Essays on Asymmetric Loss and Learning in Macroeconomic Forecasting

Sagarika Mishra
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Economics Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/796
ACKNOWLEDGEMENTS

First, I would like to convey my sincere thanks to my chair Prof. Matthew L. Higgins for his continuous support and supervision throughout this dissertation. Without his help this work would have never been completed this dissertation. I am also grateful to my other committee members, Prof. C. James Hueng and Prof. J. Kevin Corder, who helped me in this project. I would also like to thank my brother Santosh Mishra and my husband Sandip Dhole for their help. I also thank my parents, in-laws and closest friends for their emotional support.

Sagarika Mishra
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ................................................................. ii

LIST OF TABLES ........................................................................... v

LIST OF FIGURES ....................................................................... vi

CHAPTER

I. TESTING FORECAST RATIONALITY UNDER ASYMMETRIC LOSS WITH THE MINCER-ZARNOWITZ REGRESSION ............ 1

1.1 Introduction ................................................................. 1

1.2 Augmented Mincer-Zarnowitz Regression ......................... 3

1.3 Homogenous Loss Functions With Mean-Variance Dynamics ................................................................. 8

1.4 Linex Loss Function ..................................................... 13

1.5 Asymmetric Loss and U.S. Inflation Forecasts ................. 15

1.6 Conclusions ............................................................... 19

II. STATE DEPENDENT ASYMMETRIC LOSS ....................... 23

2.1 Introduction ................................................................. 23

2.2 Theoretical Results ....................................................... 25

2.2.1 Linex Loss .............................................................. 27

2.2.2 Asymmetric Power Loss ........................................... 29

2.3 Survey of Professional Forecasters ................................. 31

2.4 Data and Variables that Can Explain Bias .................... 32
Table of Contents – Continued

CHAPTER

2.5 Empirical Results ................................................. 35
2.6 Conclusions ......................................................... 38

III. DO AGENTS LEARN BY LEAST SQUARES? THE EVIDENCE PROVIDED BY CHANGES IN MONETARY POLICY .......... 44

3.1 Introduction ......................................................... 44
3.2 Theoretical Results ............................................. 47

3.2.1 Properties of Forecast Error under Least Squares Learning ................................................. 48

3.2.2 Properties of Forecast Error under Least Squares Learning in the Presence of Structural Breaks ......................... 50

3.3 The Taylor Rule and an Interest Rate Forecasting Model ......................................................... 54

3.3.1 Specification of the Models ...................... 54
3.3.2 Data ......................................................... 57

3.4 Empirical Results .................................................. 60

3.4.1 Structural Change Tests Results ................. 60
3.4.2 Test for Standardized Residuals .............. 64

3.5 Conclusions .......................................................... 67

REFERENCES ............................................................. 78
LIST OF TABLES


2. ARCH LM Test and ARCH-Sup LM Test .................................................. 20

3. Determinants of Time-varying Asymmetry Parameter in the One Quarter Ahead GDP Growth Rate, 1981:3-2007:4 .................................................. 40


5. Bai and Perron (2001) 95% Confidence Interval for the Break Dates .......... 69

6. ACF, Q-stat and P-values of the Standardized Residuals under Multiple Structural Breaks in the Mean 3-month Treasury Bill Rate ...................... 72

7. ACF, Q-stat and P-values of the Standardized Residuals under Multiple Structural Breaks in the Median 3-month Treasury Bill Rate ...................... 73

8. ACF, Q-stat and P-values of the Standardized Residuals of the Mean 3-month Treasury Bill Rate under Multiple Structural Breaks in the Taylor Rule .................................................. 74

9. ACF, Q-stat and P-values of the Standardized Residuals of the Median 3-month Treasury Bill Rate under Multiple Structural Breaks in the Taylor Rule .................................................. 75

10. ACF, Q-stat and P-values of the Standardized Residuals of the Mean 3-month Treasury Bill Rate under Multiple Structural Breaks in the Taylor Rule from Duffy and Warnick (2006) .................................................. 76

11. ACF, Q-stat and P-values of the Standardized Residuals of the Median 3-month Treasury Bill Rate under Multiple Structural Breaks in the Taylor Rule from Duffy and Warnick (2006) .................................................. 77
LIST OF FIGURES

1. Asymmetric Power Loss Function for U.S. Inflation Forecasts .............. 21
2. Linex Loss Function for U.S. Inflation Forecasts ............................. 22
3. Time-varying Asymmetry Parameter for Linex Loss .......................... 41
4. Time-varying Bias for Linex Loss .................................................. 41
5. Time-varying Asymmetry Parameter for Linlin Loss .......................... 42
6. Time-varying Bias for Linlin Loss .................................................. 42
7. Time-varying Asymmetry Parameter for Quadquad Loss ...................... 43
8. Time-varying Bias for Quadquad Loss ............................................ 43
9. Raw Forecast Error and Mean Corrected Forecast Error with Three Structural Breaks ................................................................. 70
10. Conditional Standard Deviation of the Forecast Error with Three Structural Breaks ................................................................. 71
1.1 Introduction

The rational expectations hypothesis (REH) underlies most modern macroeconomics. Assuming agents face quadratic loss, the rational forecast is the mean of a series conditional on available information. If the forecast is the conditional mean, the forecast errors should have mean zero and be orthogonal to any variable in the information set. The unbiasedness and efficiency properties implied by the REH have been frequently tested using forecasts obtained from surveys of professional forecasters. In one of the first formal tests of unbiasedness and efficiency, Mincer and Zarnowitz (1969) proposed regressing the observed series on an intercept and the forecast, and testing whether intercept is zero and the coefficient on the forecast is one. Most unbiasedness and efficiency tests are a variation of the Mincer-Zarnowitz (MZ) regression test. Although the results depend on the survey being used, the sample period and the series being considered, unbiasedness and efficiency are often rejected for the forecasts obtained from surveys. Pesaran and Weale (2006) provide an excellent survey of the conclusions from rationality tests using forecast survey data.

Rather than concluding that agents are irrational, researchers have recently proposed that the apparent bias and inefficiency in the survey forecasts may be due to the fact that forecasters face asymmetric loss. Granger (1969) and Christoffersen and Diebold (1997)
showed that under asymmetric loss, the optimal forecast is the conditional mean plus a bias that depends on the second and higher order conditional moments of the process and on the parameters in the loss function. Batchelor and Peel (1998) showed that standard regression tests can be modified to provide a valid test for forecast unbiasedness and efficiency in the presence of asymmetric loss assuming agents have linex loss and the conditional distribution of the innovation series is normal. More recently, Elliot, Komunjer and Timmerman (2005) proposed a GMM test of forecast rationality when the loss function is asymmetric power. Patton and Timmerman (2007) presented a very general test which only assumes the loss function is homogeneous and that the dynamics of the process can be represented through a time-varying conditional mean and variance.

In this paper, we build on the results of Batchelor and Peel (1998) and show that the MZ regression can be modified to provide a valid test of forecast rationality for any of the loss functions and assumed dynamic properties of the series that have been considered in the literature. In the next section, we first provide a detailed analysis of the probability limits of the coefficients in the conventional MZ regression to show how asymmetric loss and the process dynamics affect the standard test. We then show at a very general level how to augment the MZ regression to make it a valid test for rationality in the presence of asymmetric loss. In section 1.3, we demonstrate how the augmented regression can be used to test rationality when the loss function is homogenous and the dynamics of the process are captured through a time-varying conditional mean and variance. We emphasize the advantages and disadvantages of our test compared to those of Elliott, Komunjer and Timmermann (2005) and Patton and Timmerman (2007). In section 1.4, we show how the augmented regression can be used to test rationality when the loss
function is linear and there is a closed form expression for the moment generating function of the conditional distribution of the series. In section 1.5, we present an empirical example in which we use the augmented MZ regression to test whether inflation forecasts from the Livingston survey are rational allowing for asymmetric loss. Section 1.6 is a short conclusion.

1.2 Augmented Mincer-Zarnowitz Regression

Let \( y_t, t = 1, \ldots, T \) be an observed sequence of stationary random variables. Using information in period \( t \), agents forecast \( y_t \) at horizon \( h \). The series in period \( (t+h) \) can be decomposed as

\[
y_{t+h} = \mu_{t+h|t} + \epsilon_{t+h},
\]

where \( \mu_{t+h|t} \) is the mean of the process conditional on the information set available in period \( t \) and \( \epsilon_{t+h} \) is an innovation which has mean zero and is uncorrelated with the elements in the information set. The forecasted value of \( y_{t+h} \) conditioned on the information set available at time period \( t \) is denoted by \( \hat{y}_{t+h|t} \). Under quadratic loss the optimal predictor is \( \hat{y}_{t+h|t} = \mu_{t+h|t} \) and the forecast error is the innovation \( y_{t+h} - \hat{y}_{t+h|t} = \epsilon_{t+h} \). As a result, the forecasts are unbiased and the forecast errors are uncorrelated with the elements of the information set used to construct the forecast.

Assuming the loss is only a function of the forecast error, Granger (1969) and Christoffersen and Diebold (1997) showed that under asymmetric loss the optimal predictor is
\[
\hat{y}_{t+h} = \mu_{t+h} + \lambda_{t+h},
\]

(2)

where \( \lambda_{t+h} \) depends on the loss function and the conditional moments of order two and higher. The presence of asymmetric loss introduces a bias into the forecast. If agents have an asymmetric loss function, the usual properties of rational forecasts no longer hold. Apparent irrationality may be due to asymmetric loss.

The standard test of forecast rationality is the MZ regression and is conducted under the assumption of quadratic loss. The MZ regression regresses the actual value of the series on the forecasted value and tests the joint hypothesis that the constant is zero and the slope coefficient is one. The regression takes the form

\[
y_{t+h} = a_0 + a_1 \hat{y}_{t+h} + u_{t+h}.
\]

(3)

Testing \( a_0 = 0 \) provides a test of forecast unbiasedness and testing \( a_1 = 1 \) provides a test of forecast efficiency. Under quadratic loss, the optimal h-step ahead forecast is the conditional mean, so (3) is equivalent to (1) with \( a_0 = 0 \) and \( a_1 = 1 \). In contrast, when the loss function is asymmetric the MZ regression for testing rationality is inappropriate because

\[
y_{t+h} = \hat{y}_{t+h} - \lambda_{t+h,t} + \epsilon_{t+h,t}.
\]

(4)

The conventional MZ regression omits the variable \( \lambda_{t+h,t} \). The least square estimators of the coefficients are inconsistent if we omit a variable which is correlated with the other regressors. The probability limits of the regression coefficients under a general loss function are presented in the following proposition.
Proposition:- Assuming the bias is strictly a function of the forecast error, in the MZ regression (3), the probability limit of the LS estimator of the slope coefficient is

\[ \text{Plim}(\hat{a}_1) = \frac{\text{var}(\mu_{t+h,t}) + \text{cov}(\mu_{t+h,t}, \lambda_{t+h,t})}{\text{var}(\mu_{t+h,t}) + \text{var}(\lambda_{t+h,t}) + 2\text{cov}(\mu_{t+h,t}, \lambda_{t+h,t})} \]  

(5)

and the probability limit of the intercept is

\[ \text{Plim}(\hat{a}_0) = \mu_y - (\mu_y + \mu_\lambda)\text{Plim}(\hat{a}_1) \]  

(6)

where \( \mu_y = E(y_t) \) is the unconditional mean of \( y_t \) and \( \mu_\lambda = E(\lambda_t) \) is the unconditional mean of \( \lambda_t \).

Proof:- We have that

\[ \text{Plim}(\hat{a}_t) = \frac{\text{cov}(y_{t+h,t}, \hat{y}_{t+h,t})}{\text{var}(\hat{y}_{t+h,t})} \]

\[ = E\left\{ (\mu_{t+h,t} + \epsilon_{t+h,t})(\mu_{t+h,t} + \lambda_{t+h,t}) \right\} - E(\mu_{t+h,t} + \epsilon_{t+h,t})E(\mu_{t+h,t} + \lambda_{t+h,t}) \]

\[ \text{var}(\mu_{t+h,t} + \lambda_{t+h,t}) \]

(7)

where we substitute in the value of \( y_{t+h} \) from (1) and \( \hat{y}_{t+h,t} \) from (2). Using the fact that \( \epsilon_{t+h,t} \) is orthogonal to the information set in period \( t \), we can easily simplify (6) to get (4). Also

\[ \text{Plim}(\hat{a}_0) = \text{Plim}(\bar{y}) - \text{Plim}(\hat{a}_1)\text{Plim}(\bar{y}) \]

\[ = E(\mu_{t+h,t}) - \text{Plim}(\hat{a}_1)E(\mu_{t+h,t}) - \text{Plim}(\hat{a}_1)E(\lambda_{t+h,t}) \]

(8)

which is (6).
The above proposition indicates the behavior of the MZ regression depends on whether $\lambda_{t+h,t}$ is time-varying, and if so, the value of $cov(\mu_{t+h,t}, \lambda_{t+h,t})$. If $\lambda_{t+h,t}$ is time invariant, then $cov(\mu_{t+h,t}, \lambda_{t+h,t}) = 0$ and $var(\lambda_{t+h,t}) = 0$. It is then easily seen that $Plim(\hat{\lambda}_1) = 1$ and $Plim(\hat{\lambda}_0) = -\mu$. The slope coefficient is consistent but the intercept has an upward asymptotic bias if $\mu$ is negative and a downward asymptotic bias if $\mu$ is positive. In this case, the MZ regression is likely to reject rationality due to the bias in the forecast even though the bias may be rational given the loss function. When $\lambda_{t+h,t}$ is time varying but $cov(\mu_{t+h,t}, \lambda_{t+h,t}) = 0$, the estimator of the slope coefficient will have a downward asymptotic bias

$$\frac{var(\mu_{t+h,t})}{var(\mu_{t+h,t}) + var(\lambda_{t+h,t})} < 1$$

When $\mu_{t+h,t}$ and $\lambda_{t+h,t}$ are correlated, the direction of the asymptotic bias in the estimator of the slope coefficient depends on the sign of the covariance. If $cov(\mu_{t+h,t}$, $\lambda_{t+h,t}) > 0$, then

$$\frac{var(\mu_{t+h,t}) + cov(\mu_{t+h,t}, \lambda_{t+h,t})}{var(\mu_{t+h,t}) + var(\lambda_{t+h,t}) + 2cov(\mu_{t+h,t}, \lambda_{t+h,t})} < 1$$

If $cov(\mu_{t+h,t}$, $\lambda_{t+h,t}) < 0$, the sign of $Plim(\hat{\lambda}_1)$ is indeterminate. If $var(\lambda_{t+h,t}) > cov(\mu_{t+h,t}, \lambda_{t+h,t})$, then $Plim(\hat{\lambda}_1) > 1$, while if $var(\lambda_{t+h,t}) < cov(\mu_{t+h,t}, \lambda_{t+h,t})$, $Plim(\hat{\lambda}_1) < 1$. For most of the loss functions that have been considered in the literature, $\lambda_{t+h,t}$ is a simple function of second and higher order moments. Depending on the variable being forecasted, there may be prior knowledge about whether $\lambda_{t+h,t}$ is time varying and the likely sign of $cov(\mu_{t+h,t}, \lambda_{t+h,t})$. In such a case, the researcher may be
able to conjecture about the direction of the bias in $\hat{a}_t$. We will illustrate this later in the paper.

