UvA ECONOMETRICS

Discussion Paper: 2004/04

Predicting Growth and Levels in Loglinear Unit Root Models

Kees Jan van Garderen

www.fee.uva.nl/ke/UvA-Econometrics

Department of Quantitative Economics
Faculty of Economics and Econometrics
Universiteit van Amsterdam
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Predicting Growth and Levels in Loglinear Unit Root Models

Kees Jan van Garderen

Faculty of Economics and Econometrics
University of Amsterdam
Roetersstraat 11,
1018 WB, Amsterdam,
The Netherlands

December 2004

Abstract

This paper considers unbiased prediction of growth and levels when data series are modelled as a random walk with drift after taking logs. The nonlinear transformations involved cause standard symmetry arguments for showing unbiasedness of predictors not to apply. We derive exact small sample unbiased forecasts which are shown to be superior in terms of mean squared forcast error to usual methods for obtaining level forecasts. For growth forecasts there is little to be gained. Parameter estimates are highly correlated with the last observations, regardless of sample size. This causes conceptual problems in terms of conditioning on endogenous variables and we show that no conditionally unbiased estimator exists. On a practical level we show that the correlation is quantitatively more important than uncertainty in parameter estimation and future disturbances together.

JEL *classification*: C20; C53.
Keywords: loglinear unit root models, stochastic growth, unbiasedness, parameter uncertainty.

1 Introduction

Many macro- and other economic series appear to be stationary after applying a log-transformation and taking first differences. In the seminal paper of Nelson and Plosser (1982), for instance, the natural logs of all the data are taken, except for the bond yield, and they argue that with the exception of unemployment all the series could well belong to the difference stationary class. A common modelling strategy found in many parts of empirical economics is therefore to model the variables in log-difference if the hypothesis of a unit root cannot be rejected. In this paper we will assume that taking log-differences is indeed correct and renders the series stationary, but that interest is in predicting the growth and level of the original, untransformed series. This leads to a number of interesting issues not encountered in linear stationary settings, even when abstracting from the complication of testing for a unit root. We allow for exogenous regressors and the aim of the paper is to highlight the issues involved, provide solutions for the problems encountered, and investigate the quantitative importance of the results.

In order to set out the principal issues involved, we present a specific example of the model we have in mind, which is the Cobb-Douglas production function with time varying technology as proposed by Rosanna (1995):

$$\begin{aligned} Y_t &= Z_t L_t^{\alpha_1} K_t^{\alpha_2}, \\ \ln[Z_t] &\equiv z_t = \alpha_0 + \delta\, t + u_t, \\ u_t &= u_{t-1} + \varepsilon_t, \end{aligned}$$

with ε_t i.i.d. The term z_t represents technological progress which is assumed to grow deterministically over time, but also has a stochastic trend component. The model was also used in Garderen, Lee, and Pesaran (2000) in comparing various nonlinear aggregate and disaggregate models based on predictions of the output level Y.

The central issue is predicting one or more step ahead output level Y_{T+h}, and predicting growth $G_{h,T} = 100(Y_{T+h} - Y_T)/Y_T$ based on observations up to and including time T. We focus on unbiased predictors since in a prediction and forecasting setting one often starts out with a square loss function and use a Mean Squared Forecasting Error (MSFE) criterion function for determining, or comparing the quality of predictors. It is easily shown that the conditional mean given the available information at time t minimizes the MSFE. This theoretical result assumes the parameters are known, but if the parameters in this conditional mean function are estimated, the predictions are no longer unbiased.

After taking log-differences the model becomes:

$$\Delta y_t = \delta + \Delta l_t \alpha_1 + \Delta k_t \alpha_2 + \varepsilon_t,$$

where small letters indicate that logs have been taken. The transformed model is easily estimated, forecasting is standard, and inference is straightforward. The inverse transformation (exponentiation) is nonlinear and the nonstationary nature of the log variable gives rise to further complications, mainly in predicting the level Y. We want to highlight the following issues.

First, the current level y_T is highly informative about future levels, but it is also very informative about the parameter values. The nonstationarity causes the estimators associated with any variable trending in the levels to be highly correlated with the current level of the dependent variable. In a weakly dependent situation the influence of y_T on the estimator would be of order $1/T$ and dropping the last observation in a strict *i.i.d.* setting would actually make estimator independent of y_T. In the presence of a unit root this is no longer the case. Current y_T is correlated with all y_t's in the past and the covariance does not go to zero due to the stochastic trend. The covariance between y_T and the estimator is of order 1 and the correlation does not disappear asymptotically. Hence formulas like $\check{Y}_{T+1} = Y_T \exp\{\hat{\delta} + \Delta l_{T+1}\hat{\alpha}_1 + \Delta k_{T+1}\hat{\alpha}_2 + \frac{1}{2}\hat{\sigma}^2\}$, which appear reasonable at first glance and are in common use (either with or without the last term $\frac{1}{2}\hat{\sigma}^2$), are significantly biased and not even consistent predictors of the conditional expectation.

Second, the nonlinearity of the inverse transformation (taking exponentials) causes the expectation to differ from the exponential of the mean. In linear settings unbiasedness of predictors can be proved by symmetry arguments in certain cases, but these arguments do not apply in the presence of nonlinear transformations. Furthermore, the variance of the random walk component of the log-series is increasing linearly over time and increasingly affects the expected value of the levels of the original series. See also Granger and Newbold (1976) who consider forecasting series that are non-linearly transformed, including exponential transformations and consider quadratic transformations for non-stationary variables. They also compare different predictors and the loss involved, but do not consider estimation uncertainty (the optimal forecast is the conditional mean) and no explanatory variables.

Third, and related to both previous issues, is that of parameter uncertainty and the relation between the forecast horizon, h, and the number of observations T. Sampson (1991) analyses, in a standard linear setting, the way in which parameter uncertainty affects the way conditional forecast variances grow as the forecast horizon increases. He shows that parameter uncertainty causes the conditional forecast variance to increase with the square of the forecast horizon (when T increases as a multiple of h) in the unit root setting instead of rate h. He actually shows that the same holds in a trend stationary setting, where the forecast variance is bounded in the absence of parameter uncertainty. See also Clements and Hendry (1999, p111 ff.), stressing the importantance of estimation uncertainty, but show that when T and h are both allowed to increase proportionally, as in Sampson (1991) that the forecast variance of the difference stationary model outgrows that of the trend stationary model, the ratio of the two going to infinity. The exponential transformation only exacerbates the situation and the variance increases even faster. Phillips (1979) also analyzes the role of parameter estimation in forecasting from stable AR(1) models and discusses conditioning on the last observation explicitly.

The sample size itself is an important issue, particularly in macro where short time series are common place and structural breaks might lead to modelling shorter periods of time. Our analysis applies to both large samples and small samples.

The main fundamental issue in the paper concerns conditioning and unbiasedness, or how to define unbiasedness when conditioning on past observations. For prediction puposes we would like to condition on all information available at time T. Conditioning on past observations, however, means that estimators are fixed since they are deterministic functions of the conditioning variables.

Probability statements such as median- or mean unbiasedness are therefore vacuous since the distribution is degenerate: $\hat{\delta}$ is fixed and never equal to δ in the example above, irrespective of the estimator (function of the data) being used. This causes the predicament that expressions are either conditional on past observations and no statement about unbiasedness can be made, or we make unconditional predictions and average over all possible sample paths. Unconditional statements are undesirable because, for example, in a basic random walk model with zero initial value, the next observation will almost by definition be close to the last observation, whereas the unconditional prediction is always zero regardless of how far the process at time T has deviated from zero. Secondly, the unconditional variance of the process is increasing linearly with T, whereas the conditional variance does not depend on T (only on h). The issue becomes more subtle in the case of a unit root model with drift (and other exogenous variables) because of the correlation between the last observation and the estimator. If y_T is larger than the unconditional expectation, then the drift parameter will be overestimated. Using this over-estimate leads to an over-prediction of future y_{T+h} and under-prediction when y_T is small. With the exponential transformation these effects do not average out, since over-prediction of the log variables will lead to a larger contribution to the MSFE than under-estimation.

One possible solution is to condition only on those variables that enter the conditional mean. In the leading example those are the exogenous variables and the endogenous variable Y_T. This was also discussed and analyzed by Phillips (1979) for the AR(1) case with stable parameter values. The appendix gives the relevant conditional distributions for the present case with a unit root and exogenous variables. The problem remains, however, that when a deterministic linear trend is included, the conditional distribution of the parameter estimates is still degenerate. E.g. if only a trend is included then $\hat{\delta} = y_T/T$ and the difficulty is the same as when conditioning on the whole past, namely that $\hat{\delta}$ is fixed.

We show theoretically that when a linear trend is included no conditionally unbiased estimator exists at all. This interesting result holds whether conditioning on the whole past or just the last observation. We derive two exact unbiased estimators, one unconditionally unbiased estimator i.e. by directly solving the unconditional unbiasedness condition, and the second based on a conditional expression using the last observation Y_T, but taking into account the correlation between the estimator and the last observation and requiring unbiasedness unconditionally. It is interesting that, although both predictors are derived from very different perspectives, they are actually identical, as will be proved below.

Finally, a minor comment about assuming it is known that the model is a unit root with drift. This has two important consequences. First, there is no (pre-) testing for unit roots. Proper inference procedures should take the value of the test-statistic and the outcome of the tests into account, and this would obscure the issues we highlight here. Second, the drift parameter is more accurately estimated in the first difference model than in (log) levels. Estimating a trend stationary model results in an estimator of the drift term that has the same root-T order of convergence as the estimate of the constant in the first difference model, but the intercept estimator in the first difference model has a much smaller variance than estimator of the drift in the model where it is jointly estimated with the unit root of the lagged dependent variable. (In fact it has the smallest variance in the class of unbiased estimators since it achieves the Cramer-Rao lowerbound under the model assumptions entertained in this paper). The parameter uncertainty is an important factor in adjusting the predictions and using a less accurate estimator would lead to even larger effects than reported in this paper.

There are of course a number of issues the paper does not deal with, including for instance dynamic forecasts of exogenous variables, as e.g. Schmidt (1978). This additional uncertainty would increase the bias correction term and would require the exact distribution of the vector autoregressive moving average estimators, which is as yet an unsolved problem. Another issue we do not deal with is heteroskedasticity in the transformed model, which is either trivial to deal with if the exact form is known, or analytically very complicated when it involves unknown parameters.

