram of S_t with $m = 0.9$

Figure 3.1: Histogram of S_t with $m = 0.1, 0.5$ and 0.9

	m								
percentile	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
99%	7.260	8.419	9.620	10.837	12.021	12.943	13.509	14.458	14.629
95%	6.507	7.336	7.948	8.396	9.045	9.636	10.116	10.569	10.913
90%	6.122	6.673	7.034	7.503	7.961	8.313	8.562	8.929	9.220
80%	5.705	6.038	6.259	6.452	6.643	6.822	6.995	7.104	7.309
70%	5.383	5.583	5.669	5.733	5.820	5.868	5.929	5.977	6.040
60%	5.164	5.208	5.200	5.198	5.212	5.158	5.191	5.122	5.128
50%	4.742	4.604	4.468	4.361	4.224	4.021	3.885	3.817	3.767
40%	4.951	4.890	4.831	4.770	4.663	4.570	4.506	4.452	4.438
30%	4.535	4.316	4.123	3.922	3.717	3.476	3.324	3.174	3.088
20%	4.294	3.988	3.712	3.418	3.174	2.973	2.715	2.584	2.480
10%	4.017	3.539	3.189	2.922	2.602	2.320	2.086	1.927	1.746
5%	3.752	3.237	2.867	2.508	2.246	1.921	1.695	1.459	1.292
1%	3.337	2.681	2.222	1.875	1.621	1.388	1.150	0.960	0.777

Table 3.6: Critical Values for A_L Statistics

3.3.2 Size of the statistics

In this paper, the tests are implemented in out-of-sample context, the importance of parameter uncertainty will depend on the forecasting scheme and on the size of the estimation sample (R) relative to the prediction sample (P). By taking a mean value expansion of the test around the true value, we have

$$\sqrt{P}(\widehat{\alpha}_i(\widehat{\theta}, J)-\alpha_i) = \sqrt{P}(\widehat{\alpha}_i(\theta_0, J)-\alpha_i)+\frac{\sqrt{P}}{\sqrt{R}}\sqrt{R}(\widehat{\theta}-\theta_0)' \lim_{T\to\infty} E(\frac{\partial\widehat{\alpha}_i(\theta, J)}{\partial\theta}|_\theta = \theta_0)+o_p(1)$$

We assume the parameter estimator are $\sqrt{R}$-consistent estimators, i.e. $(\widehat{\theta} - \theta_0) = O_p(R^{-1/2})$ with a well-defined asymptotic distribution. In general, this condition will be easily satisfied such as the quasi maximum-likelihood (QML) estimator. Under assumption 1, $R \to \infty$, $P \to \infty$, and $P/R \to 0$ as $T \to \infty$, parameter uncertainty is asymptotically negligible. Therefore, as long as the ratio of the prediction sample to the estimation sample tends to zero as the total sample size grows to infinity, the test statistics can be applied based on estimated parameters, for which the critical values calculated above can be used directly. In situations where the condition $P/R \to 0$

is violated, we can bootstrap the tests to approximate their asymptotic distribution. The theoretical analysis are same for the fixed, recursive and rolling schemes. In this section, we perform the extensive Monte Carlo simulations to assess the finite sample properties of the tests proposed in this paper and the following Mote Carlos simulation results show that the parametric bootstrap of the asymptotic variance together with standard asymptotic critical values delivers test statistics with very good size and power properties. This is a sensible approach because the null hypothesis fully specifies the parametric data generating process. Specifically, we generate B samples of size T from $f_t^*(y_t|\hat{\theta};\Omega_{t-1})$, which is the predictive density function specified for the random process of interest. Let $\hat{\theta}_b$ denote the estimator under the fixed, rolling or recursive scheme from the bth bootstrap sample, then the statistics are calculated from $f_t^*(y_t|\hat{\theta}_b;\Omega_{t-1})$ and for each statistic, we can bootstrap the distribution. For example, the bootstrap approximation to the p-value for the $S_t^{\alpha_{i,k}}(b)$ for $b = 1, ..., B$, is given by

$$\hat{p}(S_t^{\alpha_{i,k}}) = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left(S_t^{\alpha_{i,k}}(b) > S_t^{\alpha_{i,k}}\right)$$

We consider the following data-generating process (DPG), which does not have the structural break. $y_t = \alpha_1 + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \sigma\epsilon_t$ and $x_t = \phi_1 + \phi_2 x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$, where $\phi_1 = 1.38$, $\phi_2 = 0.77$, $\alpha_1 = 1.5$, $\beta_1 = 0.5$, $\beta_2 = 0.6$, $\sigma = 1$. The information contained in $\{x_t\}$ is known in advance. We evaluate the size of the statistics with different proportion of *rol* in the whole evaluation sample, where the proportion is defined as m and $m = 1/3$, $m = 1/2$ and $m = 2/3$. The simulations are performed in the fixed, rolling and recursive context respectively. The number of Monte Carlo replications is 1000. We set T to be 150, 375 and 750 and consider P as 90, 225 and 450 respectively. The nominal size level is 5%. The set of autocontour coverage levels is given by $C = (0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99)$. Finally, we set

the number of bootstrap replications equal to 500. In Table 3.7, 3.8 and 3.9, we show the size of the test for different sample sizes $T = 150, 375$ and 750, with $m = 1/3$. We can see that the overall performance of the tests is very good. When the sample size is small as 150, the size of the statistics at some extreme contour levels (1% or 99%) are distorted slightly. However, the slight size distortion can be corrected as we increase the sample size. We can observe the similar pattern for $m = 1/2$ from Table 3.13, Table 3.14 and Table 3.15; and $m = 2/3$ from Table 3.10, Table 3.11 and Table 3.12.

T=150, R=90,P=60	$S_t^{1,1}$	$S_t^{1,2}$	$S_t^{1,3}$	$S_t^{1,4}$	$S_t^{1,5}$	$S_t^{1,6}$	$S_t^{1,7}$	$S_t^{1,8}$	$S_t^{1,9}$	$S_t^{1,10}$	$S_t^{1,11}$	$S_t^{1,12}$	$S_t^{1,13}$
fixed	0.042	0.037	0.033	0.046	0.049	0.037	0.04	0.051	0.045	0.045	0.048	0.036	0.039
rolling	0.028	0.04	0.041	0.041	0.38	0.042	0.04	0.04	0.041	0.046	0.044	0.036	0.038
recursive	0.029	0.041	0.042	0.04	0.039	0.039	0.038	0.039	0.044	0.042	0.039	0.035	0.032
	$A_t^{1,1}$	$A_t^{1,2}$	$A_t^{1,3}$	$A_t^{1,4}$	$A_t^{1,5}$	$A_t^{1,6}$	$A_t^{1,7}$	$A_t^{1,8}$	$A_t^{1,9}$	$A_t^{1,10}$	$A_t^{1,11}$	$A_t^{1,12}$	$A_t^{1,13}$
fixed	0.047	0.048	0.049	0.056	0.055	0.058	0.049	0.059	0.049	0.057	0.052	0.055	0.057
rolling	0.032	0.038	0.048	0.043	0.041	0.046	0.046	0.045	0.04	0.046	0.048	0.046	0.031
recursive	0.024	0.037	0.037	0.043	0.043	0.038	0.043	0.052	0.043	0.044	0.042	0.035	0.033
	$S_C^{1,13}$	$A_C^{1,13}$	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_C^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$			
fixed	0.059	0.051	0.06	0.06	0.052	0.045	0.063	0.063	0.058	0.055			
rolling	0.038	0.043	0.045	0.052	0.042	0.042	0.051	0.05	0.045	0.045			
recursive	0.033	0.043	0.044	0.055	0.053	0.052	0.052	0.058	0.039	0.033			
	$S_t^{1,7}$	$S_t^{2,7}$	$S_t^{3,7}$	$S_t^{4,7}$	$S_t^{5,7}$	$A_t^{1,7}$	$A_t^{2,7}$	$A_t^{3,7}$	$A_t^{4,7}$	$A_t^{5,7}$			
fixed	0.04	0.038	0.042	0.04	0.038	0.049	0.053	0.054	0.051	0.039			
rolling	0.04	0.033	0.032	0.033	0.034	0.036	0.038	0.039	0.04	0.032			
recursive	0.038	0.037	0.031	0.035	0.034	0.043	0.039	0.04	0.05	0.037			

Table 3.7: Size of the statistics:T=150 R=90 P=T-R=60 $m = 1/3$ (nominal size 5%)

T=375, R=225,P=150	$S_t^{1,1}$	$S_t^{1,2}$	$S_t^{1,3}$	$S_t^{1,4}$	$S_t^{1,5}$	$S_t^{1,6}$	$S_t^{1,7}$	$S_t^{1,8}$	$S_t^{1,9}$	$S_t^{1,10}$	$S_t^{1,11}$	$S_t^{1,12}$	$S_t^{1,13}$
fixed	0.04	0.04	0.05	0.042	0.048	0.038	0.041	0.05	0.048	0.046	0.055	0.033	0.044
rolling	0.03	0.044	0.042	0.041	0.042	0.044	0.042	0.04	0.041	0.039	0.046	0.031	0.036
recursive	0.031	0.042	0.04	0.04	0.041	0.037	0.042	0.033	0.036	0.042	0.039	0.033	0.046
	$A_t^{1,1}$	$A_t^{1,2}$	$A_t^{1,3}$	$A_t^{1,4}$	$A_t^{1,5}$	$A_t^{1,6}$	$A_t^{1,7}$	$A_t^{1,8}$	$A_t^{1,9}$	$A_t^{1,10}$	$A_t^{1,11}$	$A_t^{1,12}$	$A_t^{1,13}$
fixed	0.055	0.054	0.069	0.049	0.046	0.047	0.053	0.062	0.043	0.047	0.05	0.051	0.05
rolling	0.052	0.056	0.063	0.041	0.043	0.047	0.057	0.061	0.051	0.051	0.047	0.054	0.044
recursive	0.043	0.044	0.04	0.044	0.044	0.048	0.043	0.046	0.039	0.038	0.035	0.042	0.049
	$S_C^{1,13}$	$A_C^{1,13}$	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_C^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$			
fixed	0.046	0.041	0.049	0.047	0.04	0.046	0.054	0.048	0.045	0.051			
rolling	0.05	0.046	0.058	0.054	0.06	0.047	0.058	0.063	0.063	0.06			
recursive	0.06	0.053	0.06	0.056	0.043	0.057	0.055	0.052	0.042	0.048			
	$S_t^{1,7}$	$S_t^{2,7}$	$S_t^{3,7}$	$S_t^{4,7}$	$S_t^{5,7}$	$A_t^{1,7}$	$A_t^{2,7}$	$A_t^{3,7}$	$A_t^{4,7}$	$A_t^{5,7}$			
fixed	0.041	0.043	0.033	0.036	0.042	0.053	0.06	0.052	0.052	0.061			
rolling	0.042	0.04	0.039	0.039	0.048	0.057	0.062	0.066	0.067	0.46			
recursive	0.042	0.038	0.044	0.04	0.036	0.043	0.048	0.052	0.045	0.039			

Table 3.8: Size of the statistics:T=375 R=225 P=T-R=150 $m = 1/3$ (nominal size 5%)

T=750, R=450,P=300	$S_t^{1,1}$	$S_t^{1,2}$	$S_t^{1,3}$	$S_t^{1,4}$	$S_t^{1,5}$	$S_t^{1,6}$	$S_t^{1,7}$	$S_t^{1,8}$	$S_t^{1,9}$	$S_t^{1,10}$	$S_t^{1,11}$	$S_t^{1,12}$	$S_t^{1,13}$
fixed	0.048	0.043	0.057	0.043	0.048	0.057	0.052	0.05	0.047	0.054	0.042	0.051	0.049
rolling	0.042	0.046	0.047	0.05	0.04	0.048	0.046	0.04	0.044	0.04	0.049	0.04	0.041
recursive	0.044	0.048	0.053	0.046	0.043	0.054	0.052	0.04	0.04	0.046	0.048	0.042	0.044
	$A_t^{1,1}$	$A_t^{1,2}$	$A_t^{1,3}$	$A_t^{1,4}$	$A_t^{1,5}$	$A_t^{1,6}$	$A_t^{1,7}$	$A_t^{1,8}$	$A_t^{1,9}$	$A_t^{1,10}$	$A_t^{1,11}$	$A_t^{1,12}$	$A_t^{1,13}$
fixed	0.055	0.042	0.05	0.05	0.061	0.058	0.062	0.046	0.049	0.061	0.057	0.052	0.047
rolling	0.048	0.04	0.043	0.058	0.043	0.047	0.041	0.056	0.049	0.049	0.043	0.04	0.045
recursive	0.041	0.043	0.049	0.051	0.053	0.054	0.051	0.048	0.047	0.048	0.047	0.043	0.055
	$S_C^{1,13}$	$A_C^{1,13}$	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_C^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$			
fixed	0.044	0.045	0.043	0.044	0.05	0.045	0.051	0.044	0.043	0.045			
rolling	0.041	0.045	0.047	0.043	0.046	0.047	0.047	0.044	0.049	0.046			
recursive	0.042	0.049	0.04	0.048	0.046	0.04	0.044	0.04	0.048	0.042			
	$S_t^{1,7}$	$S_t^{2,7}$	$S_t^{3,7}$	$S_t^{4,7}$	$S_t^{5,7}$	$A_t^{1,7}$	$A_t^{2,7}$	$A_t^{3,7}$	$A_t^{4,7}$	$A_t^{5,7}$			
fixed	0.052	0.049	0.05	0.058	0.055	0.062	0.056	0.056	0.057	0.068			
rolling	0.046	0.045	0.041	0.047	0.046	0.041	0.052	0.047	0.04	0.049			
recursive	0.052	0.047	0.04	0.044	0.043	0.051	0.055	0.046	0.041	0.045			

Table 3.9: Size of the statistics:T=750 R=450 P=T-R=300 $m = 1/3$ (nominal size 5%)

T=150, R=90,P=60	$S_t^{1,1}$	$S_t^{1,2}$	$S_t^{1,3}$	$S_t^{1,4}$	$S_t^{1,5}$	$S_t^{1,6}$	$S_t^{1,7}$	$S_t^{1,8}$	$S_t^{1,9}$	$S_t^{1,10}$	$S_t^{1,11}$	$S_t^{1,12}$	$S_t^{1,13}$
fixed	0.032	0.032	0.039	0.03	0.043	0.039	0.04	0.043	0.039	0.035	0.032	0.041	0.03
rolling	0.023	0.025	0.042	0.046	0.048	0.044	0.042	0.042	0.044	0.039	0.035	0.03	0.026
recursive	0.021	0.027	0.041	0.046	0.047	0.042	0.43	0.047	0.039	0.42	0.044	0.026	0.27
	$A_t^{1,1}$	$A_t^{1,2}$	$A_t^{1,3}$	$A_t^{1,4}$	$A_t^{1,5}$	$A_t^{1,6}$	$A_t^{1,7}$	$A_t^{1,8}$	$A_t^{1,9}$	$A_t^{1,10}$	$A_t^{1,11}$	$A_t^{1,12}$	$A_t^{1,13}$
fixed	0.042	0.04	0.042	0.041	0.042	0.041	0.048	0.043	0.053	0.042	0.043	0.046	0.049
rolling	0.032	0.043	0.032	0.03	0.04	0.035	0.036	0.032	0.034	0.046	0.043	0.045	0.038
recursive	0.031	0.034	0.043	0.031	0.04	0.04	0.044	0.053	0.04	0.048	0.041	0.034	0.038
	$S_C^{1,13}$	$A_C^{1,13}$	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_C^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$			
fixed	0.044	0.048	0.053	0.044	0.049	0.05	0.05	0.044	0.046	0.05			
rolling	0.038	0.041	0.039	0.045	0.044	0.053	0.049	0.047	0.033	0.039			
recursive	0.039	0.037	0.05	0.053	0.049	0.047	0.048	0.059	0.039	0.042			
	$S_t^{1,7}$	$S_t^{2,7}$	$S_t^{3,7}$	$S_t^{4,7}$	$S_t^{5,7}$	$A_t^{1,7}$	$A_t^{2,7}$	$A_t^{3,7}$	$A_t^{4,7}$	$A_t^{5,7}$			
fixed	0.04	0.043	0.052	0.048	0.042	0.048	0.046	0.043	0.056	0.041			
rolling	0.02	0.028	0.054	0.055	0.045	0.046	0.048	0.038	0.035	0.035			
recursive	0.03	0.032	0.029	0.033	0.028	0.044	0.046	0.045	0.049	0.047			

Table 3.10: Size of the statistics:T=150 R=90 P=T-R=60 $m = 2/3$ (nominal size 5%)

