NRB Working Paper No. 52

August 2021

Inflation Forecasting in Nepal: A Univariate Time-Series Approach

Rajan Krishna Panta, PhD*

ABSTRACT

This paper uses several standard univariate time series models and out-of-sample forecasts to assess their forecasting ability with year-on-year monthly inflation in Nepal over the period mid-August 2002 to mid-March 2021, corresponding to Nepal's fiscal year. These include conventional models based on seasonal and monthly dummies, Holt-Winters method with seasonality, standard ARIMA models, autoregressive models with different error structures and more sophisticated unobserved component model. It finds that an autoregressive model with AR(1) errors consistently outperforms its competing models - based on the mean squared forecast error (MSFE) and the direction of change. The results are robust to forecast horizons, out-of-sample forecast period, different lags structures and tests of predictability.

JEL Classification: C13, C22, C53, E31

Key Words: Inflation forecasting, state-space model

* Director, Nepal Rastra Bank (email: rajanpanta@nrb.org.np). I thank Gaurab Aryal (University of Virginia) for his helpful comments. I would be grateful for your suggestions and comments on the paper.

I. INTRODUCTION

Maintaining price stability is one of the major objectives of Nepal Rastra Bank (NRB) along with maintaining favorable balance of payment. For the conduct of monetary policy, central banks need to be forward-looking due to the lag in the monetary transmission mechanism. Inflation forecasting provides the basis to target the appropriate levels of other monetary and financial variables, such as money supply and credit. Reasonably accurate inflation forecasting is, thus, a sine qua non for the successful implementation of the monetary policy.

This paper utilizes the univariate time series techniques to model and forecast inflation in Nepal. The univariate time series analysis consists of a single variable ordered sequentially over uniform time intervals. In contrast to structural models, time series models are often a-theoretical in nature which attempt to capture empirical regularities in the data. Modelling the univariate series involves the analyses of its own current and past values and the error terms that are particularly useful for forecasting purposes (Enders, 2004; Hamilton, 1994; Chatfield, 2000).

There are several studies which examine the determinants of inflation in Nepal.[1] As Nepalese currency is pegged to the Indian rupee, the major variable influencing inflation in Nepal is the inflation rate in India among others. These econometric models, though useful in explaining the past behavior of inflation, have limited use in forecasting inflation. For forecasting inflation, one needs to forecast the explanatory variables, which increases the possibility of errors as the number of variables increase. In contrast to previous studies, this article focuses on exploiting the properties of time series data on inflation. This article, to the author's best knowledge, is the first attempt to rigorously model inflation in Nepal using a more general univariate time series approach.

The article presents twelve competing models starting from simple specifications to progressively complex ones. The first set estimates the models with different seasonal dummies and time trend. The second set utilizes the Holt-Winters method with seasonality. The third set imposes different error structures on the autoregressive models. Finally, the last model uses the state-of-the-art unobserved component model. The pseudo out-of-sample forecasting exercise is done and the Mean Square Forecast Error criterion is used to select the model with best forecasting model.

The remainder of the paper is structured as follows. Section two provides the description of the data. Section three examines the alternative theoretical models and estimation method of these models. Section four presents result of the estimation and Section five conducts a brief sensitivity analysis of the results. Finally, section six concludes.

---

[1] See for example, the report Inflation in Nepal for a literature review (Nepal Rastra Bank, 2007).

II. DATA

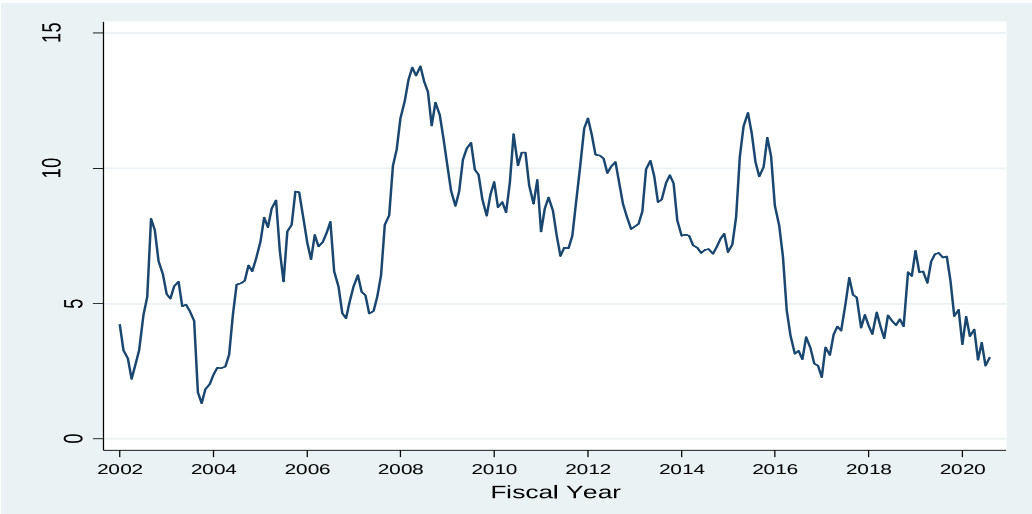
The data consist of a single time-series monthly inflation data from 2002 (mid-July) to 2021 (mid-March), corresponding to Nepalese fiscal year (that is 224 observations in total). The monthly inflation is computed from the year-on-year change in the Consumer Price Index (CPI). The CPI data series is taken from various issues of Quarterly Economic Bulletin published by the NRB.[2] For example, the inflation for first month of 2002 refers to the change in the CPI of mid-August, 2002 to the CPI of mid-August 2001. Taking year-on-year change inflation is one of the common methods to remove seasonality.[3]

The time series plot of the inflation data reveals the persistence and discernible upward trend up to mid-2008, reaching its peak (Figure1). Though the inflation declined afterwards, there does not appear a trend up to 2016, after which inflation declined sharply. Table 1 provides the major descriptive statistics of the inflation data series.