The remainder of the paper is organized as follows. The next section discusses the basic model. Section 3 deals with predicting growth and Section 4 deals with predicting future levels in this model. Section 5 compares our suggestions with common solutions and Section 6 concludes. Proofs are given in the appendices.

2 The Model

Consider the following loglinear unit root model, which includes the Cobb-Douglass production function with stochastic technology in the introduction as a special case:

$$\ln Y_t = x_t'\beta + u_t, \quad t = 1, 2, ..., \tag{1}$$

$$u_t = u_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim I.I.N(0, \sigma^2), \tag{2}$$

where x_t is a $(k \times 1)$ vector of regressors including a trend, β is a $(k \times 1)$ vector of unknown parameters, and ε_t are *i.i.d.* normal with mean zero and variance σ^2. The model implies that for the log-differences

$$\Delta y_t = \Delta x_t'\beta + \varepsilon_t, \tag{3}$$

where $y_t = \ln Y_t$,and we assume that the following conditions hold throughout:

Condition 1 $y_t = \sum_{i=1}^{t} \Delta y_i$ *and* $x_t = \sum_{i=1}^{t} \Delta x_i$

Condition 2 *The matrix* $\Delta X = (\Delta x_1, \Delta x_2, ..., \Delta x_T)'$ *has full column rank.*

Assumption 1 is West's (1988) condition (2.1) and avoids having to track the effects of the initial values on the mean, for instance. One could think of this as having subtracted the initial values from every observation, or simply as the initial values being zero in which case it is just an identity. Implicit in any case is that we condition on the initial value.

Assumption 2 ensures that the OLS estimator in (3) is uniquely defined. The assumption does imply that no constant term included in x_t, but that is related to the fact that the constant cannot be identified from the equation in first differences.

With these assumptions the parameters β and σ^2 are simply estimated using OLS in (3). Predicting Δy_T is straightforward, as is the prediction of the log-variable y_t using Goldberger (1962). Using his result it is easily shown, see Lemma (@@) in the appendix, that the optimal (minimal variance linear unbiased) predictor is given by:

$$\hat{y}^*_{T+1} = y_T + \Delta x'_{T+1}\hat{\beta}. \tag{4}$$

Exponentiation of $\hat{y}_t$ does, however, not lead to optimal predictors for levels and growth of Y_t. The first reason is that the transformation is non-linear and results in a bias. Second, y_T and $\hat{\beta}$ are highly correlated as stated in the following lemma and corollary, and this causes an additional bias term.

Lemma 3 *The covariance between y_T and $\hat{\beta}$. Let $\widehat{\beta}$ be the OLS estimator of the model in log-differences, then:*

$$Cov\left(y_T, \widehat{\beta}\right) = \sigma^2 x'_T \left(\Delta X' \Delta X\right)^{-1}. \tag{5}$$

Corollary 4 *If x_t includes a linear trend or is $I(1)$, such that $(\Delta X'\Delta X) = O_p(T)$ and $x_T = O_p(T)$, then the covariance between the covariance between the last observation and the estimator $\hat{\beta}$ does not go to zero as the sample size increases*

$$Cov\left(y_T, \widehat{\beta}\right) = O(1) \tag{6}$$

In particular if x_t consists of only a linear trend then the correlation between y_T and the drift estimate is 1:

$$Corr\left(y_T, \widehat{\delta}\right) = 1 \tag{7}$$

The problem is that y_T is very informative about the future levels y_{T+h}, while at the same time also containing much information about the parameter. If y_T is high, then the estimate is also high, and if y_T is low than the estimate is low. In a linear setting these effects may cancel out but exponentiation destroys the possible symmetry.

2.1 Predicting Growth

Growth per period in the model does not depend on the level Y. It is a function only of the disturbance term, parameters, and exogenous variables: $G_T = 100(\exp\left\{\Delta x'_{T+1}\beta + \varepsilon_{T+1}\right\} - 1)$ and has expectation

$$E\left[G_T\right] = 100(\exp\left\{\Delta x'_{T+1}\beta + \tfrac{1}{2}\sigma^2\right\} - 1). \tag{8}$$

The term $\frac{1}{2}\sigma^2$ is due to taking the expectation of a nonlinear function of ε_{T+1}. The expected growth can be estimated unbiasedly using a result in Van Garderen (2001) where an unbiased estimate of the variance is also derived. This leads to:

Theorem 5 *The **Exact Minimum Variance Unbiased Predictor** of growth in period* T *is given by*

$$\hat{G}_T = 100(\exp\{\Delta x'_{T+1}\hat{\beta}\}\,{}_0F_1(m, m\hat{\sigma}^2\tfrac{1}{2}(1-a_{T+1})) - 1) \tag{9}$$

where $\hat{\beta}$ *and* $\hat{\sigma}^2$ *are the OLS estimators of the model in first differences, and* $a_{T+1} = \Delta x'_{T+1}(\Delta X'\Delta X)^{-1}\Delta x_{T+1}$.
An unbiased estimator for the variance is given by

$$\widehat{var}(\hat{G}_T) = 100^2\exp\{2\Delta x'_{T+1}\hat{\beta}\}({}_0F_1(m;\hat{\sigma}^2 m\tfrac{1}{2}(1-a_{T+1}))^2 - {}_0F_1(m;m(1-2a_{T+1})\hat{\sigma}^2) \tag{10}$$

The proof essentially uses the result that $E[{}_0F_1(m, m\hat{\sigma}^2 z)] = \exp\{\sigma^2 z\}$, see Van Garderen (2001). We can therefore attribute the term $\frac{1}{2}\hat{\sigma}^2(1-a_{T+1})$ in $\hat{G}_T$ to the uncertainty in ε_{T+1} which leads to the term $\frac{1}{2}\sigma^2$ in the expected growth (8), and a_{T+1} to the uncertainty in the estimate $\hat{\beta}$, since $E[\exp\{\Delta x'_{T+1}\hat{\beta}\}] = \exp\{\Delta x'_{T+1}\beta + \frac{1}{2}\sigma^2 a_{T+1}\}$.

Theorem 1 is easily generalized to multi-period growth and growth at some future date as follows:

(1) predicting growth from period T to period $T+h$ by replacing

(a) $\Delta x'_{T+1}$by $\Delta_h x'_{T+h}$ where, in obvious notation, $\Delta_h x'_t = x'_{T+h} - x'_T$, and

(b) $(1-a_{T+1})$ by $(h - a_{T+h})$, and $2a_{T+1}$ by $2a_{T+h}$, where $a_{T+h} = \Delta_h x'_{T+h}(\Delta X'\Delta X)^{-1}\Delta_h x_{T+h}$.

(2) predicting growth in period $T+h$ (given information at time T) by replacing $\Delta x'_{T+1}$by $\Delta x'_{T+h+1}$.

In practice, other predictors are in use and biased alternatives are available. When T is sufficiently large, it turns out that the exact unbiased estimator and its variance estimator are closely approximated by the Approximate Unbiased predictor given below in the next definition.

A second alternative, the Consistent predictor, takes into account the uncertainty in ε_{T+1} and ignores the uncertainty in the estimation of β and σ^2, but, since this goes to zero as the sample size increases, results in a consistent estimator of the expected growth. The third alternative, the Naive predictor ignores the uncertainty in both ε_{T+1} and the estimation of β, and is biased and inconsistent but it should be noted that the effects work in opposite directions: ignoring the uncertainty in leads to underestimation and ignoring uncertainty in $\hat{\beta}$ leads to overestimation. The naive predictor can in certain circumstances be better in terms of MSFE than the other predictors.

Definition 6 *Alternative predictors*
(A) ***Approximate Unbiased Predictor*** *of Growth and its Variance:*

$$\tilde{G}_T = 100(\exp\{\Delta x'_{T+1}\hat{\beta} + \tfrac{1}{2}\hat{\sigma}^2(1-a_{T+1})\} - 1), \tag{11}$$

$$\widehat{var}(\tilde{G}_T) = 100^2\exp\{2\Delta x'_{T+1}\hat{\beta}\}(\exp\{\hat{\sigma}^2(1-a_{T+1})\} - \exp\{\hat{\sigma}^2(1-2a_{T+1})\}) \tag{12}$$

(B) ***Consistent Predictor****:*

$$\check{G}_T = 100(\exp\{\Delta x'_{T+1}\hat{\beta} + \tfrac{1}{2}\hat{\sigma}^2\} - 1). \tag{13}$$

(C) ***Naive Predictor***

$$\check{G}_T = 100(\exp\{\Delta x'_{T+1}\hat{\beta}\} - 1)$$

The approximate unbiased estimators appear much easier to calculate than the exact ones, but it should be noted that many packages, such as Mathematica, include the hypergeometric function as standard. There are exact expressions and estimators available for the variance of the consistent and naive predictors, but when using a biased estimator it is not clear why one would use an unbiased estimator of the variance or MSFE.

2.2 Predicting Levels

Predicting growth is essentially straightforward, and is only complicated by the nonlinear function involved as we have just seen. Predicting future levels is more complicated because future levels are not simply a function of parameters but also depend on the current level. This current level is highly correlated with the parameter estimates, and ignoring this dependence leads to significant bias. For this reason the obvious estimator based on the unbiased growth predictor:

$$\hat{Y}_{T+1} = Y_T(\hat{G}_T/100 - 1)$$

is not an unbiased predictor of the level. When the consistent growth estimator is used this would lead to

$$\check{Y}_{T+1} = Y_T \exp\{\Delta x'_{T+1}\hat{\beta} + \tfrac{1}{2}\hat{\sigma}^2\} \tag{14}$$

which makes it clear that the correlation between current level Y_T and $\hat{\beta}$ will cause problems. This predictor will in fact be inconsistent if x_T includes a trending variable.

Level predictors based on growth estimates are generally not unbiased for three reasons. First, the nonlinear exponential transformation, second, the parameter uncertainty, and third, the fact that Y_T and the estimator $\hat{\beta}$ are highly correlated. We will therefore have to bias-correct such predictors or consider different types of estimators, for example based directly on the expectation of Y_{T+h}.

The unconditional expectation of period $(T+h)$'s level Y_{T+h} and the conditional expectation Y_{T+h} given the current level Y_T are respectively:

$$E[Y_{T+h}] = \exp\{x'_{T+1}\beta + \tfrac{T+h}{2}\sigma^2\}, \tag{15}$$

$$E[Y_{T+h}|Y_T] = Y_T \exp\{\Delta_h x'_{T+h}\beta + \tfrac{h}{2}\sigma^2\}. \tag{16}$$

It seems therefore that there are two different ways of constructing an unbiased predictor for the level Y_T. The first is to note that the unconditional expectation of Y_{T+h} depends only on parameters and to estimate this unbiasedly using Van Garderen (2001).