T=375, R=225,P=150	$S_t^{1,1}$	$S_t^{1,2}$	$S_t^{1,3}$	$S_t^{1,4}$	$S_t^{1,5}$	$S_t^{1,6}$	$S_t^{1,7}$	$S_t^{1,8}$	$S_t^{1,9}$	$S_t^{1,10}$	$S_t^{1,11}$	$S_t^{1,12}$	$S_t^{1,13}$
fixed	0.058	0.064	0.042	0.063	0.044	0.041	0.042	0.052	0.041	0.039	0.067	0.043	0.051
rolling	0.036	0.055	0.064	0.059	0.062	0.05	0.047	0.044	0.044	0.065	0.048	0.044	0.038
recursive	0.035	0.04	0.043	0.049	0.047	0.046	0.04	0.043	0.047	0.042	0.039	0.04	0.042
	$A_t^{1,1}$	$A_t^{1,2}$	$A_t^{1,3}$	$A_t^{1,4}$	$A_t^{1,5}$	$A_t^{1,6}$	$A_t^{1,7}$	$A_t^{1,8}$	$A_t^{1,9}$	$A_t^{1,10}$	$A_t^{1,11}$	$A_t^{1,12}$	$A_t^{1,13}$
fixed	0.052	0.044	0.046	0.041	0.043	0.042	0.045	0.045	0.04	0.051	0.049	0.047	0.044
rolling	0.039	0.041	0.049	0.048	0.039	0.047	0.043	0.044	0.046	0.044	0.039	0.032	0.04
recursive	0.035	0.038	0.052	0.058	0.044	0.041	0.047	0.048	0.058	0.043	0.036	0.34	0.04
	$S_C^{1,13}$	$A_C^{1,13}$	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_C^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$			
fixed	0.047	0.05	0.062	0.055	0.052	0.048	0.055	0.054	0.064	0.047			
rolling	0.04	0.038	0.06	0.057	0.044	0.036	0.056	0.053	0.051	0.042			
recursive	0.047	0.043	0.068	0.054	0.045	0.04	0.067	0.065	0.062	0.048			
	$S_t^{1,7}$	$S_t^{2,7}$	$S_t^{3,7}$	$S_t^{4,7}$	$S_t^{5,7}$	$A_t^{1,7}$	$A_t^{2,7}$	$A_t^{3,7}$	$A_t^{4,7}$	$A_t^{5,7}$			
fixed	0.042	0.045	0.049	0.041	0.049	0.045	0.043	0.045	0.052	0.054			
rolling	0.037	0.044	0.042	0.039	0.036	0.037	0.043	0.048	0.042	0.037			
recursive	0.04	0.049	0.038	0.038	0.036	0.047	0.054	0.043	0.05	0.051			

Table 3.11: Size of the statistics:T=375 R=225 P=T-R=150 $m = 2/3$ (nominal size 5%)

T=750, R=450,P=300	$S_t^{1,1}$	$S_t^{1,2}$	$S_t^{1,3}$	$S_t^{1,4}$	$S_t^{1,5}$	$S_t^{1,6}$	$S_t^{1,7}$	$S_t^{1,8}$	$S_t^{1,9}$	$S_t^{1,10}$	$S_t^{1,11}$	$S_t^{1,12}$	$S_t^{1,13}$
fixed	0.048	0.040	0.044	0.049	0.046	0.051	0.052	0.048	0.052	0.04	0.042	0.05	0.032
rolling	0.049	0.041	0.056	0.051	0.048	0.041	0.041	0.045	0.057	0.056	0.049	0.038	0.039
recursive	0.046	0.048	0.045	0.044	0.051	0.048	0.044	0.045	0.047	0.057	0.058	0.039	0.04
	$A_t^{1,1}$	$A_t^{1,2}$	$A_t^{1,3}$	$A_t^{1,4}$	$A_t^{1,5}$	$A_t^{1,6}$	$A_t^{1,7}$	$A_t^{1,8}$	$A_t^{1,9}$	$A_t^{1,10}$	$A_t^{1,11}$	$A_t^{1,12}$	$A_t^{1,13}$
fixed	0.049	0.045	0.049	0.051	0.046	0.049	0.051	0.043	0.049	0.052	0.043	0.047	0.049
rolling	0.048	0.042	0.046	0.055	0.044	0.047	0.043	0.044	0.044	0.043	0.049	0.044	0.048
recursive	0.04	0.043	0.041	0.063	0.065	0.045	0.053	0.045	0.044	0.051	0.042	0.044	0.043
	$S_C^{1,13}$	$A_C^{1,13}$	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_C^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$			
fixed	0.045	0.049	0.047	0.044	0.053	0.05	0.049	0.042	0.046	0.051			
rolling	0.045	0.044	0.043	0.064	0.063	0.044	0.048	0.041	0.044	0.05			
recursive	0.043	0.047	0.045	0.057	0.052	0.048	0.048	0.048	0.051	0.048			
	$S_t^{1,7}$	$S_t^{2,7}$	$S_t^{3,7}$	$S_t^{4,7}$	$S_t^{5,7}$	$A_t^{1,7}$	$A_t^{2,7}$	$A_t^{3,7}$	$A_t^{4,7}$	$A_t^{5,7}$			
fixed	0.052	0.053	0.042	0.042	0.043	0.051	0.048	0.042	0.054	0.049			
rolling	0.041	0.048	0.043	0.039	0.04	0.043	0.043	0.047	0.049	0.043			
recursive	0.044	0.051	0.041	0.038	0.041	0.053	0.054	0.043	0.045	0.048			

Table 3.12: Size of the statistics:T=750 R=450 P=T-R=350 $m = 2/3$ (nominal size 5%)

T=150, R=90,P=60	$S_t^{1,1}$	$S_t^{1,2}$	$S_t^{1,3}$	$S_t^{1,4}$	$S_t^{1,5}$	$S_t^{1,6}$	$S_t^{1,7}$	$S_t^{1,8}$	$S_t^{1,9}$	$S_t^{1,10}$	$S_t^{1,11}$	$S_t^{1,12}$	$S_t^{1,13}$
fixed	0.021	0.032	0.027	0.034	0.042	0.045	0.039	0.041	0.041	0.035	0.039	0.028	0.022
rolling	0.023	0.033	0.033	0.033	0.036	0.042	0.036	0.04	0.037	0.037	0.035	0.024	0.025
recursive	0.02	0.025	0.025	0.029	0.024	0.034	0.031	0.036	0.021	0.019	0.02	0.021	0.019
	$A_t^{1,1}$	$A_t^{1,2}$	$A_t^{1,3}$	$A_t^{1,4}$	$A_t^{1,5}$	$A_t^{1,6}$	$A_t^{1,7}$	$A_t^{1,8}$	$A_t^{1,9}$	$A_t^{1,10}$	$A_t^{1,11}$	$A_t^{1,12}$	$A_t^{1,13}$
fixed	0.048	0.051	0.049	0.041	0.049	0.053	0.042	0.053	0.049	0.044	0.039	0.055	0.049
rolling	0.051	0.054	0.047	0.044	0.044	0.051	0.036	0.05	0.052	0.046	0.048	0.058	0.047
recursive	0.038	0.047	0.038	0.039	0.045	0.044	0.045	0.054	0.04	0.041	0.039	0.034	0.041
	$S_C^{1,13}$	$A_C^{1,13}$	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_C^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$			
fixed	0.046	0.053	0.047	0.042	0.059	0.058	0.048	0.052	0.051	0.053			
rolling	0.049	0.054	0.039	0.046	0.061	0.055	0.046	0.055	0.054	0.057			
recursive	0.035	0.041	0.055	0.058	0.054	0.044	0.056	0.056	0.055	0.051			
	$S_t^{1,7}$	$S_t^{2,7}$	$S_t^{3,7}$	$S_t^{4,7}$	$S_t^{5,7}$	$A_t^{1,7}$	$A_t^{2,7}$	$A_t^{3,7}$	$A_t^{4,7}$	$A_t^{5,7}$			
fixed	0.039	0.032	0.03	0.029	0.035	0.042	0.047	0.048	0.05	0.036			
rolling	0.026	0.033	0.048	0.042	0.048	0.056	0.047	0.046	0.052	0.034			
recursive	0.031	0.029	0.042	0.045	0.041	0.045	0.044	0.047	0.05	0.037			

Table 3.13: Size of the statistics:T=150 R=90 P=T-R=60 $m = 1/2$ (nominal size 5%)

3.3.3 Power of the statistics

To study the power properties of the tests we consider the following model as a null hypothesis model: $y_t = \alpha_1 + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \sigma\epsilon_t$ and $x_t = \phi_1 + \phi_2 x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$, where $\phi_1 = 1.38$, $\phi_2 = 0.77$, $\alpha_1 = 1.5$, $\beta_1 = 0.5$, $\beta_2 = 0.6$, $\sigma = 1$. The total sample size (T) is 650, $R = 350$, $m = 1/3$. We consider the break point at $R + \tau P$, where $\tau = 1/3$. The nominal size is 5%. The number of Monte Carlo replication is 1000 and bootstrapping is 500.

T=375, R=225,P=150	$S_t^{1,1}$	$S_t^{1,2}$	$S_t^{1,3}$	$S_t^{1,4}$	$S_t^{1,5}$	$S_t^{1,6}$	$S_t^{1,7}$	$S_t^{1,8}$	$S_t^{1,9}$	$S_t^{1,10}$	$S_t^{1,11}$	$S_t^{1,12}$	$S_t^{1,13}$
fixed	0.042	0.04	0.039	0.042	0.041	0.047	0.044	0.045	0.038	0.040	0.040	0.041	0.033
rolling	0.03	0.046	0.048	0.059	0.066	0.043	0.059	0.043	0.039	0.059	0.059	0.067	0.026
recursive	0.021	0.03	0.041	0.059	0.63	0.067	0.042	0.043	0.031	0.041	0.047	0.036	0.022
	$A_t^{1,1}$	$A_t^{1,2}$	$A_t^{1,3}$	$A_t^{1,4}$	$A_t^{1,5}$	$A_t^{1,6}$	$A_t^{1,7}$	$A_t^{1,8}$	$A_t^{1,9}$	$A_t^{1,10}$	$A_t^{1,11}$	$A_t^{1,12}$	$A_t^{1,13}$
fixed	0.051	0.5	0.049	0.042	0.048	0.043	0.045	0.055	0.049	0.048	0.047	0.046	0.039
rolling	0.05	0.048	0.045	0.049	0.047	0.049	0.038	0.042	0.047	0.046	0.043	0.041	0.038
recursive	0.046	0.043	0.044	0.039	0.047	0.042	0.047	0.052	0.047	0.047	0.039	0.045	0.036
	$S_C^{1,13}$	$A_C^{1,13}$	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_C^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$			
fixed	0.042	0.043	0.052	0.05	0.043	0.044	0.05	0.048	0.051	0.043			
rolling	0.041	0.048	0.051	0.05	0.044	0.042	0.049	0.047	0.052	0.043			
recursive	0.049	0.047	0.066	0.053	0.043	0.049	0.059	0.055	0.051	0.042			
	$S_t^{1,7}$	$S_t^{2,7}$	$S_t^{3,7}$	$S_t^{4,7}$	$S_t^{5,7}$	$A_t^{1,7}$	$A_t^{2,7}$	$A_t^{3,7}$	$A_t^{4,7}$	$A_t^{5,7}$			
fixed	0.044	0.041	0.043	0.039	0.044	0.045	0.048	0.052	0.052	0.049			
rolling	0.039	0.067	0.045	0.056	0.044	0.048	0.045	0.053	0.042	0.046			
recursive	0.042	0.046	0.042	0.046	0.064	0.047	0.053	0.054	0.051	0.05			

Table 3.14: Size of the statistics:T=375 R=225 P=T-R=150 $m = 1/2$ (nominal size 5%)

T=750, R=450,P=300	$S_t^{1,1}$	$S_t^{1,2}$	$S_t^{1,3}$	$S_t^{1,4}$	$S_t^{1,5}$	$S_t^{1,6}$	$S_t^{1,7}$	$S_t^{1,8}$	$S_t^{1,9}$	$S_t^{1,10}$	$S_t^{1,11}$	$S_t^{1,12}$	$S_t^{1,13}$
fixed	0.045	0.047	0.041	0.045	0.045	0.041	0.048	0.047	0.043	0.045	0.042	0.043	0.048
rolling	0.045	0.047	0.048	0.049	0.043	0.047	0.04	0.043	0.04	0.041	0.042	0.048	0.048
recursive	0.045	0.044	0.049	0.046	0.04	0.052	0.044	0.041	0.046	0.046	0.043	0.057	0.045
	$A_t^{1,1}$	$A_t^{1,2}$	$A_t^{1,3}$	$A_t^{1,4}$	$A_t^{1,5}$	$A_t^{1,6}$	$A_t^{1,7}$	$A_t^{1,8}$	$A_t^{1,9}$	$A_t^{1,10}$	$A_t^{1,11}$	$A_t^{1,12}$	$A_t^{1,13}$
fixed	0.05	0.049	0.05	0.048	0.043	0.043	0.048	0.052	0.041	0.048	0.044	0.046	0.049
rolling	0.047	0.049	0.043	0.061	0.044	0.052	0.042	0.041	0.044	0.043	0.044	0.041	0.051
recursive	0.049	0.042	0.043	0.05	0.053	0.048	0.052	0.046	0.057	0.045	0.058	0.045	0.044
	$S_C^{1,13}$	$A_C^{1,13}$	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_C^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$			
fixed	0.052	0.047	0.05	0.049	0.048	0.046	0.05	0.049	0.048	0.047			
rolling	0.046	0.047	0.053	0.062	0.05	0.043	0.046	0.043	0.045	0.045			
recursive	0.043	0.047	0.049	0.052	0.057	0.044	0.047	0.046	0.045	0.046			
	$S_t^{1,7}$	$S_t^{2,7}$	$S_t^{3,7}$	$S_t^{4,7}$	$S_t^{5,7}$	$A_t^{1,7}$	$A_t^{2,7}$	$A_t^{3,7}$	$A_t^{4,7}$	$A_t^{5,7}$			
fixed	0.048	0.054	0.047	0.047	0.042	0.048	0.049	0.048	0.048	0.051			
rolling	0.040	0.049	0.04	0.037	0.046	0.042	0.044	0.048	0.049	0.044			
recursive	0.044	0.051	0.045	0.044	0.048	0.052	0.054	0.049	0.045	0.046			

Table 3.15: Size of the statistics:T=750 R=450 P=T-R=300 $m = 1/2$ (nominal size 5%)

We consider following four data generating mechanisms:

DGP1: With pre-known structural break in intercept $y_t = \alpha_t + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \sigma\epsilon_t$, $\epsilon_t \sim N(0,1)$

$$\alpha_t = \begin{cases} \alpha_1 = 1.5 & \text{if } t < \text{break} \\ \alpha_2 = 2 & \text{otherwise} \end{cases}$$

$\beta_1 = 0.5$, $\beta_2 = 0.6$, $\sigma = 1$. In DGP1, we consider the structural break to the intercept only.

DGP2: With pre-known structural break in variance

$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \sigma_t\epsilon_t$, $\epsilon_t \sim N(0,1)$

$$\sigma_t = \begin{cases} \sigma_1 = 1.5 & \text{if } t < \text{break} \\ \sigma_2 = 1.8 & \text{otherwise} \end{cases}$$

$\alpha = 1.5$, $\beta_1 = 0.5$, $\beta_2 = 0.6$. In DGP2, we consider the structural break to the variance only.

DGP3: With pre-known structural break in slope coefficients

$y_t = \alpha + \beta_{1,t} y_{t-1} + \beta_{2,t} x_{t-1} + \sigma\epsilon_t$, $\epsilon_t \sim N(0,1)$

$$\beta_{1,t} = \begin{cases} \beta_{1,1} = 0.5 & \text{if } t < \text{break} \\ \beta_{1,2} = 0.3 & \text{otherwise} \end{cases}$$

$$\beta_{2,t} = \begin{cases} \beta_{2,1} = 0.6 & \text{if } t < \text{break} \\ \beta_{2,2} = 0.4 & \text{otherwise} \end{cases}$$

$\alpha = 1.5$, $\sigma = 1$. In DGP3, we consider the structural break to the coefficients only.

DGP4: With pre-known structural break in intercept, variance and slope coefficients

$y_t = \alpha_t + \beta_{1,t} y_{t-1} + \beta_{2,t} x_{t-1} + \sigma_t\epsilon_t$, $\epsilon_t \sim N(0,1)$

$$\alpha_t = \begin{cases} \alpha_1 = 1.5 & \text{if } t < \text{break} \\ \alpha_2 = 2 & \text{otherwise} \end{cases}$$

$$\sigma_t = \begin{cases} \sigma_1 = 1.5 & \text{if } t < \text{break} \\ \sigma_2 = 1.8 & \text{otherwise} \end{cases}$$

$$\beta_{1,t} = \begin{cases} \beta_{1,1} = 0.5 & \text{if } t < \text{break} \\ \beta_{1,2} = 0.3 & \text{otherwise} \end{cases}$$

$$\beta_{2,t} = \begin{cases} \beta_{2,1} = 0.6 & \text{if } t < \text{break} \\ \beta_{2,2} = 0.4 & \text{otherwise} \end{cases}$$

In DGP4, we consider the structural break to the intercept, coefficients and variance. For all these four cases, we analyze the behavior of the S_t and A_t statistics for 13 autocontours that span the entire uniform density and lag $l = 1, 2, 3, 4$ and 5 respectively. In addition, we present the power of the portmanteaus statistics $S_C^{l,13}$, $A_C^{l,13}$, $S_L^{7,l}$ and $A_L^{7,l}$ for several values of l. Table 3.16, Table 3.17, Table 3.18 and Table 3.19 report the power of the statistics for the above four different breaks with fixed estimation scheme respectively. We present the power of the statistics in Table 3.20, 3.21, 3.22, and 3.23. In general, the power is excellent across model and sample size under these three estimation schemes. The power of our statistics under DGP2 which specifies the break at variance is comparatively lower than the other three cases. However, the break at variance is very small and the power for the portmanteaus statistics are still substantial enough for us to reject the null. For DGP4, our tests have the strongest power overall since it accumulates all the four different types of breaks.