To test forecast rationality under possibly asymmetric loss using the MZ regression it is necessary to include $\lambda_{t+h,t}$ in the regression. Let $\alpha$ denote the parameters in loss function and let $\theta$ denote the parameters in the conditional distribution of $\varepsilon_{t+h}$. Forecast rationality can in principle be tested by estimating the augmented Mincer-Zarnowitz (AMZ) regression

$$y_{t+h} = a_0 + a_1 \hat{y}_{t+h,t} - \lambda_{t+h,t}(\alpha, \theta) + \varepsilon_{t+h,t}$$

by maximum likelihood and testing $a_0 = 0$ and $a_1 = 1$. In (9) we explicitly show the dependence of $\lambda_{t+h,t}$ on $\alpha$ and $\lambda$. Again, the exact parametric form of $\lambda_{t+h,t}(\alpha, \theta)$ will depend on the assumed loss function employed by forecasters and the specified conditional distribution of $\varepsilon_{t+h}$. Note in (9) if $\lambda_{t+h,t}(\alpha, \theta)$ is time invariant, it is not separately identifiable from $a_0$. This implies that all of the dynamics of the process are completely captured through the conditional mean, which is true for many economic time series. The AMZ regression cannot be used to determine if a bias in the forecasts is due to asymmetric loss. In the absence of time-varying higher order moments, however, is still meaningful to test forecast efficiency by testing $H_0: a_1 = 1$. If $\lambda_{t+h,t}(\alpha, \theta)$ is time varying, $\alpha$ and $\theta$ may be identified in (9). For most of the loss functions that have been proposed, $\alpha$ is a scalar and the parameters in (9) are fully identified. An advantage of using the AMZ regression to test for forecast rationality is that it provides an estimate of the parameter in the loss function. Assuming that $\alpha$ is identified, if the null hypothesis of forecast unbiasedness cannot be rejected in (8), one can then estimate
\[ e_{t+h} = -\lambda_{t+h,t}(\alpha, \theta) + \varepsilon_{t+h} \]  

(10)

where \( e_{t+h} = y_{t+h} - \hat{y}_{t+h,t} \), by maximum likelihood to obtain a fully efficient estimator of \( \alpha \). In the following subsections we present examples of the AMZ regression (8) for different loss functions that have been used in the literature on forecasting under asymmetric loss.

1.3 Homogeneous Loss Functions With Mean-Variance Dynamics

A broad class of loss functions is the homogeneous class where for any constant \( c \) and some \( k \)

\[ L[c(y_{t+h} - \hat{y}_{t+h,t})] = c^k L(y_{t+h} - \hat{y}_{t+h,t}). \]

For this class of loss functions, Granger (1999) and Patton and Timmermann (2007) showed that the optimal predictor is

\[ \hat{y}_{t+h,t} = \mu_{t+h,t} + \omega_{t+h,t} \sigma_{t+h,t}, \]

where \( \sigma_{t+h,t}^2 \) is the variance of \( y_{t+h} \), conditional on information in period \( t \) and \( \omega_{t+h,t} \) is a function of the parameters in the loss function and the third and higher order conditional moments of \( y_{t+h} \). If all of the dynamics of \( y_{t+h} \) are captured through a time-varying conditional mean and conditional variance, then \( \omega_{t+h,t} = \omega(\alpha, \theta) \) is constant and is a function of only the parameters \( \alpha \) in the loss function and the time-invariant parameters \( \theta \) in the conditional distribution of \( y_{t+h} \). The AMZ regression (8) becomes

\[ y_{t+h} = a_0 + a_t \hat{y}_{t+h,t} - \omega(\alpha, \theta) \sigma_{t+h,t} + \varepsilon_{t+h,t}. \]  

(11)
Assuming a parametric form for the loss function and the conditional distribution
of $\varepsilon_{t+h}$ will determine a parametric form for $\omega(\alpha, \theta)$. Adding a parametric model for
$\sigma_{t+h,t}$, (11) can be estimated by maximum likelihood and the joint hypothesis $a_0 = 0$ and
$a_1 = 1$ can be tested using a Wald test. The model (11) is a generalization of the Engle,

If the conditional distribution of $\varepsilon_{t+h}$ is a member of a location-scale family, such
as the normal distribution, then $\omega(\alpha, \theta) = \omega(\alpha, \theta) \equiv -a_2$ will depend only on the
parameters in the loss function. The AMZ regression (1.8) then reduces to

$$y_{t+h} = a_0 + a_1 \hat{\gamma}_{t+h,t} - a_2 \sigma_{t+h,t} + \varepsilon_{t+h,t}$$

(12)

This is a conventional ARCH-M model and can be estimated by maximum
likelihood using standard software and the hypotheses $a_0 = 0$ and $a_1 = 1$ again tested with
a Wald test. Note that the rationality test based on the AMZ regression (12) only assumes
homogeneity of the loss function, not a specific functional form. If the loss function is
assumed known, it will typically depend on a single parameter so that $\alpha \in R$. If this is the
case, by the invariance principle of the MLE, the MLE of $\alpha$ can be obtained as $\hat{\alpha} =
\omega^{-1}(-\hat{a}_2)$.

We illustrate the AMZ regression (12) assuming the agent’s loss function is a
member of the asymmetric power loss family considered by Elliot, Komunjer and
Timmermann (2005). This
family of loss functions is specified as

$$L(y_{t+h} - \hat{y}_{t+h,t}) = [\alpha + (1 - 2\alpha)I(y_{t+h} - \hat{y}_{t+h,t} < 0)] |y_{t+h} - \hat{y}_{t+h,t}|^p$$

(13)
where $\alpha \in (0, 1)$ and $I(.)$ is the indicator function that takes the value one when the condition in the argument is true. The power $p$ is assumed to be known. The loss function (13) is clearly homogenous of degree $p$. This family of loss functions includes many of the loss functions commonly used in the literature. When $\alpha = 0.5$ the loss function is symmetric and becomes the standard symmetric linear and quadratic loss functions for $p=1$ and $p=2$. When $\alpha \neq 0.5$, the loss function is asymmetric and becomes linlin loss when $p=1$ and quadquad loss when $p=2$. If in addition to (13) we also assume that $y_{t+h}$ is conditionally normal, the AMZ regression takes the form (12). Assuming $p$ is known, the AMZ regression (12) can also be used to estimate the parameter $\alpha$ in the loss function (13). Higgins (2007) showed that for a normal conditional distribution and an asymmetric power loss function with given $p$ and $\alpha$, the parameter $a_2$ in (12) satisfies the equation

\[
(1 - \alpha) \int_{-\infty}^{a_2} \phi(z)dz - a_2^{p-1} \phi(z)dz - \alpha \int_{-\infty}^{a_2} z - a_2^{p-1} \phi(z)dz = 0
\]

where $\phi(z)$ is the standard normal density. Equation (14) is the first order condition for $a_2 = -\omega(\alpha, p)$ in the optimal predictor $\hat{y}_{t+h,t} = \mu_{t+h,t} + a_2 \sigma_{t+h,t}$. In general, there is no analytical solution for $a_2$ in (14). Higgins (2007) tabulated numerical solutions for $a_2$ for $p \in \{1.00, 1.25, 1.50, 1.75, 2.00\}$ and an extensive number of values for $\alpha \in (0, 1)$. Given a chosen $p$ and a point estimate of $a_2$ from (12), Table 1 in Higgins (2007) can be used to interpolate an estimate for $\alpha$.

Assuming agents have asymmetric power loss (13) and that the series is conditionally normal, the proposition of the previous section can be use to access the direction of the bias in the standard MZ regression in which $\lambda_{t+h} = a_2 \sigma_{t+h,t}$ is omitted from (12). The nature of the bias will depend on the sign of $a_2$ and the correlation.
between $\mu_{t+h,t}$ and $\sigma_{t+h,t}$. Having specified $p$, the sign of $a_2$ is determined by $\alpha$ from (14).

To consider a concrete example, suppose we are testing the rationality of forecasts of rates of return on an asset produced by a financial analyst. Also suppose the analyst faces loss so that $p = 1$ in (13) and from (14) it is seen that $a_2 = \Phi^{-1}(\alpha)$, where $\Phi^{-1}(\cdot)$ is the standard normal CDF. For many assets, empirical evidence suggest that both $\mu_{t+h,t}$ and $\sigma_{t+h,t}$ are time-varying. Furthermore, finance theory asserts there is a positive relationship between expected return and risk. This implies $\text{cov}(\mu_{t+h,t}, \sigma_{t+h,t}) > 0$. If the analyst views under predicting the return as more costly, then $\alpha > .5$ and $a_2 > 0$. The coefficient $\hat{a}_1$ will have a downward asymptotic bias. If the analyst views overpredicting the return as more costly, so that $\alpha < .5$ and $a_2 < 0$, the asymptotic bias in $\hat{a}_1$ is indeterminate.

Other tests of forecast rationality under asymmetric loss have been proposed. Elliot, Komunjer and Timmermann (2005) presented a GMM J-test derived from the first order condition for the optimal choice of $\hat{y}_{t+h,t}$ in the asymmetric power loss family (13). The statistic is based on the orthogonality condition between the gradient of (13) and variables $z_t$ assumed to be used in the construction of the forecast. The orthogonality condition also provides an estimator of the asymmetry parameter $\alpha$. The test assumes suitable instruments $z_t$ are known and that the $z_t$'s enter linearly in the forecast function. The test is limited to the asymmetric power loss family and assumes that $p = 1$ or $p = 2$. The test advocated in this paper only requires the specification of a parametric model for the conditional variance. If the conditional variance is assumed to be ARCH, for example, no knowledge of the $z_t$ variables is required. The appropriateness of the parametric model for the conditional variance can be easily tested by testing for higher
order dependence in the standardized residuals \( \frac{\hat{\epsilon}_{t+h,t}}{\hat{\sigma}_{t+h,t}} \) from (12). Also, the AMZ regression test based on (12) only requires that the loss function is homogenous. It does not assume the loss is a member of the asymmetric loss family (13). The loss function (13) is only assumed in order to estimate \( \alpha \).

More recently, Patton and Timmermann (2007) advocated a very general test for forecast rationality which only assumed homogenous loss and mean-variance dynamics based on the representation (11). They observed that the standardized forecast errors
\[
d_{t+h,t} = \frac{y_{t+h} - y_{t+h,t}}{\sigma_{t+h,t}} = -\omega(\alpha, \theta) + u_{t+h},
\]
where
\[
u_{t+h} = \frac{\epsilon_{t+h}}{\sigma_{t+h,t}}\]
are serially uncorrelated and orthogonal to any variables \( z_t \) in the agents' information set. Given an estimator \( \hat{\sigma}_{t+h,t} \) of the conditional standard deviation, rationality can be tested by testing for serial correlation in \( \hat{d}_{t+h,t} = \frac{y_{t+h} - y_{t+h,t}}{\hat{\sigma}_{t+h,t}} \) and by testing for significance in the regression of \( \hat{d}_{t+h,t} \) on \( z_t \). This test presumes that the estimator \( \hat{\sigma}_{t+h,t} \) is available. Patton and Timmermann (2007) suggested that \( \hat{\sigma}_{t+h,t} \) could be a nonparametric estimator of the standard deviation of \( y_{t+h} \) conditional on \( z_t \). For such an estimator to be consistent, it would be necessary to correctly specify all of the variables in information set used by the forecasters. This is very difficult to do. As emphasized above, The AMZ test does not necessarily require specifying any \( z_t \) variables. Patton and Timmermann's test is a two-step procedure which first requires estimating the conditional variance. The AMZ test is presumably more powerful than their test if the parametric model is correctly specified because it is based on maximum likelihood estimation. Patton and Timmermann (2007) clearly stated that their test procedure cannot be extended to nonhomogenous loss.
functions. In contrast the AMZ test can be applied to nonhomogenous loss functions as we show in the next section.

1.4 Linex Loss Function

The AMZ rationality test can also be applied to test if agents optimally forecast under linex loss. Varian (1974) introduced the linex loss function,

$$L(y_{t+h} - \hat{y}_{t+h,t}) = \exp\{\alpha(y_{t+h} - \hat{y}_{t+h,t})\} - \alpha(y_{t+h} - \hat{y}_{t+h,t}) - 1$$

when $$\alpha > 0$$, the loss is approximately exponential for positive forecast errors and approximately linear for negative forecast errors, whereas, if $$\alpha < 0$$, the loss is approximately linear for positive errors and approximately exponential for negative errors. Zellner (1986) showed that the optimal predictor is

$$\hat{y}_{t+h,t} = \mu_{t+h,t} + \frac{1}{\alpha} \log (E_t[\exp\{\alpha(y_{t+h} - \hat{y}_{t+h,t})]\])$$

(15)

In (15), $$E_t[\exp\{\alpha(y_{t+h} - \hat{y}_{t+h,t})]\}]$$ is the centered conditional moment generating function (MGF) of $$y_{t+h}$$, and hence, the precise expression for the optimal predictor will depend on the conditional distribution of $$y_{t+h}$$. Notice that the linex loss function is not homogenous, and therefore, the tests of Patton and Timmermann (2007) described in the previous section cannot be used to test for forecast rationality. In contrast, the AMZ rationality test has a particularly simple form.

Assuming normality, the optimal predictor under linex loss is $$\hat{y}_{t+h,t} = \mu_{t+h,t} + \alpha \sigma_{t+h,t}^2 / 2$$ and the AMZ regression becomes
\[ y_{t+h} = a_0 + a_1 \hat{y}_{t+h,t} + a_2 \sigma_{t+h,t}^2 + \varepsilon_{t+h} \]  

(16)

where \( a_2 = -\alpha/2 \). Like (12), (16) is a standard ARCH-M model. We can estimate (16) by maximum likelihood and test \( a_0 = 0 \) and \( a_1 = 1 \). The MLE of the parameter in the loss function is immediately given by \( \hat{\alpha} = 2\hat{\alpha}_2 \). The forecast rationality test based on the regression (16) is essentially the one originally proposed by Batchelor and Peel (1998).

Notice that in spite of the differences between linex and the asymmetric power loss function, the AMZ regression (16) differs from (12) only in that (16) adds the conditional variance whereas (12) adds the conditional standard deviation. As we did in the previous section, we can use the proposition to analyze the consequences of omitting the augmenting variable from the regression (16). If \( \sigma_{t+h,t}^2 \) is constant, the slope coefficient is consistent but the estimator of the intercept is asymptotically biased, the size and the direction of the bias depends on the magnitude and sign of \( \alpha \). When \( \sigma_{t+h,t}^2 \) is time varying, \( \text{Plim}(\hat{\alpha}_1) < 1 \) if \( \text{cov}(\mu_{t+h}, \sigma_{t+h,t}^2) = 0 \). If \( \text{cov}(\mu_{t+h}, \sigma_{t+h,t}^2) > 0 \) and \( \alpha > 0 \), then \( \text{Plim}(\hat{\alpha}_1) < 1 \). But if \( \alpha < 0 \), then the sign of the \( \text{Plim}(\hat{\alpha}_1) - 1 \) will be indeterminate. Similarly if \( \text{cov}(\mu_{t+h}, \sigma_{t+h,t}^2) < 0 \) and \( \alpha < 0 \), then \( \text{Plim}(\hat{\alpha}_1) < 1 \). But if \( \alpha > 0 \), the sign of \( \text{Plim}(\hat{\alpha}_1) - 1 \) will be indeterminate.