The second is to estimate the conditional expectation of Y_{T+h} given Y_T and adjust for the bias caused by the fact that $\hat{\beta}$ and Y_T are correlated, to obtain a predictor that is unbiased. This leads to two predictors which are both unbiased:

Proposition 7 *The unconditional level predictor*

$$F_{T+h} = \exp\{x'_{T+h}\hat{\beta}\}\, {}_0F_1(m; m\hat{\sigma}^2\, z_{T+h})$$

with

$$z_{T+h} = \tfrac{1}{2}(T + h - x'_{T+h}\,(\Delta X'\Delta X)^{-1}\, x_{T+h}),$$

is unbiased.

Proposition 8 *The conditional level predictor*

$$F_{T+h|T} = Y_T \exp\{\Delta_h x'_{T+h}\hat{\beta}\}\, {}_0F_1(m; m\hat{\sigma}^2 z_{T+h|T})$$

with

$$z_{T+h|T} = \tfrac{1}{2}(h - 2x'_T\,(\Delta X'\Delta X)^{-1}\,\Delta_h x'_{T+h} - \Delta_h x'_{T+h}(\Delta X'\Delta X)^{-1}\Delta_h x_{T+h}),$$

is unbiased

The terms in z_{T+h} and $z_{T+h|T}$ are easily attributed to sources of uncertainty. In z_{T+h}, the term $(T+h)$ can be attributed to the total number of disturbances up to and including period $T + h$, and the second term is correcting for the parameter uncertainty when estimating $x'_{T+h}\beta$. In $z_{T+h|T}$: the term h derives from the h disturbance terms ε_t in the future between time T and $T + h$, the second term derives from the covariance between y_T and $\hat{\beta}$, and the third term $x'_{T+h}\,(\Delta X'\Delta X)^{-1}\, x_{T+h}$ is correcting for parameter uncertainty in the estimate $\Delta_h x'_{T+h}\hat{\beta}$.

Although the predictors look principally different and are derived from very different perspectives, they are very closely related. They are in fact identical if a drift term is included as we prove in the appendix and state here as:

Proposition 9 *If the model includes a deterministic trend, such that ΔX includes a constant term, then*

$$F_{T+h} = F_{T+h|T}$$

This paradoxical equality between the conditional and unconditional predictors can be explained in the two different ways.

The first explanation is by noting that the unconditional predictor F_{T+h} equals Y_T for the limiting case $h = 0$.[1] So, although we are averaging over all possible sample paths for $\{Y_t\}$, each prediction based on any realized sample path still goes through (when varying h) the last observation Y_T. The only difference is that the conditional predictor $F_{T+h|T}$ is an explicit function of Y_T, but the unconditional predictor only depends implicitly on Y_T, but it behaves exactly the same.

The second explanation is via the unbiasedness condition imposed. Expectations can only be evaluated by integrating over all possible sample paths, and hence by varying Y_T, since conditioning on all past observations would leave the estimators fixed: e.g. $\hat{\beta} - \beta$ would be a fixed number different from 0 (with probability 1). So the unbiasedness condition is essentially an unconditional

[1] $x'_{T+0}\hat{\beta} = \imath'\Delta X(\Delta X'\Delta X)^{-1}\Delta X'\Delta y = \imath'\Delta y = y_T$ and hence $exp\{x'_T\hat{\beta}\} = Y_T$. $z_T = \frac{1}{2}(T - \imath'\Delta X(\Delta X'\Delta X)^{-1}\Delta X'\imath)) = 0$, and ${}_0F_1(m, 0) = 1$ and hence $F_{T+0} = Y_T$

requirement. The model density is complete with complete sufficient statistics $(\hat{\beta}, \hat{\sigma}^2)$, (see e.g. Lehmann and Casella (1998, p.42)). This means that there is a unique function of $(\hat{\beta}, \hat{\sigma}^2)$ which has expectation equal to a function of (β, σ^2) for all parameter values (see Appendix) and $F_{T+h|T}$ must equal F_{T+h}.

It seems undesirable to average over all possible sample paths that Y can take. Given that the process goes through Y_T we should want to condition on the fact that, at time T, the process goes through Y_T. The appeal of the first explanation is that the predictions do in fact go through Y_T, although we do not condition on this outcome.

An alternative approach is to condition only on the terms that enter the conditional expectation, in this case Y_T. In this approach we consider the conditional distribution given only Y_T and do not condition on previous values $\{Y_1, ..., Y_{T-1}\}$. The problem is, however, that the conditional distribution of $\hat{\beta}$ is still degenerate given only Y_T. For example, in the log-difference model with only a constant term, the estimated drift parameter is simply Y_T/T and hence a deterministic function of Y_T and it is impossible to find for a predictor that is conditionally unbiased given Y_T.

Proposition 10 *If the model includes a deterministic trend, such that ΔX includes a constant term, then no conditionally unbiased predictor of Y_{T+h} exists given either (a) $\{Y_1, ..., Y_T\}$ or (b) $\{Y_T\}$.*

The proof is given in the Appendix, but (a) is obvious since the estimates are constants given $\{Y_1, ..., Y_T\}$ and not equal to the conditional expectation of Y_{T+h} with probability 1. Part (b) also follows from a degeneracy in the conditional distribution of $\hat{\beta}$.

Other predictors for Y_{T+h} are of course available and, although they will be biased, they need not necessarily be worse in terms of MSFE. Alternatives include:

Definition 11 *Alternative predictors:*
(A) ***Approximate unbiased predictor:***

$$F_{T+h}^{au} = Y_T \exp\{\Delta_h x_{T+h}' \hat{\beta} + \hat{\sigma}^2 z_{T+h|T}\}, \tag{17}$$

with

$$z_{T+h|T} = \tfrac{1}{2}(h - 2x_T' (\Delta X' \Delta X)^{-1} \Delta_h x_{T+h}' - \Delta_h x_{T+h}' (\Delta X' \Delta X)^{-1} \Delta_h x_{T+h}).$$

(B) ***Consistent predictor:***

$$F_{T+h}^{c} = \exp\{x_{T+h}' \hat{\beta} + \frac{T+h}{2} \hat{\sigma}^2\} \tag{18}$$

(C) ***Naive predictor:***

$$F_{T+h}^{naiv} = Y_T \exp\{\Delta_h x_{T+h}' \hat{\beta}\} \tag{19}$$

The consistent predictor is a consistent estimator of the (unconditional) mean of Y_{T+h}. It was used in Garderen, Lee, and Pesaran (2000) in their prediction based criterion for deciding between aggregate and disaggregate nonlinear models. This choice turns out not to be optimal as we will see below. The approximate unbiased predictor is constructed by approximating the exact unbiased predictor, i.e. using $z_{T+h|T}$ which corrects for uncertainty in Y_{T+h}, $\hat{\beta}$ and the correlation between Y_T and $\hat{\beta}$. The naive predictor ignores all the uncertainty and, similar to Proposition (@@), if a time trend is included in x_t then the conditional $F_{T+h}^{naiv} = \exp\{x'_{T+h}\hat{\beta}\}$, which can be regarded as the unconditional version.

3 Comparing the Predictors: Sectoral Production Forecasts

In this section we make a numerical comparison of the various predictors of growth and levels. This serves not only as an illustration but is motivated by two reasons. First, the unbiasedness was partly motivated by the choice of MSFE as optimality criterion. Although the conditional mean minimizes the MSFE when the parameters are known, this is not necessarily true if the parameters are estimated, or if an exact unbiased estimator of the mean is found, that the resulting predictors are best in terms of MSFE. Second, although there may be important theoretical differences between the various predictors, it might be that the differences in practice are not important.

In order to show the difference we have applied the different predictors to the same data set as in Garderen, Lee, and Pesaran (2000). The models include general and sector specific productivity dummies for oil price shocks, major strikes, etc. For a full explanation see the original article where also various specification tests are reported including tests for normality and functional form which do not reject the model. We then simulated the model with the estimated parameters. For observations after 1995 the regressors where dynamically generated and Y was then generated according to the model.

We also use a simulated a log-linear unit root model with a trend, a random walk with drift, and a stationary variable as explanatory variables with slightly different parameter values similar to estimates based on the data in Nelson and Plosser (1982) or Sampson (1991).

Results on growth.

The first two tables report the average actual and predicted values based on the various predictors, the bias, and the MSFE .The individual sectoral results are reported in the appendix. Note that the average value of the drift term, and the average value of the standard deviation over the eight sectors are both very small.

$T = 25$ 1956-1980	*Actual Growth*	*Exact Unbiased*	*Approximate unbiased*	*Consistent*	*Naive*
Mean	11.39	11.38	11.38	11.61	11.38
Bias		-0.01	-0.01	0.21	-0.01
MSFE		101.02	101.02	101.51	101.2

$T=40$ 1956-1995	*Actual Growth*	*Exact Unbiased*	*Approximate unbiased*	*Consistent*	*Naive*
Mean	22.89	22.87	22.87	23.21	22.42
Bias		-0.02	-0.02	0.26	-0.47
MSFE		236.1	236.1	236.6	235.9

Number of replications:100.000, $h=5$, $\bar{\delta}=0.014$ $\bar{\beta}_{\ln L}=0.241$,$\bar{\beta}_{\ln K}=0.430$,$\bar{\sigma}=0.027$

Number of replications:100.000, $h=10$, $\bar{\delta}=0.021$ $\bar{\beta}_{\ln L}=0.296$,$\bar{\beta}_{\ln K}=0.173$,$\bar{\sigma}=0.035$

The results show very little difference between the various predictors for growth. There is very little to choose between them in terms of bias or in terms of MSFE. This is not unexpected because the approximate relation between growth and logs cause the loglinear model to give a reasonable answers when growth is not too large.

The following tables give the results for predicting growth over 4 periods using parameter values similar to those of Sampson (1991) or based on the data set used by Nelson and Plosser (1982). The MSFE is based on the actual difference between the realized growth and the predictions.