Fixed Scheme	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.14	0.45	0.765	0.956	0.979	0.987	0.995	0.992	0.987	0.976	0.938	0.867	0.589
$l=2$	0.14	0.464	0.789	0.955	0.981	0.99	0.993	0.993	0.988	0.974	0.943	0.864	0.593
$l=3$	0.14	0.419	0.792	0.965	0.985	0.997	0.997	0.994	0.985	0.971	0.937	0.871	0.594
$l=4$	0.11	0.439	0.797	0.953	0.986	0.993	0.994	0.994	0.99	0.979	0.938	0.866	0.595
$l=5$	0.12	0.453	0.796	0.952	0.984	0.992	0.994	0.993	0.989	0.971	0.938	0.867	0.598
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.08	0.244	0.698	0.898	0.948	0.969	0.981	0.978	0.968	0.945	0.904	0.812	0.588
$l=2$	0.09	0.277	0.741	0.896	0.952	0.965	0.979	0.983	0.972	0.947	0.898	0.815	0.585
$l=3$	0.07	0.276	0.72	0.893	0.941	0.969	0.98	0.984	0.967	0.943	0.906	0.815	0.588
$l=4$	0.07	0.319	0.702	0.887	0.957	0.971	0.979	0.981	0.971	0.944	0.904	0.817	0.586
$l=5$	0.08	0.339	0.722	0.906	0.951	0.968	0.985	0.976	0.972	0.944	0.903	0.816	0.584
	$S_C^{13,l}$	$A_C^{13,1}$											
$l=1$	0.961	0.921											
$l=2$	0.965	0.924											
$l=3$	0.967	0.923											
$l=4$	0.963	0.921											
$l=5$	0.965	0.921											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C=7$	0.989	0.985	0.976	0.964	0.98	0.969	0.956	0.942					

Table 3.16: Power for DGP1: Fixed Scheme

Fixed Scheme	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.292	0.271	0.191	0.084	0.073	0.109	0.281	0.278	0.391	0.512	0.638	0.638	0.518
$l=2$	0.26	0.239	0.171	0.079	0.075	0.106	0.282	0.292	0.418	0.518	0.637	0.639	0.52
$l=3$	0.285	0.265	0.175	0.089	0.075	0.12	0.283	0.29	0.404	0.53	0.635	0.639	0.523
$l=4$	0.311	0.26	0.162	0.088	0.063	0.096	0.282	0.284	0.399	0.512	0.624	0.636	0.526
$l=5$	0.273	0.253	0.158	0.068	0.068	0.108	0.282	0.288	0.408	0.539	0.645	0.638	0.531
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.345	0.242	0.162	0.09	0.078	0.127	0.275	0.271	0.361	0.472	0.561	0.604	0.527
$l=2$	0.327	0.225	0.16	0.081	0.077	0.12	0.269	0.277	0.377	0.467	0.56	0.606	0.522
$l=3$	0.347	0.239	0.167	0.089	0.087	0.123	0.272	0.286	0.374	0.468	0.558	0.6	0.527
$l=4$	0.368	0.246	0.159	0.09	0.069	0.111	0.272	0.277	0.345	0.455	0.55	0.592	0.523
$l=5$	0.317	0.222	0.146	0.086	0.072	0.12	0.273	0.288	0.37	0.469	0.564	0.6	0.523
	$S_C^{l,13}$	$A_C^{1,13}$											
$l=1$	0.659	0.595											
$l=2$	0.651	0.604											
$l=3$	0.666	0.625											
$l=4$	0.644	0.625											
$l=5$	0.652	0.611											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C=7$	0.286	0.285	0.281	0.285	0.271	0.273	0.276	0.275					

Table 3.17: Power for DGP2: Fixed Scheme

Fixed Scheme	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.95	0.998	0.999	0.999	0.999	1	0.97	0.985	0.961	0.849	0.449	0.33	0.29
$l=2$	0.96	0.998	0.999	1	0.999	0.997	0.982	0.981	0.96	0.859	0.444	0.16	0.19
$l=3$	0.97	0.997	0.998	1	0.999	0.998	0.979	0.982	0.956	0.85	0.436	0.35	0.31
$l=4$	0.96	0.997	0.998	1	0.998	0.998	0.98	0.984	0.965	0.853	0.446	0.36	0.13
$l=5$	0.96	0.997	1	0.999	1	0.999	0.982	0.983	0.96	0.854	0.447	0.2	0.23
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.93	0.951	1	0.999	0.999	0.995	0.95	0.973	0.914	0.475	0.38	0.25	0.16
$l=2$	0.931	0.967	0.998	1	0.999	0.994	0.952	0.974	0.896	0.518	0.34	0.22	0.17
$l=3$	0.937	0.98	0.998	0.999	0.997	0.995	0.95	0.973	0.883	0.555	0.36	0.24	0.19
$l=4$	0.935	0.96	1	0.998	0.997	0.994	0.95	0.968	0.908	0.487	0.31	0.32	0.22
$l=5$	0.939	0.95	0.999	0.998	0.997	0.995	0.95	0.962	0.875	0.545	0.31	0.12	0.23
	$S_C^{l,13}$	$A_C^{1,13}$											
$l=1$	1	0.97											
$l=2$	1	0.98											
$l=3$	0.999	0.97											
$l=4$	1	0.999											
$l=5$	1	1											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C=7$	0.98	0.98	0.98	0.98	0.97	0.969	0.97	0.97					

Table 3.18: Power for DGP3: Fixed Scheme

Fixed Scheme	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	1	1	0.999	1	1	1	0.999	0.999	0.989	0.984	0.969	0.93	0.79
$l=2$	0.999	1	0.999	0.996	1	1	1	1	0.98	0.983	0.967	0.901	0.78
$l=3$	0.998	1	1	1	1	0.999	1	1	0.996	0.98	0.965	0.912	0.781
$l=4$	1	0.999	1	1	0.998	1	0.999	1	0.99	0.98	0.962	0.923	0.744
$l=5$	1	0.996	1	1	1	1	1	0.998	1	0.98	0.966	0.917	0.775
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.989	1	0.995	1	1	1	0.98	0.998	0.989	0.96	0.914	0.812	0.639
$l=2$	0.989	1	0.995	0.995	0.997	1	0.988	0.989	0.986	0.967	0.934	0.815	0.633
$l=3$	0.99	1	1	1	1	0.997	0.985	0.99	0.986	0.966	0.914	0.815	0.622
$l=4$	1	0.996	1	1	0.996	0..	0.989	0.991	0.986	0.967	0.954	0.817	0.644
$l=5$	1	0.996	1	1	1	1	0.989	0.996	0.986	0.97	0.914	0.816	0.675
	$S_C^{l,13}$	$A_C^{1,13}$											
$l=1$	1	1											
$l=2$	1	1											
$l=3$	1	1											
$l=4$	1	1											
$l=5$	1	1											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C=7$	1	1	1	1	1	1	1	1					

Table 3.19: Power for DGP4: Fixed Scheme

Rolling Scheme	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.18	0.12	0.308	0.416	0.459	0.394	0.42	0.264	0.21	0.08	0.119	0.092	0.08
$l=2$	0.11	0.11	0.295	0.324	0.323	0.431	0.433	0.333	0.193	0.084	0.123	0.093	0.08
$l=3$	0.11	0.11	0.258	0.402	0.386	0.39	0.472	0.325	0.207	0.096	0.123	0.094	0.09
$l=4$	0.13	0.11	0.289	0.439	0.41	0.413	0.378	0.325	0.173	0.093	0.121	0.092	0.082
$l=5$	0.16	0.11	0.262	0.445	0.418	0.423	0.438	0.318	0.171	0.091	0.121	0.092	0.083
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.25	0.318	0.559	0.783	0.738	0.710	0.534	0.634	0.418	0.18	0.21	0.102	0.091
$l=2$	0.22	0.292	0.527	0.733	0.71	0.714	0.56	0.621	0.438	0.17	0.216	0.100	0.092
$l=3$	0.24	0.275	0.521	0.768	0.756	0.744	0.499	0.638	0.442	0.155	0.224	0.103	0.092
$l=4$	0.25	0.28	0.519	0.806	0.775	0.759	0.611	0.638	0.421	0.133	0.21	0.101	0.092
$l=5$	0.25	0.305	0.516	0.797	0.781	0.769	0.533	0.626	0.392	0.121	0.24	0.101	0.093
	$S_C^{l,13}$	$A_C^{1,13}$											
$l=1$	0.638	0.749											
$l=2$	0.521	0.733											
$l=3$	0.619	0.801											
$l=4$	0.641	0.822											
$l=5$	0.637	0.803											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C=7$	0.414	0.419	0.408	0.432	0.563	0.59	0.578	0.532					

Table 3.20: Power for DGP1: Rolling Scheme

Rolling Scheme	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.116	0.104	0.088	0.063	0.05	0.056	0.112	0.11	0.171	0.341	0.302	0.264	0.261
$l=2$	0.114	0.09	0.086	0.063	0.054	0.051	0.118	0.113	0.172	0.318	0.296	0.264	0.262
$l=3$	0.107	0.094	0.076	0.065	0.06	0.053	0.108	0.126	0.189	0.335	0.28	0.273	0.266
$l=4$	0.113	0.104	0.072	0.054	0.058	0.049	0.175	0.114	0.167	0.329	0.287	0.266	0.267
$l=5$	0.109	0.093	0.08	0.053	0.05	0.053	0.16	0.121	0.184	0.337	0.274	0.265	0.267
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.179	0.129	0.099	0.078	0.066	0.075	0.19	0.151	0.204	0.36	0.344	0.35	0.378
$l=2$	0.169	0.112	0.082	0.075	0.066	0.076	0.18	0.145	0.205	0.386	0.345	0.343	0.379
$l=3$	0.175	0.139	0.092	0.07	0.067	0.077	0.196	0.159	0.228	0.378	0.352	0.348	0.378
$l=4$	0.183	0.143	0.083	0.063	0.066	0.077	0.192	0.144	0.201	0.358	0.342	0.352	0.375
$l=5$	0.159	0.129	0.09	0.063	0.066	0.065	0.191	0.151	0.208	0.371	0.347	0.35	0.377
	$S_C^{l,13}$	$A_C^{1,13}$											
$l=1$	0.409	0.414											
$l=2$	0.363	0.394											
$l=3$	0.381	0.427											
$l=4$	0.398	0.431											
$l=5$	0.306	0.318											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C=7$	0.168	0.115	0.113	0.112	0.129	0.189	0.191	0.182					

Table 3.21: Power for DGP2: Rolling Scheme

Rolling Scheme	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.46	0.461	0.395	0.184	0.26	0.329	0.396	0.457	0.401	0.134	0.212	0.092	0.08
$l=2$	0.358	0.32	0.338	0.381	0.421	0.475	0.511	0.508	0.39	0.121	0.241	0.093	0.08
$l=3$	0.437	0.398	0.389	0.442	0.465	0.524	0.524	0.518	0.385	0.099	0.223	0.094	0.09
$l=4$	0.445	0.425	0.444	0.481	0.502	0.544	0.54	0.512	0.345	0.095	0.232	0.092	0.082
$l=5$	0.448	0.409	0.425	0.468	0.511	0.54	0.546	0.506	0.361	0.075	0.232	0.092	0.083
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.189	0.219	0.332	0.473	0.584	0.678	0.726	0.739	0.706	0.395	0.163	0.082	0.06
$l=2$	0.585	0.628	0.686	0.788	0.808	0.834	0.826	0.815	0.716	0.356	0.171	0.081	0.06
$l=3$	0.68	0.711	0.728	0.796	0.832	0.839	0.832	0.811	0.7	0.334	0.163	0.081	0.07
$l=4$	0.684	0.733	0.765	0.832	0.851	0.861	0.846	0.815	0.698	0.308	0.162	0.081	0.062
$l=5$	0.674	0.717	0.762	0.82	0.846	0.86	0.848	0.814	0.669	0.291	0.161	0.081	0.063
	$S_C^{l,13}$	$A_C^{1,13}$											
$l=1$	0.57	0.817											
$l=2$	0.539	0.845											
$l=3$	0.53	0.814											
$l=4$	0.532	0.78											
$l=5$	0.534	0.782											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
	0.477	0.496	0.48	0.421	0.728	0.75	0.749	0.703					

Table 3.22: Power for DGP3: Rolling Scheme

Rolling Scheme	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.569	0.461	0.486	0.589	0.696	0.746	0.76	0.729	0.69	0.643	0.481	0.331	0.123
$l=2$	0.44	0.387	0.358	0.577	0.664	0.73	0.744	0.733	0.703	0.629	0.464	0.332	0.12
$l=3$	0.541	0.465	0.421	0.528	0.649	0.695	0.705	0.71	0.691	0.624	0.445	0.332	0.127
$l=4$	0.586	0.514	0.48	0.525	0.639	0.698	0.703	0.694	0.663	0.618	0.447	0.319	0.131
$l=5$	0.571	0.495	0.484	0.525	0.643	0.724	0.722	0.723	0.678	0.627	0.448	0.318	0.135
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.25	0.318	0.559	0.761	0.831	0.861	0.858	0.839	0.819	0.763	0.606	0.461	0.236
$l=2$	0.22	0.292	0.527	0.739	0.794	0.835	0.857	0.852	0.82	0.767	0.601	0.46	0.242
$l=3$	0.24	0.275	0.521	0.705	0.789	0.835	0.85	0.854	0.819	0.767	0.609	0.45	0.237
$l=4$	0.25	0.28	0.519	0.699	0.795	0.833	0.86	0.842	0.8	0.736	0.606	0.443	0.24
$l=5$	0.25	0.305	0.516	0.704	0.789	0.837	0.846	0.848	0.823	0.75	0.588	0.446	0.242
	$S_C^{l,13}$	$A_C^{1,13}$											
$l=1$	0.75	0.85											
$l=2$	0.74	0.85											
$l=3$	0.76	0.87											
$l=4$	0.76	0.83											
$l=5$	0.771	0.81											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C=7$	0.656	0.558	0.558	0.56	0.801	0.724	0.656	0.605					

Table 3.23: Power for DGP4: Rolling Scheme

Recursive Scheme	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.071	0.211	0.368	0.711	0.827	0.881	0.895	0.87	0.854	0.807	0.676	0.497	0.203
$l=2$	0.069	0.223	0.371	0.71	0.814	0.864	0.88	0.886	0.848	0.804	0.665	0.501	0.203
$l=3$	0.071	0.211	0.351	0.673	0.785	0.844	0.872	0.881	0.863	0.804	0.651	0.498	0.209
$l=4$	0.073	0.242	0.388	0.68	0.795	0.846	0.863	0.875	0.845	0.803	0.65	0.493	0.209
$l=5$	0.074	0.201	0.369	0.662	0.787	0.854	0.884	0.872	0.833	0.799	0.647	0.495	0.206
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.06	0.367	0.642	0.817	0.894	0.921	0.923	0.914	0.904	0.874	0.725	0.585	0.305
$l=2$	0.058	0.332	0.611	0.817	0.881	0.908	0.935	0.929	0.9	0.871	0.732	0.59	0.3
$l=3$	0.062	0.315	0.596	0.789	0.877	0.906	0.926	0.93	0.915	0.882	0.737	0.597	0.306
$l=4$	0.064	0.325	0.605	0.783	0.881	0.908	0.921	0.917	0.902	0.872	0.724	0.585	0.308
$l=5$	0.062	0.348	0.601	0.781	0.883	0.909	0.929	0.925	0.908	0.871	0.727	0.587	0.3
	$S_C^{l,13}$	$A_C^{l,13}$											
$l=1$	0.665	0.718											
$l=2$	0.657	0.722											
$l=3$	0.646	0.706											
$l=4$	0.657	0.72											
$l=5$	0.655	0.709											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C=7$	0.814	0.735	0.671	0.624	0.905	0.843	0.787	0.759					

Table 3.24: Power for DGP1: Recursive Scheme

Recursive Scheme	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.144	0.115	0.094	0.053	0.062	0.055	0.112	0.142	0.212	0.31	0.368	0.36	0.229
$l=2$	0.122	0.118	0.09	0.053	0.061	0.061	0.105	0.135	0.223	0.29	0.362	0.357	0.229
$l=3$	0.135	0.118	0.087	0.056	0.053	0.068	0.114	0.152	0.225	0.306	0.367	0.359	0.229
$l=4$	0.152	0.142	0.081	0.054	0.54	0.056	0.105	0.152	0.214	0.292	0.366	0.364	0.228
$l=5$	0.135	0.115	0.088	0.059	0.057	0.058	0.112	0.145	0.219	0.303	0.37	0.355	0.229
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.206	0.156	0.111	0.069	0.057	0.076	0.126	0.175	0.228	0.311	0.402	0.415	0.336
$l=2$	0.192	0.126	0.094	0.064	0.06	0.081	0.116	0.17	0.247	0.318	0.407	0.418	0.332
$l=3$	0.205	0.153	0.099	0.062	0.055	0.082	0.128	0.183	0.253	0.319	0.41	0.409	0.332
$l=4$	0.221	0.164	0.098	0.054	0.051	0.067	0.123	0.164	0.222	0.318	0.41	0.42	0.333
$l=5$	0.193	0.154	0.089	0.051	0.055	0.072	0.126	0.185	0.235	0.336	0.407	0.42	0.336
	$S_C^{l,13}$	$A_C^{1,13}$											
$l=1$	0.395	0.401											
$l=2$	0.382	0.375											
$l=3$	0.4	0.39											
$l=4$	0.415	0.406											
$l=5$	0.401	0.371											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C=7$	0.109	0.108	0.105	0.106	0.124	0.126	0.129	0.127					

