I am interested in a test that determines that an out of time k-step forecast of an ARX model is a "good" forecast (i.e., residuals are as close to zero as possible).

Let's say we have a $ARX(1)$ model developed on m observations and it has n co-variates.

$$Y_t = \alpha Y_{t-1} + \sum_{i=1}^{n} \beta_i X_{it} + \epsilon_t \tag{1}$$

Reference to Eqn1 goes here (1). The error terms are i.i.d. normal $N(0, \sigma)$ I go on and build a k step forecast over a data period that follows the original m observations (i.e., out of time forecast). I attempt to evaluate residuals recursively.

The first forecasted point is

$$\hat{Y}_1 = \alpha Y_0 + \sum_{i=1}^{n} \beta_i X_{i1}$$

the corresponding residual will be

$$r_1 = Y_1 - \hat{Y}_1 = \epsilon_1$$

which follows that $r_1 \sim N(0, \sigma)$. Moving on to the second forecasted point (Note this uses the forecast from the first step.)

$$\hat{Y}_2 = \alpha \hat{Y}_1 + \sum_{i=1}^{n} \beta_i X_{i1}$$

Since $\hat{Y}_1 = Y_1 - \epsilon_1$

$$r_2 = Y_2 - \hat{Y}_2 = Y2 - \alpha Y_1 - \sum_{i=1}^{n} \beta_i X_{i2} + \alpha \epsilon_1 = \epsilon_2 + \alpha \epsilon_1$$

The distribution of

$$r_2 \sim N(0, \sigma \sqrt{(}1 + \alpha^2))$$

repeating the process we get that

$$r_j = \sum_{k=0}^{j-1} \epsilon_{j-k} \alpha^k$$

and the distribution of

$$r_j \sim N\left((0, \sigma^2 \sqrt{\sum_{k=0}^{j-1} \alpha^k}\right)$$

Now, I am interested in answering a question if the k-step forecast is good and the model I have does not need to be redeveloped. The statistic I am think that would work for a potential test is sum of normalize residual squares.

$$Stat = \sum_{i=1}^{k} \frac{r_i^2}{\sigma^2 \sum_{l=0}^{i-1} \alpha^l}$$

the above Stat will be $\chi^2(k)$. I am thinking I should be able to use this Stat to test a null hypothesis $r_i = 0 \ \forall i = 1...k$ v.s. alternative $r_i \neq 0 \ \forall i = 1...k$. Leads to questions.

1) Is this a worthwhile test to evaluate performance of a model?

2) Have you ever seen anyone do something like this? I am basing this off Chow test, Chow, Gregory C. (1960). "Tests of Equality Between Sets of Coefficients in Two Linear Regressions". Econometrica . The difficulty with applying the test in that paper is that its for OLS and the residuals are all i.i.d. normal, whereas in time series model the k-step forecast residuals have changing distribution.