Table 1: Summary Statistics (in percent)

Variable	Mean	Standard Deviation	Minimum	Maximum
Inflation	7.09	2.82	1.31	13.77

Figure 1: Monthly year-on-year inflation



Note: The monthly inflation refers to the year-on-year change in the Urban CPI from mid-August, 2002 to Mid-March, 2021 according to the Nepalese fiscal year, which starts from around Mid-July.

[2] The Quarterly Economic Bulletin published by Nepal Rastra Bank contains various quarterly and monthly CPI series. For the present analysis, I use the National Urban CPI to compute the monthly inflation series.

[3] I also tried the annualized data which better reflects the subsequent changes in monthly inflation. However, the annualized data showed a significant degree of variation, which is not appropriate to forecast inflation.

III. MODEL AND ESTIMATION METHOD

This section presents competing models starting from a simple specification with the linear time trend and different seasonal dummies. The second model is based on Holt-Winters method with seasonality. The third core set comprise of autoregressive models with four different error structures. Finally, the results of these models are compared with the integrated moving average models. A one-step ahead recursive pseudo out-of-sample forecast is done for all models beginning from T0= 48 months.

The mean square (forecast) error (MSFE) is computed for all models. The mean square error (MSE)[4] is one of the popular criteria for model selection (Diebold, 2004).

The model with best forecast performance is selected based on the minimum MSFE.

3.1 Seasonal Dummy variables model

The first set of following four models are used for the forecasting exercises: Model (1) with the time trend and the fourth quarter dummy, Model (2) with the time trend and twelfth month (month of Ashad) dummy, Model (3) with time trend and four quarterly dummies, Model (4) with time trend and twelve monthly dummies.

$$y_t = a_0 + a_1 t + \alpha_4 D_{4t} + \varepsilon_t \quad \text{.......... (1)}$$

$$y_t = a_0 + a_1 t + \alpha_4 M_{12t} + \varepsilon_t \quad \text{.......... (2)}$$

$$y_t = a_1 t + \sum_{i=1}^{4} \alpha_i D_{it} + \varepsilon_t \quad \text{.......... (3)}$$

$$y_t = a_1 t + \sum_{i=1}^{12} \beta_i M_{it} + \varepsilon_t, \quad \text{.......... (4)}$$

where M_{it} is dummy variable that denotes month of the year, i.e. $M_{it} = 1$ if period t is the i-th month and $M_{it} = 0$ otherwise. Similarly, $D_{it} = 1$ if t is in the i-th quarter and $D_{it} = 0$ otherwise. Also $E\varepsilon_t = 0$ for all t = 1, …, T.

These models are estimated using the recursive Ordinary Least Squares (OLS) method.

3.2 Holt-Winters method with seasonality

The second set of alternative specification involves a more refined method using the Holt-Winters method with seasonality. Holt-Winters method extends the single exponential smoothing to linear exponential smoothing which is appropriate to forecast data with trends (Montgomery, Jennings, & Kulahci, 2015). The simple additive smoothing is suitable for data that do not seem to have a deterministic trend.

---

4 The mean square error, $= \frac{\sum_{i=1}^{T} e_t^2}{T}$, where T is the sample size and e_t is the forecast error.

Assuming that the data has not only a slowly evolving local level, but also a trend with a slowly evolving local slope, the standard Holt-Winters method with seasonality can be expressed as a system of four equations (Montgomery, Jennings, & Kulahci, 2015):

$$y_t = \mu_t + \lambda_t t + \varepsilon_t , \qquad \varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2)$$

$$\mu_t = ì_{t-1} + \eta_t , \qquad \eta_t \sim \mathrm{N}\left(0, \sigma_\eta^2\right) \qquad \ldots\ldots\ldots (5)$$

$$\lambda_t = \lambda_{t-1} + \nu_t , \qquad \nu_t \sim \mathrm{N}\ (0, ó_\nu^2\ )$$

$$\hat{y}_{t|t-1} = L_{t-1} + b_{t-1} + S_{t-s} ,$$

Where $\hat{y}_{t|t-1}$ is the point forecast, s = periodicity of seasonality (s = 12 for our monthly data) and the error terms are independent of each other at all leads and lags.

The level, slope (trend) and seasonality are updated, respectively, as follows:

$$L_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1})$$
$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$$
$$S_t = \gamma(\gamma_t - L_t) + (1-\gamma)S_{t-s}$$

The three components are initialized at t=s as follows:

$$L_s = \frac{1}{s}\ (y_1 + \cdots + y_s),$$
$$b_s = 0,$$
$$S_i = \frac{y_i}{L_s}\ for\ i = 1, \ldots., s.$$

Then the one-step-ahead forecast is given by

$$\hat{y}_{T+1|T} = L_T + b_T + S_{T+1-s} .$$

In general, the h-step ahead forecast is given by

$$\hat{y}_{T+h|T} = L_T + hb_T + S_{T+h-s} .$$

The specifications above assume that the values of the smoothing parameters (α,β, γ) are given. One way to estimate the model is to assume some initial values of the parameters and then choose the values that minimizes the MSFE. Alternatively, the optimal values of smoothing parameters can be obtained by numerical optimization method.

3.3 Autoregressive models with different error structures

The next set of models consist of the autoregressive models with different error structures. The motivation for this choice comes from the fact the series depicts a high degree of persistence. Moreover, the errors from the simple AR(1) models reveal that the errors are autocorrelated providing an additional source of information to improve the forecasting performance of the model. The following four set of AR(1) models with different error structures are considered: Model (6) has the standard assumption that the errors are normally distributed, Model (7) assumes that the errors have t-distribution.

Further, due to high persistence in the data, we assume the errors are autocorrelated with an AR(1) process (Model 8), Model (9) assumes that the errors follow an MA(1) process, and finally Model (10) is based on the state-space representation or the unobserved component model. The likelihood estimation approach is adopted to estimate the models. The concentrated log- likelihood function is derived and then the maximum likelihood estimates (MLE) for the coefficients are estimated using the using the simplex optimization method (Davidson & MacKinnon, 2003).

For estimation of Model (7) and Model (8), the log-likelihood function is derived using the standard normal distribution and t-distribution respectively.