Table 3.25: Power for DGP2: Recursive Scheme

Recursive Scheme	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.352	0.472	0.591	0.752	0.817	0.851	0.901	0.897	0.765	0.211	0.096	0.09	0.08
$l=2$	0.749	0.839	0.879	0.911	0.922	0.924	0.94	0.91	0.719	0.192	0.085	0.1	0.059
$l=3$	0.834	0.872	0.907	0.934	0.929	0.929	0.946	0.915	0.716	0.129	0.078	0.09	0.08
$l=4$	0.856	0.901	0.925	0.936	0.94	0.946	0.944	0.91	0.664	0.118	0.101	0.101	0.07
$l=5$	0.844	0.907	0.926	0.938	0.951	0.939	0.946	0.901	0.651	0.09	0.112	0.092	0.07
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.769	0.885	0.936	0.972	0.982	0.983	0.987	0.984	0.952	0.599	0.191	0.13	0.098
$l=2$	0.967	0.988	0.997	0.999	0.995	0.998	0.996	0.991	0.944	0.537	0.185	0.121	0.095
$l=3$	0.977	0.996	0.994	0.999	0.997	0.997	0.997	0.989	0.937	0.481	0.173	0.108	0.088
$l=4$	0.983	0.995	0.996	0.996	0.996	0.998	0.998	0.992	0.915	0.445	0.173	0.113	0.091
$l=5$	0.986	0.996	0.996	0.998	0.996	0.997	0.997	0.987	0.904	0.414	0.21	0.122	0.092
	$S_C^{l,13}$	$A_C^{1,13}$											
$l=1$	0.835	0.819											
$l=2$	0.834	0.979											
$l=3$	0.891	0.99											
$l=4$	0.918	0.989											
$l=5$	0.918	0.994											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C=7$	0.919	0.932	0.923	0.896	0.987	0.989	0.99	0.985					

Table 3.26: Power for DGP3: Recursive Scheme

Recursive Scheme	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.897	0.889	0.847	0.838	0.882	0.729	0.849	0.732	0.504	0.276	0.16	0.12	0.08
$l=2$	0.867	0.875	0.88	0.852	0.84	0.848	0.828	0.777	0.473	0.269	0.15	0.1	0.059
$l=3$	0.9	0.899	0.904	0.873	0.875	0.872	0.85	0.771	0.448	0.179	0.145	0.09	0.08
$l=4$	0.924	0.927	0.927	0.902	0.882	0.882	0.853	0.771	0.397	0.179	0.171	0.11	0.07
$l=5$	0.933	0.935	0.93	0.913	0.895	0.883	0.859	0.749	0.391	0.179	0.162	0.132	0.07
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.825	0.879	0.916	0.939	0.944	0.948	0.943	0.92	0.785	0.343	0.29	0.201	0.182
$l=2$	0.987	0.993	0.99	0.99	0.988	0.981	0.978	0.955	0.789	0.304	0.27	0.201	0.172
$l=3$	0.994	0.996	0.996	0.996	0.989	0.982	0.981	0.955	0.755	0.268	0.241	0.191	0.173
$l=4$	0.994	0.997	0.994	0.993	0.991	0.988	0.985	0.951	0.745	0.234	0.224	0.181	0.166
$l=5$	0.997	0.995	0.998	0.995	0.993	0.987	0.978	0.949	0.717	0.22	0.235	0.181	0.164
	$S_C^{l,13}$	$A_C^{1,13}$											
$l=1$	0.958	0.982											
$l=2$	0.903	0.99											
$l=3$	0.94	0.995											
$l=4$	0.955	0.998											
$l=5$	0.965	0.997											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C=7$	0.897	0.804	0.785	0.781	0.947	0.951	0.936	0.912					

Table 3.27: Power for DGP4: Recursive Scheme

3.3.4 Detection of break location

In this session, we will show how to use our method to detect the location of the structural break point under three estimation schemes (Rolling, Recursive and Fixed). To illustrate the idea, we consider two different data generating processes (DGPs), which have break in the intercept and the variance respectively. Given the rolling window size of rol in the evaluation sample, in total, we can have $T - rol + 1$ statistics and each of those statistics reflects the performance of the predictive models in corresponding periods with rol observations. For every subsample in the evaluation sample, we assume that it contains the break point. If the beak in contained in a specific period, the statistics (single t or C) calculated using the observation in this period will be higher than the critical values. With the information provided by our statistics , we can always find the narrowest interval which contains the break point. In addition, when the interval for which we calculated the t and C statistics delivers the S_t and S_C simultaneously, we can identify the position for the break point, which is the first observation contained in this interval since when we move the window over this point, the value of the statistic will be lower or at most equal to the maximum value. First, we consider how to detect the break in the intercept and the DGP is specified as follows

DGP1: With pre-known structural break in intercept $y_t = \alpha_t + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \sigma\epsilon_t$, $\epsilon_t \sim N(0,1)$

$$\alpha_t = \begin{cases} \alpha_1 = 1.5 & \text{if } t < \text{break} \\ \alpha_2 = 0.1 & \text{otherwise} \end{cases}$$

Model under the null: $y_t = \alpha + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \sigma\epsilon_t$, $\epsilon_t \sim N(0,1)$, where $\beta_1 = 0.5$, $\beta_2 = 0.6$, $\sigma = 1$, T=600, R=350, P=250, break=480, rol=100.

Rolling scheme: The critical value at 5% for $S_C = 50.66$ with lag=1 and

for $S_t = 3.047$ with lag=1,contour level=50%. (Since the DGP1 does not have the misspecification of dynamics, we can just choose lag=1 to show how to detect the break point). Given the size of evaluation sample, we will have $151(T - rol + 1 - R)$ t and C statistics in total. Each of the statistics can show performance of the model under the null the at every single period with sample size 100. To detect the location of the break points we might need to follow these procedures (1) With bootstrapping the whole distribution, we can obtain the critical value for our statistics. We use the critical values for Sup type statistics at 5% level as the threshold line to locate the first significant interval that should contain the break point. (2) we need to check the points that deliver the biggest values of the statistics because those points might be the break points.

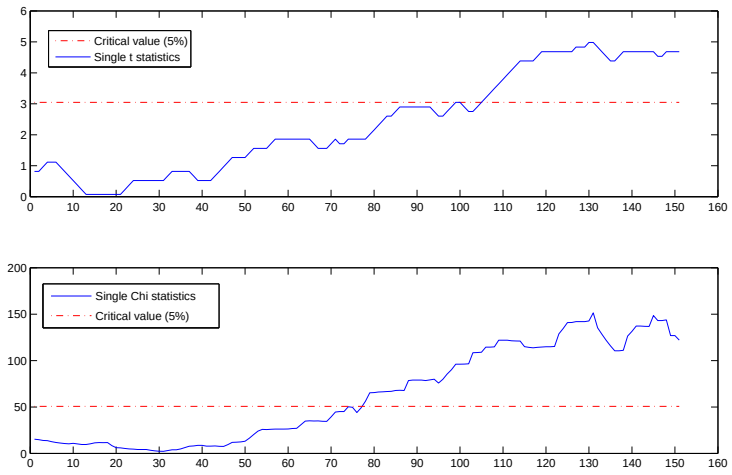


Figure 3.7: Plots of t and C Statistics under Rolling Scheme: DGP1

From Figure 3.7, we can see that for the single t statistics, after the 105th statistic, all the statistics are above the threshold line. Therefore from the t statistics, we can infer that the break point must be contained in the time interval [455 554].

As for the single C statistics, which accumulate the information of all contour levels thus the can detect the break faster than t statistics do. All the statistics after the 78th one are above the threshold. Therefore, from the C statistics, we can infer that the break point must contained in the time interval [428 527]. Take the intersection of these two intervals,we can narrow down the time intervals that contain the break point, which is [455 527].

Furthermore, We can observe that for both t and C statistics, 130th statistics have the highest value among all these 151 statistics. (151.36 and 4.98)It indicates that the location of the break point is at 480, which is corresponding to the 130th statistics. When we keep rolling the evaluation sample over the 480th observation (or the 130th statistics), which is indeed the break point designed in the DGP1, the value of the statistics declines and thus indicates that the break happens at the 480th observation.

Recursive scheme: The critical value at 5% for $S_C = 44.24$ with lag=1 and for $S_t = 3.05$ with lag=1,contour level=50%.

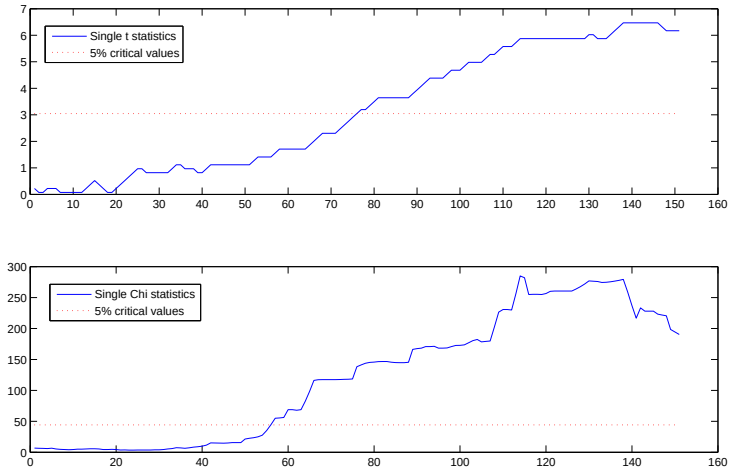


Figure 3.8: Plots of t and C Statistics under Recursive Scheme: DGP1

From Figure 3.8, we can see that for the single t statistics, after the 77th statistic, all the statistics are above the threshold line. Therefore from the t statistics, we can infer that the break point must be contained in the time interval [427 526]. As for the single C statistics, which accumulate the information of all contour levels thus the can detect the break faster than t statistics do. All the statistics after the 56th one are above the threshold. Therefore, from the C statistics, we can infer that the break point must contained in the time interval [406 505]. Take the intersection of these two intervals,we can narrow down the time intervals that contain the break point, which is [427 505].

Furthermore, We can observe that for both t and C statistics, 130th statistics have the highest value among all these 151 statistics (276.93 and 6.02). It indicates that the location of the break point is at 480, which is corresponding to the 130th statistics. When we keep rolling the evaluation sample over the 480th observation (or the 130th statistics), which is indeed the break point designed in the DGP1, the value of the statistics declines and thus indicates that the break happens at the 480th observation.

Fixed scheme: The critical value at 5% for $S_C = 32.7261$ with lag=1 and for $S_t = 2.7501$ with lag=1,contour level=50%.

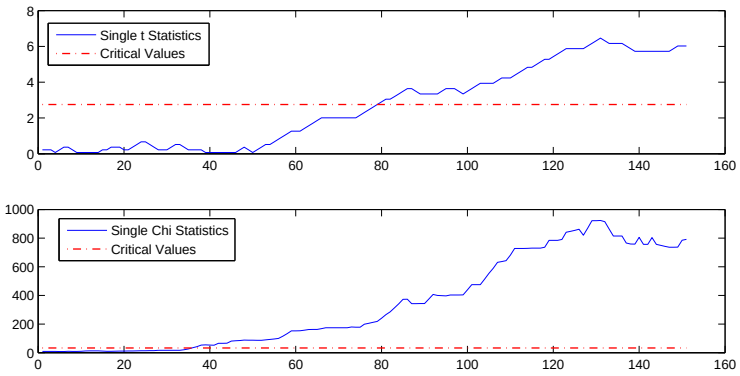


Figure 3.9: Plots of t and C Statistics under Fixed Scheme: DGP1

From Figure 3.9, we can see that for the single t statistics, after the 80th statistic, all the statistics are above the threshold line. Therefore from the t statistics, we can infer that the break point must be contained in the time interval [430 529]. As for the single C statistics, which accumulate the information of all contour levels thus the can detect the break faster than t statistics do. All the statistics after the 36th one are above the threshold. Therefore, from the C statistics, we can infer that the break point must contained in the time interval [386 485]. Take the intersection of these two intervals,we can narrow down the time intervals that contain the break point, which is [430 485]. Furthermore, We can observe that for both t and C statistics, 130th statistics have the highest value among all these 151 statistics (923.23 and 6.47). It indicates that the location of the break point is at 480, which is corresponding to the 130th statistics. When we keep rolling the evaluation sample over the 480th observation (or the 130th statistics), which is indeed the break point designed in the DGP1, the value of the statistics declines and thus indicates that the break happens at the 480th observation.

Then, we investigate how to detect the break in the variance and the model is specified as follows:

DGP2: With pre-known structural break in variance $y_t = \alpha+\beta_1 y_{t-1}+\beta_2 x_{t-1}+ \sigma_t\epsilon_t$, $\epsilon_t \sim N(0,1)$

$$\sigma_t = \begin{cases} \sigma_1 = 2 & \text{if } t < \text{break} \\ \sigma_2 = 0.5 & \text{otherwise} \end{cases}$$

Model under the null: $y_t = \alpha + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \sigma\epsilon_t$, $\epsilon_t \sim N(0,1)$ $\alpha = 1.5$, $\beta_1 = 0.5$, $\beta_2 = 0.6$, $\sigma_1 = 2$, $\sigma_2 = 0.5$ T=600, R=350, P=250, break=480, r=100

Rolling scheme: The critical value at 5% for $S_C = 32.088$ with lag=1 and for $S_t = 2.6014$ with lag=1,contour level=50%. (Since the DGP2 does not have the misspecification of dynamics, we can just choose lag=1 to show how to detect the break point). Given the size of evaluation sample, we will have $151(T - r + 1 - R)$ t and C statistics in total. Each of the statistics can show performance of the model under the null the at every single period with sample size 100. To detect the location of the break points we might need to follow these procedures (1) With bootstrapping the whole distribution, we can obtain the critical value for our statistics. We use the critical values for Sup type statistics at 5% level as the threshold line to locate the first significant interval that should contain the break point. (2) we need to check the points that deliver the biggest values of the statistics because those points might be the break points.

From Figure 3.10, we can see that for the single t statistics, after the 85th statistic, all the statistics are above the threshold line. Therefore from the t statistics, we can infer that the break point must be contained in the time interval [435 534]. As for the single C statistics, which accumulate the information of all contour levels thus the can detect the break faster than t statistics do. All the statistics after the 79th one are above the threshold. Therefore, from the C statistics, we can infer that the break point must contained in the time interval [429 528]. Take the intersection of

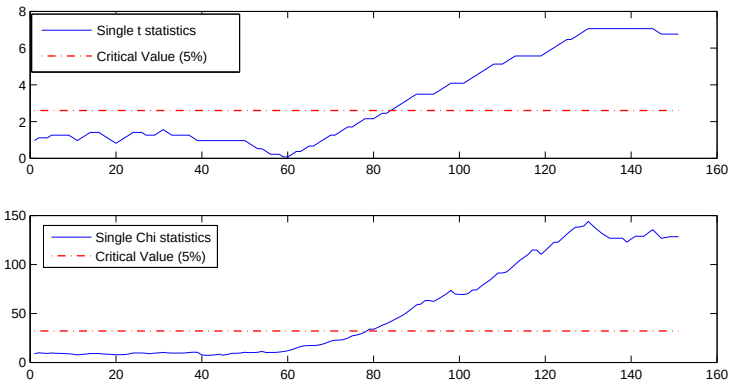


Figure 3.10: Plots of t and C Statistics under Rolling Scheme: DGP2

these two intervals,we can narrow down the time intervals that contain the break point, which is [435 528]. Furthermore, We can observe that for both t and C statistics, 130th statistics have the highest value among all these 151 statistics. (143.91 and 7.0610)It indicates that the location of the break point is at 480, which is corresponding to the 130th statistics. When we keep rolling the evaluation sample over the 480th observation (or the 130th statistics), which is indeed the break point designed in the DGP2, the value of the statistics declines and thus indicates that the break happens at the 480th observation.

Recursive scheme: The critical value at 5% for $S_C = 31.4583$ with lag=1 and for $S_t = 2.7501$ with lag=1,contour level=50%.