$T=25$	*Actual Growth*	*Exact Unbiased*	*Approximate unbiased*	*Consistent*	*Naive*
Mean	45.84	46.05	46.05	45.37	46.18
Bias		0.2086	0.2088	-0.4706	0.3406
MSFE		271.1	271.1	270.9	271.2

$T=50$	*Actual Growth*	*Exact Unbiased*	*Approximate unbiased*	*Consistent*	*Naive*
Mean	42.22	42.42	42.42	41.77	42.56
Bias		0.1914	0.1915	-0.4591	0.3297
MSFE		262.9	262.9	262.7	263.0

Number of replications:10.000, $\beta=\{\ 0.04, 0.6, 0.2\}$, $\sigma=0.05267$. $h=4$

Results on levels

The following tables give the results for the level forecasts of sectorial production based on the same model as for growth. The average results over all 8 sectors are reported and the detailed results per sector are given in the appendix. The conditional unbiased predictor and unconditional unbiased predictor are identical because a time trend is included in the model, and results are given in column 3. For the growth based estimator $Y_T\hat{G}_{unb}$ the exact unbiased growth predictor is used. This estimator ignores the correlation between Y_T and the estimator. The last row gives the percentage difference in MSFE from the minimum MSFE.

$T=25$ 1956-1980	*Actual* Y_{T+h}	*Exact Unbiased*	*Approximate unbiased*	$Y_T\hat{G}_{unb}$	*Naive*	*Consistent*
Mean	2.629	2.629	2.629	2.641	2.640	2.646
Bias		0.000	0.000	0.012	0.011	0.017
MSFE		0.095	0.095	0.097	0.097	0.098
%diff from minimum		0%	0.0%	1.0%	0.98%	1.6%

Number of replications:100.000, $h = 5$, $\bar{\delta} = 0.014$ $\bar{\beta}_{\ln L} = 0.241$,$\bar{\beta}_{\ln K} = 0.430$,$\bar{\sigma} = 0.027$

$T = 40$ 1956-1995	*Actual* Y_{T+h}	*Exact Unbiased*	*Approximate unbiased*	$Y_T\hat{G}_{unb}$	*Naive*	*Consistent*
Mean	4.638	4.636	4.636	4.682	4.670	4.693
Bias		−0.001	−0.001	0.032	0.055	0.044
MSFE		0.476	0.476	0.490	0.487	0.496
%diff from minimum		0	0.001%	3.5%	2.2%	4.4%

Number of replications:100.000, $h = 10$, $\bar{\delta} = 0.014$ $\bar{\beta}_{\ln L} = 0.241$,$\bar{\beta}_{\ln K} = 0.430$,$\bar{\sigma} = 0.027$

The small differences between the prediction methods is due to small values of the variance and the non-linearity of the transformation is therefore less important, and secondly the small value of the drift term. The following tables show the results from simulations using parameter values comparable with estimates based on the data in Nelson and Plosser (1982) or Sampson (1991).

$T = 25$	*Actual* Y_{T+h}	*Exact Unbiased*	*Approximate unbiased*	$Y_T\hat{G}_{unb}$	*Naive*	*Consistent*
Mean	16.82	16.85	16.74	17.04	16.96	17.05
Bias		0.02	−0.08	0.21	0.13	0.23
MSFE		3.973	3.929	4.115	4.045	4.131
%diff from minimum		1.1%	0	4.7%	2.9%	5.1%

$T = 50$	*Actual* Y_{T+h}	*Exact Unbiased*	*Approximate unbiased*	$Y_T\hat{G}_{unb}$	*Naive*	*Consistent*
Mean	289.4	289.6	287.8	292.9	291.5	293.2
Bias		0.2	−1.6	3.5	2.1	3.8
MSFE		1273.	1260.	1314.	1294.	1319.
%diff from minimum		0.1%	0	4.2%	2.6%	4.6%

$T = 100$	*Actual* Y_{T+h}	*Exact Unbiased*	*Approximate unbiased*	$Y_T\hat{G}_{unb}$	*Naive*	*Consistent*
Mean	3.337e4	3.335e4	3.316e4	3.372e4	3.354e4	3.373e4
Bias		−21.80	−213.7	349.4	171.2	357.4
MSFE		1.907e7	1.895e7	1.960e7	1.930e7	1.962e7
%diff from minimum		0.6%	0	3.4%	1.8%	3.5%

Number of replications:10.000, $\beta = \{\ 0.04, 0.6, 0.2\}$, $\sigma = 0.05267$. $h = 4$

The results show that although the approximate unbiased predictor has some bias, it actually has lower MSFE than the exact unbiased predictor due to its lower variance. The difference in MSFE is actually not more than 0.6%. The difference with predictors that ignore the correlation between Y_T and $\hat{\beta}$ like the growth based predictors and the consistent predictor is considerably more. There is an appreciable difference of up to 5.1%. The naive predictor is performing surprisingly better. It ignores the randomness in ε, which increases the expectation, and ignores the randomness in the estimator and correlation between Y_T and $\hat{\beta}$ which increase the expectation of the prediction. These two effects partly cancel. This can also be seen from the $z_{T+h|T}$ term which is used in the unbiased and approximate unbiased estimators. The numeric values for the constituent parts of $z_{T+h|T}$ when $T = 50$ are for the uncertainty in $\varepsilon : h/2 = 2$, for the parameter uncertainty $-\Delta_h x'_{T+h}(\Delta X'\Delta X)^{-1}\Delta_h x_{T+h}/2 = -0.3$, and for the correlation between Y_T and $\hat{\beta}$ $-2x'_T\,(\Delta X'\Delta X)^{-1}\,\Delta_h x'_{T+h}/2 = -4$, giving a value of approximately $z_{T+h|T} = -2.3$ as cumulative term for correction of the bias due to uncertainty and the nonlinear transformation. The correlation term is the largest term and, together with the term $h/2$, does not decrease as the number of observations T increases. Only the parameter uncertainty term goes to zero.

4 Conclusion

This paper shows that intuitive predictors for the level of macro- and other economic series can be seriously biased in a loglinear unit root model. Not only the nonlinearity of the transforms causes bias, but more importantly, there is high correlation between the last observation and parameter estimates and this correlation persists even asymptotically. We derive exact unbiased predictors for growth and for levels that address both issues. We show that the two exact predictors, one based on the conditional expectation but correcting for covariance between the last observation and the estimator, and the other based on the unconditional expectation, are identical if a linear trend is included in the model. We have shown that a conditionally unbiased predictor does not exist.We also derive a simple approximation to the exact unbiased estimator that is for parameter values used here virtually indistiguishable from the exact unbiased predictor.

The results showed that there is little difference in the various predictors for growth, but for predicting the levels our exact unbiased predictor can be significantly more accurate. The paper has investigated the effect of parameter estimation and one of the contributions of the paper is to highlight the correlation between the level of the last observation and parameter estimates. It is shown that this correlation is far more important than parameter uncertainty. It is even more important than the forecast uncertainty itself in the sense that the contribution to the mean correcting term $z_{T+h|T}$, is much larger.

5 Appendix

5.1 Notation

Given initial values $y_0 = 0$ and $x_0 = 0$

$$\begin{aligned} y_t &= x'_t\beta + u_t, \qquad t = 1, 2, ..., T \\ u_t &= u_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim IIN(0, \sigma^2) \end{aligned} \tag{20}$$

Let $\imath_T = (1, ..., 1)'$, a $T \times 1$ vector of ones, and let L be the first differencing matrix,

$$L = \begin{pmatrix} 1 & & & 0 \\ -1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & -1 & 1 \end{pmatrix}$$

and $y_{(t)} = (y_1, y_2, ..., y_t)'$, i.e. the $t \times 1$ column vector of observations from 1 up to t and $X_{(t)}$ the associated $t \times k$ matrix of regressors. The full sample quantities are written as y and X. Then $Ly = LX\beta + Lu$, and we have

$$y \sim N(X\beta, \sigma^2 L^{-1} L'^{-1})$$

$$\sigma^2 L^{-1} L'^{-1} = \sigma^2 \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & \cdots & T \end{pmatrix}$$

$$\begin{aligned} y_t &= \textstyle\sum_{s=1}^{t} \Delta x_s' \beta + \sum_{s=1}^{t} \varepsilon_s, \qquad (21) \\ &= x_t' \beta + S_t, \\ S_t &= \textstyle\sum_{s=1}^{t} \varepsilon_s = \imath_t' \varepsilon_{(t)}. \end{aligned}$$

5.2 Proofs

Covariance between y_T and $\hat{\beta}$. Let $\widehat{\beta}$ be the OLS estimator of the model in log-differences, then:

$\mathrm{Cov}\left(y_T, \widehat{\beta}\right) = \sigma^2 x_T' \left(\Delta X' \Delta X\right)^{-1}.$

Lemma 12 ***Proof.***

$$\begin{aligned} \widehat{\beta} &= \left(\Delta X' \Delta X\right)^{-1} \Delta X' \Delta y \\ &= \beta + \left(\Delta X' \Delta X\right)^{-1} \Delta X' \varepsilon \sim N\left(\beta, \sigma^2 \left(\Delta X' \Delta X\right)^{-1}\right) \\ y_T &= \sum_{s=1}^{T} \Delta x_s' \beta + \sum_{s=1}^{T} \varepsilon_s \\ &= \underbrace{x_T' \beta}_{E[y_T]} + \underbrace{S_T}_{\imath_T' \varepsilon} \end{aligned}$$

Hence

$$\begin{aligned} Cov\left(y_T, \widehat{\beta}\right) &= E\left[S_T \left(\left(\Delta X' \Delta X\right)^{-1} \Delta X' \varepsilon\right)'\right] \\ &= E\left[\imath_T' \varepsilon \varepsilon' \Delta X' \left(\Delta X' \Delta X\right)^{-1}\right] \\ &= \sigma^2 \imath_T' \Delta X' \left(\Delta X' \Delta X\right)^{-1} \\ &= \sigma^2 x_T' \left(\Delta X' \Delta X\right)^{-1} \end{aligned}$$

■

Covariance between y_T and $\Delta x_{T+1}' \widehat{\beta}$

$$Cov\left(y_T, \Delta x_{T+1}' \widehat{\beta}\right) = \sigma^2 x_T' \left(\Delta X' \Delta X\right)^{-1} \Delta x_{T+1}$$

Proof. of $F_{T+h} = F_{T+h|T}$ if a time trend is included. According to Assumption 1 we have $x_T = \imath_T' \Delta X$. Now let $P_{\Delta X} = \Delta X \left(\Delta X' \Delta X\right)^{-1} \Delta X'$ be the projection matrix onto the column space of ΔX, then, $\imath_T' \Delta X \left(\Delta X' \Delta X\right)^{-1} \Delta X' \imath_T =$

$\imath_T' P_{\Delta X} \imath_T = \imath_T' \imath_T = T$ since ΔX includes a column of ones (i.e.$\imath_T$) associated with the constant term (the drift term).