3.4 AR(1) model with normal and t-distributed errors

Consider the AR(1) model with drift

$$y_t = \mu + \rho y_{t-1} + \varepsilon_t , \qquad \dots\dots\dots (6)$$

where ε_t are iid $N(0, \sigma^2)$.

Similarly, consider alternative AR(1) model with t-distributed errors η_t.

$$y_t = \mu + \rho_1 y_{t-1} + \eta_t , \qquad \dots\dots\dots (7)$$

The derivation of the log-likelihood function is relatively straightforward (see, Hamilton (1994) for the derivation).

3.5 AR(1) model with AR(1) errors

The derivation of the log-likelihood function for the AR(1) model with AR(1) errors can be obtained as follows (Davidson & MacKinnon, 2003; Hamilton, 1994):

Consider the AR(1) model with drift

$$y_t = \mu + \rho y_{t-1} + \varepsilon_t , \qquad \dots\dots\dots (8)$$

Where the errors follow an AR(1) process:

$\varepsilon_t = \phi\varepsilon_{t-1} + u_t$,

For t=1, …, T, $\varepsilon_0 = 0$, and u_t are iid $N(0, \sigma^2)$.

Given the observations $= (y_{1,} \dots, y_T)'$, and the initial y_0 , the MLE for ϕ, μ, ρ and σ^2 can be obtained as follows:

Write the model 8 as a linear regression:

$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\beta} = (\mu, \rho\,)'$,

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix},\ \boldsymbol{X} = \begin{pmatrix} 1 & y_0 \\ 1 & y_1 \\ \vdots & \vdots \\ 1 & y_{T-1} \end{pmatrix},\ \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}.$$

To derive the log-likelihood, we need the joint distribution of . Rewriting the system of errors in matrix form as :

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi & 1 & 0 & \cdots & 0 \\ 0 & -\phi & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & -\phi & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_T \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_T \end{pmatrix},$$

i.e. $\mathbf{H}_\phi \boldsymbol{\varepsilon} = \mathbf{u}$. Since $|\mathbf{H}_\phi| = 1$, $\mathbf{H}_\phi$ is invertible. Hence,

$\boldsymbol{\varepsilon} \sim N\left(0, \sigma^2 \left(\mathbf{H}_\phi' \mathbf{H}_\phi\right)^{-1}\right)$.

Therefore, it follows that

$(\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma^2) \sim N\left(\boldsymbol{X\beta}, \sigma^2 \left(\mathbf{H}_\phi' \mathbf{H}_\phi\right)^{-1}\right)$.

So the log-likelihood is

$$\begin{aligned} l(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}) &= -\frac{1}{2}\log|2\pi\sigma^2 \left(\mathbf{H}_\phi' \mathbf{H}_\phi\right)^{-1}| - \frac{1}{2\sigma^2}\ (\boldsymbol{y} - \boldsymbol{X\beta})' \mathbf{H}_\phi' \mathbf{H}_\phi (\boldsymbol{y} - \boldsymbol{X\beta}) \\ &= -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\ ((\boldsymbol{y} - \boldsymbol{X\beta})' \mathbf{H}_\phi' \mathbf{H}_\phi (\boldsymbol{y} - \boldsymbol{X\beta}). \end{aligned}$$

Differentiating the log-likelihood with respect to $\boldsymbol{\beta}$ and solving the first order condition for $\boldsymbol{\beta}$, we get

$$\widehat{\boldsymbol{\beta}}_\phi = (\mathbf{X}' \mathbf{H}_\phi' \mathbf{H}_\phi \mathbf{X})^{-1} \mathbf{X}' \mathbf{H}_\phi' \mathbf{H}_\phi \mathbf{y}\ , \qquad (\dots)$$

Similarly, differentiating the log-likelihood with respect to σ^2 and solving the first order condition, we obtain

$$\hat{\sigma}^2_{\phi} = \frac{1}{T}\left(y - X\widehat{\boldsymbol{\beta}}_{\phi}\right)'\mathbf{H}'_{\phi}\boldsymbol{H}_{\phi}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{\phi})$$

The MLE for the parameters is obtained by using the unconstrained nonlinear optimization method.[5] Given the MLE $(\hat{\mu},\ \hat{\rho}, \hat{\phi})$ of the model parameters we can produce the one-step-ahead forecast $\hat{y}_{t+1}$ as

$$\hat{y}_{t+1} = \hat{\mu} + \hat{\rho}y_t + \hat{\phi}\hat{\varepsilon}_t\ ,$$

As before, the recursive pseudo out-of-sample one-step-ahead forecasting is done beginning at T0 =48 and MSFE is computed.

3.6 AR(1) model with MA(1) errors

The third model in the set assumes that the errors in the AR(1) follows an MA(2) process. The log-likelihood function for the model can be obtained as follows:

Consider the following AR(1) model with MA(1) errors:

$$y_t = \mu + \rho y_{t-1} + \varepsilon_t \qquad \text{………. (9)}$$

$$\varepsilon_t = u_t + \psi u_{t-1}\ ,$$

For t=1, …, T, $u_0 = 0$, and u_t are iid $N(0, \sigma^2)$ for t ≥ 1.

Given the observations $= (y_{1,} \ldots, y_T)'$, and the initial y_0 , the log-likelihood function can be derived by similar method to the model with AR (1) errors, the only difference being the distribution of error terms ε. Rewriting the system of errors in matrix form as:

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_T \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \psi & 1 & 0 & \cdots & 0 \\ 0 & \psi & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & \psi & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_T \end{pmatrix},$$

i.e., $\boldsymbol{\varepsilon} = \mathbf{H}_{\psi}\mathbf{u}$ and $\boldsymbol{\varepsilon} \sim N\big(0, \sigma^2(\mathbf{H}_{\boldsymbol{\psi}}\mathbf{H}'_{\emptyset}\ \big)$. Therefore, it follows that

$$(\boldsymbol{y}\ |\psi, \boldsymbol{\beta}, \sigma^2) \sim N(X\beta, \sigma^2\mathbf{H}_{\boldsymbol{\psi}}\mathbf{H}'_{\psi}),$$

and the log-likelihood is (since $|\mathbf{H}_{\emptyset}| = 1$):

$$l(\psi, \boldsymbol{\beta}, \sigma^2\ |\boldsymbol{y}) = -\frac{1}{2}\log|2\pi\sigma^2\ \mathbf{H}_{\boldsymbol{\psi}}\mathbf{H}'_{\psi}| - \frac{1}{2\sigma^2}\ (\boldsymbol{y} - \boldsymbol{X\beta})'(\mathbf{H}_{\boldsymbol{\psi}}\mathbf{H}'_{\psi})^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})$$