From Figure 3.11, we can see that for the single t statistics, after the 73th statistic, all the statistics are above the threshold line. Therefore from the t statistics, we can infer that the break point must be contained in the time interval [423 522]. As for the single C statistics, which accumulate the information of all contour levels thus the can detect the break faster than t statistics do. All the statistics after the 73th one are above the threshold. Therefore, from the C statistics, we can infer that

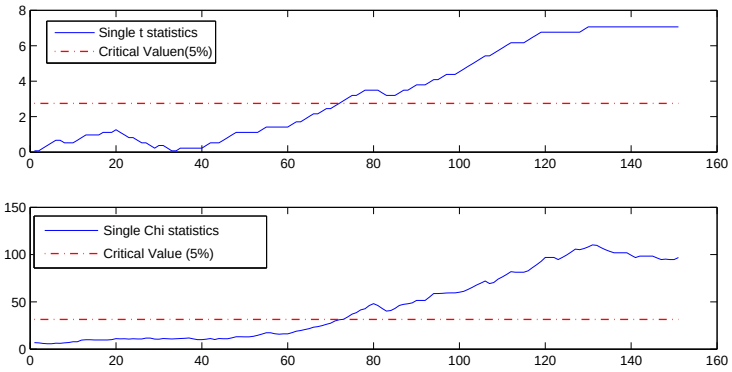


Figure 3.11: Plots of t and C Statistics under Recursive Scheme: DGP2

the break point must contained in the time interval [423 522]. Take the intersection of these two intervals,we can narrow down the time intervals that contain the break point, which is [423 522]. Furthermore, We can observe that for both t and C statistics, 130th statistics have the highest value among all these 151 statistics (110.19 and 7.0610). It indicates that the location of the break point is at 480, which is corresponding to the 130th statistics. When we keep rolling the evaluation sample over the 480th observation (or the 130th statistics), which is indeed the break point designed in the DGP2, the value of the statistics declines and thus indicates that the break happens at the 480th observation.

Fixed scheme: The critical value at 5% for $S_C = 34.6505$ with lag=1 and for $S_t = 2.8987$ with lag=1,contour level=50%.

From Figure 3.12, we can see that for the single t statistics, after the 61th statistic, all the statistics are above the threshold line. Therefore from the t statistics, we can infer that the break point must be contained in the time interval [411 520]. As for the single C statistics, which accumulate the information of all contour levels thus the can detect the break faster than t statistics do. All the statistics after the 74th one

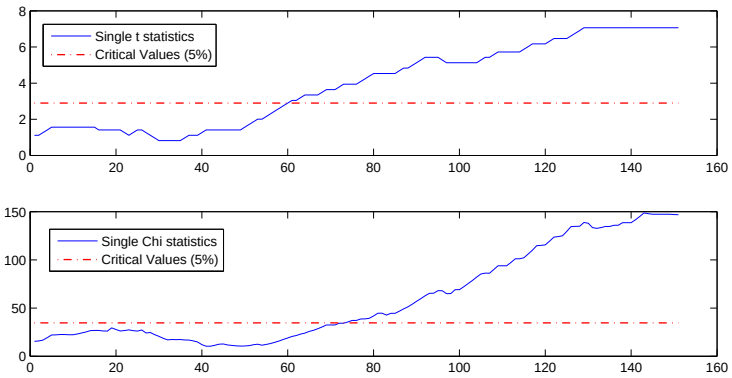


Figure 3.12: Plots of t and C Statistics under Fixed Scheme: DGP2

are above the threshold. Therefore, from the C statistics, we can infer that the break point must contained in the time interval [424 523]. Take the intersection of these two intervals,we can narrow down the time intervals that contain the break point, which is [423 520]. Furthermore, We can observe that for t, the 129th statistics has the highest value among all these 151 statistics (7.0610) and that for C statistic, the 143th statistic has the highest value (148.67). In this case, with the fixed scheme, we cannot detect the exact location for the break point, since the maximum values of t and C statistics do not happen at the exact same point. However, we can conclude the break point must be contained in the interval of [479 482], which is the narrowest interval containing the break point (the 480th observation).

3.4 Application

The usefulness of Phillips curve (PC) at forecasting inflation is always a heated debate among researchers. The essence of PC is to relate inflation to some measure of the level of real activity, in most cases the unemployment rate. There is widespread empirical evidence on the existence of parameter instability in forecasting inflation such

as the works by Stock and Watson (2003), and Clark and McCracken (2005) etc. Thus, researchers are interested in the evaluation of the absolute or the relative forecasting performance of PC. Stock and Watson (1999) found some empirical evidence in favor of the Phillips curve as a forecasting tool and demonstrated that inflation forecasts produced by the Phillips curve generally are more accurate than forecasts based on simple autoregressive or multivariate models. They also found the existence of parameter instabilities across different subsample. Stock and Watson (2007) have documented a change in the forecast stability in output growth as well as inflation over time. Rossi and Skehposyan (2010) adopted Giacomini and Rossi (2010a)'s Fluctuation test based on the same model used by Stock and Watson (2003), and Clark and McCracken (2005) to investigate the relative forecasting performance of inflation. They show that the predictive power of the Phillips curve disappeared around the time of the Great Moderation. Follow Stock and Watson (1999)'s work, Amisano annd Giacomini proposed a test for comparing the out-of-sample density forecast accuracy of U.S inflation by linear and Markov-switching Phillips curve models estimated by either maximum likelihood or Bayesian methods and claimed that a Markov-switching Phillips curve estimated by maximum likelihood produces the best density forecasts of inflation. Their tests are based on scoring rules, which are loss functions defined over the density forecast and the realizations of the variable. They restrict attention to the logarithmic scoring rule and propose an out-of-sample weighted likelihood ratio test that compares weighted averages of the scores for the competing forecasts. Their comparison is based on the average forecasting performance of competing models across time and thus can not take care of the instabilities that widely documented in the literature. In our paper, we will also evaluate the predictive density forecasting performance of linear and Markov-switching Phillips curve models. We are focusing on the absolute density forecast performance in

the presence of instability. Same as Amisano annd Giacomini (2012), we follow Stock and Watson (1999)'s linear PC model, in which changes of inflation depend on their lags and on lags of the unemployment rate. We take the model specification chosen in Amisano annd Giacomini (2012):

$$\Delta\pi_t = \alpha_1 + \beta_1\Delta\pi_{t-1} + \beta_2\Delta\pi_{t-2} + \beta_{12}\Delta\pi_{t-12} + \gamma u_{t-1} + \sigma\epsilon_t$$

where $\epsilon_t \sim N(0,1)$, CPI_t is the consumer price index and u_{t-1} is the unemployment rate. This specification is consistent with the natural rate hypothesis, because the natural rate of unemployment (NAIRU) is $u^* = \frac{-\alpha}{\gamma(1)}$. $\Delta\pi_t$ and u_t do not have a unit root. $\pi_t = 100ln(CPI_t/CPI_{t12})$. Our test is based on the above. We use the monthly data spanning the period 1958M012012:M01 obtained from FRED. CPI_t is the Consumer Price Index for All Urban Consumers: All Items and u_{t-1} is the Civilian Unemployment Rate. Both series are seasonally adjusted. The series runs from 1958M01-2012M01, which updated the data set (1958M01-2004M07) used in Amisano annd Giacomini (2012). The estimation sample is from 1958M01 to 1988M12, which is same estimation periods used in Amisano annd Giacomini (2012). The rest of the series are used as evaluation sample. The subsample of the evaluation sample (rol) is 200.

3.4.1 Evaluation of Linear Phillips Curve

We present the evaluation results for linear Phillips Curve in Table 3.28, Table 3.29 and 3.30 under fixed, rolling and recursive schemes. The linear Phillips Curve has been rejected under all of these three estimation scheme. The result is consistent with the empirical evidence that the predictive forecasting performance of the Philip Curve is time varying.

Fixed Scheme	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.017	0.529	0.523	0.471	0.268	0.002	0.008	0.045	0.136	0.091	0.085	0.031	0.0001
$l=2$	0.015	0.285	0.55	0.434	0.481	0.002	0.005	0.041	0.157	0.086	0.09	0.027	0.0001
$l=3$	0.028	0.584	0.614	0.651	0.419	0.006	0.005	0.03	0.049	0.088	0.073	0.013	0.0001
$l=4$	0.013	0.104	0.807	0.285	0.209	0.009	0.003	0.027	0.098	0.088	0.069	0.012	0.0001
$l=5$	0.031	0.221	0.212	0.302	0.206	0.009	0.006	0.034	0.103	0.088	0.057	0.006	0.0001
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.342	0.671	0.759	0.405	0.163	0.042	0.013	0.248	0.243	0.243	0.412	0.199	0.042
$l=2$	0.335	0.453	0.621	0.352	0.972	0.043	0.022	0.336	0.297	0.223	0.332	0.25	0.04
$l=3$	0.494	0.579	0.316	0.567	0.309	0.041	0.016	0.216	0.225	0.235	0.359	0.197	0.049
$l=4$	0.338	0.348	0.764	0.319	0.152	0.033	0.024	0.144	0.264	0.223	0.327	0.226	0.042
$l=5$	0.138	0.595	0.637	0.238	0.142	0.032	0.024	0.218	0.273	0.222	0.308	0.184	0.042
	$S_C^{13,l}$	$A_C^{13,1}$											
$l=1$	0.004	0.022											
$l=2$	0.001	0.019											
$l=3$	0.003	0.016											
$l=4$	0.001	0.025											
$l=5$	0.001	0.025											
	$S_C^{2,7}$	$S_C^{3,7}$	$S_C^{4,7}$	$S_C^{5,7}$	$A_C^{2,7}$	$A_C^{3,7}$	$A_C^{4,7}$	$A_C^{5,7}$					
	0.006	0.009	0.012	0.01	0.027	0.018	0.027	0.029					

Table 3.28: P-value for linear PC: Fixed Scheme

Rolling Scheme	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.528	0.46	0.843	0.472	0.142	0.002	0.001	0.007	0.008	0.093	0.281	0.026	0.0002
$l=2$	0.258	0.35	0.58	0.251	0.229	0.007	0.004	0.009	0.002	0.081	0.24	0.024	0.0003
$l=3$	0.559	0.31	0.741	0.437	0.16	0.002	0.001	0.002	0.002	0.066	0.288	0.013	0.0003
$l=4$	0.38	0.31	0.479	0.463	0.17	0.001	0.001	0.003	0.008	0.088	0.275	0.021	0.0002
$l=5$	0.285	0.249	0.297	0.225	0.12	0.007	0.001	0.003	0.003	0.063	0.327	0.019	0.0002
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.789	0.801	0.853	0.656	0.143	0.001	0.023	0.033	0.032	0.145	0.297	0.271	0.024
$l=2$	0.267	0.55	0.812	0.244	0.568	0.002	0.018	0.024	0.045	0.127	0.283	0.337	0.037
$l=3$	0.751	0.918	0.923	0.723	0.144	0.007	0.015	0.032	0.047	0.114	0.288	0.24	0.023
$l=4$	0.358	0.476	0.575	0.448	0.108	0.001	0.015	0.031	0.031	0.142	0.313	0.271	0.034
$l=5$	0.305	0.655	0.576	0.251	0.129	0.009	0.018	0.031	0.039	0.13	0.286	0.288	0.033
	$S_C^{13,l}$	$A_C^{13,1}$											
$l=1$	0.001	0.009											
$l=2$	0.0003	0.004											
$l=3$	0.0003	0.01											
$l=4$	0.0003	0.005											
$l=5$	0.002	0.004											
	$S_C^{2,7}$	$S_C^{3,7}$	$S_C^{4,7}$	$S_C^{5,7}$	$A_C^{2,7}$	$A_C^{3,7}$	$A_C^{4,7}$	$A_C^{5,7}$					
	0.008	0.008	0.006	0.009	0.038	0.025	0.019	0.031					

Table 3.29: P-value for linear PC: Rolling Scheme

Recursive Scheme	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.08	0.171	0.571	0.555	0.254	0.003	0.001	0.019	0.038	0.132	0.184	0.001	0.0003
$l=2$	0.038	0.133	0.646	0.406	0.416	0.007	0.005	0.015	0.021	0.113	0.185	0.008	0.0001
$l=3$	0.042	0.168	0.926	0.589	0.544	0.001	0.006	0.012	0.017	0.094	0.105	0.005	0.0001
$l=4$	0.023	0.103	0.683	0.272	0.46	0.003	0.003	0.009	0.035	0.119	0.098	0.008	0.0001
$l=5$	0.045	0.158	0.489	0.233	0.64	0.009	0.004	0.016	0.023	0.095	0.093	0.007	0.0001
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.282	0.649	0.794	0.522	0.285	0.033	0.034	0.16	0.238	0.24	0.355	0.283	0.061
$l=2$	0.287	0.65	0.79	0.259	0.849	0.032	0.037	0.242	0.219	0.21	0.296	0.356	0.059
$l=3$	0.76	0.864	0.855	0.747	0.582	0.014	0.043	0.141	0.112	0.187	0.337	0.242	0.059
$l=4$	0.45	0.58	0.775	0.249	0.121	0.012	0.037	0.139	0.163	0.209	0.282	0.25	0.056
$l=5$	0.24	0.447	0.733	0.193	0.115	0.022	0.021	0.144	0.142	0.204	0.263	0.271	0.06
	$S_C^{13,l}$	$A_C^{13,1}$											
$l=1$	0.009	0.034											
$l=2$	0.009	0.034											
$l=3$	0.004	0.029											
$l=4$	0.003	0.05											
$l=5$	0.008	0.049											
	$S_C^{2,7}$	$S_C^{3,7}$	$S_C^{4,7}$	$S_C^{5,7}$	$A_C^{2,7}$	$A_C^{3,7}$	$A_C^{4,7}$	$A_C^{5,7}$					
	0.007	0.016	0.011	0.015	0.037	0.44	0.04	0.049					

Table 3.30: P-value for linear PC: Recursive Scheme

3.4.2 Evaluation of Markov-switching Phillips Curve

From the above session, we can observe that the simple linear Philips Curve fails to provide a good predictive performance in the presence of instabilities. Then we evaluate the performance of the Markov-Switching Philips Curve , which allows for non-Gaussian density forecasts. We assume that some of the parameters of the linear model vary depending on the value of an unobserved discrete variable s_t following a finite Markov chain. To evaluate this mode, we use the fixed estimation scheme. We consider the two-state Markov Switching model with ϵ_t follows a standard normal distribution and

$$\Delta\pi_t = \alpha_1 + \beta_1\Delta\pi_{t-1} + \beta_2\Delta\pi_{t-2} + \beta_{12}\Delta\pi_{t-12} + \gamma u_{t-1} + \sigma\epsilon_t$$

.

$$s_t = \begin{cases} 1 \\ \\ 2 \end{cases} \qquad Pr(s_t = j|s_{t-1} = i) = p_{ij}$$

First, we consider the model (Model 1) that all the model's parameters depend on the hidden state s_t. Table 3.31 presents the p-values of the statistics based on Model 1. Since the p-value of the statistics at every contour and the statistics accumulates all contours or lags are very large, we fail to reject Model 1. It means that if the model has high flexibility across time, we are tend to accept the model. To further investigate the key factor which characterize the flexibility of the mode, we consider the second model specification. The specification for the second model (Model 2) imposes that the intercept of the model does not depend on s_t but allowing all the other parameters to switch. In this specification, we are interested in investigating the role of the intercept on the instability of the unconditional mean for the inflation. Table 3.32 present the testing results for Model2 and the large p-value indicates that even if the intercept does not depend on the state, the other parameters which are allowed to change across time

can capture the instabilities of the model. Then, we follow the MS2 model specified in Amisano annd Giacomini (2012) to impose the constraints that the intercept and the coefficient on lagged unemployment are constant across states:$\gamma^{s_t} = \gamma$, $\alpha^{s_t} = \alpha$. This imposition induces constancy of the NAIRU across states ($u^* = \frac{-\alpha}{\gamma(1)}$). In Table 3.33, we can observe that although we still fail to reject the model overall, since the p-values for $C_{max}^{13,l}$, $C_{mean}^{13,l}$ (stacking all contours) and $C_{max}^{7,k}$, $C_{mean}^{7,k}$ (stacking 5 lags) are large. However, for contour levels: 90%, 95%, 99% the p values are smaller than 5% and thus we can reject the model at these contour levels. It indicates that the restriction on γ will lower the model's ability to fully capture the instability of Philips Curve. Finally, we impose the constraints on the coefficients on lagged unemployment and allow α^{s_t}, γ^{s_t} and σ_{s_t} to vary across time. From Table 3.34 we can find that the p-values at every contour level are very big and thus we fail to reject Model 4. Therefore, the results indicate that the predictive performance of the Philips Curve is unstable over time and the traditional linear Philips Curve is unable to forecast the inflation in the presence of instability. The introduction of Markov-Switching Philips Curve can deliver much better performance than its linear competitor. In addition, it is more important to characterize the instability of the effect of the lag on unemployment rate than the other factor such as the lag of the inflation.

Model 1: $\Delta\pi_t = \alpha^{s_t} + \beta_1^{s_t}\Delta\pi_{t-1} + \beta_2^{s_t}\Delta\pi_{t-2} + \beta_{12}^{s_t}\Delta\pi_{t-12} + \gamma^{s_t}u_{t-1} + \sigma_{s_t}\epsilon_t$, $\epsilon_t \sim N(0,1)$.