Depending on the variable being forecasted, the normality assumption which leads to the AMZ regression (16) may not be appropriate. For example, if the forecasts are for rates of returns on financial assets, it is well known that such returns have tails heavier than the normal distribution. The t-distribution is often used to model the distribution of asset returns. The t-distribution, however, does not have a finite MGF. The optimal predictor (15) is not defined, and therefore, there is no meaningful AMZ
rationality test. But this reflects the fact that when the predictive density lacks higher
order moments, linex loss is probably not an appropriate loss function and agents don’t
employ it. As an alternative heavy-tailed distribution, it may be reasonable to assume that
the innovations have a conditional Laplace distribution with conditional density

$$f(\varepsilon_{t+h}) = \frac{1}{\sqrt{2\sigma_{t+h,t}}} \exp \left( \frac{-|\varepsilon_{t+h,t}|}{\sigma_{t+h,t}} \right)$$

The optimal predictor is

$$y_{t+h} = \beta_{t+h,t} - \alpha^{-1} \log \left( 1 - \frac{\alpha^2}{2} \sigma_{t+h,t}^2 \right)$$

and the AMZ regression becomes

$$y_{t+h} = a_0 + a_t y_{t+h,t} + \alpha^{-1} \log \left( 1 - \frac{\alpha^2}{2} \sigma_{t+h,t}^2 \right) + \varepsilon_{t+h}$$

(17)

Although not a standard ARCH-M model, (17) can be estimated by maximum
likelihood. Again the MLE’s can be used to test $a_0 = 0$ and $a_t = 1$. In other contexts
different distributional assumptions may be appropriate. For example, if the variable
being forecasted is non-negative, it may be plausible to assume that the predictive density
is Gamma or Weibull. Both of these distributions have closed form parametric
expressions for their MGF’s and it is straightforward to write down the AMZ regression
for testing forecast rationality.

1.5 Asymmetric Loss and U.S. Inflation Forecasts

To illustrate the tests described in the preceding sections, we examine whether the
apparent bias in inflation forecasts may be due to forecasters having an asymmetric loss
function. As shown above, the AMZ regression can be used to test for bias and efficiency in the presence of asymmetric loss when second and higher order conditional moments are time-varying and correlated with the conditional mean of the series being forecasted. Freidman (1977) conjectured that high inflation is associated with high inflation uncertainty, suggesting that the conditional mean and condition variance of inflation are positively correlated. Beginning with Engle (1982) and Engle (1983), a long line of research has established that inflation does have a time-varying conditional variance.

We use the Livingston survey of forecasters because it is the longest running survey which contains data on inflation forecasts. The semi-annual survey started in 1946 and is published in June and December of each year. Among other things, respondents are asked to predict the level of the CPI six months ahead. We start our sample in 1954 to avoid the period of U.S. price controls during the Korean conflict. We end our sample in December, 2003 because after that survey respondents were asked to forecast the seasonally adjusted CPI rather then the unadjusted CPI. We use the median consensus forecast of the CPI. The six month horizon inflation rate and inflation forecasts are calculated following Carlson's (1977) method from the realized CPI and the forecasted CPI series.

In the first column of Table 1, we present the MZ regression for the inflation forecasts. Both individual t-tests and the joint F-test reject forecast unbiasedness and efficiency at the 5 percent significance level. For these tests, the null hypothesis implicitly maintains that the forecasts are rational under quadratic loss. To test rationality in the presence of possibly asymmetric power loss or linex loss using the AMZ regressions (12) and (16), we must demonstrate that the inflation innovation has time-
varying conditional variance. We assume time variation in the condition variance can be represented by a G/ARCH process. For the conditional variance specification

\[ \sigma^2_{t+h,t} = \alpha_0 + \alpha_1 \epsilon^2_t + \cdots + \alpha_p \epsilon^2_{t-p+1}. \]

We want to test the null hypothesis \( H_0: \alpha_1 = \cdots = \alpha_p = 0 \) without also imposing quadratic loss under the null. The LM test for ARCH in the ARCH-m model is nonstandard because the parameter in \( \alpha_2 \) in (12) and (16) is not identified under the null. Bera and Ra (1995) suggested the Sup LM(\( \alpha_2 \)) statistic be used and a conservative critical value computed by the method of Davies (1987). They showed the test has better power than the conventional LM test that imposes \( \alpha_0 = 0 \). In Table 2, we report the conventional LM test and the test for ARCH in the AMZ regression (12). The reported supremum of the LM statistic is calculated for \( \alpha_2 \) in the interval (-3,3). Experimentation shows that the results are not sensitive to the choice of the interval. As seen in the table, the conventional LM statistics are not significant at any order \( p \). In contrast, the Sup LM statistics are highly significant at all orders and suggest that inflation has time-varying second moments.

In Table 1, we also present the AMZ regressions (12) and (16) which assume the loss functions are asymmetric power and linex. We model the time-varying conditional variance with a GARCH(1,1) process \( \sigma^2_{t+1,t} = \alpha_0 + \alpha_2 \epsilon^2_t + \beta_1 \sigma^2_{t,t-1}. \) Under linex loss, the parametric form of the augmenting term in (16) depends strongly on the assumption the innovation \( \epsilon_t \) has a conditional normal distribution. Under asymmetric power loss the parametric form of the augmenting term in (12) depends only on the assumption the innovation distribution is a member of a location-scale family. A normality assumption is required to use (12) to estimate the asymmetry parameter \( \alpha \). The Jarque-Bera normality
statistics computed with the standardized residuals $\hat{e}_t / \hat{\sigma}_t$ are insignificant for all of the models in Table 1. The statistics are insignificant and suggest the normality assumption is not inappropriate. Q-statistics at order 4 and 8 computed with the squared standardized residuals are also insignificant for all of the models, indicating the GARCH(1,1) model adequately captures the dependence in the second moments of inflation. For the asymmetric power loss and linex loss AMZ regressions report in Table 1, the individual t-tests and the joint Wald tests for $H_0: a_0 = 0, a_1 = 1$ are not significant. Although rationality cannot be rejected, the augmenting terms in both the asymmetric power and linex augmenting regressions are insignificant. This may be due to multi-collinearity between the forecast $\hat{y}_{t+1,t}$ and the augmenting terms in (12) and (16). Recall it is the correlation between the forecast and the augmenting term which invalidates the MZ regression under asymmetric loss. To examine this possibility, we re-estimate (12) and (16) imposing $a_0 = 0$ and $a_1 = 1$. The results are in the last two columns of Table 1. When rationality is imposed, the augmenting terms become significant.

These final two regressions in Table 1 can be used to obtain fully efficient estimators of the asymmetry parameter $\alpha$. Under asymmetric power loss, assuming $p$ in (13) is 1, 1.5 or 2, we use Table 1 in Higgins (2007) to determine the implied values for $\hat{\alpha}$ are respectively .4, .35, and .325. Assuming the loss is linex, we immediately have $\hat{\alpha} = .16$. The estimated loss functions for asymmetric power are shown in Figure 1 and the estimated loss function for linex is shown in Figure 2. The loss functions are plotted over the range of the forecast errors. Both specifications of the loss function indicate forecasters attach larger costs to over predicting the inflation rate than to under predicting the inflation rate by the same magnitude.
1.6 Conclusions

The MZ regression is a standard procedure for testing forecast unbiasedness and efficiency. If forecasts are rational, but constructed under asymmetric loss, the MZ regression may reject both unbiasedness and efficiency. We show that an appropriate test can be obtained by augmenting the MZ regression with a variable that depends on the loss function and the time-varying second and possibly higher order moments of the process being forecasted. We demonstrate that the AMZ regression is easily constructed for all of the standard loss functions that have been assumed in the literature, including for example linlin, quadquad and linex. The AMZ regression often requires assuming a parametric model for the conditional variance, but the appropriateness of the parametric model can easily be tested from the data. If the conditional variance is assumed to be an ARCH process, the AMZ regression can be estimated using standard econometric software. If the loss function depends on a single parameter, the AMZ regression often provides an estimator of that parameter. We illustrate the use of the AMZ regression by testing the rationality of inflation forecasts from the Livingston survey. The conventional MZ regression rejects both unbiasedness and efficiency. Under the assumption of homogenous loss or linex loss, rationality cannot be rejected with the AMZ regression. The results indicate that agents view over predicting inflation as more costly than under predicting inflation.
Table 1


<table>
<thead>
<tr>
<th>Coefficients</th>
<th>MZ Regression</th>
<th>Linlin</th>
<th>Linex</th>
<th>Linlin ((a_0=0, a_1=1))</th>
<th>Linex ((a_0=0, a_1=1))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{a}_0)</td>
<td>0.088**</td>
<td>0.012</td>
<td>0.085</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.625)</td>
<td>(0.324)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\hat{a}_1)</td>
<td>1.139**</td>
<td>1.093</td>
<td>1.09</td>
<td>-0.28**</td>
<td>-0.16**</td>
</tr>
<tr>
<td></td>
<td>(0.066)</td>
<td>(0.067)</td>
<td>(0.067)</td>
<td>(0.075)</td>
<td>(0.046)</td>
</tr>
<tr>
<td>(\hat{a}_2)</td>
<td>-0.103</td>
<td>-0.033</td>
<td>-0.28**</td>
<td>-0.16**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.520)</td>
<td>(0.179)</td>
<td>(0.075)</td>
<td>(0.046)</td>
<td></td>
</tr>
<tr>
<td>(\hat{a}_3)</td>
<td>0.16</td>
<td>0.17</td>
<td>0.13</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.194)</td>
<td>(0.19)</td>
<td>(0.139)</td>
<td>(0.167)</td>
<td></td>
</tr>
<tr>
<td>(\hat{a}_4)</td>
<td>0.2*</td>
<td>0.2*</td>
<td>0.22**</td>
<td>0.22**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.111)</td>
<td>(0.11)</td>
<td>(0.110)</td>
<td>(0.104)</td>
<td></td>
</tr>
<tr>
<td>(\hat{\beta})</td>
<td>0.73**</td>
<td>0.72**</td>
<td>0.72**</td>
<td>0.7**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.138)</td>
<td>(0.139)</td>
<td>(0.118)</td>
<td>(0.124)</td>
<td></td>
</tr>
</tbody>
</table>

F-statistics are reported for the Wald test and the associated probabilities are in the parenthesis. ** denotes significance at 5% and * denotes significance at 10%.

Table 2

ARCH LM Test and ARCH-Sup LM Test

<table>
<thead>
<tr>
<th>Order</th>
<th>ARCH LM P-value</th>
<th>ARCH Sup LM P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>0.45</td>
<td>1.45</td>
</tr>
<tr>
<td></td>
<td>(0.5)</td>
<td>(0.23)</td>
</tr>
<tr>
<td></td>
<td>17.24**</td>
<td>20.47**</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0001)</td>
</tr>
</tbody>
</table>

** Denotes significance at 5%.
Figure 1

Asymmetric Power Loss Function for U.S. Inflation Forecasts
Figure 2

Linex Loss Function for U.S. Inflation Forecasts
CHAPTER II

STATE DEPENDENT ASYMMETRIC LOSS

2.1 Introduction

Forecasts for the macro economy are important because they guide the decisions of individuals, businesses and policy-makers. The purpose of a forecasting exercise is to predict the future accurately. However, in spite of significant improvement in the linear/non-linear modeling of time series, forecasters tend to produce biased forecasts. Among others, Carlson (1977), Urich and Wachtel (1984), Caskey (1985), Zarnowitz (1985), Frankel and Froot (1987), Croushore (1993, 1997), Jeong and Maddala (1996) and Souleles (2001) found that forecast reported by professional forecasters are biased. If we assume forecasters are rational, the above research implies forecasters intentionally underpredict or overpredict the series being forecasted. This decision depends on the cost associated with underprediction and overprediction. If the cost of underprediction is higher than overprediction then agents will tend to overpredict and vice versa.

In essence, whenever agents have asymmetric loss bias is introduced into their forecasts. Bias due to asymmetric loss in agents' forecasts has been documented in recent studies by Batchelor and Peel (1998), Ruge (2004), Nobay and Peel (2003), Elliot, Komunjer and Timmermann (2005, 2008), Capistran (2006). The main idea of these papers is that when we apply standard tests of rationality, rationality is rejected because
of bias. But when we allow for asymmetric loss, the forecasts are no longer found to be irrational.

Recently authors have suggested that agents' loss function not only depends on the forecast error but also on the state of the economy. Fildes and Stekler (2002) documented that there is an element of judgment in agents' forecasts implying agents do not strictly follow formal forecasting models. Using individual level data, Batchelor (2007) documented the presence of systematic bias in the real GDP forecasts in the G7 economies in the years 1990-2005. He conjectured why individual agents bias their forecasts. He suggested factors like optimism, pessimism, conservatism and herding are responsible for the bias. Although several studies have suggested that agents have state dependent time-varying asymmetric loss, Patton and Timmermann (2007) is the only study which has presented a formal test. They showed using a set of quantile tests and a non-parametric test that agents' loss function not only depend on the forecast error but also on the level of the variable being forecasted. They argued that overprediction of GDP growth rate is costlier than underprediction especially during episodes of low growth.

In this paper we introduce time-varying asymmetric loss in a parametric framework. First we show that under asymmetric loss the time-varying bias can be explained either by the time-varying asymmetry parameter or by the time-varying higher order conditional moments or by both. However, in the absence of time-varying higher order moments, the time-varying bias can only be explained by the time-varying asymmetry parameter. We model the time-varying asymmetry parameter as a linear combination of variables that represent the state of the economy. We consider factors
such as duration of business cycle, uncertainty of forecasts and type of government, which may cause agents to intentionally bias their forecasts. We show that traditional efficiency regression can be used to test for time-varying asymmetric loss under linex and asymmetric power loss. In the empirical analysis, we use one quarter ahead GDP growth rate forecasts from the SPF to show that state of the economy plays an important role in agents' forecasting behavior. The paper proceeds as follows. In section 2.2 we present the theoretical results. In section 2.3 and 2.4 we describe the data and variables. In section 2.5 we analyze the empirical results. Finally, section 2.6 concludes the paper.