---

5 The optimization procedures presents the danger of being trapped into a local maximum depending on the initial values assigned to the log-likelihood function. In order to ensure the maximum is the global one, we construct a rough grid using grid search method to obtain good starting values. Moreover, the concentrated likelihood graph will also be plotted for visual inspection.

$$= -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\,(\boldsymbol{y} - \boldsymbol{X\beta})'(\mathbf{H}_\psi\mathbf{H}_\psi')^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}).$$

To obtain the MLE, first concentrate the log-likelihood. Given ψ, the MLE for $\boldsymbol{\beta}$ and σ^2 can be obtained analytically through first order conditions:

$$\widehat{\boldsymbol{\beta}}_\psi = (\mathbf{X}'(\mathbf{H}_\psi\mathbf{H}_\psi')^{-1}\mathbf{X})^{-1}\mathbf{X}'(\mathbf{H}_\psi\mathbf{H}_\psi')^{-1}\mathbf{y},$$

$$\hat{\sigma}_\psi^2 = \frac{1}{T}\left(y - X\widehat{\boldsymbol{\beta}}_\psi\right)'(\mathbf{H}_\psi\mathbf{H}_\psi')^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_\psi)$$

The parameters here also are obtained using the numerical optimization method similar to Model 8. The one-step ahead forecast $\hat{y}_{t+1}$ can be obtained as follows:

$$\hat{y}_{t+1} = \hat{\mu} + \hat{\rho}y_t + \hat{\psi}\hat{u}_t\,.$$

3.7 The Unobserved Component Model

In this model, I estimate a state space model allowing an additional channel for persistence. Consider the following state space model:

$$y_t - \tau_t = \rho(y_{t-1} - \tau_{t-1}) + \varepsilon_t \quad \varepsilon_t \sim N(0,\sigma^2) \qquad \text{.........} (10)$$

$$\tau_t = \tau_{t-1} + u_t \qquad u_t \sim N(0,\omega^2)\,,$$

where the state equation is initialized with $\tau_1 \sim N(0,10)$, $\omega^2 = 0.5^2$ and $\tau_0 = 0$ without loss of generality. Suppose we observe $y_{0,}\, y_1 \dots, y_T$. For later reference, let $\boldsymbol{y} = \left(y_{1,} \dots, y_T\right)'$. Then the full sample is used to compute the MLE for τ, ρ and σ^2 as follows:

Equation 10 can be written as

$$y_t - \rho y_{t-1} = \tau_t - \rho\tau_{t-1} + \varepsilon_t$$

For $t = 1,\ y_1 = \tau_1 + \varepsilon_t$

For $t = 2,\ y_2 - \rho y_1 = \tau_2 - \rho\tau_1 + \varepsilon_2$, and so on. Therefore, Equation (9) can be written in matrix form as

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & -\rho & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & & & \ddots & 0 \\ 0 & 0 & \dots & -\rho & 1 \end{bmatrix}\begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \vdots \\ \tau_T \end{bmatrix} + \begin{bmatrix} \rho(y_0) \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

$\mathbf{H}_\rho\mathbf{y} = \mathbf{H}_\rho\boldsymbol{\tau} + \widetilde{\boldsymbol{\alpha}} + \boldsymbol{\varepsilon}$, where the notations denote the respective matrices. Therefore,

$$\mathbf{y} = \boldsymbol{\tau} + \mathbf{H}_\rho{}^{-1}\widetilde{\boldsymbol{\alpha}} + \mathbf{H}_\rho{}^{-1}\boldsymbol{\varepsilon}\,,$$

$\mathbf{y} = \boldsymbol{\alpha} + \boldsymbol{\tau} + \mathbf{H_\rho}^{-1}\boldsymbol{\varepsilon}$, where $\boldsymbol{\alpha} = \mathbf{H_\rho}^{-1}\widetilde{\boldsymbol{\alpha}}$

The MLE for $\boldsymbol{\tau}$ can be computed analytically given ρ and σ^2 as (set $\tau_0 = 0$:

$$\hat{\tau} = \frac{1}{\sigma^2}\mathbf{K}^{-1}\mathbf{H}_\rho'\mathbf{H}_\rho(\boldsymbol{y} - \boldsymbol{\alpha}),$$

where $\mathbf{K} = \mathbf{H}_\rho'\mathbf{H}_\rho/\sigma^2 + \mathbf{H}'\boldsymbol{\Omega}^{-1}\mathbf{H}, \boldsymbol{\Omega} = \mathrm{diag}(V_\tau, \omega^2, \dots, \omega^2), \boldsymbol{\alpha} = \mathbf{H_\rho}^{-1}\widetilde{\boldsymbol{\alpha}},$

$$\mathrm{H}=\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix}, \quad \mathbf{H_\rho} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & -\rho & 1 \end{bmatrix}, \quad \widetilde{\boldsymbol{\alpha}}= \begin{pmatrix} \rho y_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