$$\begin{aligned}
z_{T+h} &= \frac{1}{2}\left(T + h - \left(x_{T+h} - x_T + \imath_T' \Delta X\right)' \left(\Delta X' \Delta X\right)^{-1} \left(x_{T+h} - x_T + \imath_T' \Delta X\right)\right), \\
&= \frac{1}{2}\left(T + h - \left(x_{T+h} - x_T\right) \left(\Delta X' \Delta X\right)^{-1} \left(x_{T+h} - x_T\right) - 2\imath_T' \Delta X \left(\Delta X' \Delta X\right)^{-1} \left(x_{T+h} - x_T\right) - \imath_T' \Delta X \left(\right.\right. \\
&= \frac{1}{2}\left(h - \left(x_{T+h} - x_T\right) \left(\Delta X' \Delta X\right)^{-1} \left(x_{T+h} - x_T\right) - 2\imath_T' \Delta X \left(\Delta X' \Delta X\right)^{-1} \left(x_{T+h} - x_T\right)\right), \\
&= z_{T+h|T}.
\end{aligned}$$

Remains to show that $x_{t+h}' \hat{\beta} = \ln Y_T + \Delta_h x_{T+h}' \hat{\beta}$. By Assumption 1 $x_T' = \imath_T' \Delta X$ and by definition of Δ_h we have $x_{T+h}' = x_T' + \Delta_h x_{T+h}'$ and hence

$$\begin{aligned}
x_{T+h}' \hat{\beta} &= x_T' \hat{\beta} + \Delta_h x_{T+h}' \hat{\beta}, && (22) \\
&= \imath_T' \Delta X \ \left(\Delta X' \Delta X\right)^{-1} \Delta X' \Delta y + \Delta_h x_{T+h}' \hat{\beta}, && (23) \\
&= y_T + \Delta_h x_{T+h}' \hat{\beta}, && (24)
\end{aligned}$$

since $\imath_T' \Delta X \left(\Delta X' \Delta X\right)^{-1} \Delta X' = \imath_T'$. ■

Proof. Of Lemma @@.

$$\begin{aligned}
E[F_{T+h}] &= E_{\hat{\sigma}^2} E_{\hat{\beta}|\hat{\sigma}^2} \left[\exp\{x_{T+h}' \hat{\beta}\} \, {}_0F_1(m; \tfrac{1}{2} m \hat{\sigma}^2 \, z_{T+h}) | \hat{\sigma}^2\right], \\
&= E_{\hat{\sigma}^2} \left[\exp\{x_{T+h}' \beta + \frac{\sigma^2}{2} x_{T+h}' \left(\Delta X' \Delta X\right)^{-1} x_{T+h}\} \, {}_0F_1(m; \tfrac{1}{2} m \hat{\sigma}^2 \, z_{T+h})\right], \\
&= \exp\{x_{T+h}' \beta + \frac{\sigma^2}{2} x_{T+h}' \left(\Delta X' \Delta X\right)^{-1} x_{T+h}\} \exp\left\{\frac{\sigma^2}{2}(T + h - x_{T+h}' \left(\Delta X' \Delta X\right)^{-1} x_{T+h})\right\}, \\
&= \exp\left\{x_{T+h}' \beta + \frac{T+h}{2} \sigma^2\right\}, \\
&= E\left[Y_{T+h}\right],
\end{aligned}$$

Hence F_{T+h} is unconditionally unbiased. ■

Proof. The conditional expectation, given Y_T, of the conditional predictor can be derived using the results on conditional distributions given in the next subsection of this appendix. We obtain:

$$\begin{aligned}
E[F_{T+h|T} | Y_T] &= E_{\hat{\sigma}^2|Y_T}[E_{\hat{\beta}|\hat{\sigma}^2, Y_T} \left[Y_T \exp\{\Delta_h x_{T+h}' \hat{\beta}\} \, {}_0F_1(m; m\hat{\sigma}^2 z_{T+h|T}) | \hat{\sigma}^2, Y_T\right] | Y_T], \\
&= Y_T \ E_{\hat{\sigma}^2|Y_T}[\exp\{\Delta_h x_{T+h}'(\beta + \left(\Delta X' \Delta X\right)^{-1} x_T \frac{(y_T - x_T' \beta)}{T}) \\
&\quad + \tfrac{1}{2} \sigma^2 \Delta_h x_{T+h}' \left(\left(\Delta X' \Delta X\right)^{-1} \Delta X' M_\imath \Delta X \left(\Delta X' \Delta X\right)^{-1}\right) \Delta_h x_{T+h}\} \\
&\quad \times {}_0F_1(m; m\hat{\sigma}^2 z_{T+h|T}) \mid Y_T] \\
&= Y_T \ \exp\{\Delta_h x_{T+h}' \beta + \Delta_h x_{T+h}' \left(\Delta X' \Delta X\right)^{-1} x_T \frac{(y_T - x_T' \beta)}{T} + \\
&\quad + \tfrac{1}{2} \sigma^2 \Delta_h x_{T+h}' \left(\Delta X' \Delta X\right)^{-1} \Delta_h x_{T+h} \\
&\quad - \frac{1}{2T} \sigma^2 \Delta_h x_{T+h}' \left(\Delta X' \Delta X\right)^{-1} x_T x_T{}' \left(\Delta X' \Delta X\right)^{-1} \Delta_h x_{T+h} + \\
&\quad + \sigma^2 \tfrac{1}{2} (h - 2x_T' \left(\Delta X' \Delta X\right)^{-1} \Delta_h x_{T+h}' - \Delta_h x_{T+h}' (\Delta X' \Delta X)^{-1} \Delta_h x_{T+h})\}
\end{aligned}$$

$$
\begin{aligned}
&= Y_T \ \exp\{\ \Delta_h x'_{T+h}\beta + \sigma^2 \tfrac{h}{2} + \Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} x_T \frac{(y_T - x'_T\beta)}{T} + \\
&\quad -\frac{1}{2T}\sigma^2 \Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} x_T x'_T \left(\Delta X' \Delta X\right)^{-1} \Delta_h x_{T+h} - \sigma^2 x'_T \left(\Delta X' \Delta X\right)^{-1} \Delta_h x'_{T+h} \ \}
\end{aligned}
$$

Note that this expectation is not equal to the conditional expectation $E[Y_{T+h}|Y_Y]$.
Now, using the fact that $(y_T - x'_T\beta) \sim N(0, T\sigma^2)$ it follows that

$$
E[\exp\{\Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} \frac{x_T}{T}(y_T - x'_T\beta)\} = \exp\{\frac{1}{2}\frac{T\sigma^2}{T^2}\Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} x_T x'_T \left(\Delta X' \Delta X\right)^{-1} \Delta_h x_{T+h}
$$

and

$$
\begin{aligned}
E[Y_T] &= E[\exp\{y_T\}] \\
&= \exp\{x'_T\beta + \frac{T}{2}\sigma^2\}
\end{aligned}
$$

Hence

$$
\begin{aligned}
&E[\exp\{y_T + \Delta_h x'_{T+h}\beta + \sigma^2 \tfrac{h}{2} + \Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} x_T \frac{(y_T - \ldots}{T} \\
&E[\exp\left\{(y_T - x'_T\beta)\left(1 + \Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} \frac{x_T}{T}\right) + \underbrace{x'_T\beta + \Delta_h x'_{T+h}\beta} + \sigma \ldots\right. \\
&\exp\left\{\frac{T\sigma^2}{2}\left(1 + \Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} \frac{x_T}{T}\right)\left(1 + \Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} \frac{x_T}{T}\right) + x'_{T+h}\beta + \ldots\right. \\
&\exp\left\{\frac{T\sigma^2}{2}\left(1 + 2\Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} \frac{x_T}{T} + \Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} \frac{x_T}{T}\frac{x'_T}{T} \left(\Delta X' \Delta X\right)^{-1} \Delta_h x_{T+h}\right) + x'_{T+h}\beta + \ldots\right. \\
&\exp\left\{\frac{\sigma^2}{2}\left(T + 2\Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} x_T + \frac{1}{T}\Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} x_T x'_T \left(\Delta X' \Delta X\right)^{-1} \Delta_h x_{T+h}\right) + x'_{T+h}\beta + \ldots\right. \\
&\exp\left\{x'_{T+h}\beta + \sigma^2 \tfrac{T+h}{2} + \frac{1}{T}\Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} x_T x'_T \left(\Delta X' \Delta X\right)^{-1} \Delta_h x_{T+h}\right) + \frac{\sigma^2}{2}\left(2\Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} \ldots\right.
\end{aligned}
$$

consequently

$$
\begin{aligned}
E[F_{T+h|T}] &= E[\ Y_T \ \exp\{\ \Delta_h x'_{T+h}\beta + \sigma^2 \tfrac{h}{2} + \Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} x_T \frac{(y_T - x'_T\beta)}{T} + \\
&\quad -\frac{1}{2T}\sigma^2 \Delta_h x'_{T+h} \left(\Delta X' \Delta X\right)^{-1} x_T x'_T \left(\Delta X' \Delta X\right)^{-1} \Delta_h x_{T+h} - \sigma^2 x'_T \left(\Delta X' \Delta X\right)^{-1} \Delta_h x'_{T+h} \ \}\ ] \\
&= \exp\{x'_{T+h}\beta + \frac{T+h}{2}\sigma^2\}
\end{aligned}
$$

and $F_{T+h|T}$ is unconditionally unbiased. ■

5.3 Conditional Distributions given Y_T

Lemma 13 *(a) Conditional distribution of y_{T+h} given y_T*

$$
y_{T+h}|\, y_T \sim N\left(y_T + \Delta_h x'_{T+h}\beta; h\sigma^2\right)
$$

(b) Conditional expectation of Y_{T+h} given Y_T

$$
E[Y_{T+h}|Y_T] = Y_T \exp\{\Delta_h x'_{T+h}\beta + \frac{h}{2}\sigma^2\}
$$

(c) Unconditional expectation of Y_{T+h}

$$E[Y_{T+h}] = \exp\{x'_{T+h}\beta + \frac{T+h}{2}\sigma^2\}$$

Proof. *(a) follows from, see Section 7.1,*

$$\begin{aligned} y_{T+h} &= \sum_{s=1}^{T+h} \Delta x'_s\beta + \sum_{s=1}^{T+h} \varepsilon_s \\ &= y_T + \sum_{s=1}^{h} \Delta x'_{T+s}\beta + \sum_{s=1}^{h} \varepsilon_{T+s}, \\ &= y_T + \Delta_h x'_{T+h}\beta + \sum_{s=1}^{h} \varepsilon_{T+s}, \end{aligned}$$

and calculating the mean and variance. Part (b) and (c) follow from the moment generating function of a normal distribution.