2.2 Theoretical Results

For any observed time series $y_t$, the series in period $t+1$ can be decomposed as

$$y_{t+1} = \mu_{t+1,t} + \varepsilon_{t+1},$$

where $\mu_{t+1,t}$ is the mean in period $t+1$ conditional on the information set available in time period $t$ and $\varepsilon_{t+1}$ is an innovation which has mean zero and is uncorrelated with elements in the information set. Let $\tilde{y}_{t+1,t}$ be the one period ahead forecast of $y_t$ made using all information available in time period $t$. For simplicity we assume a one period horizon but our results can be generalized to any horizon. When forecasters have quadratic loss the optimal predictor is $\tilde{y}_{t+1,t} = \mu_{t+1,t}$. So the forecast error $y_{t+1} - \tilde{y}_{t+1,t}$ in time period $t+1$ is nothing but the innovation. However it is possible that agents give unequal weight to positive and negative forecast errors. Granger (1969) and Christoffersen and Diebold (1997) showed that if agents have asymmetry in their loss function then the optimal predictor is the conditional mean plus a systematic bias.
\[ \hat{y}_{t+1,t} = \mu_{t+1,t} + \lambda_{t+1,t}, \]

where the bias \( \lambda_{t+1,t} \) depends on the loss function and conditional moments of order two or higher.

The systematic bias under asymmetric loss has different interpretation. Under quadratic loss, the forecast errors are uncorrelated with information agents use to forecast. However, under asymmetric loss it is not true. When agents have asymmetric loss, rational forecast errors may be correlated with variables in agents' information set. Traditionally the efficiency test is done by the following regression,

\[ y_{t+1} - \hat{y}_{t+1,t} = \epsilon_{t+1} = \lambda_{t+1,t} + \epsilon_{t+1} \]

(18)

where \( \lambda_{t+1,t} \) contains variables in agents' information set. Under quadratic loss \( \lambda_{t+1,t} \) is zero because forecast errors are uncorrelated with agents' information set so the forecast error is nothing but the innovation. However, when agents have asymmetric loss \( \lambda_{t+1,t} \) is not zero but a function of the loss function parameter and conditional higher order moments of the series being forecasted. Thus, \( \lambda_{t+1,t} \) contains variables from agents' information set. Hence under asymmetric loss we can test the efficiency of agents' forecasts using the traditional efficiency regression by adding variables from agents' information set.

Rationality and asymmetric loss have been discussed by Batchelor and Peel (1998), Elliot, Komunjer and Timmermann (2005, 2008), Capistran (2006) and Timmermann and Patton (2007). The bias \( \lambda_{t+1,t} \) is a function of the parameters in the loss function and the conditional moments of the series being forecasted. All the above studies assumed the parameters in the loss function to be time-invariant and the conditional
moment parameter to be time-varying. However recent studies by Krane (2003), Batchelor (2007) and Timmermann and Patton (2007) suggested that the loss function not only depends on the forecast error but also on the state of the economy. In that case, the bias in agents' forecasts can be explained either by the time-varying asymmetry parameter or by time-varying higher order moments or by both. There are many variables in the survey forecasts, such as GDP growth rate, that do not display time-varying higher order moment dynamics. In that case the time-varying bias in the forecasts can only be explained by the time-varying asymmetry parameter. We introduce a time-varying asymmetry parameter into two commonly used asymmetric loss functions.

2.2.1 Linex Loss

In this section we introduce a time-varying asymmetry parameter into the linex loss function introduced by Varian (1974). If the asymmetry parameter in the loss function is time-varying then the linex loss function can be written as

\[ L(y_{t+1} - \gamma_{t+1,t}) = \exp(\alpha_t(y_{t+1} - \gamma_{t+1,t})) - \alpha_t(y_{t+1} - \gamma_{t+1,t}) - 1 \]

where \( \alpha_t \) can be any real number. When \( \alpha_t > 0 \), and agents underpredict, the loss function is exponential. It is, however, linear when agents overpredict. On the other hand, if \( \alpha_t < 0 \), the loss is approximately linear for underprediction and approximately exponential for overprediction. We put a \( t \) subscript on \( \alpha \) because we assume \( \alpha \) as time-varying. We model \( \alpha_t \) as a linear combination of a set of state variables \( z_t \) so that \( \alpha_t = z_t' \gamma \). The purpose of including these variables is to capture the state of the economy and quantify the parameter of asymmetry through \( \gamma \). Since the costs of overprediction and
underprediction are different, the optimal prediction under time-varying linex loss will reflect this bias.

Under conditional normality, the optimal predictor under time-varying linex loss is,

$$\hat{y}_{t+1,t} = \mu_{t+1,t} + \sigma_{t+1,t}^2 \alpha_t / 2.$$  
So the forecast error becomes

$$e_{t+1} = \frac{\sigma_{t+1,t}^2 (z_t' \gamma)}{2} + \epsilon_{t+1}$$

$$= -\frac{\sigma_{t+1,t}^2 \alpha_t}{2} + \epsilon_{t+1}.$$  

As we can see from the above equation, under linex loss the time-varying bias in the forecast can be due to the time-varying asymmetry parameter $\alpha_t$ or due to the conditional variance $\sigma_{t+1,t}^2$ of the series being forecasted or due to both. In the absence of time-varying higher order moments, the bias can only be explained by the time-varying asymmetry parameter. As we mentioned in the previous section, empirically it is found that one quarter ahead GDP growth rate forecasts do not have higher order moment dynamics. In that case, time-varying variance cannot explain the bias. However, the bias could only be explained by the time-varying asymmetry parameter $\alpha_t = z_t' \gamma$. Under constant variance the forecast error becomes

$$e_{t+1} = \frac{\sigma_{t+1,t}^2 (z_t' \gamma)}{2} + \epsilon_{t+1}$$

$$= z_t' \gamma^* + \epsilon_{t+1}$$  

(19)

Under conditional normality, the maximum likelihood estimator of $\gamma^*$ can be computed by the least squares regression of $e_{t+1}$ on $z_t$. We can estimate the time-varying bias with the negative of the predictions from this regression $\hat{\lambda}_{t+1,t} = -z_t' \gamma^*$. We can
estimate the time-varying asymmetry parameter by $\alpha_t = 2\lambda_{t+1,t}/\sigma^2$, where $\hat{\sigma}^2_\varepsilon$ is the LS estimator of the variance of $\varepsilon_{t+1}$ in (19). From the above analysis we see that traditional efficiency regression can be interpreted as a regression to estimate the time-varying asymmetric loss parameter.

2.2.2 Asymmetric Power Loss

In this section we introduce a time-varying asymmetry parameter into the asymmetric power loss function used by Diebold and Christoffersen (1996) and Elliot, Komunjer and Timmermann (2005). When $\alpha$ is time-varying, the loss function $L$ can be written as,

$$L(y_{t+1,t}, \alpha_t) = [\alpha_t + (1 - 2\alpha_t)I(y_{t+1} - \hat{y}_{t+1,t} < 0)]|y_{t+1} - \hat{y}_{t+1,t}|^p$$

(20)

where $0 < \alpha_t < 1$ is the asymmetry parameter and $I(.)$ is the indicator function that takes the value 1 when the condition in the argument is true. Normally, the power $p$ is assumed to be known. The loss function (20) is clearly homogenous of degree $p$. This family of loss functions includes many of the loss functions commonly used in the literature. When $\alpha_t \neq 0.5$ and $p = 1$, the loss function is asymmetric linlin loss. When $\alpha_t \neq 0.5$ and $p = 2$ the loss function is asymmetric quadquad. When $\alpha_t = 0.5$ and $p = 1$, the loss function is the standard symmetric linear loss function and for $p = 2$ the loss function is the standard symmetric quadratic loss function. Because the asymmetry parameter is required to be in the interval $(0, 1)$, we assume $\alpha_t = \Phi(z_t'q')$, where $\Phi(.)$ is the standard normal
cumulative distribution function and $z_t$ are again state variables. For general asymmetric power loss function with a time-varying asymmetry parameter, the optimal predictor is

$$\hat{y}_{t+1,t} = \mu_{t+1,t} + \Phi^{-1}(\alpha_t)\sigma$$

Under a conditional normal distribution, for a given $p$ and $\alpha_t$, the parameter $\omega_t = \omega(\alpha_t, p)$ satisfies the equation

$$(1 - \alpha_t) \int_{-\omega_t}^{\omega_t} |z - \omega_t|^p \phi(z) dz - \alpha_t \int_{-\omega_t}^{\omega_t} |z - \omega_t|^{p-1} \phi(z) dz = 0$$

(21)

where $\phi(z)$ is the standard normal density [see Diebold and Christoffersen (1996)]. In the absence of time-varying standard deviation the time-varying bias can only be explained by the asymmetry parameter $\alpha_t$.

Like linex loss, the estimation of the time-varying asymmetry parameter and bias under general asymmetric power loss can be done by the method of maximum likelihood. Under constant standard deviation and time-varying asymmetry parameter, the forecast error becomes

$$e_{t+1} = -\omega(\alpha_t, p)\sigma + \epsilon_{t+1}.$$ 

The normal log-likelihood function for the entire sample can be written as

$$l_T(\gamma, \sigma) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \frac{1}{2} \sum_{t=1}^{T} (e_{t+1} + \omega(\Phi(z'_t \gamma), p)\sigma)^2$$

where $\omega(\Phi(z'_t \gamma), p)$ can be solved from (21). Given the MLE's of $\gamma$ and $\sigma$, we can estimate the time-varying asymmetry parameter by $\hat{\alpha}_t = \Phi(z'_t \hat{\gamma})$. The time-varying bias $\hat{\lambda}_{t+1,t} = \Phi(\hat{\alpha}_t, p)\hat{\sigma}$ can be computed in each time period from (21).

Now let us consider a specific loss function in this general class. When $\alpha_t \neq 0.5$ and $p = 1$ the loss function becomes asymmetric linlin. Assuming conditional normality,
the optimal predictor under linlin loss in the absence of time-varying standard deviation is \( \hat{y}_{t+1,t} = \Phi^{-1}(\alpha_t)\sigma \). Solving for \( \hat{y}_{t+1,t} \) yields
\[
\hat{y}_{t+1,t} = \mu_{t+1,t} + \Phi^{-1}(\alpha_t)\sigma
\]
Because we assume \( \hat{a}_t = \Phi(z'_t\hat{\gamma}) \) to insure that \( \hat{a}_t \in (0, 1) \), the forecast error becomes
\[
e_{t+1} = -\sigma z'_t \gamma + \epsilon_{t+1} = z'_t \hat{\gamma} + \epsilon_{t+1}
\]
(22)
In this case maximum likelihood estimation is equivalent to linear regression. Thus, similar to the linex case under conditional normality, we can estimate the parameters in the loss function by linear regression. We can estimate the time-varying bias with the negative of the predictions from this regression \( \hat{\alpha}_{t+1,t} = -z'_t \hat{\gamma} \). We can estimate the time-varying asymmetry parameter with \( \hat{a}_t = \Phi(\hat{\alpha}_{t+1,t})/\hat{\sigma}^2 \), where \( \hat{\sigma} \) is the LS estimator of the standard deviation of \( \epsilon_{t+1} \) in (22). From the above exercise we can see that traditional efficiency regression can be interpreted as testing for either asymmetric linex or linlin loss. The regressions are observationally equivalent. The only difference lies in the parameterizations \( \gamma^* = -\frac{\sigma^2\gamma}{2} \) and \( \bar{\gamma} = -\sigma \gamma \).

2.3 Survey of Professional Forecasters

We obtain the one quarter ahead GDP growth rate forecasts from the Survey of Professional Forecasters (SPF). Starting with the first quarter of 1968, the Federal Reserve Bank of Philadelphia has been conducting surveys of forecasts of several important economic variables produced by private sector economists. These forecasted
variables include output, inflation, and interest rates. It is important that anonymity of the forecasters be maintained here so that the economists cannot be held liable for inaccurate forecasts and neither can they claim credit for accurate forecasts. Croushore (1993) provides a very detailed description. This survey was initially conducted by the American Statistical Association, together with the National Bureau of Economic Research, starting with the fourth quarter of 1968. In the early days of the survey there were 50 participants in each quarter. However, the number of respondents reduced to 20 later. In the third quarter of 1981, the scope of the survey was expanded to also include forecasts of one quarter ahead GDP growth rate, the 3-month T-bill rates and the rate of inflation. The Federal Reserve Bank of Philadelphia revived the survey in 1990 after it was discontinued by the ASA and the NBER. For the purpose of this study, the sample period begins with the third quarter of 1981 and closes with the fourth quarter of 2007.

2.4 Data and Variables that Can Explain Bias

In this section we propose variables that explain why agents knowingly introduce bias into their forecasts. There can be many reasons why agents report different forecasts from what their econometric model actually produce. Batchelor (2007) suggested few subjective factors that can explain the bias in agents' forecasts. These are optimism, pessimism and conservatism. Batchelor and Dua (1990) suggested that forecasters find it beneficial to develop a reputation as optimists or pessimists. Findings by McNicholas and O'Brien (1997), Butler and Lang (1991), Francis, Hanna and Philbrick (1997), Francis
and Philbrick (1993), Lim (2001) and Dechow, Hutton and Sloan (2000) are similar in this line.

In the literature authors have suggested that agents want to account for recoveries and downturns in the economy when they forecast as these are important events. Forecasters usually do not want to miss these important events while forecasting because it is a matter of reputation for them. They may intentionally bias their forecasts depending on whether the economy is in expansion or contraction [see Loungni and Trehan (2002), Zarnowitz and Braun (1992)]. We test whether expansion/contraction in the economy can cause bias in agents' forecasts by using a expansion/contraction dummy (RECDM) in our regression. In addition to the current phase of the business cycle, duration of the current phase may also cause bias. As the length of the expansion becomes longer agents may become increasingly optimistic and bias their forecasts upward. Similarly, as the length of a contraction becomes longer, agents may become increasingly pessimistic and bias their forecast downward. So we include duration (DURATION) measured as the number of quarters since the last turning point as a variable in our regression. We also include an interaction between recession and DURATION (REC_DUR) to allow duration to have an asymmetric effect during expansion and contraction.

The dummy variables for recession and the duration of business cycle are constructed from the NBER's recession dating procedure. The NBER's business cycle dating committee maintains a chronology of the U.S business cycles. According to the NBER, recession is not a decline in real GDP for two consecutive quarters but a significant decline in economic activity lasting more than few months. According to the
NBER, the indicator of recession is not only decline in real GDP but also a decline in real income, employment, industrial productivity and wholesale-retail-sales. According to the NBER dating committee the period from a peak to a trough is a recession and the period from a trough to a peak is an expansion. Our dummy variable takes the value 1 if the economy is experiencing a recession and zero otherwise. The duration of business cycle is a count variable. Suppose the economy is going from a trough to a peak then the count variable takes the value 1 right after the trough and keeps increasing till it reaches a peak. That is the duration of expansion. We follow similar rules for calculating the duration of contraction as well.

Sometimes forecasters have preference for the incumbent government as suggested by Ulan, Dewald and Bullard (1995). There is a large literature that documents the difference in political parties and inflation/output outcome in the economy [see Alesina et al. (1993), Hibbs (1977, 1986) and Snowbergs, Wolfers and Zitewitz (2007)]. The former studies suggest that a democratic party is good for stimulating growth and the later study suggest that a Republican party is good for business so agents will be more optimistic if a Republican is in office. Hence the political party in office might explain the bias so we have a political party dummy (POLIDM) variable. The dummy variable for type of government takes the value 1 if a Republican is in office and zero otherwise.