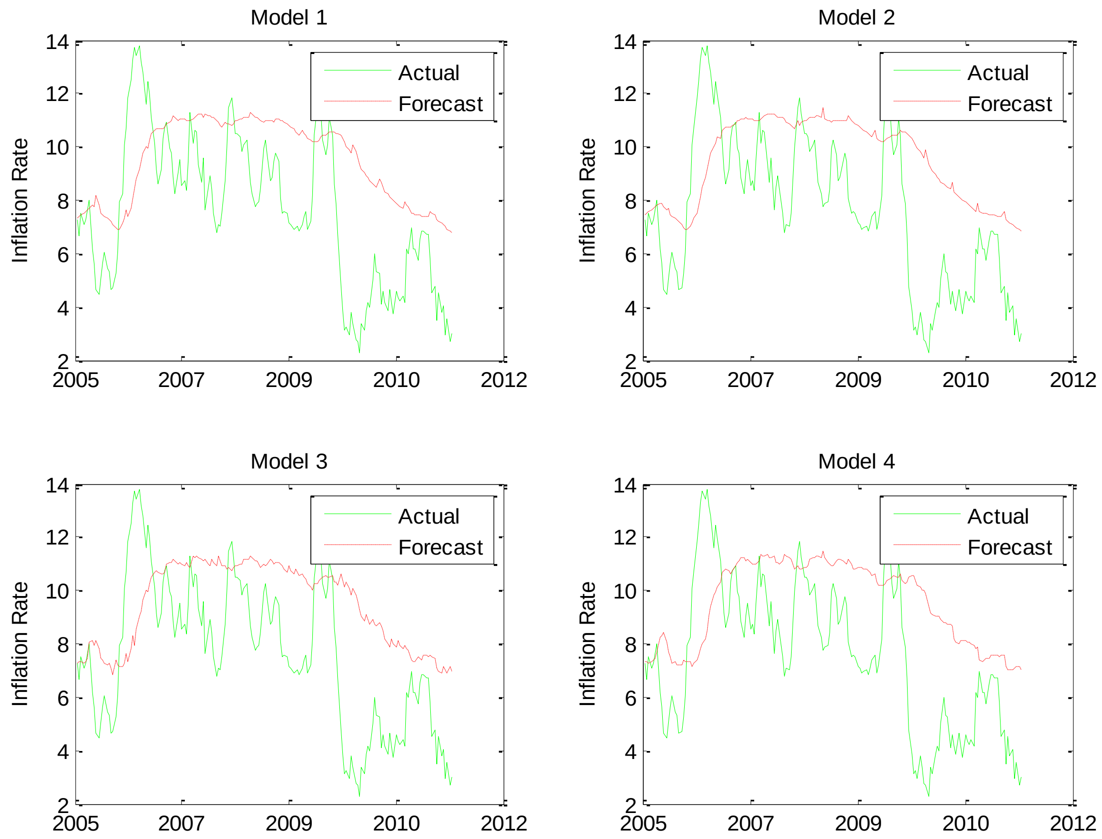
The one-step-ahead point forecast can be obtained as

$\hat{y}_{t+1} = \hat{\tau}_T + \hat{\rho}(y_T - \hat{\tau}_T).$

IV. RESULTS

This section presents the estimation of the different models mentioned in section three. The first set of four models are simple with a time trend and various seasonal dummies. Model (1) has a time trend and a fourth quarter dummy; Model (2) has a time trend and last month dummy; Model (3) has a time trend and four quarterly dummies; finally, Model (4) has a trend and twelve monthly dummies. Figure 3 plots the graphs of the one-step ahead out-of-sample forecasts beginning at for the models. The graphs reveal that all four models perform poorly in terms of mimicking the actual data. Particularly, the models fail to predict the downturns during 2010–2012 and slightly increasing trend upward from 2013 onwards.

Figure 3: Out-of-sample forecasts under four models



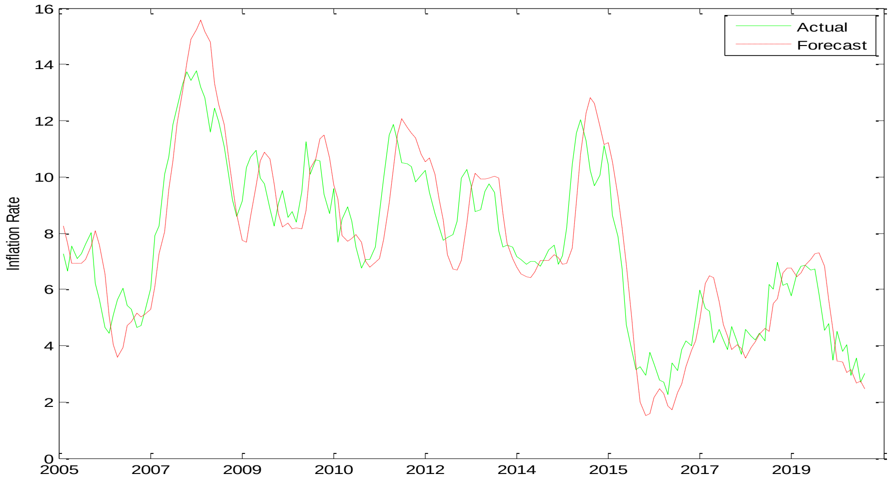
The performance of the models is compared in terms of MSFE for the out-of-sample forecasting. The MSFEs of the models with one-step ahead forecast are given in Table 2.

Table 2: Mean-Square Forecast Errors for different models

Specification	Model 1 (time trend and a fourth quarter dummy)	Model 2 (time trend and twelfth month dummy)	Model 3 (time trend quarterly dummies)	Model 4 (time trend and monthly dummies)
MSFE	9.4536	9.4580	9.5893	10.1821

Second, in an attempt to improve the forecasting model, the Holt-Winters method with seasonality is estimated (Model 5). The optimal smoothing parameters for the trend, seasonal and cyclical components (namely alpha, beta and gamma) are obtained through optimization method minimizing the MSFE. The Holt-Winters with seasonality performs significantly better than the seasonal specifications with MSFE of 1.53 (Figure 4). The out-of-sample forecast closely mimics the actual price. In particular, the method is able to capture the cyclical movements of CPI, specially 2010 onwards.

Figure 4: Out-of-sample forecast with Holt-Winters with Seasonality[6]



Note: The optimal values of the parameters that minimizes the MSFE are alpha=0.4; beta=0.3; gamma=0.1.