$$E[e^{r'z}] = \exp\{\, r'E[z] + \frac{1}{2}r'Cov(z)\,r\}$$

■

Lemma 14 *Conditional distribution of* $y_{(T-1)}$ *given* y_T

$$y_{(T-1)} \mid y_T \sim N\left(X_{(T-1)}\beta + \begin{pmatrix} 1 \\ 2 \\ \vdots \\ T-1 \end{pmatrix} \tfrac{1}{T}\,(y_T - x'_T\beta)\,; \sigma^2\Sigma \right)$$

with $\Sigma_{ij} = \sigma^2 \left(\min\{i,j\} - \frac{1}{T} i\ j\right)$;

$$\text{or} \quad \Sigma = \sigma^2 \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & \cdots & T-1 \end{pmatrix} - \sigma^2\ \frac{1}{T} \begin{pmatrix} 1 \\ 2 \\ \vdots \\ T-1 \end{pmatrix} \begin{pmatrix} 1 & 2 & \cdots & T-1 \end{pmatrix}$$

Proof. *The proof uses the following standard result on conditional multivariate normal distributions (e.g. Muirhead (1982) Theorem 1.2.11)*

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

where Y_1 *is* $(n_1 \times 1)$, Y_2 *is* $(n_2 \times 1)$, μ *and* Σ *partitioned accordingly, then*

$$Y_1 | \, Y_2 \sim N\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y_2 - \mu_2), \quad \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right).$$

Since $y_{(T)} \sim N(X'_{(T)}\beta; \sigma^2\Omega_{(T)})$ *with*

$$\Omega_{(T)} \equiv \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2 & \cdots & 2 \\ 1 & 2 & \ddots & & \vdots \\ \vdots & \vdots & & T-1 & T-1 \\ 1 & 2 & \cdots & T-1 & T \end{pmatrix},$$

we have $\Sigma_{21} = \Sigma_{12}' = \sigma^2 (1, 2, \cdots, T-1)$, $\Sigma_{22} = \sigma^2 T$, *and hence*

$$\Sigma_{12}\Sigma_{22}^{-1}(Y_2 - \mu_2) = \begin{pmatrix} 1 \\ \vdots \\ T-1 \end{pmatrix} \frac{y_T - x_T'\beta}{T}; \tag{25}$$

$$\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \frac{1}{T}\sigma^2 \begin{pmatrix} 1 \\ \vdots \\ T-1 \end{pmatrix} \begin{pmatrix} 1 & \cdots & T-1 \end{pmatrix} \tag{26}$$

Finally, since $\Sigma_{11} = \sigma^2 \Omega_{(T-1)}$, *the result follows.* ■

Lemma 15 *The conditional distribution of* Δy *given* y_T *is a degenerate normal distribution:*

$$\Delta y |\, y_T \sim N_{\text{deg}} \left(\Delta X \; \beta + \imath_T \frac{1}{T} \left(y_T - x_T'\beta \right), \; \sigma^2 M_{\imath_T} \right), \tag{27}$$

with $M_{\imath_T} = (I_T - \frac{1}{T}\imath_T\imath_T')$

Proof. Using Lemma (@@) the conditional distribution of $y_{(T)}$ given y_T is a degenerate normal distribution

$$y \mid y_T \sim N_{\text{deg}} \left(\left(\begin{array}{c} X_{(T-1)}\beta + \begin{pmatrix} 1 \\ \vdots \\ T-1 \end{pmatrix} \frac{1}{T} \left(y_T - x_T'\beta \right) \\ y_T \end{array} \right); \begin{pmatrix} \Sigma_{(T-1)} & 0 \\ 0 & 0 \end{pmatrix} \right). \tag{28}$$

Since $\Delta y_{(T)} = L_{(T)} \; y_{(T)}$, we have that $\Delta y_{(T)}$ is normally distributed with mean

$$L_{(T)} E[y \mid y_T] = \Delta X \; \beta + \imath_T \frac{1}{T} \left(y_T - x_T'\beta \right), \tag{29}$$

$$L_{(T)} \begin{pmatrix} \Sigma_{(T-1)} & 0 \\ 0 & 0 \end{pmatrix} L_{(T)}' = (I_T - \frac{1}{T}\imath_T\imath_T') = M_{\imath_T}. \tag{30}$$

■

Note that the distribution of $\Delta y_{(T)} |\, y_T$ is degenerate since by assumption $\imath_T'\Delta y_{(T)} = y_T$.

Theorem 16 *Let* $\widehat{\beta} = \left(\Delta X'\Delta X\right)^{-1} \Delta X'\Delta y$ *and* $\widehat{\sigma}^2 = \dfrac{\Delta y' M_{\Delta X} \Delta y}{n-k}$ *then, the conditional distributions of* $\widehat{\beta}$ *and* $\widehat{\sigma}^2$ *given* Y_T *are*
(A) $\quad \widehat{\beta} \mid Y_T \sim N(\mu, \Sigma)$ *with*

$$\mu = E\left[\widehat{\beta}|Y_T\right] = \beta + \frac{y_T - x_T'\beta}{T} \left(\Delta X'\Delta X\right)^{-1} x_T,$$

$$\Sigma = Var(\widehat{\beta}|Y_T) = \sigma^2 \left(\left(\Delta X'\Delta X\right)^{-1} - \frac{1}{T} \left(\Delta X'\Delta X\right)^{-1} x_T x_T' \left(\Delta X'\Delta X\right)^{-1} \right),$$

(B) $\quad (T-k)\dfrac{\widehat{\sigma}^2}{\sigma^2} |\, Y_T \; \sim \chi^2_{T-k}$,
(C) $\quad \widehat{\sigma}^2$ *and* $\hat{\beta}$ *are conditionally and unconditionally independent.*

Proof. (A) $\widehat{\beta} = (\Delta X'\Delta X)^{-1} \Delta X'\Delta y$. The distribution of Δy given y_T (or Y_T) is given in the previous lemma. The result follows by noting that $\Delta X'\imath_T = x_T$ and the mean and variance follow by basic matrix multiplication as follows:

$$\begin{aligned} E\left[\widehat{\beta}|Y_T\right] &= (\Delta X'\Delta X)^{-1} \Delta X' E\left[\Delta y|Y_T\right], \\ &= \beta - (\Delta X'\Delta X)^{-1} \Delta X'\imath_T \frac{(y_T - x_T'\beta)}{T}, \\ &= \beta - (\Delta X'\Delta X)^{-1} x_T \frac{(y_T - x_T'\beta)}{T}. \end{aligned}$$

$$\begin{aligned} Var(\widehat{\beta}|Y_T) &= \sigma^2 (\Delta X'\Delta X)^{-1} \Delta X'(I_T - \frac{1}{T}\imath_T\imath_T')\Delta X (\Delta X'\Delta X)^{-1}, \\ &= \sigma^2 \left((\Delta X'\Delta X)^{-1} - \frac{1}{T} (\Delta X'\Delta X)^{-1} x_T x_T' (\Delta X'\Delta X)^{-1} \right), \end{aligned}$$

since $x_T = \imath_T'\Delta X$ by Assumption 1.
(B) and (C). We will use a conditional version of a theorem by Ogasawara and Takahashi (1951), see Muirhead (1982) which states that if $z \sim N(0, \Sigma)$, with Σ possibly singular, then $z'Az \sim \chi^2_{rank(A\Sigma)}$ if and only if $A\Sigma A\Sigma = A\Sigma$. Using the conditional distribution of $\Delta y_{(T)}|\, y_T$, we see that $M_{\Delta X} E[\Delta y_{(T)}|\, y_T] = 0$, and for the (degenerate) covariance matrix for $M_{\Delta X}\Delta y_{(T)} = \sigma^2 M_{\Delta X}\ M_{\imath_T} M_{\Delta X} = \sigma^2 M_{\Delta X}$, since $M_{\Delta X} M_{\imath_T} = M_{\Delta X}$. We have therefore

$$M_{\Delta X}\Delta y_{(T)}|\, y_T \sim N_{\text{deg}} \left(0,\ \ \sigma^2\ M_{\Delta X} \right)$$

and hence with $A = M_{\Delta X} = \Sigma$, we have $rank(A\Sigma) = T - k$

$$\Delta y' M_{\Delta X} \Delta y \sim \chi^2_{T-k}$$

(C) Follows since $\hat{\beta} = \left[(\Delta X'\Delta X)^{-1}\Delta X'\right] \Delta y$ and $M_{\Delta X}\Delta y$ are both linear functions of the conditionally and unconditionally normally distributed Δy and uncorrelated since $M_{\Delta X}\Delta X(\Delta X'\Delta X)^{-1} = 0$, and therefore independent. Hence $\hat{\beta}$ is also independent of $\Delta y' M_{\Delta X}\Delta y/(T-k)$, conditionally and unconditionally.
■

Remark 17 *If the model only includes a time trend such that* $x_t = t$, *and therefore* $\Delta X = \imath_T$, *then* $\mu = \frac{y_T}{T}$. *No* β *appears in the expectation and the conditional variance is* 0.

Remark 18 *alternative:* $M_{\imath_T} = (I_T - \frac{1}{T}\imath_T\imath_T') = (I_T - \imath_T(\imath_T'\imath_T)^{-1}\imath_T')$, *is a projection matrix with* $M_{\imath_T}\imath_T = 0$. *This implies that if the model only includes a time trend, and therefore* $\Delta X = \imath_T$, *that the conditional variance of* $\widehat{\beta}|Y_T$ *is* 0, *which is as it should be since in that case* $\widehat{\beta} = (\imath_T'\imath_T)^{-1}\imath_T'\Delta y = \frac{1}{T}y_T$, *and fixed for given* Y_T.

Corollary 19 $Y_T \exp\{C\widehat{\beta}\}$ *is conditionally independent of* $\widehat{\sigma}^2$, *given* Y_T, *for any arbitrary, fixed matrix* C.

This is important because it allows us to take expectations with respect to $\hat{\beta}$ first, before evaluating the expectation of a function of $\widehat{\sigma}^2$.