In addition to the above mentioned variables we suggest uncertainty may affect agents' bias. Fildes and Stekler (2002) mentioned from the perspective of bias in agents' forecast that uncertainty is important. When there is uncertainty the bias can be upward or downward. It is also important to see if uncertainty has an asymmetric effect on agents' forecast depending on the state of the economy. When there is expansion and there is lack
of consensus among forecasters, agents may want to build a reputation as optimist or pessimist. So they may bias their forecast upward or downward. Similarly in the presence of recession, uncertainty may or may not have any asymmetric effect in agents forecast. So we include uncertainty (UNCERT) and an interaction between RECDM and UNCERT (REC_UNCER) as state variables in our analysis. Uncertainty is calculated as the standard deviation of individual level one quarter ahead GDP growth rate forecasts for a given time period. The GDP growth rate uncertainty is calculated from the individual level data from the SPF.

2.5 Empirical Results

We begin our empirical analysis by testing for time-varying second order moments in the forecast error of GDP growth rate. We conduct an ARCH-LM test up to lag 5 and fail to reject the null hypothesis of no ARCH. To check if linex and linlin loss have time-varying second order moments we estimate a GARCH(1,1)-M model for both the loss functions. If $\alpha$ is time-invariant then we can test for the significance of time-varying second order moments in linex loss by including the variance of the forecast error in GARCH(1,1) conditional mean specification. Similarly in case of linlin, when $\alpha$ is time-invariant we can test the significance of the time-varying second order moments by including the standard deviation of the forecast error in the GARCH(1,1) conditional mean specification. For both linex and linlin we do not find the coefficient of the conditional variance/standard deviation in the GARCH(1,1) conditional mean equation to be significant. Also in the conditional variance equation we find the coefficient of the
lagged squared error to be insignificant. The above tests confirm that GDP growth rate has no time-varying higher order moments.

We begin by estimating a general model with all variables and all possible interaction terms in it except for an interaction between RECDM and POLIDM. We cannot include this interaction in our analysis because this variable is perfectly correlated with RECDM. In our sample there are three brief periods of recession and during these periods Republicans were in office. We do not have observations to test for the impact of recession under a Democratic government. We present the results in Table 3. In the first column of Table 3 we present the general result for linex and linlin loss. We find the coefficient of the intercept and the coefficient of RECDM to be insignificant. Even though it is argued in the literature that agents bias their forecast during expansion and contraction our results suggest that agents do not bias their forecast during expansion or contraction. We also find the coefficients of DURATION and REC_DUR to be insignificant. This indicates that there is no optimism nor pessimism that accumulates as the current phase of the business cycle continues. We also find the coefficient of POLIDM to be insignificant. Agents do not appear to be optimistic or pessimistic based on which party is in office.

The coefficients on UNCERT and REC_UNCER are significant. The sign of the coefficient on UNCERT is negative and the sign of the coefficient on REC_UNCER is positive. Furthermore, the coefficient on UNCERT is less in absolute value than the coefficient on REC_UNCER. This implies that when the economy is in recession, increasing uncertainty causes forecasters to introduce a positive bias, whereas, when the economy is in expansion, increasing uncertainty causes forecasters to introduce a
negative bias. Uncertainty causes forecasters to take a conservative approach. As uncertainty increases, forecasters bias the forecast towards its historical average. Unlike the finding by McNees (1976, 1988, 1992), McNees and Reis (1983) and Zarnowitz (1992) that during recession agents overpredict, our results suggest that in the presence of recession, increasing uncertainty cause agents to overpredict.

To simplify the model, we sequentially remove the least significant variables and re-estimate the model. When we exclude all insignificant variables the variables that remain significant are UNCERT and REC_UNCER. We present our final model under linex and linlin loss in the second column of Table 3. We follow the same steps for quadquad loss and get similar results. The general model that we estimate under quadquad loss is presented in the third column of Table 3. Like linex and linlin case, under quadquad loss UNCERT and REC_UNCER are significant in the general model and in the final model under quadquad loss both these variables remain significant. The final model for quadquad loss is presented in the fourth column. This means irrespective of the functional form of the loss function we find UNCERT to be negative and significant and REC_UNCER to be positive and significant. For both linex/linlin and quadquad we present the SIC criteria. In both the cases the SIC is minimized in case of the final model. So the final model is our best model. To estimate our models we assume normality. To test for normality, we present the Jarque-Bera statistics. For all models the Jarque-Bera statistics are insignificant which means normality assumption is appropriate. Q-statistics of order 4 and 8 are computed with the residuals and squared residuals. The statistics are insignificant for all models which suggest that our models are properly specified.
In Figures 3, 4, 5, 6, 7 and 8 we plot the time-varying asymmetry parameter and the time-varying bias for linex, linlin and quadquad loss. Notice, the pattern of variation of the time-varying bias is almost same for all three loss functions. For linex and linlin the time-varying bias is a multiple of \(a_t\). In linex the time-varying bias is \(z't\gamma\sigma^2 = 2\) and for linlin it is \(z't\gamma\sigma\). We find the time-varying bias does not depend upon the specification of the loss function. This implies our estimation is robust with respect to any of these three loss functions. As we can see the plots of the time-varying bias, there are three peaks that correspond to three brief periods of recession over the last two decades. During early 80s recession the bias was about 5% whereas during early 90s and 2000 recessions the bias was a little below 2%. These are the time periods when agents overpredicted GDP growth rate due to uncertainty. In all other time periods the bias was about -1%.

2.6 Conclusions

Although there have been important advances in time series modeling, agents still produce biased forecasts. Agents do not rely on econometric modeling completely but use their judgment to some extent and this element of judgment reflects the state of the economy. In this paper we explain why agents knowingly introduce bias into their forecasts. We use one quarter ahead GDP growth rate forecasts from the SPF to answer this question. Many authors suggested that agents' loss function depend on the forecast error and the state of the economy. Under asymmetric loss the time-varying bias can be explained by either a time-varying asymmetry parameter or by time-varying higher order moments or by both. However, in the absence of time-varying higher order moments the
time-varying bias can only be due to the time-varying asymmetry parameter. Since GDP growth rate has no time-varying higher order moments so the time-varying bias can only be explained by the time-varying asymmetry parameter. To quantify the time-varying asymmetry parameter and the time-varying bias in the time-varying asymmetric loss we introduce time-varying loss explicitly into a parametric model. We model the time-varying asymmetry parameter as a linear combination of different state variables. For the empirical analysis we use linex and asymmetric power loss. Using these loss functions we estimate the time-varying bias and the asymmetry parameter by maximum likelihood estimation. The factors that are responsible for agents' biased forecasts are uncertainty of forecasts and uncertainty in the presence of recession.
Table 3
Determinants of Time-varying Asymmetry Parameter in the One Quarter Ahead GDP Growth Rate, 1981:3-2007:4

<table>
<thead>
<tr>
<th>Variables</th>
<th>Linex/Linlin Loss</th>
<th>Quadquad Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>General</td>
<td>Final</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.49</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(0.78)</td>
<td>(0.36)</td>
</tr>
<tr>
<td>RECDM</td>
<td>1.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.50)</td>
<td></td>
</tr>
<tr>
<td>POLIDM</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.46)</td>
<td></td>
</tr>
<tr>
<td>UNCERT</td>
<td>-0.63**</td>
<td>-0.57**</td>
</tr>
<tr>
<td></td>
<td>(0.33)</td>
<td>(0.29)</td>
</tr>
<tr>
<td>DURATION</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td>REC_UNCER</td>
<td>3.43**</td>
<td>1.70**</td>
</tr>
<tr>
<td></td>
<td>(1.79)</td>
<td>(0.40)</td>
</tr>
<tr>
<td>REC_DUR</td>
<td>-1.61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.30)</td>
<td></td>
</tr>
<tr>
<td>Adj Rsq</td>
<td>0.21</td>
<td>0.20</td>
</tr>
<tr>
<td>SIC</td>
<td>4.43</td>
<td>4.30</td>
</tr>
<tr>
<td>Q(4)</td>
<td>0.161</td>
<td>0.123</td>
</tr>
<tr>
<td>Q(8)</td>
<td>0.443</td>
<td>0.264</td>
</tr>
<tr>
<td>Q²(4)</td>
<td>0.899</td>
<td>0.852</td>
</tr>
<tr>
<td>Q²(8)</td>
<td>0.950</td>
<td>0.676</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>0.499</td>
<td>0.854</td>
</tr>
</tbody>
</table>

Standard errors are in parenthesis. ** denotes significant at 5% level and * denotes significant at 10% level. For residuals, squared residuals and normality test p-values are given.
Figure 3
Time-varying Asymmetry Parameter for Linex Loss

Figure 4
Time-varying Bias for Linex Loss
Figure 5
Time-varying Asymmetry Parameter for Linlin Loss

Figure 6
Time-varying Bias for Linlin Loss
Figure 7

Time-varying Asymmetry Parameter for Quadquad Loss

Figure 8

Time-varying Bias for Quadquad Loss
CHAPTER III

DO AGENTS LEARN BY LEAST SQUARES? THE EVIDENCE PROVIDED BY CHANGES IN MONETARY POLICY

3.1 Introduction

The rational expectations hypothesis (REH) assumes that agents acquire and process information rationally. When agents have complete knowledge of the structure of the model, forecast errors have mean zero, constant variance and no serial correlation. Empirical tests based on surveys of expectations often reject the REH. Among others, Carlson (1977), Urich and Wachtel (1984), Caskey (1985), Zarnowitz (1985), Frankel and Froot (1987), Croushore (1993, 1997), Jeong and Maddala (1996) and Souleles (2001) all rejected the REH. The REH is rejected in the above studies because survey forecast errors are serially correlated. This implies that the forecasts are not efficient in that they omit relevant information available to forecasters.

Due to the accumulated evidence rejecting the REH, researchers have proposed that agents may not have access to all information about the structure of the economy but that they learn about it over time. This assumption is referred to as “adaptive learning”. Agents behave like econometricians when making forecasts and adjust their forecast rule as new data becomes available over time. When agents estimate parameters by least squares using the available data, it is known as “least squares learning”. Early studies such as Lucas (1978), Grandmont (1985) and Woodford (1991) demonstrated the stability of steady states under least squares learning. Convergence of least squares learning to
rational expectation (RE) equilibrium in linear models was proved by Bray (1982), Bray and Savin (1986), Marcet and Sargent (1988, 1989a, 1989b), Evans and Honkapohja (1994a) and in nonlinear models the convergence was shown by Fuchs (1979), Grandmont (1985) Grandmont and Laroque (1991), Margaritis (1987) and Evans and Honkapohja (1994b). The key idea of these papers that RE is not necessary to achieve market clearing conditions. It can be achieved by implementing least squares learning rules. An excellent survey of the adaptive learning literature is provided by Evans and Honkapohja (2001).

More recent research focuses on the design and stability of monetary policy in the presence of least squares learning by private agents. Evans and Honkapohja (2001, 2003a, 2003b, 2006); MacGough, Rudebusch, and Williams (2005); and Preston (2006) showed the stability of Taylor style rules in different stylized macro models when agents use linear forecasting rules to form expectations. These studies also show that monetary policy under least squares learning converges to a rational expectation equilibrium under certain conditions. Preston (2006) showed that policy makers will be efficient in setting policies if they have more information about the determinants of agents’ expectation. Orphanides and Williams (2003, 2004, and 2005) showed that policies that would be efficient under RE can perform poorly when agents are learning. Bullard and Cho (2005) showed that in a macroeconomic system where monetary authority employs a Taylor-type rule it is possible to avoid liquidity trap episodes under certain conditions. Bullard and Eusepi (2005) considered an economy where agents use adaptive learning to form expectations and the monetary policymaker is committed to using a Taylor-type policy rule. They showed that the observed increase in inflation during the 1970s can be
attributed to the unexpected change in the trend productivity growth rate under learning. In most cases the above-mentioned studies use simulation and calibration to show their results.

Although there is a growing number of studies that use least squares learning in theoretical macro models, there are practically no studies that conduct empirical tests of whether agents show learning behavior. The one exception is Branch and Evans (2006) who compared the performance of alternative recursive forecasting models for inflation and output growth. Their study used different models and estimation techniques to mimic the Survey of Professional Forecasters (SPF). They found evidence that a forecasting model which implements a constant gain learning algorithm best fits the SPF. Their study conducts an indirect test of whether agents show learning behavior by using different learning algorithms to try and reproduce the SPF’s forecasts.

In this study, we conduct a formal direct test of whether agents’ forecasts of the 3-month Treasury bill rate are consistent with least squares learning when there are changes in monetary policy. We specifically examine the one quarter ahead mean and median 3-month Treasury bill forecasts from the SPF. We first derive the theoretical conditional mean, variance and autocovariances of the forecast error of the short term interest rate from the SPF's assuming there are discrete shifts in policy and least squares learning. We show that when the optimal weights associated with the parameters of the forecasting model changes due to structural shifts, under least squares learning the forecast error will not have zero mean and constant variance. Then we apply the Bai and Perron (1998, 2001, and 2003) test to identify the number and dates of the breaks in a Taylor rule and in the optimal weights for the information set of the short term interest rate forecasting
model. Next we estimates the optimal weights for agents information set in each policy regime and empirically estimate the conditional mean and variance of the short term interest rate forecast error. Then we standardize the mean and median forecast error from the SPF with the estimated mean and variance. Finally we test the standardized residuals for serial correlation. If agents have same variables in their information set that we have and use least squares learning then standardizing their forecast error will give residuals which will produce zero mean and constant variance. Our result shows that the standardized residuals are serially uncorrelated.

The paper proceeds as follows. Theoretical results and an econometric methodology are discussed in section 3.2, section 3.3 describes data, specification of the Taylor rule and an interest rate forecasting model, empirical results are discussed in section 3.4 and finally section 3.5 concludes.

3.2 Theoretical Results

We assume agents construct their forecasts based on the model,

$$y_t = x_t \beta + u_t$$

$$u_t \sim IID(0, \sigma^2)$$

(23)

where $y_t$ is a scalar of observed dependent variable at time $t$. The $p \times 1$ vector $x_t$ contains conditioning variables known to agents at time $t-1$ and are used to forecast $y_t$. The conditioning variables include lagged values of $y_t$ itself. If agents are rational, (23) represents the reduced form equation for and is derived from the structural model that describes how $y_t$ is actually generated in the economy. Assuming rationality $\beta$ contains the true reduced form parameters. If agents are not rational, then (23) is simply the
forecasting model used by agent, where \( x_t \) contains the variables in their information set and \( \beta \) contains the optimal weights in the projection of \( y_t \) on \( x_t \). If agents are rational, the properties of their forecast errors are very easy to derive. Under rationality, agents know the reduced form model (23). The forecast in period \( t + 1 \) is \( y_{t+1}^* = x_{t+1}'\beta \) and the forecast error is \( v_{t+1}^* = y_{t+1} - y_{t+1}^* = u_{t+1} \). Therefore the mean, the variance and the autocovariances of the observed forecast errors, conditioned on the past information are

\[
E(v_{t+1}^*|\Phi_t) = E(u_{t+1}|\Phi_t) = 0
\]

\[
\text{var}(v_{t+1}^*|\Phi_t) = \text{var}(u_{t+1}|\Phi_t) = \sigma^2
\]

and

\[
\text{cov}(v_{s}^*, v_{k}^*) = \text{cov}(u_{s}, u_{k}) = 0
\]

for \( s \neq k \). where \( \Phi_t \) is the information set at time \( t \). Under rationality, the forecast error is identical to the reduced form error and inherits its properties. The one period forecast errors should have mean zero and be serially uncorrelated. If the mean is not zero, there is a systematic bias in the forecast. If the forecast errors are serially correlated, the forecasts are not efficient and omit relevant information contained in the information set. A violation of either condition contradicts the assumption rationality. As described in the previous section, test biased on surveys of forecast often reject rationality because forecast errors are serially correlated.