Model 1	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.904	0.798	0.93	0.89	0.742	0.726	0.766	0.942	0.98	0.968	0.818	0.146	0.164
$l=2$	0.902	0.76	0.834	0.914	0.946	0.914	0.864	0.938	0.934	0.936	0.822	0.142	0.134
$l=3$	0.896	0.792	0.892	0.98	0.88	0.768	0.804	0.98	0.992	0.986	0.78	0.134	0.134
$l=4$	0.972	0.788	0.924	0.882	0.67	0.654	0.762	0.952	0.998	0.954	0.79	0.142	0.158
$l=5$	0.744	0.624	0.948	0.922	0.756	0.724	0.78	0.932	0.994	0.964	0.786	0.138	0.134
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.894	0.794	0.922	0.874	0.728	0.74	0.766	0.952	0.98	0.996	0.83	0.152	0.152
$l=2$	0.89	0.748	0.824	0.908	0.94	0.926	0.874	0.958	0.946	0.946	0.828	0.152	0.132
$l=3$	0.882	0.786	0.89	0.97	0.852	0.768	0.796	0.98	0.994	0.99	0.796	0.14	0.13
$l=4$	0.972	0.81	0.926	0.886	0.67	0.658	0.748	0.954	0.998	0.946	0.79	0.142	0.15
$l=5$	0.732	0.61	0.946	0.926	0.762	0.728	0.776	0.948	0.988	0.96	0.79	0.136	0.144
	$S_C^{13,l}$	$A_C^{13,1}$											
$l=1$	0.796	0.792											
$l=2$	0.798	0.792											
$l=3$	0.766	0.76											
$l=4$	0.728	0.722											
$l=5$	0.734	0.728											
	$S_C^{2,7}$	$S_C^{3,7}$	$S_C^{4,7}$	$S_C^{5,7}$	$A_C^{2,7}$	$A_C^{3,7}$	$A_C^{4,7}$	$A_C^{5,7}$					
$C=7$	0.768	0.824	0.778	0.796	0.772	0.812	0.772	0.79					

Table 3.31: P Value of The Statistics:M1

Model 2: $\Delta\pi_t = \alpha + \beta_1^{s_t}\Delta\pi_{t-1} + \beta_2^{s_t}\Delta\pi_{t-2} + \beta_{12}^{s_t}\Delta\pi_{t-12} + \gamma^{s_t}u_{t-1} + \sigma_{s_t}\epsilon_t$, $\epsilon_t \sim N(0,1)$. In this model, we assume that coefficients and conditional variance depend on the hidden state s_t, but the intercept does not depend on the hidden state.

Model 2	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.816	0.754	0.634	0.652	0.606	0.58	0.52	0.612	0.614	0.62	0.436	0.224	0.132
$l=2$	0.544	0.464	0.428	0.414	0.4	0.488	0.402	0.568	0.552	0.608	0.4	0.244	0.162
$l=3$	0.962	0.602	0.514	0.592	0.57	0.578	0.498	0.584	0.584	0.614	0.43	0.264	0.134
$l=4$	0.844	0.802	0.732	0.61	0.564	0.588	0.526	0.618	0.616	0.626	0.436	0.266	0.146
$l=5$	0.862	0.634	0.62	0.488	0.534	0.554	0.5	0.608	0.614	0.642	0.426	0.254	0.152
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.8	0.742	0.626	0.644	0.634	0.62	0.54	0.638	0.636	0.638	0.452	0.252	0.154
$l=2$	0.536	0.458	0.42	0.412	0.422	0.538	0.46	0.6	0.582	0.632	0.412	0.274	0.186
$l=3$	0.962	0.592	0.562	0.588	0.576	0.61	0.52	0.622	0.588	0.62	0.44	0.278	0.152
$l=4$	0.836	0.834	0.744	0.612	0.57	0.598	0.548	0.638	0.628	0.636	0.444	0.274	0.148
$l=5$	0.85	0.622	0.614	0.484	0.538	0.572	0.524	0.634	0.616	0.658	0.432	0.272	0.152
	$S_C^{13,l}$	$A_C^{13,1}$											
$l=1$	0.36	0.384											
$l=2$	0.342	0.374											
$l=3$	0.362	0.382											
$l=4$	0.388	0.396											
$l=5$	0.36	0.362											
	$S_C^{2,7}$	$S_C^{3,7}$	$S_C^{4,7}$	$S_C^{5,7}$	$A_C^{2,7}$	$A_C^{3,7}$	$A_C^{4,7}$	$A_C^{5,7}$					
$C=7$	0.462	0.44	0.462	0.47	0.464	0.47	0.478	0.494					

Table 3.32: P Value of The Statistics:M2

Model 3: $\Delta\pi_t = \alpha + \beta_1^{s_t}\Delta\pi_{t-1} + \beta_2^{s_t}\Delta\pi_{t-2} + \beta_{12}^{s_t}\Delta\pi_{t-12} + \gamma u_{t-1} + \sigma\epsilon_t,\ \epsilon_t \sim N(0,1)$.

Model 3	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.542	0.682	0.306	0.254	0.18	0.196	0.178	0.26	0.4	0.188	0.044	0.026	0.036
$l=2$	0.478	0.32	0.226	0.15	0.102	0.118	0.102	0.116	0.228	0.088	0.034	0.024	0.032
$l=3$	0.802	0.394	0.188	0.236	0.172	0.176	0.146	0.234	0.274	0.192	0.038	0.024	0.032
$l=4$	0.496	0.65	0.524	0.296	0.188	0.2	0.19	0.284	0.422	0.21	0.042	0.024	0.034
$l=5$	0.794	0.19	0.286	0.186	0.136	0.148	0.164	0.296	0.386	0.166	0.032	0.024	0.032
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.496	0.628	0.296	0.264	0.176	0.198	0.178	0.288	0.424	0.256	0.046	0.028	0.036
$l=2$	0.438	0.298	0.216	0.152	0.102	0.116	0.104	0.16	0.27	0.128	0.044	0.026	0.032
$l=3$	0.372	0.374	0.298	0.24	0.168	0.17	0.148	0.246	0.288	0.22	0.046	0.026	0.032
$l=4$	0.442	0.416	0.366	0.33	0.21	0.2	0.21	0.31	0.434	0.24	0.046	0.026	0.034
$l=5$	0.362	0.22	0.332	0.192	0.138	0.15	0.17	0.32	0.402	0.232	0.036	0.024	0.032
	$S_C^{13,l}$	$A_C^{13,1}$											
$l=1$	0.328	0.332											
$l=2$	0.18	0.18											
$l=3$	0.284	0.272											
$l=4$	0.32	0.328											
$l=5$	0.266	0.268											
	$S_C^{2,7}$	$S_C^{3,7}$	$S_C^{4,7}$	$S_C^{5,7}$	$A_C^{2,7}$	$A_C^{3,7}$	$A_C^{4,7}$	$A_C^{5,7}$					
$C=7$	0.112	0.12	0.12	0.116	0.118	0.124	0.12	0.122					

Table 3.33: P Value of The Statistics:M3

Model 4: $\Delta\pi_t = \alpha^{s_t} + \beta_1\Delta\pi_{t-1} + \beta_2\Delta\pi_{t-2} + \beta_{12}\Delta\pi_{t-12} + \gamma^{s_t}u_{t-1} + \sigma_{s_t}\epsilon_t$,

$\epsilon_t \sim N(0,1)$.

Model 4	$S_t^{l,1}$	$S_t^{l,2}$	$S_t^{l,3}$	$S_t^{l,4}$	$S_t^{l,5}$	$S_t^{l,6}$	$S_t^{l,7}$	$S_t^{l,8}$	$S_t^{l,9}$	$S_t^{l,10}$	$S_t^{l,11}$	$S_t^{l,12}$	$S_t^{l,13}$
$l=1$	0.452	0.662	0.454	0.406	0.466	0.694	0.702	0.814	0.788	0.57	0.41	0.254	0.296
$l=2$	0.428	0.552	0.218	0.464	0.436	0.678	0.652	0.816	0.684	0.55	0.41	0.276	0.296
$l=3$	0.476	0.784	0.454	0.424	0.466	0.66	0.602	0.874	0.782	0.582	0.43	0.206	0.296
$l=4$	0.656	0.738	0.676	0.55	0.584	0.72	0.807	0.884	0.82	0.584	0.386	0.224	0.296
$l=5$	0.652	0.538	0.422	0.444	0.536	0.78	0.836	0.844	0.842	0.578	0.416	0.244	0.296
	$A_t^{l,1}$	$A_t^{l,2}$	$A_t^{l,3}$	$A_t^{l,4}$	$A_t^{l,5}$	$A_t^{l,6}$	$A_t^{l,7}$	$A_t^{l,8}$	$A_t^{l,9}$	$A_t^{l,10}$	$A_t^{l,11}$	$A_t^{l,12}$	$A_t^{l,13}$
$l=1$	0.442	0.644	0.43	0.39	0.568	0.682	0.856	0.822	0.818	0.596	0.45	0.278	0.294
$l=2$	0.402	0.446	0.318	0.362	0.326	0.488	0.674	0.834	0.72	0.57	0.446	0.308	0.296
$l=3$	0.606	0.748	0.446	0.388	0.454	0.558	0.804	0.868	0.79	0.606	0.434	0.23	0.294
$l=4$	0.638	0.768	0.666	0.348	0.58	0.722	0.86	0.874	0.824	0.584	0.4	0.226	0.296
$l=5$	0.632	0.514	0.404	0.432	0.522	0.722	0.83	0.852	0.86	0.582	0.412	0.246	0.39
	$S_C^{13,l}$	$A_C^{13,1}$											
$l=1$	0.614	0.596											
$l=2$	0.48	0.454											
$l=3$	0.494	0.482											
$l=4$	0.51	0.478											
$l=5$	0.488	0.472											
	$S_C^{2,7}$	$S_C^{3,7}$	$S_C^{4,7}$	$S_C^{5,7}$	$A_C^{2,7}$	$A_C^{3,7}$	$A_C^{4,7}$	$A_C^{5,7}$					
$C=7$	0.518	0.566	0.548	0.598	0.534	0.578	0.564	0.618					

Table 3.34: P Value of The Statistics:M4

3.5 Conclusion

Evaluation of out-of-sample forecasting performance in the presence of instabilities is major concern of researchers. The impact of structural instability for forecast evaluation has not been formally investigated until very recently. However, in order to fill the gap in the literature, most of the literatures are just focused on point forecast evaluation. In this paper, we propose a robust out-of-sample density forecasting evaluation method in the presence of the instabilities based on Generalized Autocontour Method. We define the instabilities as as time variation in the density function of a stochastic process. These variations include changes of mean, variance and/or the functional form of the underlying density function. To take care of the instability, we evaluate one subsample (with size rol) of the evaluation sample, using data from $t-rol+1$ up to t, where $t = R+rol, \cdots, T$, to evaluate the assumed predictive density. According to Generalized Autocontour, for one subsample, we can obtain three different types of statistics: t, C and L. With the rolling window evaluation method, we can obtain a collection of $T - rol - R + 1$ statistics of for t, C and L statistics based on which Sup type and Avg type statistics are constructed. The asymptotic distributions of the statistics constructed in this paper are functional of standard Brownian motions and have very good finite sample properties. We have applied our tests to evaluate the density forecast performance of U.S. inflation produced by linear and Markov-switching Philips Curve. Our tests show that the linear Philips Curve can not deliver a good density forecast in the presence of instabilities. However, the Marokv switching Philips Curve can provide a good out-of-sample density forecast for U.S. inflation in the presence of instabilities as long as the model can characterize the instability relationship between the unemployment rate and inflation.

Bibliography

[1] Andrews, D.W. (1993), "Tests for Parameter Instability and Structural Change With Unknown Change Point", *Econometrica*, Vol. 61, No. 4, pp. 821-856.

[2] Andrews, D.W. and Ploberger, W. (1994), "Optimal TestsWhen a Nuisance Parameter is Present Only Under the Alternative", *Econometrica* Vol. 62, No. 6, pp. 1383-1414.

[3] Aiol, M. and Timmermann, A. (2006), "Persistence of Forecasting Performance and Conditional Combination Strategies",*Journal of Econometrics*, Vol. 135, pp. 31-53.

[4] Chen, Y., K. Rogoff and B. Rossi (2010), "Can Exchange Rates Forecast Commodity Prices?",*Quarterly Journal of Economics*, Vol. 125, No. 3, 1145-1194.

[5] Chow, Gregory C. (1960), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions", *Econometrica*, Vol. 28, No. 3, pp. 591-605.

[6] Clements, M.P. and D.F. Hendry (1996), "Intercept Corrections and Structural Change",*Journal of Applied Econometrics* Vol. 11, pp. 475-494.

[7] Cogley, T. and T. Sargent (2005), "Drifts and Volatilities: Monetary Policies and Outcomes in the post World War II U.S.",*Review of Economic Dynamics*, Vol.8, pp. 262-302.

[8] Clark, T.E. and M.W. McCracken (2010), "Averaging Forecasts from VARs with Uncertain Instabilities",*Journal of Applied Econometrics*, Vol. 25, No. 1, pp. 5-29.

[9] Diebold, F.X. and Mariano, R.S. (1995)," Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, Vol. 13, No. 3, pp. 253-263.

[10] Elliott, G. and Muller, U (2007), "Condence Sets for the Date of a Single Break in Linear Time Series Regressions", *Journal of Econometrics*, Vol.141, pp. 1196-1218.

[11] Giacomini, R. and White, H. (2006), "Tests of Conditional Predictive Ability", *Econometrica*, Vol. 74, pp. 1545-1578.

[12] Giacomini, R. and Rossi, B. (2006), "How Stable is the Forecasting Performance of the Yield Curve for Output Growth?", *Oxford Bulletin of Economics and Statistics*, Vol. 68, pp. 783-795.

[13] Giacomini, R. and Rossi, B. (2009), "Detecting and Predicting Forecast Breakdowns", *Review of Economic Studies*, Vol. 76, No. 2.

[14] Giacomini, R. and Rossi, B. (2010a), "Forecast Comparisons in Unstable Environments", *Journal of Applied Econometrics*, Vol. 25, No. 4, pp. 595-620.

[15] Giacomini, R. and Rossi, B. (2010b), "Model Comparisons in Unstable Environments", mimeo.

[16] Hansen, B. (1992) "Tests for Parameter Instability in Regression with I(1) Processes", *Journal of Buisness and Economic Statistics*, Vol. 10, No. 3, pp. 321-335.

[17] Inoue, A. and Rossi, B. (2010)," Out of Sample Forecast Tests Robust to the Window Size Choice", Duke University ERID Working Paper 94.

[18] Kozicki, S. and P.A. Tinsley (2001), "Shifting Endpoints in the Term Structure of Interest Rates",*Journal of Monetary Economics*, Vol. 47, pp. 613-652.

[19] Koop, G. and S.M. Potter (2007), "Estimation and Forecasting inModels withMultiple Breaks",*Review of Economic Studies*, Vol. 74, pp. 763-789.

[20] Pesaran, M.H. and Timmermann, A. (2002), "Market Timing and Return Prediction Under Model Instability", *Journal of Empirical Finance*, Vol. 9, No. 5, pp. 495-510.

[21] Pesaran, M.H. and A. Timmermann (2005), "Small Sample Properties of Forecasts from Autoregressive Models under Structural Breaks", *Journal of Econometrics*, Vol. 129, pp. 183-217.

[22] Pesaran, M.H. and A. Timmermann (2007), "Selection of Estimation Window in the Presence of Breaks", *Journal of Econometrics*, Vol. 137, No. 1, pp. 134-161.

[23] Pesaran, M. H., Pettenuzzo, D. and Timmermann, A. (2006), "Forecasting Time Series Subject to Multiple Structural Breaks", *Review of Economic Studies*, Vol. 73, pp. 1057-1084.

[24] Rapach, D.E. and Wohar, M.E. (2006), "Structural Breaks and Predictive Regression Models of Aggregate U.S. Stock Returns", *Journal of Financial Econometrics*, Vol. 4, No. 20, pp. 238-274.

[25] Rossi, B. (2005), "Optimal Tests for Nested Model Selections With Underlying Parameter Instabilities", *Econometric Theory*, Vol. 21, No. 5, pp. 962-990.

[26] Rossi, B. (2006), "Are Exchange Rates Really RandomWalks? Some Evidence Robust to Parameter Instability",*Macroeconomic Dynamics*, Vol. 10, No. 1, pp. 20-38.

[27] Rossi, B. (2011), "Advances in Forecasting Under Instability",*Macroeconomic Dynamics*, working paper, Duke University.

[28] Rossi, B. and Sekhposyan,T. (2010), "Have economic models forecasting performance for US output growth and inflation changed over time, and when?", *International Journal of Forecasting*, Vol. 26, No.4, pp. 808835.

[29] Rossi, B. and Sekhposyan, T. (2011a), "Understanding Models Forecasting Performance", *Journal of Econometrics*, Vol. 164, pp. 158-172.

[30] Rossi, B. and Sekhposyan, T. (2011b), "Forecast Optimality Tests in the Presence of Instabilities",mimeo, Duke University.

[31] Stock, J.H. and M.W. Watson (2003), "Forecasting Output and Ination: The Role of Asset Prices", *Journal of Economic Literature XLI*, pp. 788-829.

[32] Stock, J.H. and M.W. Watson (2004), "Combination Forecasts of Output Growth in a Seven Country Data Set", *Journal of Forecasting*, Vol. 23, pp. 405-430.