Lemma 20 *If ΔX includes a constant term, then no conditionally unbiased predictor based on the complete sufficient statistic $\left(\hat{\beta}, \hat{\sigma}^2\right)$ of Y_{T+h} exists given either (a) $\{Y_1, ..., Y_T\}$ or (b) $\{Y_T\}$.*

Proof. The conditional expectation $E[Y_{T+h}|Y_T] = Y_T \exp\{\Delta_h x'_{T+h}\beta + \frac{h}{2}\sigma^2\}$. The statistic $\left(\hat{\beta}, \hat{\sigma}^2\right)$ given Y_T is a complete sufficient statistic for the distribution of the distribution of $Y_{(T-1)}|Y_T, X$, which implies that any function $f\left(\hat{\beta}, \hat{\sigma}^2\right)$ with expectation $g\left(\beta, \sigma^2\right)$ is essentially unique: if $\tilde{f}\left(\hat{\beta}, \hat{\sigma}^2\right)$ is a function with the same expectation function $g\left(\beta, \sigma^2\right)$, then $E_{\beta,\sigma^2}\left[f\left(\hat{\beta}, \hat{\sigma}^2\right) - \tilde{f}\left(\hat{\beta}, \hat{\sigma}^2\right)\right] = 0$ for all $\beta, \sigma \in \Re^k \times \Re^+$, but the completeness of $\left(\hat{\beta}, \hat{\sigma}^2\right)$ means by definition that $E_{\beta,\sigma^2}\left[h\left(\hat{\beta}, \hat{\sigma}^2\right)\right] = 0$ for all parameter values, implies that $h\left(\hat{\beta}, \hat{\sigma}^2\right) = 0$ a.e. and hence $f = \tilde{f}$ a.e. Since $\widehat{\beta} \mid Y_T \sim N(\mu, \Sigma)$ and for a fixed vector a we have $E[exp\{a\prime\widehat{\beta}\}|Y_T] = exp\{a'\beta + \frac{1}{2}a'\Sigma a\}$, we know that the function must be proportional to $exp\{a\prime\widehat{\beta}\}$ since the expectation function would otherwise not be log-linear in β. Hence, to find a conditional unbiased estimator we first need to solve

$$E_{\hat{\beta}|Y_T}\left[\exp\{a'\hat{\beta}\}\,|Y_T\right] \propto \exp\left\{\Delta_h x'_{T+h}\beta\right\} \tag{31}$$

note that terms involving Y_T and σ^2 are imaterial at this stage, since they, by the independence of $\hat{\sigma}^2$ and $\hat{\beta}$, can be taken account of via the ${}_0F_1$-function. Using the conditional distribution of $\hat{\beta}$ we have

$$E_{\hat{\beta}|Y_T}\left[\exp\{a'\hat{\beta}\}\,|Y_T\right] = \exp\left\{a'\beta + a'\left(\Delta X'\Delta X\right)^{-1} x_T \frac{(-x'_T)}{T}\beta\right\} \times \textit{terms not involving } \beta \tag{32}$$

Hence, we need to solve

$$a'(I_k - \left(\Delta X'\Delta X\right)^{-1} x_T x'_T \frac{1}{T})\beta = \Delta_h x'_{T+h}\beta, \quad \textit{for all } \beta \in \Re^k \tag{33}$$

Let $B = (I_k - \left(\Delta X'\Delta X\right)^{-1} x_T x'_T \frac{1}{T})$ and let B^+ denote its Moore-Penrose generalized inverse (explicit expressions for B and B^+ are given in the Lemma below), then

$$a'(I_k - \left(\Delta X'\Delta X\right)^{-1} x_T x'_T \frac{1}{T}) = \Delta_h x'_{T+h}$$

$$(I_k - \left(\Delta X'\Delta X\right)^{-1} x_T x'_T \frac{1}{T})'a = \Delta_h x_{T+h}$$

has general solution (as derived by Penrose, see e.g. Magnus and Neudecker '87, p37)

$$a = B'^{+}\Delta_h x_{T+h} + (I_k - B'^{+}B')q$$

with q an arbitrary $(k \times 1)$ vector, if and only if:

$$B'B'^{+}\Delta_h x_{T+h} = \Delta_h x_{T+h} \tag{34}$$

Lemma (@@) below shows that $B'B'^{+} = \begin{pmatrix} 0 & 0 \\ 0 & I_{k-1} \end{pmatrix}$. Since the first element of $\Delta_h x_{T+h}$ equals h, and not 0, the consistency condition (34) cannot be satisfied and no a exists such that $a'B = \Delta_h x'_{T+h}$. This implies that no conditional unbiased predictor based on the complete sufficient statistics exists. ■

Lemma 21 (BB+) *The matrix B equals*

$$B = \begin{pmatrix} 0 & b' \\ 0 & I_{k-1} \end{pmatrix}$$

with $b = -\frac{1}{T}x_{T,2:k}$ where $x_T = (T,\, x'_{T,2:k})'$ and has Moore-Penrose inverse

$$B^+ = \frac{1}{1+b'b}\begin{pmatrix} 0 & 0 \\ b & (1+b'b)I_{k-1} - bb' \end{pmatrix}$$

Proof. First note that $\Delta X\left(\Delta X'\Delta X\right)^{-1}\Delta X'\iota = \iota$ since $P_{\Delta X} = \Delta X\left(\Delta X'\Delta X\right)^{-1}\Delta X'$ is a projection matrix and ι is the first column of ΔX. This implies that $\left(\Delta X'\Delta X\right)^{-1}\Delta X'\iota = (1, 0, ..., 0)'$ since if we write

$$\begin{aligned} \left(\Delta X'\Delta X\right)^{-1}\Delta X'\iota &= \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \Longrightarrow \\ \Delta X\left(\Delta X'\Delta X\right)^{-1}\Delta X'\iota &= \begin{pmatrix} a_1 + \Delta x'_{1,2:k}a_2 \\ \vdots \\ a_1 + \Delta x'_{T,2:k}a_2 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \quad \forall \Delta x_{t,2:k} \in \mathbb{R}^{k-1}, \end{aligned}$$

which implies that $a_1 = 1$ and $a_2 = 0$. Hence $B = I_k - (1, 0, ..., 0)'x'_T\dfrac{1}{T} = I_k - (1, 0, ..., 0)'(1, \frac{1}{T}x_{T,2:k})$. For the second part of the Lemma the four conditions of the Moore-Penrose inverse are easily verified: (a) $B^+B = \begin{pmatrix} 0 & 0 \\ 0 & I_{k-1} \end{pmatrix}$ and hence symmetric, (b) BB^+ is symmetric, (c) $BB^+B = B$, and (d) $B^+BB^+ = B^+$. ■

Lemma 22

$$\begin{aligned} & E\left[Y_T \exp\{\Delta x'_{T+1}\widehat{\beta}\} \,|\, Y_T\right] \\ = \;& Y_T \exp\{\Delta x'_{T+1}\beta - \Delta x'_{T+1}\left(\Delta X'\Delta X\right)^{-1} x_T \frac{y_T - x_T\beta}{T} + \\ & \tfrac{1}{2}\,\sigma^2 \Delta x'_{T+1}(\left(\Delta X'\Delta X\right)^{-1} - \frac{1}{T}\left(\Delta X'\Delta X\right)^{-1} x_T x'_T \left(\Delta X'\Delta X\right)^{-1})\Delta x_{T+1}\} \end{aligned}$$

Lemma 23 *Goldberger (1962). Let*

$$\begin{pmatrix} y_{(T)} \\ y_{T+1} \end{pmatrix} = \begin{pmatrix} X_{(T)} \\ x'_{T+1} \end{pmatrix}\beta + \begin{pmatrix} \varepsilon_{(T)} \\ \varepsilon_{T+1} \end{pmatrix}, \begin{pmatrix} \varepsilon_{(T)} \\ \varepsilon_{T+1} \end{pmatrix} \sim N(0, \begin{pmatrix} \Omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}$$

then the Best Linear Unbiased Predictor of y_{T+1} is given by

$$\hat{y}^*_{T+1} = x'_{T+1}\hat{\beta}_{GLS} + \omega_{21}\Omega_{11}^{-1}e_{GLS,T}$$

Best in Goldberger (1962) means that the predictor minimizes $E\left[\left(\hat{y}^*_{T+1} - y_{T+1}\right)\left(\hat{y}^*_{T+1} - y_{T+1}\right)'\right]$, subject to (i) linearity $\hat{y}^*_{T+1} = C'y_{(T)}$ and (ii) unbiasedness $E\left[\hat{y}^*_{T+1} - y_{T+1}\right] = 0$.

Since in the unit root case $\omega_{21}\Omega_{11}^{-1} = 1$ we have for Model 1:

$$\hat{y}^*_{T+1} = x'_{T+1}\hat{\beta} + 1.(y_T - x'_T\hat{\beta})$$

and hence

Corollary 24 *In Model 1, the Best Linear Unbiased Predictor of y_{T+1} is given by*

$$\hat{y}^*_{T+1} = y_T + \Delta x'_{T+1}\hat{\beta}$$

5.4 Expectations involving exponentials and hypergeometric functions

Definition 25 *The hypergeometric function can be defined as the infinite sum:*

$$\begin{aligned} {}_0F_1(m, x) &\equiv \sum_{i=0}^{\infty} \frac{x^i}{i!(m)_i}, \\ (m)_0 &= 1 \text{ and } (m)_i = m(m+1)\ldots(m+i-1) \end{aligned}$$

See e.g. Abadir (2001) who reviews the use of hypergeometric functions in economics or Van Garderen (2001) who further proves:

Lemma 26 *Let $m\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_m$, then for any real constant z we have,*

$$\begin{aligned} E\left[{}_0F_1(\frac{m}{2}; z\frac{m}{2}\hat{\sigma}^2\right] &= \exp\left\{\ z\ \sigma^2\right\} \\ E\left[\left\{{}_0F_1(\frac{m}{2}; z\frac{m}{2}\hat{\sigma}^2)\right\}^2\right] &= \exp\left\{2\ z\ \sigma^2\right\}_0 F_1(\frac{m}{2}; z^2\sigma^4) \end{aligned}$$

5.5 Sectorial Results

	drift	dd72	d74	dd74	dd75	dd76	d80	dd84	dlL	dlK	$\hat{\sigma}$
Agriculture	0.029			-0.063		-0.135			-0.025	-0.147	0.0
Mining	0.002	-0.127		-0.116					0.329	0.262	0.0
Manufacturing	-0.014								0.100	1.157	0.0
Energy	0.036								-0.319	0.715	0.0
Construction	0.006		-0.132				-0.061		0.518	0.387	0.0
Transport	0.020						-0.066		0.009	0.429	0.0
Communication	0.039								0.707	0.055	0.0
Other Services	-0.004						-0.053		0.610	0.582	0.0

==

Startdate = 1956 Enddate = 1980 Tobs = 25.000 Periods ahead: h = 5 Number of MC replications:100.000