To improve the forecasting, three autoregressive models with different error structures are estimated. One of the prerequisites to fit autoregressive models is that the data should be stationary. The Augmented Dickey fuller test with drift and maximum lags of 12 months show that the series is stationary at 5 percent level of significance. Moreover, the correlogram plot also exhibit the autocorrelations that die out fairly quickly as number of lags increases.

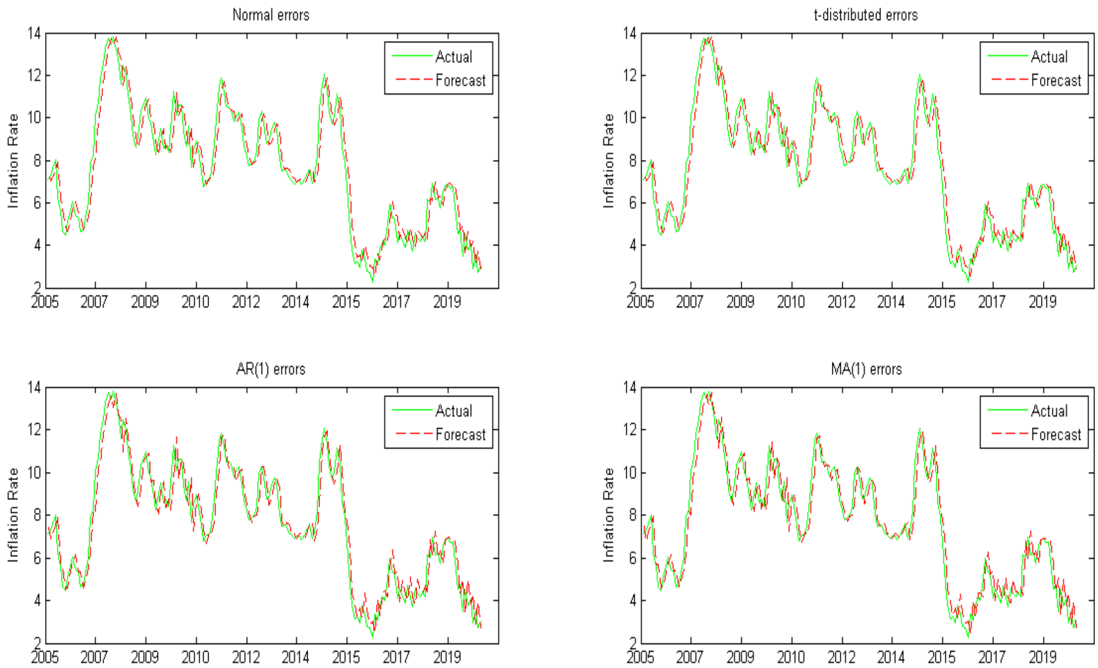
The first specification assumes that the errors follow white noise (Model 6). The maximum likelihood estimates for this model μ, ρ, σ^2 are respectively 0.3949, 0.9489 and 0.6638. Similarly, the maximum likelihood estimates for the model with t-distributed errors (Model 7) μ, ρ, σ^2 are respectively 0.3833, 0.9485 and 0.5381. There is not much difference between these models in terms of forecasting performance which is evident from their plots and similar values of MSFE (Table 3).

The third model assumes that the error follow AR(1) process without drift (Model 8). First, we find the value of ϕ that maximizes the concentrated log-likelihood function $l_c(\phi|\boldsymbol{y}) = l(\phi, \widehat{\boldsymbol{\beta}}_\phi, \hat{\sigma}_\phi^2|\boldsymbol{y})$ with respect to ϕ. The parameters of the models are estimated using the unconstrained optimization procedure as mentioned in Section 3. The maximum likelihood estimates for μ, ρ, σ^2 turns out to be 0.6547, 0.9132 and 0.6153 respectively. The corresponding log-likelihood value is -239.0973. The optimization procedures presents the danger of being trapped into a local maximum depending on the initial values assigned to the log-likelihood function. In order to ascertain the global maximum, the concentrated log-likelihood function is plotted against the possible values of ϕ (Appendix 2).

6 The optimal values of the parameters that minimizes the MSFE are alpha=0.4; beta=0.3; gamma=0.1.

The fourth specification assumes that the errors are MA (1) process (Model 9). Similar to Model 8, we find the value of ψ that maximizes the concentrated log-likelihood defined in Section 3.6. The values of the parameters are $\hat{\psi} = 0.1903$, $\hat{\mu} = 0.4282$, $\hat{\rho} = 0.9405$ and $\hat{\sigma}^2 = 0.6324$. In order to ensure the maximum is the global one, concentrated likelihood graph is plotted for the Models 8 (Appendix 3). The out-of-sample forecasts using AR models with different error structures are given in Figure 5.

Figure 5 : Out-of-sample forecasts using AR models with different error structures



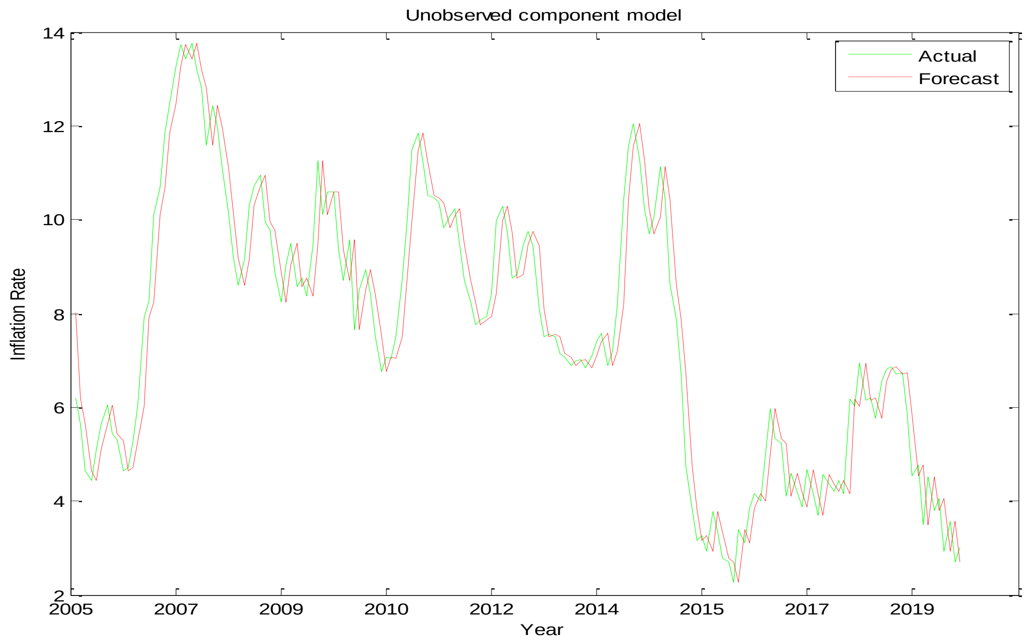
The out-of-sample forecasts of different AR models show that there is a high degree of persistence in the series. As a result, all three AR models forecast the CPI series very well. Out of these, model 8, namely the model with AR(1) errors (without drift) mimics the data well with the lowest MSFE (Table 3).