[33] Stock, J.H. and M.W. Watson (2007), "Why Has U.S. Inflation Become Harder to Forecast?", *Journal of Money, Credit and Banking*, Vol. 39, No. 1, pp. 3-33.

[34] White, H. (2001), "Asymptotic Theory for Econometricans, Revised Edition, Academic Press.

[35] West, K.D. (1996), "Asymptotic Inference about Predictive Ability", *Econometrica*, Vol. 64, pp. 1067-1084.

[36] Wright, J.H. (2008), "Bayesian Model Averaging and Exchange Rate Forecasts", *Journal of Econometrics*, Vol.146, No. 2, pp. 329-341.

Appendix A

Mathematical Proofs

At the core of all proofs is the indicator function I_t^{k,α_i}, which is a Bernoulli random variable with the following moments: $E(I_t^{k,\alpha_i}) = \alpha_i$, $Var(I_t^{k,\alpha_i}) = \alpha_i(1-\alpha_i)$ and covariance

$$r_h^{\alpha_i} \equiv cov(I_t^{k,\alpha_i}, I_{t-h}^{k,\alpha_i}) = \begin{cases} 0 & \text{if } h \neq k \\ \alpha_i^{3/2}(1-\alpha_i^{1/2}) & \text{if } h = k \end{cases}$$

When $h \neq k$ there is not common information between the indicators I_t^{k,α_i} and I_{t-h}^{k,α_i}, and since u_t's are i.i.d., their covariance is zero. When $h = k$, $cov(I_t^{k,\alpha_i}, I_{t-k}^{k,\alpha_i}) = E(I_t^{k,\alpha_i} \times I_{t-k}^{k,\alpha_i}) - \alpha_i^2 = \alpha_i^{3/2} - \alpha_i^2$.

Proof of Proposition 1:

$var(\widehat{\alpha_i}) = var(\frac{\sum_{t=k+1}^{T} I_t^{k,\alpha_i}}{T-k}) = \frac{1}{(T-k)}(var(I_t^{k,\alpha_i}) + 2r_k^{\alpha_i}) + o(1)$. Since $var(I_t^{k,\alpha_i}) = \alpha_i(1-\alpha_i)$ and $r_k^{\alpha_i} = \alpha_i^{3/2}(1-\alpha_i^{1/2})$

Then:

$$\sigma_{k,\alpha_i}^2 = var(I_t^{k,\alpha_i}) + 2r_k^{\alpha_i} = \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2})$$

Asymptotic normality of the test follows as in González-Rivera *et al.* (2011). The asymptotic variance of the test follows directly from the first and second moments of

the indicator function stated above.

Proof of Proposition 2:

For $j = k$,by proposition 1

$\lambda_{j,k} = cov(\ell_{j,\alpha_i}, \ell_{k,\alpha_i}) = var(\sqrt{T-k}(\widehat{\alpha}_i - \alpha_i)) = \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2})$

For $j > k$,

$$\begin{aligned} cov(\ell_{k,\alpha_i}, \ell_{j,\alpha_i}) &= cov(I_t^{k,\alpha_i}, I_t^{j,\alpha_i}) + cov(I_t^{k,\alpha_i}, I_{t-k}^{j,\alpha_i}) && \text{(A.1)} \\ &+ cov(I_{t-j}^{k,\alpha_i}, I_t^{j,\alpha_i}) + cov(I_t^{j,\alpha_i}, I_{t-j+k}^{k,\alpha_i}) + o(1) && \text{(A.2)} \end{aligned}$$

$$\begin{aligned} cov(I_t^{k,\alpha_i}, I_t^{j,\alpha_i}) &= E((I_t^{k,\alpha_i} - \alpha_i)(I_t^{j,\alpha_i} - \alpha_i)) && \text{(A.3)} \\ &= E(I_t^{k,\alpha_i} I_t^{j,\alpha_i}) - \alpha_i^2 = \alpha_i^{3/2} - \alpha_i^2 = \alpha_i^{3/2}(1-\alpha_i^{1/2}) && \text{(A.4)} \end{aligned}$$

Similarly

$$cov(I_t^{k,\alpha_i}, I_{t-k}^{j,\alpha_i}) = cov(I_{t-j}^{k,\alpha_i}, I_t^{j,\alpha_i}) = cov(I_{t-j+k}^{k,\alpha_i}, I_t^{j,\alpha_i}) = \alpha_i^{3/2}(1-\alpha_i^{1/2})$$

$$\begin{aligned} \lambda_{j,k} &= cov(I_t^{k,\alpha_i}, I_t^{j,\alpha_i}) + cov(I_t^{k,\alpha_i}, I_{t-k}^{j,\alpha_i}) + cov(I_{t-j}^{k,\alpha_i}, I_t^{j,\alpha_i}) + cov(I_t^{j,\alpha_i}, I_{t-j+k}^{k,\alpha_i}) \\ &= 4\alpha_i^{3/2}(1-\alpha_i^{1/2}) \end{aligned}$$

Due to the symmetric of covariance matrix, when $k > j$, $\lambda_{j,k} = 4\alpha_i^{3/2}(1-\alpha_i^{1/2})$.

The proof of asymptotic normality follows the method provided in Gonzalez-Rivera, et al.,'s paper (2011)

Proof of Proposition 3:

$cov(c_{k,i}, c_{k,j}) = cov(I_t^{k,\alpha_i}, I_t^{k,\alpha_j}) + cov(I_t^{k,\alpha_i}, I_{t-k}^{k,\alpha_j}) + cov(I_{t-k}^{k,\alpha_i}, I_t^{k,\alpha_j}) + o(1)$

If $i = j, \omega_{i,i} = var(c_{k,i}) = var(\sqrt{T-k}(\widehat{\alpha}_i - \alpha_i)) = \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2})$

If $i < j$,

$cov(I_t^{k,\alpha_i}, I_t^{j,\alpha_i}) = E(I_t^{k,\alpha_i} I_t^{k,\alpha_j}) - \alpha_i * \alpha_j = \alpha_i(1-\alpha_j)$

$cov(I_t^{k,\alpha_i}, I_{t-k}^{k,\alpha_i}) = E(I_t^{k,\alpha_i} I_{t-k}^{k,\alpha_j}) - \alpha_i * \alpha_j = \alpha_i * \sqrt{\alpha_j} - \alpha_i * \alpha_j$

$cov(I_{t-k}^{k,\alpha_i}, I_t^{k,\alpha_i}) = E(I_{t-k}^{k,\alpha_i} I_t^{k,\alpha_j}) - \alpha_i * \alpha_j = \alpha_i * \sqrt{\alpha_j} - \alpha_i * \alpha_j$

$\omega_{i,j} = cov(I_t^{k,\alpha_i}, I_t^{k,\alpha_j}) + cov(I_t^{k,\alpha_i}, I_{t-k}^{k,\alpha_j}) + cov(I_{t-k}^{k,\alpha_i}, I_t^{k,\alpha_j}) = \alpha_i(1-\alpha_j) + 2\alpha_i * \sqrt{\alpha_j}(1-\sqrt{\alpha_j})$

Similarly when $i > j$:

$\omega_{i,j} = cov(c_{k,i}, c_{k,j}) = \alpha_j(1-\alpha_i) + 2\alpha_j * \sqrt{\alpha_i}(1-\sqrt{\alpha_i})$

The proof of asymptotic normality follows the method provided in Gonzalez-Rivera, et al.,'s paper (2011)

Proof of Proposition 5: Now, we are considering about location scale model: $y_t = \mu + \sigma\varepsilon_t$, ε_t is $i.i.d$

Assumption 1:

$$\sqrt{T}(\widehat{\theta}_T - \theta_0) \to N(0, A^{-1}BA^{-1}) \tag{A.5}$$

$$A = -E(H(\theta_0)), H(\theta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_t(\theta_0) \text{ is the Hessian} \tag{A.6}$$

$$B = E(S(\theta_0)S(\theta_0)), S(\theta_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_t s_t(\theta_0) \text{ is the score vector} \tag{A.7}$$

Assumption 2:

$$P = \lim_{T\to\infty} E(\frac{\partial\widehat{\alpha}_i(\theta)}{\partial\theta}|_{\theta=\theta_0}) = \lim_{T\to\infty} E(\frac{\partial(\widehat{\alpha}_i(\varepsilon_t(\theta)))}{\partial u_t}.\frac{\partial u_t}{\partial\varepsilon_t}.\frac{\partial\varepsilon_t}{\partial\theta}|_{\theta=\theta_0}) \tag{A.8}$$

P is bounded for all its elements.

Assumption 3:

$$cov(I_t^{k,\alpha_i}, s_t(\theta_0)) < \infty, cov(I_t^{k,\alpha_i}, s_{t-k}(\theta_0)) < \infty \tag{A.9}$$

Proposition 4: under assumption 1 to 3:

$$\sqrt{T}(\widehat{\alpha}_i(\widehat{\theta}_T) - \alpha_i) \to N(0, \xi^2_{k,i}) \tag{A.10}$$

$$\xi^2_{k,i} = \sigma^2_{k,i} + PA^{-1}BA^{-1}P + 2E[\sqrt{T}(\widehat{\alpha}_i(\theta_0) - \alpha_i)S^{'}(\theta_0)]A^{-1}P \tag{A.11}$$

The first assumption is a general assumption for consistency of QLME estimators. The second assumption ensures a bound gradient for the statistic, which can be relaxed. Assumption 3 states that the correlation is bounded. The proof of this proposition can follow the similar method used to prove proposition 4 in GSY's paper(2010).

Proposition 5: for Gaussian location –scale model: $y_t = \mu_0 + \sigma_0\varepsilon_t$,where $\varepsilon_t \sim$ *i.i.d* $N(0,1)$

$$P = [\frac{2f(z_{\sqrt{\alpha_i}})\sqrt{\alpha_i}}{\sigma_0}, \frac{f(z_{\sqrt{\alpha_i}})z_{\sqrt{\alpha_i}}\sqrt{\alpha_i}}{\sigma^2_0}]^{'} \tag{A.12}$$

$$E[\sqrt{T}(\widehat{\alpha}_i(\theta_0) - \alpha_i)S^{'}(\theta_0)] = E\begin{bmatrix} \frac{I_t^{k.\alpha_i}\varepsilon_t}{\sigma_0} \\ \frac{I_t^{k.\alpha_i}(\varepsilon_t^2-1)}{\sigma_0^2} \end{bmatrix}^{'} + E\begin{bmatrix} \frac{I_t^{k.\alpha_i}\varepsilon_{t-k}}{\sigma_0} \\ \frac{I_t^{k.\alpha_i}(\varepsilon_{t-k}^2-1)}{\sigma_0^2} \end{bmatrix}^{'} \tag{A.13}$$

$$E[I_t^{k.\alpha_i}\varepsilon_t] = -\sqrt{\alpha_i}f(F^{-1}(\sqrt{\alpha_i})) \tag{A.14}$$

Similarly:

$$E[I_t^{k.\alpha_i}\varepsilon_{t-k}] = -\sqrt{\alpha_i}f(F^{-1}(\sqrt{\alpha_i})) \tag{A.15}$$

$$E[I_t^{k.\alpha_i}\varepsilon_t^2] = \int_{-\infty}^{F^{-1}(\sqrt{\alpha_i})} f(\varepsilon_{t-k})d\varepsilon_{t-k} \int_{-\infty}^{F^{-1}(\sqrt{\alpha_i})} \varepsilon_t^2 f(\varepsilon_t)d\varepsilon_{t.} \tag{A.16}$$

$$E[I_t^{k.\alpha_i}\varepsilon_{t-k}^2]] = \int_{-\infty}^{F^{-1}(\sqrt{\alpha_i})} f(\varepsilon_t)d\varepsilon_t \int_{-\infty}^{F^{-1}(\sqrt{\alpha_i})} \varepsilon_{t-k}^2 f(\varepsilon_{t-k})d\varepsilon_{t-k.} \tag{A.17}$$

Note: Here F denotes the CDF of standard normal distribution and f denotes the PDF of the standard normal distribution. Step 1

$$\begin{aligned} E[\sqrt{T}(\widehat{\alpha_i}(\theta_0)-\alpha_i)S^{'}(\theta_0)))] &= \lim_{T\to 0} E[\frac{1}{T}\sum_{t=1}^{T}(I_t^{k.\alpha_i}-\alpha_i)\sum_{t=1}^{T}s_t^{'}(\theta_0) \\ &= E(I_t^{k.\alpha_i}s_t^{'}(\theta_0))+E[I_t^{k.\alpha_i}s_{t-k}^{'}(\theta_0)]+o(1) \\ &= E\begin{bmatrix} \frac{I_t^{k.\alpha_i}\varepsilon_t}{\sigma_0} \\ \frac{I_t^{k.\alpha_i}(\varepsilon_t^2-1)}{\sigma_0^2} \end{bmatrix}^{'} + E\begin{bmatrix} \frac{I_t^{k.\alpha_i}\varepsilon_{t-k}}{\sigma_0} \\ \frac{I_t^{k.\alpha_i}(\varepsilon_{t-k}^2-1)}{\sigma_0^2} \end{bmatrix}^{'} + o(1) \end{aligned}$$

$$\begin{aligned} E[I_t^{k.\alpha_i}\varepsilon_t] &= E[1(0\le u_t\le\sqrt{\alpha_i}, 0\le u_{t-k}\le\sqrt{\alpha_i})\varepsilon_t] && \text{(A.18)} \\ &= \iint\limits_{0\le u_t\le\sqrt{\alpha_i}, 0\le u_t\le\sqrt{\alpha_i}} \varepsilon_t f(\varepsilon_t)f(\varepsilon_{t-k})d\varepsilon_t d\varepsilon_{t-k} && \text{(A.19)} \\ &= \int_{-\infty}^{F^{-1}(\sqrt{\alpha_i})} f(\varepsilon_{t-k})d\varepsilon_{t-k}\int_{-\infty}^{F^{-1}(\sqrt{\alpha_i})}\varepsilon_t f(\varepsilon_t)d\varepsilon_t && \text{(A.20)} \\ &= -\sqrt{\alpha_i}f(F^{-1}(\sqrt{\alpha_i})) && \text{(A.21)} \end{aligned}$$

$E[I_t^{k.\alpha_i}\varepsilon_{t-k}] = -\sqrt{\alpha_i}f(F^{-1}(\sqrt{\alpha_i}))$

$$\begin{aligned} E[I_t^{k.\alpha_i}\varepsilon_t^2] &= E[1(0\le u_t\le\sqrt{\alpha_i}, 0\le u_{t-k}\le\sqrt{\alpha_i})\varepsilon_t^2] && \text{(A.22)} \\ &= \int_{-\infty}^{F^{-1}(\sqrt{\alpha_i})} f(\varepsilon_{t-k})d\varepsilon_{t-k}\int_{-\infty}^{F^{-1}(\sqrt{\alpha_i})}\varepsilon_t^2 f(\varepsilon_t)d\varepsilon_{t.} && \text{(A.23)} \end{aligned}$$

$E[I_t^{k.\alpha_i}\varepsilon_{t-k}^2] = \int_{-\infty}^{F^{-1}(\sqrt{\alpha_i})} f(\varepsilon_t)d\varepsilon_t \int_{-\infty}^{F^{-1}(\sqrt{\alpha_i})}\varepsilon_{t-k}^2 f(\varepsilon_{t-k})d\varepsilon_{t-k.}$

Step 2

Properties of dirac delta function: $\delta(x)$

$$\int_{-\infty}^{+\infty}\delta(x)g(x)dx = g(0) \qquad \text{(A.24)}$$

$$\int_{-\infty}^{+\infty} \delta(x-a)g(x)dx = g(a) \tag{A.25}$$

$$H_a^{'}(x) = \delta(x-a) \tag{A.26}$$

Where $H_a(x)$ is the Heaviside function s.t:

$$H_a(x) = \{ \begin{array}{l} 0 \text{ if } x \leq a \\ \\ 1 \text{ if } x \geq a \end{array}$$

Define $\chi_{[0,a]*[0,a]}$ to be a binary indication function s.t.:

$$\chi_{[0,a]*[0,a]}(x,y) = \{ \begin{array}{ll} 1 \text{ if } 0 \leq x \leq a \text{ and } 0 \leq y \leq a \\ \\ 0 \quad \text{otherwise} \end{array}$$

The partial derivative of $\chi_{[0,a]*[0,a]}(x,y)$ are:

$$\frac{\partial \chi_{[0,a]*[0,a]}(x,y)}{\partial x} = \frac{\partial \chi_{[0,a]}(x)}{\partial x} * \chi_{[0,a]}(y) = (\delta(x) - \delta(x-a)) * \chi_{[0,a]}(y) \tag{A.27}$$

$$\frac{\partial \chi_{[0,a]*[0,a]}(x,y)}{\partial y} = \frac{\partial \chi_{[0,a]}(y)}{\partial y} * \chi_{[0,a]}(x) = (\delta(y) - \delta(y-a)) * \chi_{[0,a]}(x) \tag{A.28}$$

$$P = \lim_{T\to\infty} E(\frac{\partial \widehat{\alpha_i}(\theta)}{\partial \theta}|_{\theta=\theta_0}) = \lim_{T\to\infty} E[\begin{array}{c} \frac{\partial \widehat{\alpha_i}(\theta)}{\partial \mu_0} \\ \\ \frac{\partial \widehat{\alpha_i}(\theta)}{\partial \sigma_0^2} \end{array}] = \begin{pmatrix} P_1 \\ \\ P_2 \end{pmatrix}$$

$$\begin{aligned} P_1 &= \lim_{T\to\infty} E[\frac{1}{T} \frac{\sum_{t=1}^{T} \partial I_t^{k.\alpha_i}(\theta)}{\partial \mu_0}] \\ &= \lim_{T\to\infty} E[\frac{1}{T} \frac{\sum_{t=1}^{T} \partial\{1 \cdot (0 \leq u_t \leq \sqrt{\alpha_i}, 0 \leq u_{t-k} \leq \sqrt{\alpha_i})\}}{\partial \mu_0}] \end{aligned}$$