### 3.2.1 Properties of Forecast Error under Least Squares Learning

Next we consider the properties of forecast errors under least squares learning. Let \( X^t = (x_1, ..., x_t)' \) be a \( t \times p \) matrix containing the conditioning variables starting at the
beginning of the sample and running through time period \( t \). Under least squares learning, the forecast in period \( t + 1 \) is

\[
\hat{y}_{t+1} = x_{t+1}' \beta
\]

where

\[
b_t = b_{t-1} + [(X^{t-1})'X^{t-1}]^{-1} x_t (y_t - x_t'b_{t-1}) / f_t
\]

is the recursive least squares estimate of \( \beta \) using data through time period \( t \) and

\[
f_t = 1 + x_t'[(X^{t-1})'X^{t-1}]^{-1} x_t
\]

The mean, variance and autocovariance of the recursive least squares forecast errors conditioned on the past information are, [see, for example, Harvey (1990)],

\[
E(\hat{v}_{t+1} | \Phi_t) = E(x_{t+1}' \beta + u_{t+1} - x_{t+1}' b_t)
\]

\[
= x_{t+1}'(\beta - E(b_t)) = 0
\]

\[
var(\hat{v}_{t+1} | \Phi_t) = \sigma^2 [1 + x_{t+1}'[(X')'X']^{-1} x_{t+1}]
\]

and

\[
cov(\hat{v}_k, \hat{v}_s) = 0
\]

for \( k \neq s \) In a stable structural environment, the assumption of least squares learning has little empirical content to distinguish it from an assumption of RE. If agents learn by least squares, the forecast errors still have a zero mean and zero covariances. The only distinction under learning is that the forecast error variance becomes heteroscedastic. The variance of the forecast error under least squares learning has two components. The first component is due to the variance of the error in the DGP and the other component is due to the parameter uncertainty. Although the conditional variance is heteroscedastic, if the structure of the economy remains the same, as the forecaster obtains more data over time the parameter uncertainty disappears because

\[
\lim_{t \to \infty} var(\hat{v}_{t+1} | \Phi_t) = \lim_{t \to \infty} \sigma^2 [1 + x_{t+1}'[(X')'X']^{-1} x_{t+1}]
\]

49
\[ a = 1, \quad r_l = \frac{\lim_{t \to \infty}}{\sum_{i=1}^{\infty} \frac{1}{i}}(X^t)'X^t x_{t+1} = \sigma^2 \]

If the economy is stable and the forecaster has a large amount of data, the observed forecast error will have the same properties as rational forecast errors. In such a scenario, it is difficult to test for the presence of learning.

3.2.2 Properties of Forecast Error under Least Squares Learning in the Presence of Structural Breaks

For least squares learning to be empirically identified, there must be structural change in the economy. In the post-war period of U.S. history, there is ample evidence of structural shifts not only in monetary policy but also in other important macroeconomic relationships. The structural breaks are in time periods \( T_i, ..., T_m \). The forecast model (23) can be written in the matrix form as

\[
Y_i = X_i \beta_i + U_i \quad i = 1, ..., m
\]

(24)

where \( Y_i \) is a vector containing the variable to be forecasted in regime \( i \), \( X_i \) is a \((T_{i+1} - 1 - T_i) \times p\) matrix of conditioning variables, \( \beta_i \) is a vector of optimal weights in regime \( i \) and \( U_i \) is the vector of forecast errors with mean vector 0 and variance-covariance matrix \( \Sigma^2 I_{T_{i+1} - 1 - T_i} \). Each regime has \( T_{i+1} - 1 - T_i \) observations. The coefficient vectors \( \beta_i \) and the error variances \( \sigma^2 \) are allowed to vary across regimes.

We now derive the properties of the forecast errors under least squares learning when there are regime shifts as described in (24). The forecast error in period \( T_j + s + 1 \) is

\[
\tilde{\sigma}_{T_j+s+1} = y_{T_j+s+1} - \tilde{y}_{T_j+s+1}
\]

(25)
\[ \hat{\beta}_{T_j+s+1} = x'_{T_j+s+1} (\beta_j - b_{T_j+s+1}) + u_{T_j+s+1} \] (26)

where period \( T_j + s + 1 \) denotes the \((s + 1)\)th observation in period \( T_j \). When there are regime shifts, we can show that

\[ b_{T_j+s+1} = \left[ (X^{T_j+s})' X^{T_j+s} \right]^{-1} \sum_{i=1}^{j-1} (X'_i X_i \beta_i + X'_i U_i) \]

and therefore,

\[ E(b_{T_j+s+1}) = \left[ (X^{T_j+s})' X^{T_j+s} \right]^{-1} \sum_{i=1}^{j-1} (X'_i X_i \beta_i) \]

We see that \( E(b_{T_m+s}) \) depends on all previous values of the conditioning variables and is a matrix weighted average of the coefficients \( \beta_i \) in the current and all previous regimes. Using the expression for \( E(b_{T_m+s}) \) in (26), we can show that

\[ E(\hat{\beta}_{T_j+s+1}) = \sum_{i=1}^{j-1} x'_{T_j+s+1} \left[ (X^{T_j+s})' X^{T_j+s} \right]^{-1} (X'_i X_i) (\beta_j - \beta_i) \]

(27)

When there are structural shifts under least squares learning, forecast error no longer have zero mean. The mean of the forecast error is a function of all the conditioning variables through time \( T_j + s \) and the difference between the parameters in the current regime and the past regimes. When there are no structural shifts so that \( \beta_j = \beta_i \) for all \( i \), the bias vanishes and the forecast errors have zero mean. The variance of the forecast error in period \( T_j + s + 1 \) is

\[ \text{var}(\hat{\beta}_{T_j+s+1} | \Phi_{T_j+s}) = E \left( u_{T_j+s+1} - x'_{T_j+s+1} \left[ (X^{T_j+s})' X^{T_j+s} \right]^{-1} (X^{T_j+s})' U^{T_j+s} \right)^2 
\]

\[ = \sigma^2 + x'_{T_j+s+1} \left[ (X^{T_j+s})' X^{T_j+s} \right]^{-1} (X^{T_j+s})' \Sigma^{T_j+s} X^{T_j+s} \left[ (X^{T_j+s})' X^{T_j+s} \right]^{-1} x_{T_j+s+1} \]

(28)
where $\Sigma_{T_j+s}^T$ is a block diagonal variance-covariance matrix of $T_j + s$ observations. The first $j - 1$ blocks have $T_i+1 - 1 - T_i$ elements $\sigma_i^2$ for $i = 1, \ldots, j - 1$ and the last block has $s$ elements of $\sigma_j^2$. The variance is same within the block but varies across blocks due to varying regimes. Similarly variance of the forecast error is heteroscedastic. The variance of the forecast error does not depend on the $\beta_i$'s but does depend on the variance of the error in each of $m$ regimes. If structural breaks are present to continue in the future, the heteroscedasticity does not disappear as the sample increases. If the error variances are the same across regimes $\sigma_1^2 = \cdots = \sigma_m^2$, the variance of the forecast error in (28) reduces to the conventional formula for the variance of the recursive residual. We also derive the autocovariances of the forecast error. Consider the forecast error in any two periods $\tilde{\epsilon}_{T_i+k}$ and $\tilde{\epsilon}_{T_j+s}$, where $T_i + k < T_j + s$. Note that $T_i + k$ and $T_j + s$ may be in the same regime where $T_i = T_j$ or may be in different regimes where $T_i < T_j$.

The autocovariance is

\[
\text{cov} \left( \tilde{\epsilon}_{T_i+k}, \tilde{\epsilon}_{T_j+s} \right) = E \left[ \left( u_{T_i+k} - x_{T_i+k} \left( X_{T_i+k-1}^T X_{T_i+k-1} \right)^{-1} \left( X_{T_i+k-1} U_{T_i+k-1} \right) \right) \right.
\]

\[
\left. \left( u_{T_j+s} - x_{T_j+s} \left( X_{T_j+s-1}^T X_{T_j+s-1} \right)^{-1} \left( X_{T_j+s-1} U_{T_j+s-1} \right) \right) \right]
\]

\[
= -E \left[ u_{T_i+k} (U_{T_i+k-1})^T X_{T_i+s-1} \left( (X_{T_i+s-1})^T X_{T_i+s-1} \right)^{-1} x_{T_i+s} \right]
\]

\[
+ E \left[ x_{T_i+k} \left( (X_{T_i+k-1})^T X_{T_i+k-1} \right)^{-1} (X_{T_i+k-1})^T U_{T_i+k-1} (U_{T_i+k-1})^T X_{T_i+k-1} \right]
\]

\[
\left( (X_{T_i+s-1})^T X_{T_i+s-1} \right)^{-1} x_{T_i+s} = 0.
\]

(29)

Although the expressions for the mean and variance of the forecast error change when there is least squares learning and the model has structural shifts, in the absence of
structural shifts it reduce down to the standard results under least square learning. The covariance, when there are structural shifts, is 0 if the forecast errors are bias corrected otherwise the bias in the forecast error will show up as serial correlation in the autocovariance. Least squares learning with structural shifts give a clear idea about how and why forecast errors have non-zero mean and heteroscedastic variance.

The preceding results provide a method of testing whether observed forecast errors are consistent with least squares learning. Although agents may not be aware of structural breaks when they occur, we can identify them ex-post with a sufficient amount of data. We can then estimate optimal weights $\beta_i$ and the error variances $\sigma_i^2$ in (28) in each regime by least squares using data over the entire sample period. The estimates of the $\beta_i$ and the $\sigma_i^2$ can be substituted into (27) and (28). Let the estimator of $E(\delta_t | \Phi_{t-1})$ given in (27) be $\delta_t$ and the estimator of $\text{var}(\delta_t | \Phi_{t-1})$ given in (28) be $f_t$. Then for each $t$ we can standardize the forecast error,

$$\bar{\delta}_t = (\delta_t - \delta_t)/\sqrt{f_t}$$

and construct a $Q$-Statistic,

$$Q^{LS} = \sum_{i=1}^{p} (\bar{r}^i_0)^2 \sim \chi^2$$

where $\bar{r}^i_0$ is the ith autocorrelation coefficient of $\delta$. The $Q^{LS}$ will be distributed as $\chi^2$ with $p$ degree of freedom where $p$ is the number of regressors in the model. If agents do least squares learning then the autocorrelation function of $Q^{LS}$ should be iid.
3.3 The Taylor Rule and an Interest Rate Forecasting Model

3.3.1 Specification of the Models

We use the short term interest rate forecasting model to test whether agents learn by least squares. The mean and median one quarter ahead short term interest rate forecasts are used from the SPFs to conduct the above test. It is a known fact that the U.S. economy has passed through infrequent structural shifts in its post war history. Bai and Perron (1998, 2001, and 2003) provide a test that identifies number and dates of multiple structural breaks in a linear model. Thus we apply the Bai and Perron test to both the Taylor rule and the information set of the interest rate forecasting model to identify the number of breaks and break dates in the policy rule. We do this to match up the break dates of the Taylor rule and the information set of the short term interest rate forecasting model. Given that Fed rate and short term interest rates are highly correlated, it is very likely that the break dates in the policy rule and the break dates in the interest rate forecasting model coincide. The break dates that we find in the policy rule and in the interest rate model are very similar.

Our theoretical results suggest that the bias of the forecast errors when there is least squares learning and structural break depend on the $\beta_i$s in (27) which means that when there are structural shifts in the policy rule, variables are assigned different weights in different regimes. When SPFs forecast the short term interest rate in real time they do not observe these shifts and they cannot incorporate these shifts while forecasting short term interest rate. Thus their forecast errors are biased. However, ex-post we observe
these break dates. So we use all the data and estimate those optimal weights ($\beta_i$'s) in different regimes. Since ex-post these shift dates are observed, we incorporate those break dates and forecast short term interest rate. However the forecasting model is not meant to forecast short term interest rate but to estimate the bias and variance in the forecast error. Then we standardize the SPF's short term interest rate forecast errors with the estimated mean and variance of the forecast error. If the information set is properly specified (i.e. what SPF's use to forecast short term rate) then the standardized residuals should be iid. This result demonstrates least squares learning behavior.

Over the last few decades the Taylor rule (Taylor 1993) has attracted increased attention. A typical Taylor rule is specified as follows,

$$r_t = \pi_t + \tau^* + a_y y_t + a_{\pi} (\pi_t - 2) + a_r r_{t-1}$$

where $r_t$ is the Federal funds rate, $\pi_t$ is the four quarter moving average GDP deflator, $\tau^*$ is the equilibrium interest rate, $y_t$ is output gap, 2% is the target inflation rate and $r_{t-1}$ is one period lagged Fed funds rate. The Taylor rule recommends a setting for the level of nominal Federal funds rate that depends on four factors. The first factor is the current inflation rate. The second factor is the equilibrium real interest rate. Both these factors added together to provide the benchmark nominal federal funds rate. The benchmark Federal funds rate (the sum of first and second factor) raises one to one with the current inflation rate. If the current inflation rate is high, it recommends a high benchmark rate as well keeping everything else constant. The use of an equilibrium real interest rate emphasizes the importance of real interest rate in formulating monetary policy. More precisely the rule says that the real interest rate will be increased above equilibrium if the inflation is above it's target or the real GDP is above it's potential. The third factor is an inflation gap adjustment based on the difference between the inflation rate and a given
target for inflation. The target inflation rate is taken as 2%. This factor recommends raising the Federal funds rate above the benchmark rate if inflation is above its target rate and vice versa. The inflation gap adjustment also captures the long-run goal of the Central Bank. The fourth factor is an output gap adjustment based on the difference between real GDP and potential real GDP. The Fed funds rate is raised above the benchmark rate if real GDP is above potential real GDP and vice versa. The output gap adjustment also captures the short-run goal. Although monetary policy rule can be characterized by small number of variables such as inflation gap and output gap, it can’t capture the complexities of policy process. Many variants of the Taylor rule have been estimated in the literature. A key finding is that a lagged value of interest rate is highly significant in the estimated policy rule. Examples are Sack (1998), Amato and Laubach (1999), Clarida, Gali and Gertler (2002), Rudebusch (2002) English, Nelson and Sack (2003). In addition to policy inertia it also captures serially correlated errors which could reflect various episodic factors from the Taylor rule (Rudebusch 2002). Further English, Nelson and Sack (2003) have shown that even after allowing for serially correlated errors the lagged interest rate in the policy rule remains significant and plays an important role in the dynamics of the short term interest rate.