Table 3: Mean-Square Forecast Errors for different AR models

Specification	Model 6 (with normal errors)	Model 7 (with t-distributed errors)	Model 8 (with AR(1) errors)	Model 9 (with MA(1) errors)
MSFE	0.6302	0.6296	0.6040	0.6162

Finally, further complex model, namely the unobserved component model or the state space form (Model 9) is estimated. The out-of-sample forecast graph of the model is given in Figure 6.

Figure 6: Out-of-sample forecasts using the unobserved component model



The forecast value mimics the actual data very well. The MSFE for the unobserved component model is 0.6331, similar to AR models. Allowing for an additional channel for persistence through $\rho(y_{t-1} - \tau_{t-1})$ does not seem to improve the forecast performance, the MSFE for the unobserved component model is higher than that AR model with AR(1).

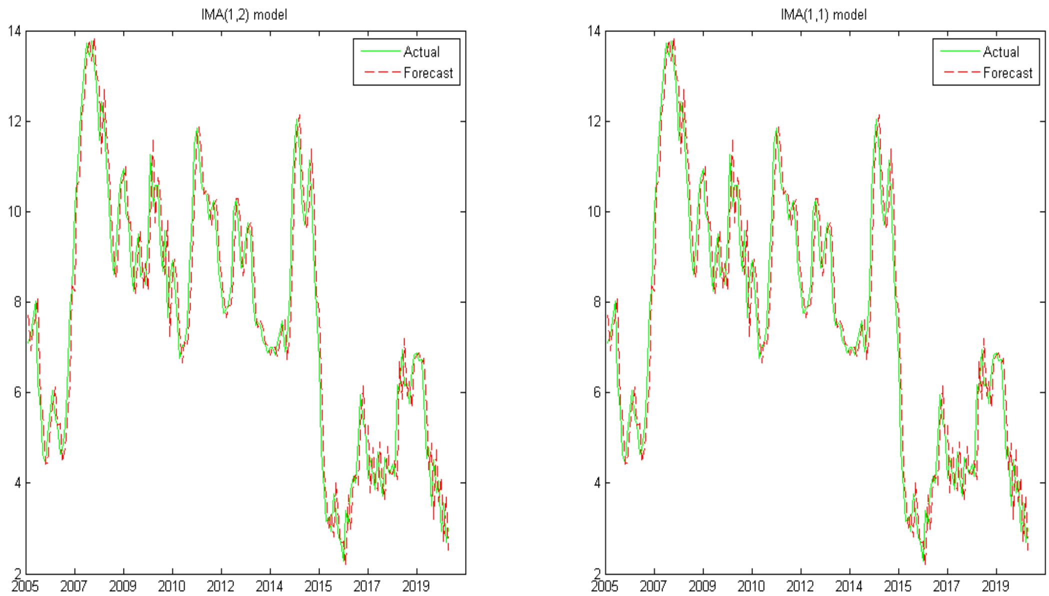
V. SENSITIVITY ANALYSIS

The robustness or the sensitivity of the forecast results is tested using different out-of-sample forecasting period and also forecast horizon. The result remains robust using a shorter out-of-sample period and also using different forecast horizon. Additionally, the forecast performance from the standard ARIMA models are also compared.

In particular, the apparent unit root in the inflation series,7 and the negative first-order autocorrelations, and generally small higher-order autocorrelations, of the inflation suggest that the inflation process might be well described by the IMA(1,1) process or the IMA(1,2). Stock and Watson have shown that the IMA(1,1) process has the best ability to forecast inflation in the U.S. (Stock & Watson, 2007). The MSFE for IMA(1,1) and IMA(1,2) are respectively 0.6089 and 0.5950, which are quite close to the MSFEs of the autoregressive models (Figure 7).

[7] I estimated the standard ARIMA(p,d,q) model of various orders using the Box-Jenkins approach, that is model identification, estimation and diagnostic checking (Box & Jenkins, 1976). Specifically, after inspecting the autocorrelogram and partial autocorrelograms, ARIMA(1,1,1), ARIMA(1,1,2). However, the coefficients are not significant.

Figure 7: The out-of-sample forecast for the IMA models



VI. CONCLUSION

The paper presents competing and progressively complex univariate time series models to forecast inflation in Nepal. The paper finds that there is a high degree persistence of inflation in Nepal implying that it takes longer for the series to return to its mean after a shock. The forecasting performance of the model is tested in terms of Mean Square (Forecast) Errors for the out-of-sample forecast period. The autoregressive model with AR(1) errors performs best compared to more sophisticated models in terms of mean squared error, which also turns out be a parsimonious model. This model is quite general and can reliably be used to forecast inflation in Nepal.

REFERENCES

Box, G.E. and G.M. Jenkins. 1976. “Time Series Analysis: Forecasting and Control.” Hoboken, NJ:Wiley.

Chatfield, C. 2000. Time Series Forecasting. Chapman and Hall/CRC.

Davidson, R. and J.G. MacKinnon. 2003. “Econometric Theory and Methods.” Oxford University Press.