By (30) and (31)

$$\begin{aligned}\frac{\partial I_t^{k.\alpha_i}(\theta)}{\partial \mu_0} &= \frac{\partial I_t^{k.\alpha_i}(\theta)}{\partial u_t}\cdot\frac{\partial u_t}{\partial \mu_0}+\frac{\partial I_t^{k.\alpha_i}(\theta)}{\partial u_t}\cdot\frac{\partial u_t}{\partial \mu_0}\\ = \;& (-\frac{1}{\sigma_0})\{[\delta(u_t)-\delta(u_t-\sqrt{\alpha_i})]\chi_{[0,\sqrt{\alpha_i}]}^{u_t-k}f(\varepsilon_t)\\ & +[\delta(u_{t-k})-\delta(u_{t-k}-\sqrt{\alpha_i})]\chi_{[0,\sqrt{\alpha_i}]}^{u_t}f(\varepsilon_{t-k})\}\end{aligned}$$

Use (27) and (28),

$P_1 = E[\frac{\partial I_t^{k.\alpha_i}(\theta)}{\partial \mu_0}] = \frac{2}{\sigma_0}\sqrt{\alpha_i}f(z_{\sqrt{\alpha_i}}), z_{\sqrt{\alpha_i}}\ is$ quantile of the standard normal distribution

$P_2 = \lim_{T\to\infty} E[\frac{1}{T}\frac{\sum_{t=1}^{T}\partial I_t^{k.\alpha_i}(\theta)}{\partial \sigma_0}] = \lim_{T\to\infty} E[\frac{1}{T}\frac{\sum_{t=1}^{T}\partial\{1\cdot(0\le u_t\le\sqrt{\alpha_i},0\le u_t\le\sqrt{\alpha_i})\}}{\partial \sigma_0^2}\cdot \frac{1}{2\sigma_0}]$

By (30) and (31)

$$\frac{\partial I_t^{k.\alpha_i}}{\partial \sigma_0^2} = \frac{1}{2\sigma_0}[\frac{\partial I_t^{k.\alpha_i}}{\partial u_t}\cdot\frac{\partial u_t}{\partial \sigma_0}+\frac{\partial I_t^{k.\alpha_i}}{\partial u_t}\cdot\frac{\partial u_t}{\partial \sigma_0}] \tag{A.29}$$

$$= -\frac{1}{2\sigma_0^2}\{[\delta(u_t)-\delta(u_t-\sqrt{\alpha_i})]\chi_{[0,\sqrt{\alpha_i}]}^{u_t-k}f(\varepsilon_t)\varepsilon_t \tag{A.30}$$

$$+ \;[\delta(u_{t-k})-\delta(u_{t-k}-\sqrt{\alpha_i})]\chi_{[0,\sqrt{\alpha_i}]}^{u_t})f(\varepsilon_{t-k})\varepsilon_{t-k}\} \tag{A.31}$$

Use (27) ,(28) and the fact the $\varepsilon_t(u_t) = F^{-1}(u_t)$, $P_2 = E[\frac{I_t^{k.\alpha_i}}{\partial \sigma_0^2}] = \frac{1}{\sigma_0^2}\sqrt{\alpha_i}f(z_{\sqrt{\alpha_i}})z_{\sqrt{\alpha_i}}$

$$P = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = \begin{pmatrix} \frac{2f(z_{\sqrt{\alpha_i}})\sqrt{\alpha_i}}{\sigma_0} \\ \frac{f(z_{\sqrt{\alpha_i}})z_{\sqrt{\alpha_i}}\sqrt{\alpha_i}}{\sigma_0^2} \end{pmatrix}$$

Appendix B

Procedures to Take PIT for Bivariate Student-t Distribution

Bivariate Student-t Distributions:

$$f(x) = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v}{2})(v\pi)}|\Sigma|^{-\frac{1}{2}}[1 + \frac{1}{v}x^t\Sigma^{-1}x]^{-\frac{v+2}{2}}$$

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

$$|\Sigma| = 1 - \rho^2$$

$$\Sigma^{-1} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \frac{1}{1-\rho^2}$$

$$f(x_1) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{v\pi}}[1 + \frac{x_1^2}{v}]^{-\frac{v+1}{2}}$$

$$f(x_2|x_1) = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v+1}{2})\sqrt{v\pi}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \frac{[1+\frac{x_1^2}{v}]^{\frac{v+1}{2}}}{[1+\frac{1}{v}x^t\Sigma^{-1}x]^{\frac{v+2}{2}}}$$

In order to simplify the formula, we can decompose $x^t\Sigma^{-1}x$ first in following way:

$$x^t\Sigma^{-1}x = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{bmatrix} \frac{1}{1-\rho^2} & -\frac{\rho}{1-\rho^2} \\ -\frac{\rho}{1-\rho^2} & \frac{1}{1-\rho^2} \end{bmatrix} * \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \frac{x_1^2}{1-\rho^2} - \frac{2\rho}{1-\rho^2}x_1x_2 + \frac{x_2^2}{1-\rho^2}$$

$= x_1^2 + \frac{1}{1-\rho^2}(x_2 - \rho x_1)^2$

Then $f(x_2|x_1)$ can be rewritten into following format:

$$\begin{aligned} f(x_2|x_1) &= \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v+1}{2})\sqrt{v\pi}} \frac{1}{\sqrt{1-\rho^2}} \frac{[(1+\frac{1}{v}x_1^2) + \frac{1}{v}\frac{1}{1-\rho^2}(x_2-\rho x_1)^2]^{-\frac{v+2}{2}}}{[1+\frac{x_1^2}{v}]^{-\frac{v+1}{2}}} \\ &= \frac{\Gamma(\frac{v+2}{v})}{\Gamma(\frac{v+1}{2})\sqrt{v\pi}} \frac{1}{\sqrt{1-\rho^2}} \left[\frac{[(1+\frac{1}{v}x_1^2) + \frac{1}{v}\frac{1}{1-\rho^2}(x_2-\rho x_1)^2]}{(1+\frac{x_1^2}{v})}\right]^{-\frac{v+2}{2}} \frac{1}{[1+\frac{x_1^2}{v}]^{\frac{1}{2}}} \\ &= = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v+1}{2})\sqrt{(v+x_1^2)\pi}\sqrt{1-\rho^2}} \left[1 + \frac{1}{1-\rho^2}\frac{(x_2-\rho x_1)^2]}{(v+x_1^2)})\right]^{-\frac{v+2}{2}} \end{aligned}$$

If $f(x_2|x_1)$ is student-t distribution, the condition:$v+x_1^2 = v+1$ should be hold, i.e. $x_1 = 1$. If $\rho \neq 0, |x_1| = 1, f(x_2|x_1) = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v+1}{2})\sqrt{(v+1)\pi}\sqrt{1-\rho^2}} \left[1 + \frac{1}{1-\rho^2}\frac{(x_2-\rho x_1)^2]}{(v+1)})\right]^{-\frac{v+2}{2}}$, which is univariate student-t distribution with degree freedom $v+1$ and correlation $1-\rho^2$.

If $\rho = 0, |x_1| = 1, f(x_2|x_1) = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v+1}{2})\sqrt{(v+1)\pi}} \left[1 + \frac{(x_2)^2}{(v+1)})\right]^{-\frac{v+2}{2}}$,which is univariate student-t distribution with degree freedom $v+1$.

In addition, we can observe that by using change of variables, *i.e*:

$\omega = \frac{1}{\sqrt{1-\rho^2}}\frac{(x_2-\rho x_1)}{\sqrt{1+\frac{1}{v}{x_1}^2}}\sqrt{\frac{v+1}{v}}$,we can obtain a variable, say, ω with univariate student-t distribution with degree of freedom$(v+1)$

$f(x_2|x_1) = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v+1}{2})\sqrt{(v+x_1^2)\pi}\sqrt{1-\rho^2}} \left[1 + \frac{1}{1-\rho^2}\frac{(x_2-\rho x_1)^2]}{(v+x_1^2)})\right]^{-\frac{v+2}{2}}$

$\omega = \frac{1}{\sqrt{1-\rho^2}}\frac{(x_2-\rho x_1)}{\sqrt{1+\frac{1}{v}{x_1}^2}}\sqrt{\frac{v+1}{v}}$.

$f(\omega) = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v+1}{2})\sqrt{(v+1)\pi}} \left[1 + \frac{\omega^2}{(v+1)}\right]^{-\frac{v+2}{2}}$

Methods to generate bivariate Student-t distribution and PIT based on it:

Method 1: Since the bivariate Student-t distribution is characterized by its correlation matrix, we can generate bivariate student t distribuion directly using its correlation matrix. Example one:

Given $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, $\rho = 0.4$; we can generate a random vector Y with bivariate student-t(v) distribution and Σ as its correlation matrix.

Step one: we can generate two independent random variables w_1 and w_2 with univariate student-t(5) respectively.

Let $A = chol(\Sigma)$, and $Y = A \begin{bmatrix} w_1 \\ \\ w_2 \end{bmatrix}$, which follows a bivariate student-t distribution, with correlation matrix Σ.

The reference can be found in:

> Sameul Kotz and Saralees Nadarajah:" Multiavariate t distributions and their application" 1.9 Distribution of a linear function P15

$i.e$:

$$f(y) = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v}{2})(v\pi)}|\Sigma|^{-\frac{1}{2}}[1 + \frac{1}{v}y^t\Sigma^{-1}y]^{-\frac{v+2}{2}}, \Sigma = \begin{bmatrix} 1 & 0.4 \\ \\ 0.4 & 1 \end{bmatrix}$$

Then the marginal distribution is

$f(y_1) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{v\pi}}[1 + \frac{y_1^2}{v}]^{-\frac{v+1}{2}}$

Conditional distribution for y_2 given y_1 is

$f(y_2|y_1) = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v+1}{2})\sqrt{v\pi}}\frac{1}{|\Sigma|^{\frac{1}{2}}}\frac{[1+\frac{y_1^2}{v}]^{\frac{v+1}{2}}}{[1+\frac{1}{v}y^t\Sigma^{-1}y]^{\frac{v+2}{2}}}$

We can apply PIT with respect to $f(y_1)$ and $f(y_2|y_1)$ respectively to obtain u^m and u^c :

$u^m = \int_{-\infty}^{y_1} f(\varphi)d\varphi$

$u^c = \int_{-\infty}^{y_2} f(\varphi)d\varphi$

Method two:

We can generate random vector Y, with a bivairate Student-t distribution by specifying its covariance matrix. As result of that, to obtain the correct distribution function for Y we need to tranform its covarince matrix into the correlation matirx.

Example two:

Given $\Omega = \begin{bmatrix} 0.64 & 0.4 \\ 0.4 & 0.64 \end{bmatrix}$, we can generate a random vector Y with bivariate student-t(v) distribution and Ω as its covariance matrix.

Step one: we can generate two independent random variables w_1 and w_2 with univariate student-t(5) respectively.

Let $A = chol(\Omega)$, and $Y = A \begin{bmatrix} \sqrt{\frac{v-2}{v}} w_1 \\ \sqrt{\frac{v-2}{v}} w_2 \end{bmatrix}$, which follows a bivariate student-t distribution, with covariance matix Ω.

We still use Σ to denote the correlation matrix of Y and $\Omega = \frac{v}{v-2} * \Sigma$,

Then the distribution function for Y is :

$$f(y) = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v}{2})(v\pi)} |\Sigma|^{-\frac{1}{2}} [1 + \frac{1}{v} y^t \Sigma^{-1} y]^{-\frac{v+2}{2}}, \ \Sigma = \frac{v}{v-2} * \begin{bmatrix} 0.64 & 0.4 \\ 0.4 & 0.64 \end{bmatrix}$$

Then the marginal and conditional distribution are slightly different from that in example one:

$$f(y_1) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{\Sigma_{1,1}}\sqrt{v\pi}} [1 + \frac{y_1^2}{v * \Sigma_{1,1}}]^{-\frac{v+1}{2}}$$

$$f(y_2|y_1) = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v+1}{2})\sqrt{v\pi}} \frac{\sqrt{\Sigma_{1,1}}}{|\Sigma|^{\frac{1}{2}}} \frac{[1 + \frac{y_1^2}{v * \Sigma_{1,1}}]^{\frac{v+1}{2}}}{[1 + \frac{1}{v} y^t \Sigma^{-1} y]^{\frac{v+2}{2}}}$$

Note: Σ is not the same as the true correlation matrix with all diagnoal elements equal to one. It should be viewed as a scale parameter in the bivariate student-t distribution.

We can apply PIT with respect to $f(y_1)$ and $f(y_2|y_1)$ respectively to obtain u^m and u^c :

$u^m = \int_{-\infty}^{y_1} f(\varphi)d\varphi$

$u^c = \int_{-\infty}^{y_2} f(\varphi)d\varphi$

Method three: Using change of variables to obtain two random variables with univariate student t distribution and then perform PIT

Suppose we have generated a random vector Y following exactly the same procedures in method two.

i.e

$$f(y) = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v}{2})(v\pi)}|\Sigma|^{-\frac{1}{2}}[1 + \frac{1}{v}y^t\Sigma^{-1}y]^{-\frac{v+2}{2}},\ \Sigma = \frac{v}{v-2} * \begin{bmatrix} 0.64 & 0.4 \\ 0.4 & 0.64 \end{bmatrix}$$

The marginal distribution keeps the same:

$$f(y_1) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{\Sigma_{1,1}}\sqrt{v\pi}}[1 + \frac{y_1^2}{v*\Sigma_{1,1}}]^{-\frac{v+1}{2}}$$

Then we apply changes of variables with respect to the conditional distribution:

Let $\xi = \frac{1}{\sqrt{\Sigma_{2,2}-\frac{\Sigma_{2,1}^2}{\Sigma_{1,1}}}} \frac{(y_2-\frac{\Sigma_{2,1}}{\Sigma_{1,1}}*y_1)}{\sqrt{1+\frac{1}{v*\Sigma_{1,1}}{y_1}^2}} \sqrt{\frac{v+1}{v}}$, then y_1 and ξ are independent, ξ has the univariate student-t distribution with degree of freedom $v+1$.

$$f(\xi) = \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v+1}{2})\sqrt{(v+1)\pi}} \left[1+\frac{\xi^2}{(v+1)}\right]^{-\frac{v+2}{2}}$$

Then we can apply PIT to y_1 and ξ respectively to otain u^m and u^c :

$$u^m = \int_{-\infty}^{y_1} f(\varphi)d\varphi = \int_{-\infty}^{y_1} \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{\Sigma_{1,1}}\sqrt{v\pi}}[1+\frac{\varphi^2}{v*\Sigma_{1,1}}]^{-\frac{v+1}{2}}d\varphi 1$$

$$u^c = \int_{-\infty}^{\xi} f(\varphi)d\varphi = \int_{-\infty}^{\xi} \frac{\Gamma(\frac{v+2}{2})}{\Gamma(\frac{v+1}{2})\sqrt{(v+1)\pi}} \left[1+\frac{\varphi^2}{(v+1)}\right]^{-\frac{v+2}{2}} d\varphi$$

I use first two methods to generate bivariate Student-t distribution, whereas in the paper "Autocontour-based Evaluation of Multivariate Predictive Densities". (Gloria Gonz alez-Rivera and Emre Yoldas 2010), method three is used to generate bivariate Student-t distribuion. Actually, method two and method three are equivalent to each other

Appendix C

Mixture Inverse Gumbel/Clayton

The estimated weight for the mixture is $p = 0.333$ on the inverse Gumbel, and $1 - p = 0.667$ on Clayton. The estimated dependence is $\hat{\theta} = 1.936$ for inverse Gumbel, and $\hat{\theta} = 0.444$ for Clayton.

Table C.1: Autocontour-based t-tests:

Negative Binomial Model with Contemporaneous Correlation

	Inverse Gumbel/Clayton		
α_i	lag 1	lag 2	lag 3
1%	-1.42	-1.15	-1.78
5%	1.23	1.53	1.66
10%	3.26	2.55	3.13
20%	3.95	3.65	4.07
30%	4.60	4.01	4.68
40%	4.23	3.98	4.23
50%	3.72	3.78	3.94
60%	2.48	2.83	2.32
70%	1.34	1.01	1.39
80%	-0.95	-1.26	-1.04
90%	-3.42	-3.78	-3.57
95%	-8.95	-8.32	-8.03
99%	-14.24	-14.47	-14.36

t-tests critical values:
$\pm$ 1.96 (5% level),
$\pm$ 2.58 (1%), $\pm$ 3.5 (0.05%).