We also need to specify information set that might plausibly represent what the SPF use to forecast the short term interest rate. Agents are well aware of the Federal Reserve’s two legislated goals: price stability and full employment stated in the Full Employment and Balanced Growth Act 1978. Keeping the Federal Reserve mandate in mind it is reasonable to assume that agents will use the output gap and the inflation gap variable to forecast the short term interest rate. Thus, we have included the output gap
and inflation gap variables in the information set of our forecasting model. We also have lagged short term interest rate in our information set. As mentioned above lagged short term rate has both theoretical and empirical significance. So it is reasonable to assume that agents use lagged short term rate to forecast future short term rate. Agents may also use information that is not explicitly in the Taylor rule. An example would be the slope of the yield curve. The rational expectation theory of the term structure says that the long term rate is the average of the expected future short term rates. If the spread between the long term rate and short term rate is positive then the market expects the short term rate to go up in the future. Early studies by Fama (1984), Hardouvelis (1988), Mishkin (1988) and Cambell and Shiller (1983, 1991) found that the yield spread between short term and long term rates do help to forecast the change in the short term rate. More recent studies, such as Ahrens (2002), Ang, Piazzesi and Wei (2006), Feroli (2004), Bernake (2006) Estrella (2005a, 2005b), confirmed that the slope or the term structure has a positive relationship with the future short term interest rate. Business economists are very aware of the rational expectation theory of the term structure and its empirical support. Therefore we include lagged six month Treasury bill rates in the information set.

3.3.2 Data

The one quarter ahead mean and median forecasts of the 3-month Treasury bill rate were obtained from the SPF. Every three months, starting first quarter of 1968, the Federal Reserve Bank of Philadelphia takes a survey of forecasts, prepared by private sector economists of twenty or more economic variables, including output, inflation, and
interest rates. The respondents are professional forecasters who produce regular forecasts of economic variables. One important feature of the Survey of Professional Forecasters is anonymity of the forecasters. Anonymity is important because forecasters can't claim credit for good forecasts, nor can they be held responsible for bad forecasts. For details see Crousher (1993). The American Statistical Association together with National Bureau of Economic Research began conducting the survey in fourth quarter of 1968. Initially there were 50 participants in each quarter in its early days but later on the number reduced down to 20. By 1990 the ASA and the NBER decided to discontinue the Survey. Later on in 1990 the Federal Reserve Bank of Philadelphia revived the survey by inviting participants. In the third quarter of 1981, the survey was expanded to include one quarter ahead forecasts of the 3-month Treasury bill rate. Our sample for the analysis of the survey forecasts error will therefore start in the third quarter of 1981.

The Federal funds rate, 3-month Treasury bill rate (secondary market rate), 6-month Treasury bill rate and the GDP deflator have been obtained from Federal Reserve Bank of St. Louis. There are various measures of output gap used in estimated policy rules. Taylor (1993) fitted a log linear trend to real GDP to measure potential GDP. In some studies, a quadratic trend is fitted for the potential GDP measure [Clarida, Gali and Gertler (1997a, 1997b)]. OECD measure of potential GDP [De Masi (1997)] and the IMF measure of potential GDP are used in other studies. However, the Congressional Budget Office (CBO) measure of potential GDP is most common in the policy rule literature [Rudebusch (1998), Clarida, Gali and Gertler (2000)]. For our analysis we use Congressional Budget Office measure of potential GDP. The CBO measure of potential GDP is not a simple fitted GDP trend, but is estimated in terms of a relationship with
future inflation similar to the way a time-varying NAIRU is estimated within the context of a Philips curve (Rudebusch 1998). The output gap is calculated by taking the log difference of real GDP and potential GDP, multiplied by 100. The 3-month Treasury bill and 6-month Treasury bill are converted from a monthly annualized to a quarterly annualized rate. The four-quarter moving average of the annualized price inflation of GDP deflator is used.

We estimate the Taylor rule and the Treasury bill forecasting rule using quarterly data that starts in the first quarter of 1960 and ends in the first quarter of 2007. The information set for the Taylor rule has a constant, lagged Fed rate, price inflation of GDP deflator minus the target GDP deflator (taken as 2%) and output gap. The information set for the 3-month Treasury bill forecasts has a constant, lagged 3-month Treasury bill rates, lagged 6-month Treasury bill rates, lagged output gaps and lagged price inflation of GDP deflator gaps. Although we have data since 1960 for the variables in our information set, the mean and median 3-month Treasury bill survey forecasts are observed only since third quarter of 1981. The survey forecast error can only be constructed since 1981. However, to estimate the mean and variance of the forecast error since 1981, we need all the data back from 1960 because the bias depends on the parameter estimates of past regimes (27) and the variance depends on the variance of past regimes (28).
3.4 Empirical Results

3.4.1 Structural Change Tests Results

First we test for multiple regime shifts in the Taylor rule and in the information set of the short term interest rate. In order to identify the number of breaks and the break dates we apply the Bai and Perron (1998, 2001, and 2003) test for structural change to the Taylor rule and the forecast model for 3-month Treasury bill rate. The Bai and Perron test is nice because it explicitly treats the breakpoints as unknown, and estimates of the breakpoints are generated using the least square principle. Bai and Perron (1998) developed procedures to identify the number of structural breaks in an equation. They developed two statistics, which they called “double maximum” statistics, for testing the null hypothesis of no structural breaks against the alternative hypothesis of an unknown number of breaks not to exceed an upper bound M. The first double maximum statistic is given by

\[ UD_{\text{max}} = \max_{1 \leq m \leq M} \text{Sup } F_T(m) \]

where \( F_T(m) \) is the \( F \) statistic for testing the null hypothesis of no structural breaks against the alternative of \( m \) breaks. The second double maximum statistic, \( WD^{\text{max}} \), applies different weights to the individual \( \text{Sup } F_T(m) \) statistics so that marginal \( p \) - values are equal across values of \( m \) [see Bai and Perron (1998), p. 59 for details]. Finally, Bai and Perron (1998) specified what they labeled the \( \text{Sup } F_T(l + 1|l) \) statistics to test the null hypothesis of \( l \) breaks against \( l+1 \) breaks. It begins with the global minimized sum of squared residuals for a model with \( l \) breaks. Each of the intervals defined by the \( l \) breaks
is then analyzed for an additional structural break. From all of the intervals, the partition allowing for an additional break that model with \( l + 1 \) breaks. The \( \text{Sup} F(l + 1|l) \) statistics is used to test whether the additional break leads to a significant reduction in the sum of squared residuals. Bai and Perron (2001) developed an efficient algorithm for the minimization of least squares problem based on the principle of dynamic programming. They also developed a method of forming confidence intervals for the break dates under various hypotheses about the structure of the data and error segments.

Table 4 reports Bai and Perron (1998, 2001 and 2003) statistics for the tests of structural change in the Taylor rule and the forecast model for the 3-month Treasury bill rate. In the Taylor rule specification the Federal Funds rate is regressed on a constant, lagged Fed rate, price inflation GDP deflator gap and output gap. To test breaks in the short term rate, the short term rate is regressed on its own lag, a constant, lagged GDP deflator gap, lagged output gap and lagged 6-month Treasury bill rate. For the forecasting model, the common lag length was selected to minimize the SIC criterion. The minimizing lag was two. We allow up to 5 breaks, hence each segment has at least 15 observations. We also allow for heteroscedasticity in the equation errors across regimes. The first issue to be considered is the determination of the number of breaks. For both the Taylor rule and the short term interest rate we find three breaks. The \( WD_{\text{max}} \) test, \( UD_{\text{max}} \) tests and \( \text{SupF} \) tests of sequential break dates are significant at conventional significance level. This result shows strong evidence of structural change in the Taylor rule and 3-month Treasury bill rate. Other than \( F(5|4) \) and \( F(4|3) \), all other sequential test statistic such as \( F(1|0) \), \( F(2|1) \), and \( F(3|2) \) are highly significant for both the cases. The sequential procedure using 5\% level of significance selects 3 breaks simply comparing the fitted...
models, the BIC and the modified Schwarz criterion of Liu et al. (1997) selects models with two breaks. Given the documented facts that the information criteria are biased downward and that the sequential procedure and the $F_{T}(l + 1|l)$ perform better in this case (Bai and Perron 2003). We conclude in favor of the presence of three breaks.

Table 5 reports the dates for the structural breaks in the Taylor rule and the 3-month Treasury bill rate and their 95% confidence intervals. Both the Taylor rule and the 3-month Treasury bill rate have structural breaks during late 1960s, 1970s and mid 1980s. For the Taylor rule the three breaks are first quarter of 1969, third quarter of 1979 and first quarter of 1987. The 95% confidence intervals for these three dates are first quarter 1967 and third quarter 1969, first quarter 1978 and second quarter 1979, second quarter 1986 and second quarter 1989 respectively. For the 3-month Treasury bill rate the break dates are second quarter of 1969, third quarter of 1979 and first quarter of 1987 which are very close to the breaks in the Taylor rule. The confidence intervals are third quarter 1967 and third quarter 1969, second quarter 1979 and fourth quarter 1979, fourth quarter 1986 and second quarter 1988 respectively. Compared to the first and third break dates the second break has more tight intervals in both the Taylor rule specification and 3-month Treasury bill rate specification. Our paper is based on the assumption that changes in monetary policy, as reflected by shifts in the Taylor rule, will change the optimal weights on the variables agents use to forecast the short term interest rate. It is very reassuring that the tests identify basically the same dates for the structural change in both the Taylor rule and the forecasting model for the interest rate. The estimation of the forecast error mean and variance depends on correctly specifying the variables in the information set and identifying the structural change dates. If either is incorrect, the
estimation of the bias and variance of the forecast error will be biased and the standardization of the survey forecasts won’t produce white noise. Our dates are also consistent with the dates of other papers that look for U.S. interest rate. Duffy and Warnick (2006) used a piecewise-linear classification and regression tree method to test for multiple regimes in U.S. monetary policy. They applied the non-parametric test to two specification of the Taylor rule. In the first specification the break dates were the first quarter 1968, the third quarter 1979 and the fourth quarter 1986. In the second specification the break dates were the first quarter of 1969, the third quarter of 1980, the third quarter of 1987 and the third quarter of 1996 which are more or less similar with our studies. Bai and Perron (2003) applied their own test and found three breaks in a univariate model of 3-month Treasury bill rate deflated by the inflation rate. The break dates for their studies were the fourth quarter 1966, the third quarter 1972 and the third quarter 1980. Caporale and Grier (2000) applied the Bai and Perron test and used the same interest rate definition as Bai and Perron. They found three breaks in the first quarter 1967, the fourth quarter 1972, the second quarter 1980 and the second quarter 1986. Rapach and Wohar (2005) confirmed these findings in U.S. and show that these real rate regime shifts also occur in many other industrialized countries. They applied the Bai and Perron test to a tax adjusted real rate using long term government bond rates. They found three breaks at first quarter 1966, first quarter 1973, second quarter 1981 and fourth quarter 1986.

Although there is a consensus on when monetary policy changed, the actual nature of the change is not as clearly understood. Duffy and Warnick (2006) argued that there are two major efforts at disinflationary policy, defined by increases in the weight
attached to the inflation gap variable. The first attempt begins after the first quarter of 1968 to counteract the inflationary financing of Vietnam War. Duffy and Warnick (2006), however, provide little evidence to document the nature of the change. The second break date is associated with the start of Paul Volker’s Fed chairmanship and strong disinflationary policy. During that period between the fourth quarter 1979 to third quarter 1980, the Fed switched from using the Federal Funds rate to using non-borrowed reserves as its operational target and set desired growth rates for the money aggregates M1 and M2 which lasts from third quarter 1979 to third quarter 1982. This policy change has been extensively studied and well understood. The cause of the third break date in 1986 is not very clear and no convincing argument has been made to explain it. After decades of research, academic economists have some understanding of when and how monetary policy shifted. It is implausible to imagine that forecasters identified these shifts and instantly modified their forecasting rule to accurately reflect the structural change in the economy. In the next section we examine whether the policy shifts were learned empirically through a recursive least squares learning algorithm.

3.4.2 Test for Standardized Residuals

We now test whether the mean forecast of the 3-month Treasury bill from the SPF is consistent with least squares learning. The Bai and Perron (2003) test identifies 3 structural breaks in the monetary policy rule and the optimal weights for the information set used for forecasts. First the break dates in the forecasting model are considered. Using the interest rate forecasting model the mean (27) and variance (28) of the forecast error
are estimated separately for $m=1$, $m=2$ and when $m = 3$. Then the forecast errors from the SPF are standardized using the estimated mean and variance. Assuming the information set what we specify for the interest rate forecasting model corresponds to what forecasters actually use, if agents engage in least square learning then the standardized of the forecast error should be serially uncorrelated process.

Table 6 presents the autocorrelation function (ACF), $Q^L$ statistics and their $p$-values for the mean raw forecast error and the standardized residuals. The first column shows the order of the autocorrelations and $Q^L$ statistics. The second column presents the ACF, QLS statistics and their respective $p$-values of the raw forecast error. As we can see they are correlated in low lags and clearly reject rationality. The third column shows the ACF, $Q^L$ statistics and the $p$-values allowing for the single most significant break in the third quarter 1979. The standardized residuals are highly significant. This result shows that one break and least squares learning is not enough to remove the bias from the forecast error. Column 4 shows the result for the standardized residuals incorporating two breaks, the break during third quarter 1979 and the break during first quarter 1987. When we incorporate two breaks the results change dramatically. The standardized residuals are now insignificant. Even though the theoretical mean and variance of the forecast errors are mis-specified, the optimal weights in the information set does not change significantly due to the omitted break in the second quarter of 1969. Hence the omission of the third break does not affect the unbiasedness of the standardized residuals significantly. This suggests that even though the Bai and Perron test identifies 3 breaks in the 3-month Treasury bill rate, the breaks in the third quarter of 1979 and the first quarter of 1987 are more important than the break in first quarter of 1969. Column 5 shows the
standardized residuals when we incorporate 3 breaks. In this case the results are not very different from what we have in column 4. This confirms our conjecture about the break in the first quarter of 1969 is not so important. This suggests that the change in the optimal weights of the parameters during the 1970s and 1980s are more significant than the change during the 1960s.

The analysis is repeated with the median short term interest rate forecast error from the SPF's. The same steps are used as for the mean response to test whether agents show learning behavior. The results are shown in Table 7. The median raw forecast errors are highly correlated up to 13 lags. When we incorporate 1 break and standardize the raw forecast error, there is not much improvement in the results. However when 2 or 3 breaks are incorporated the standardized residuals appear to be pure white noise. This result further confirms that the brakes during late 1970s and mid 1980s are more significant than the break during 1960s. The above analyses provide the evidence that agents learn by least squares.