Diebold, F.X. 2004. “Elements of Forecasting.” Thomson South-Western.

Enders, W. 2004. “Applied Econometric Time Series.” New Delhi: Wiley India (P.) Ltd.

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

Montgomery, D.C., Jennings, C.L. and M. Kulahci. 2015. “Time Series Analysis and Forecasting.” New Jersey: John Wiley & Sons.

Nepal Rastra Bank. 2007. “Inflation in Nepal.”

Nepal Rastra Bank. (Various Issues) Quarterly Economic Bulletin.

Stock, J.H. and M.W. Watson. 2007. “Why has U.S. Inflation Become Harder to Predict?” Journal of Money, Credit and Banking. Supplement to Vol. 39 (1).

Appendix 1

Monthly Inflation (year-on-year change in CPI) series

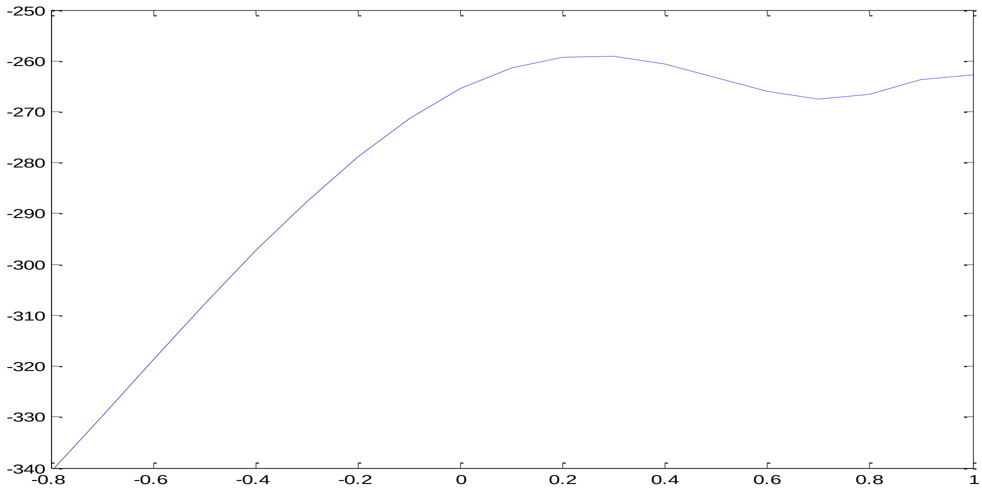
Year / Month	Aug	Sept	Oct	Nov	Dec	Jan	Feb	March	April	May	Jun	July
2002	4.24	3.27	2.98	2.21	2.73	3.27	4.59	5.24	8.14	7.72	6.58	6.09
2003	5.36	5.19	5.65	5.81	4.90	4.96	4.73	4.36	1.72	1.31	1.84	2.02
2004	2.38	2.63	2.61	2.68	3.11	4.59	5.70	5.75	5.84	6.42	6.19	6.65
2005	7.29	8.18	7.82	8.52	8.81	6.96	5.82	7.66	7.91	9.15	9.11	8.27
2006	7.26	6.64	7.54	7.11	7.28	7.62	8.02	6.20	5.63	4.64	4.45	5.09
2007	5.63	6.05	5.43	5.31	4.63	4.73	5.27	6.05	7.92	8.27	10.10	10.69
2008	11.85	12.49	13.28	13.73	13.43	13.77	13.20	12.82	11.58	12.44	11.98	11.09
2009	10.13	9.18	8.61	9.15	10.32	10.72	10.95	9.96	9.77	8.87	8.25	9.03
2010	9.50	8.57	8.76	8.38	9.47	11.26	10.11	10.59	10.58	9.37	8.69	9.58
2011	7.66	8.51	8.92	8.43	7.52	6.77	7.06	7.05	7.51	8.74	9.95	11.48
2012	11.85	11.24	10.51	10.47	10.37	9.82	10.07	10.24	9.46	8.69	8.23	7.75
2013	7.86	7.94	8.41	9.97	10.28	9.72	8.76	8.84	9.47	9.75	9.45	8.08
2014	7.50	7.56	7.50	7.15	7.06	6.88	6.99	7.01	6.84	7.09	7.39	7.58
2015	6.90	7.19	8.20	10.44	11.57	12.05	11.29	10.24	9.70	10.05	11.13	10.44
2016	8.62	7.89	6.73	4.75	3.83	3.15	3.26	2.94	3.77	3.35	2.78	2.71
2017	2.28	3.39	3.10	3.85	4.16	4.00	4.99	5.96	5.32	5.23	4.11	4.58
2018	4.19	3.87	4.68	4.15	3.71	4.57	4.35	4.20	4.43	4.16	6.17	6.02
2019	6.95	6.16	6.20	5.76	6.55	6.82	6.87	6.70	6.74	5.83	4.54	4.78
2020	3.49	4.52	3.79	4.05	2.93	3.56	2.70	3.02				

Note: The year refers to Nepalese fiscal year. For example, 2002 corresponds to the FY 2002/2003.

Source: Quarterly Economic Bulletins of Nepal Rastra Bank.

Appendix 2

Concentrated log-likelihood function $l_c(\phi|y)$ for AR model with AR(1) errors (Model 8)



The graph reveals that the concentrated log-likelihood function has two local maxima. If we use the initial or starting value 0, $\hat{\phi} = 0.3031$, whereas the coefficient is 0.9737 if the starting value is 0.9. We need to determine the global maximum between these values. The log-likelihood values and the parameters associated with these two values of $\hat{\phi}$ are given below:

Two local maxima of $l_c(\phi|y)$

Log-likelihood value	$\hat{\phi}$	$\hat{\mu}$	$\hat{\rho}$	$\hat{\sigma}^2$
-258.9987	0.2578	0.6547	0.9132	0.6153
-261.5367	0.9759	2.5844	0.2309	0.6380

From the table it is evident that the first maximum ($\hat{\phi} = 0.2578$) is the global one.

Appendix 3

Concentrated log-likelihood function $l_c(\phi|y)$ for AR model with MA(1) errors (Model 9)