In Figure 9 and Figure 10, we show the forecast errors and the conditional standard deviation of the mean corrected forecast error. As seen in Figure 9, from the beginning of the sample in the third quarter 1982 through fourth quarter 1989, there are large differences between the raw forecast errors and the mean corrected forecast errors. The errors are often times very different in magnitude and sometimes move in opposite directions. This is consistent with the structural breaks in the third quarter 1979 and fourth quarter 1986 introducing biases into the forecast errors. Following this episode, during which monetary policy was apparently stable, the errors begin to increasingly move together and by the end of sample are almost identical. In figure 10, the dramatic
increase in parameter uncertainty caused by the structural breaks in the late 1970's and mid-1980's is indicated by the sharp rise in the conditional standard deviation falls almost monotonically and converges. The series displayed in both figures are consistent with agents learning the new structural parameters as data accrues over time in the new policy regime.

The break dates in the Taylor rule and the break dates of Duffy and Warnick (2006) are also in the analysis and the results remain the same. The results are presented in Table 8, 9, 10 and 11.

3.5 Conclusions

In this paper we conduct a direct test to see if one quarter ahead 3-month Treasury bill rate forecasts from the SPF are consistent with least squares learning. It is widely believed that the U.S. monetary policy has undergone structural changes in the last few decades. To conduct the test we use the Taylor rule and the short term interest rate forecasting model. The purpose of using the Taylor rule specification is to match up the structural break dates in the interest rate model with the structural break dates in the monetary policy rule. The mean and median one quarter ahead 3-month Treasury bill forecast errors from the SPFs are used. First we derive the properties of the forecast error of a linear model when there are structural breaks and least squares learning. We show that when there are structural shifts in a linear model, under least squares learning the forecast errors are biased. This is likely the case for the SPFs since they do not observe the shifts in real time when they forecast. In the next step we apply the Bai and Perron
(1998, 2001, and 2003) test to identify number and dates of structural breaks in the Taylor rule as well as in the interest rate forecasting model. We find very similar break dates in the Taylor rule and in the interest rate model. Then incorporating those break dates we estimate the optimal weights in the information set of the interest rate model in each policy regime. Using these estimated optimal weights, least square learning we estimate the conditional mean and variance of the short term interest rate forecast error. Then the estimated conditional mean is subtracted from the mean and median 3-month Treasury bill survey forecast error to remove the bias. If the information set of the interest rate forecasting model is properly specified, the bias corrected standardized residuals will be serially uncorrelated. Before the bias correction and standardization the survey forecast errors were serially correlated but after suitable modification the survey forecast error the standardized residuals are serially uncorrelated.
Table 4

Bai and Perron (1998) Double Maximum and SupF_{T(l+1|l)} Statistics for Tests of Multiple Structural Breaks in the Taylor Rule and 3-month Treasury Bill Rate

<table>
<thead>
<tr>
<th>Parameters: $Z_t {q_{Tbill}, q_{Taylor}}$</th>
<th>Specification $q={q_{Tbill}, q_{Taylor}}$</th>
<th>$p=0$</th>
<th>$h=.15$</th>
<th>$M=5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tests: SupF_{T(2</td>
<td>1)} SupF_{T(3</td>
<td>2)} SupF_{T(4</td>
<td>3)} SupF_{T(5</td>
<td>4)} WD_{max} UD_{max}</td>
</tr>
<tr>
<td>T-bill: 49.06** 35.53** 25.59 0.00 54.58** 54.58**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5

Bai and Perron (2001) 95% Confidence Interval for the Break Dates

<table>
<thead>
<tr>
<th>Breaks</th>
<th>First break</th>
<th>Second break</th>
<th>Third break</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-bill: 1969:2</td>
<td>1979:3</td>
<td>1987:1</td>
<td></td>
</tr>
</tbody>
</table>
Figure 9
Raw Forecast Error and Mean Corrected Forecast Error with Three Structural Breaks
Figure 10

Conditional Standard Deviation of the Forecast Error with Three Structural Breaks
Table 6

ACF, Q-stat and P-values of the Standardized Residuals under Multiple Structural Breaks in the Mean 3-month Treasury Bill Rate

<table>
<thead>
<tr>
<th>Lags</th>
<th>Forecast error (mean)</th>
<th>1 break</th>
<th>2 breaks</th>
<th>3 breaks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACF</td>
<td>Q-stat</td>
<td>P-value</td>
<td>ACF</td>
</tr>
<tr>
<td>1</td>
<td>0.210</td>
<td>4.6383</td>
<td>0.031</td>
<td>0.338</td>
</tr>
<tr>
<td>2</td>
<td>-0.077</td>
<td>5.2648</td>
<td>0.072</td>
<td>0.088</td>
</tr>
<tr>
<td>3</td>
<td>-0.124</td>
<td>6.9231</td>
<td>0.074</td>
<td>-0.063</td>
</tr>
<tr>
<td>4</td>
<td>0.089</td>
<td>7.7785</td>
<td>0.100</td>
<td>0.038</td>
</tr>
<tr>
<td>5</td>
<td>0.070</td>
<td>8.3177</td>
<td>0.140</td>
<td>0.101</td>
</tr>
<tr>
<td>6</td>
<td>-0.131</td>
<td>10.206</td>
<td>0.116</td>
<td>-0.121</td>
</tr>
<tr>
<td>7</td>
<td>-0.188</td>
<td>14.154</td>
<td>0.049</td>
<td>-0.104</td>
</tr>
<tr>
<td>8</td>
<td>-0.074</td>
<td>14.779</td>
<td>0.064</td>
<td>-0.029</td>
</tr>
<tr>
<td>9</td>
<td>0.050</td>
<td>15.067</td>
<td>0.089</td>
<td>0.129</td>
</tr>
<tr>
<td>10</td>
<td>-0.024</td>
<td>15.133</td>
<td>0.127</td>
<td>0.064</td>
</tr>
<tr>
<td>11</td>
<td>-0.137</td>
<td>17.319</td>
<td>0.099</td>
<td>-0.031</td>
</tr>
<tr>
<td>12</td>
<td>-0.079</td>
<td>18.048</td>
<td>0.114</td>
<td>-0.025</td>
</tr>
</tbody>
</table>
Table 7

ACF, Q-stat and P-values of the Standardized Residuals under Multiple Structural Breaks in the Median 3-month Treasury Bill Rate

<table>
<thead>
<tr>
<th>Lags</th>
<th>Forecast error (median)</th>
<th>1 break</th>
<th>2 breaks</th>
<th>3 breaks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACF</td>
<td>Q-stat</td>
<td>P-value</td>
<td>ACF</td>
</tr>
<tr>
<td>1</td>
<td>0.185</td>
<td>3.6118</td>
<td>0.057</td>
<td>0.310</td>
</tr>
<tr>
<td>2</td>
<td>-0.133</td>
<td>5.4747</td>
<td>0.065</td>
<td>0.064</td>
</tr>
<tr>
<td>3</td>
<td>-0.124</td>
<td>7.1293</td>
<td>0.068</td>
<td>-0.050</td>
</tr>
<tr>
<td>4</td>
<td>0.096</td>
<td>8.1321</td>
<td>0.087</td>
<td>0.034</td>
</tr>
<tr>
<td>5</td>
<td>0.108</td>
<td>9.3996</td>
<td>0.094</td>
<td>0.121</td>
</tr>
<tr>
<td>6</td>
<td>-0.128</td>
<td>11.213</td>
<td>0.082</td>
<td>-0.130</td>
</tr>
<tr>
<td>7</td>
<td>-0.195</td>
<td>15.475</td>
<td>0.030</td>
<td>-0.104</td>
</tr>
<tr>
<td>8</td>
<td>-0.053</td>
<td>15.786</td>
<td>0.046</td>
<td>-0.025</td>
</tr>
<tr>
<td>9</td>
<td>0.049</td>
<td>16.057</td>
<td>0.066</td>
<td>0.112</td>
</tr>
<tr>
<td>10</td>
<td>-0.025</td>
<td>16.128</td>
<td>0.096</td>
<td>0.041</td>
</tr>
<tr>
<td>11</td>
<td>-0.158</td>
<td>19.026</td>
<td>0.061</td>
<td>-0.041</td>
</tr>
<tr>
<td>12</td>
<td>-0.074</td>
<td>19.675</td>
<td>0.073</td>
<td>-0.025</td>
</tr>
</tbody>
</table>
Table 8

ACF, Q-stat and P-values of the Standardized Residuals of the Mean 3-month Treasury Bill Rate under Multiple Structural Breaks in the Taylor Rule

<table>
<thead>
<tr>
<th>Lags</th>
<th>Forecast error (mean)</th>
<th>1 break</th>
<th>2 breaks</th>
<th>3 breaks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACF</td>
<td>Q-stat</td>
<td>P-value</td>
<td>ACF</td>
</tr>
<tr>
<td>1</td>
<td>0.210</td>
<td>4.6383</td>
<td>0.031</td>
<td>0.338</td>
</tr>
<tr>
<td>2</td>
<td>-0.077</td>
<td>5.2648</td>
<td>0.072</td>
<td>0.088</td>
</tr>
<tr>
<td>3</td>
<td>-0.124</td>
<td>6.9231</td>
<td>0.074</td>
<td>-0.063</td>
</tr>
<tr>
<td>4</td>
<td>0.089</td>
<td>7.7785</td>
<td>0.100</td>
<td>0.038</td>
</tr>
<tr>
<td>5</td>
<td>0.070</td>
<td>8.3177</td>
<td>0.140</td>
<td>0.101</td>
</tr>
<tr>
<td>6</td>
<td>-0.131</td>
<td>10.206</td>
<td>0.116</td>
<td>-0.121</td>
</tr>
<tr>
<td>7</td>
<td>-0.188</td>
<td>14.154</td>
<td>0.049</td>
<td>-0.104</td>
</tr>
<tr>
<td>8</td>
<td>-0.074</td>
<td>14.779</td>
<td>0.064</td>
<td>-0.029</td>
</tr>
<tr>
<td>9</td>
<td>0.050</td>
<td>15.067</td>
<td>0.089</td>
<td>0.129</td>
</tr>
<tr>
<td>10</td>
<td>-0.024</td>
<td>15.133</td>
<td>0.127</td>
<td>0.064</td>
</tr>
<tr>
<td>11</td>
<td>-0.137</td>
<td>17.319</td>
<td>0.099</td>
<td>-0.031</td>
</tr>
<tr>
<td>12</td>
<td>-0.079</td>
<td>18.048</td>
<td>0.114</td>
<td>-0.025</td>
</tr>
<tr>
<td>Lags</td>
<td>ACF</td>
<td>Q-stat</td>
<td>P-value</td>
<td>ACF</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>--------</td>
<td>---------</td>
<td>-----</td>
</tr>
<tr>
<td>1</td>
<td>0.185</td>
<td>3.6118</td>
<td>0.001</td>
<td>0.310</td>
</tr>
<tr>
<td>2</td>
<td>-0.133</td>
<td>5.4747</td>
<td>0.005</td>
<td>0.064</td>
</tr>
<tr>
<td>3</td>
<td>-0.124</td>
<td>7.1293</td>
<td>0.005</td>
<td>-0.050</td>
</tr>
<tr>
<td>4</td>
<td>0.096</td>
<td>8.1321</td>
<td>0.005</td>
<td>-0.054</td>
</tr>
<tr>
<td>5</td>
<td>0.108</td>
<td>9.3996</td>
<td>0.005</td>
<td>-0.050</td>
</tr>
<tr>
<td>6</td>
<td>-0.128</td>
<td>11.213</td>
<td>0.005</td>
<td>0.082</td>
</tr>
<tr>
<td>7</td>
<td>-0.195</td>
<td>15.475</td>
<td>0.005</td>
<td>-0.104</td>
</tr>
<tr>
<td>8</td>
<td>-0.053</td>
<td>15.786</td>
<td>0.005</td>
<td>0.046</td>
</tr>
<tr>
<td>9</td>
<td>-0.025</td>
<td>16.057</td>
<td>0.005</td>
<td>0.006</td>
</tr>
<tr>
<td>10</td>
<td>-0.158</td>
<td>19.026</td>
<td>0.005</td>
<td>0.061</td>
</tr>
<tr>
<td>11</td>
<td>-0.074</td>
<td>19.675</td>
<td>0.005</td>
<td>-0.025</td>
</tr>
<tr>
<td>12</td>
<td>-0.074</td>
<td>19.675</td>
<td>0.005</td>
<td>-0.025</td>
</tr>
</tbody>
</table>
Table 10

ACF, Q-stat and P-values of the Standardized Residuals of the Mean 3-month Treasury Bill Rate under Multiple Structural Breaks in the Taylor Rule from Duffy and Warnick (2006)

<table>
<thead>
<tr>
<th>Lags</th>
<th>Forecast error (mean)</th>
<th>1 break</th>
<th>2 breaks</th>
<th>3 breaks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ACF</td>
<td>Q-stat</td>
<td>P-value</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.210</td>
<td>4.6383</td>
<td>0.031</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>-0.077</td>
<td>5.2648</td>
<td>0.072</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>-0.124</td>
<td>6.9231</td>
<td>0.074</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.089</td>
<td>7.7785</td>
<td>0.100</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.070</td>
<td>8.3177</td>
<td>0.140</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>-0.131</td>
<td>10.206</td>
<td>0.116</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>-0.188</td>
<td>14.154</td>
<td>0.049</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>-0.074</td>
<td>14.779</td>
<td>0.064</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>0.050</td>
<td>15.067</td>
<td>0.089</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>-0.024</td>
<td>15.133</td>
<td>0.127</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>-0.137</td>
<td>17.319</td>
<td>0.099</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>-0.079</td>
<td>18.048</td>
<td>0.114</td>
</tr>
</tbody>
</table>
Table 11

ACF, Q-stat and P-values of the Standardized Residuals of the Median 3-month Treasury Bill Rate under Multiple Structural Breaks in the Taylor Rule from Duffy and Warnick (2006)

<table>
<thead>
<tr>
<th>Lags</th>
<th>Forecast error (median)</th>
<th>1 break</th>
<th>2 breaks</th>
<th>3 breaks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACF</td>
<td>Q-stat</td>
<td>P-value</td>
<td>ACF</td>
</tr>
<tr>
<td>1</td>
<td>0.185</td>
<td>3.6118</td>
<td>0.057</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.133</td>
<td>5.4747</td>
<td>0.065</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.124</td>
<td>7.1293</td>
<td>0.068</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.096</td>
<td>8.1321</td>
<td>0.087</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.108</td>
<td>9.3996</td>
<td>0.094</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-0.128</td>
<td>11.213</td>
<td>0.082</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-0.195</td>
<td>15.475</td>
<td>0.030</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-0.053</td>
<td>15.786</td>
<td>0.046</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.049</td>
<td>16.057</td>
<td>0.066</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>-0.025</td>
<td>16.128</td>
<td>0.096</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>-0.158</td>
<td>19.026</td>
<td>0.061</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>-0.074</td>
<td>19.675</td>
<td>0.073</td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


78


Capistran, C. (2006), “Bias in Federal Reserve inflation forecasts: Is the federal reserve irrational or just cautious?” Manuscript, Banco De Mexico,


Davies, R. B. (1987), “Hypothesis testing when a nuisance parameter is present only under the alternative”, *Biometrika*, 74(1), 33-43.


81