==

Average Actual and predicted GROWTHs
Actual exactunb apprunb Naive consist bias: exactunb apprunb Naive consist
15.998 15.966 15.966 15.802 16.153 0.000 -0.032 -0.032 -0.196 0.155
-5.395 -5.404 -5.404 -5.540 -5.330 0.000 -0.008 -0.008 -0.144 0.065
2.524 2.554 2.554 2.486 2.582 0.000 0.030 0.030 -0.038 0.058
30.465 30.411 30.411 30.263 30.783 0.000 -0.053 -0.053 -0.202 0.319
-0.258 -0.273 -0.273 -0.147 0.039 0.000 -0.015 -0.015 0.111 0.297
5.013 5.048 5.048 5.540 5.744 0.000 0.035 0.035 0.527 0.731
22.050 22.038 22.038 21.975 22.114 0.000 -0.012 -0.012 -0.075 0.064
20.744 20.725 20.725 20.660 20.751 0.000 -0.019 -0.019 -0.084 0.006
on average over all sectors : ==
11.393 11.383 11.383 11.380 11.605 0.000 -0.010 -0.009 -0.013 0.212
==
Average Actual and predicted LEVELs
Actual exactunb apprunb Naive consist grow-based bias exactunb apprunb Naive consist grow-based
2.286 2.286 2.286 2.296 2.303 2.300 0.000 -0.001 -0.001 0.010 0.017 0.013
0.822 0.822 0.822 0.824 0.826 0.825 0.000 0.000 0.000 0.002 0.004 0.004
1.876 1.877 1.877 1.879 1.881 1.880 0.000 0.001 0.001 0.003 0.005 0.004
6.163 6.161 6.161 6.203 6.228 6.210 0.000 -0.002 -0.002 0.040 0.065 0.047
1.673 1.672 1.672 1.681 1.684 1.679 0.000 0.000 0.000 0.008 0.011 0.006
1.719 1.719 1.719 1.734 1.737 1.726 0.000 0.000 0.000 0.015 0.019 0.007
3.984 3.984 3.984 3.991 3.995 3.993 0.000 0.000 0.000 0.007 0.011 0.009
2.512 2.512 2.512 2.514 2.516 2.516 0.000 0.000 0.000 0.002 0.004 0.003
on average over all sectors : ==
2.629 2.629 2.629 2.640 2.646 2.641 0.000 0.000 0.000 0.011 0.017 0.012
==
======= MSE =======================
==
MSE GROWTH predictions
exactunb apprunb Naive consist
125.564 125.564 125.472 125.742
53.621 53.621 53.596 53.651
25.602 25.602 25.596 25.608
233.089 233.089 232.906 233.761
99.316 99.316 99.487 99.804
190.350 190.350 191.996 192.835
52.619 52.619 52.604 52.649
28.003 28.003 28.002 28.005
on average over all sectors : ==
101.020 101.020 101.207 101.507
MSE Y LEVEL predictions
exactunb apprunb Naive consist grow-based
0.051 0.051 0.051 0.052 0.052
0.004 0.004 0.004 0.004 0.004
0.009 0.009 0.009 0.009 0.009
0.548 0.548 0.557 0.564 0.559
0.029 0.029 0.029 0.029 0.029
0.052 0.052 0.053 0.054 0.053
0.057 0.057 0.057 0.057 0.057

0.012 0.012 0.012 0.012 0.012
on average over all sectors : ==
0.095 0.095 0.097 0.098 0.097
==
=== RELATIVE MSE ======================
==
Relative: MSE GROWTH percentage relative to best
exactunb apprunb Naive consist
0.073 0.073 0.000 0.215
0.046 0.046 0.000 0.103
0.025 0.025 0.000 0.047
0.079 0.079 0.000 0.367
0.000 0.000 0.172 0.491
0.000 0.000 0.865 1.305
0.029 0.029 0.000 0.086
0.002 0.002 0.000 0.012
========= Average values over all sectors ==================================
0.032 0.032 0.130 0.328
==
Relative: MSE Y LEVEL percentage relative to best
exactunb apprunb Naive consist grow-based
0.000 0.000 1.121 2.133 1.566
0.000 0.000 0.749 1.519 1.227
0.000 0.000 0.313 0.644 0.537
0.000 0.000 1.658 2.983 2.009
0.000 0.000 1.210 1.807 0.851
0.000 0.001 2.134 2.751 0.845
0.000 0.000 0.429 0.808 0.590
0.000 0.000 0.223 0.471 0.395
========= Average values over all sectors ==================================
0.000 0.000 0.980 1.640 1.003
==

now 100k 40 obs h = 10
Startdate = 1956 Enddate = 1995 Tobs = 40.000 Periods ahead: h = 10.000
Number of MC replications: 100.000
parameter estimates per sector
inpt dd72 d74 dd74 dd75 dd76 d80 dd84 dlli dlki sigm
0.014 0.000 0.000 0.000 -0.072 -0.140 0.000 0.133 -0.378 -0.003 0.039
0.002 -0.144 0.000 -0.125 0.000 0.000 0.000 -0.480 0.261 0.028 0.053
0.050 0.000 0.000 0.000 0.000 0.000 0.000 -0.035 0.704 -0.706 0.038
0.028 0.000 0.000 0.000 0.000 0.000 0.000 -0.205 0.083 1.026 0.035
0.018 0.000 -0.119 0.000 0.000 0.000 -0.044 0.000 0.509 0.248 0.037
0.026 0.000 0.000 0.000 0.000 0.000 -0.056 0.000 0.303 -0.008 0.029
0.042 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.395 0.172 0.025
-0.009 0.000 0.000 0.000 0.000 0.000 -0.026 0.000 0.487 0.626 0.021
Average over all sectors
0.021 -0.018 -0.015 -0.016 -0.009 -0.018 -0.016 -0.073 0.296 0.173 0.035
==
Average Actual and predicted GROWTHs

Actual exactunb apprunb Naive consist bias: exactunb apprunb Naive consist
22.934 22.958 22.958 22.294 23.210 0.000 0.023 0.023 -0.641 0.275
-18.703 -18.677 -18.677 -19.495 -18.345 0.000 0.025 0.026 -0.792 0.358
12.315 12.320 12.320 11.806 12.626 0.000 0.005 0.005 -0.509 0.311
21.137 20.999 20.999 20.912 21.672 0.000 -0.138 -0.138 -0.224 0.535
26.528 26.470 26.470 25.832 26.708 0.000 -0.059 -0.059 -0.696 0.180
26.164 26.189 26.189 25.829 26.345 0.000 0.025 0.025 -0.334 0.182
59.313 59.306 59.306 58.956 59.435 0.000 -0.006 -0.006 -0.357 0.122
33.416 33.412 33.412 33.229 33.521 0.000 -0.004 -0.004 -0.187 0.106
on average over all sectors : ==
22.888 22.872 22.872 22.420 23.146 0.000 -0.016 -0.016 -0.468 0.259
==
Average Actual and predicted LEVELs
Actual exactunb apprunb Naive consist grow-based bias exactunb apprunb Naive consist grow-based
3.055 3.055 3.055 3.084 3.107 3.101 0.000 0.000 0.000 0.029 0.052 0.046
0.403 0.403 0.403 0.410 0.416 0.415 0.000 0.000 0.000 0.008 0.013 0.012
2.725 2.725 2.725 2.753 2.773 2.766 0.000 0.001 0.001 0.028 0.048 0.041
9.076 9.067 9.067 9.174 9.232 9.181 0.000 -0.010 -0.010 0.098 0.156 0.105
2.948 2.946 2.946 2.973 2.993 2.988 0.000 -0.001 -0.001 0.025 0.045 0.040
3.017 3.018 3.018 3.034 3.047 3.043 0.000 0.001 0.001 0.017 0.029 0.026
11.075 11.074 11.074 11.116 11.150 11.141 0.000 -0.001 -0.001 0.041 0.075 0.065
4.801 4.801 4.801 4.816 4.826 4.822 0.000 0.000 0.000 0.014 0.025 0.021
on average over all sectors : ==
4.638 4.636 4.636 4.670 4.693 4.682 0.000 -0.001 -0.001 0.032 0.055 0.044
==
======= MSE =====================
==
MSE GROWTH predictions
exactunb apprunb Naive consist
287.942 287.942 287.685 288.285
245.229 245.230 244.732 245.840
253.951 253.951 253.567 254.446
346.668 346.668 346.465 348.796
283.509 283.510 283.352 283.790
170.841 170.841 170.721 170.979
193.266 193.266 193.203 193.355
107.424 107.424 107.373 107.489
on average over all sectors : ==
236.104 236.104 235.887 236.622
MSE Y LEVEL predictions
exactunb apprunb Naive consist grow-based
0.193 0.193 0.198 0.203 0.201
0.007 0.007 0.007 0.008 0.008
0.161 0.161 0.165 0.169 0.167
2.079 2.079 2.139 2.182 2.143
0.166 0.166 0.169 0.173 0.172
0.102 0.102 0.104 0.105 0.105
0.962 0.962 0.971 0.981 0.978
0.142 0.142 0.143 0.144 0.144

on average over all sectors : ==
0.476 0.476 0.487 0.496 0.490
===
=== RELATIVE MSE ======================
===
Relative: MSE GROWTH percentage relative to best
exactunb apprunb Naive consist
0.089 0.089 0.000 0.209
0.203 0.203 0.000 0.453
0.151 0.151 0.000 0.346
0.059 0.059 0.000 0.673
0.055 0.055 0.000 0.154
0.071 0.071 0.000 0.151
0.033 0.033 0.000 0.079
0.048 0.048 0.000 0.109
========= Average values over all sectors ===================================
0.089 0.089 0.000 0.272
===
Relative: MSE Y LEVEL percentage relative to best
exactunb apprunb Naive consist grow-based
0.000 0.000 2.319 4.880 4.111
0.000 0.002 4.603 9.652 8.063
0.000 0.001 2.551 5.120 4.092
0.000 0.001 2.875 4.941 3.091
0.000 0.000 2.141 4.512 3.807
0.000 0.000 1.346 2.771 2.304
0.000 0.000 0.937 1.968 1.666
0.000 0.000 0.756 1.509 1.207
========= Average values over all sectors ===================================
0.000 0.001 2.191 4.419 3.543
===

6 References

Abadir, K.M. (1999). An introduction to hypergeometric functions for economists. Econometric Reviews, **18**, 287-330.

Diebold, F. X., Mariano, R. S. (1995) "Comparing Predictive Accuracy", Journal of Business and Economic Statistics, 13, 253-263.

Clements, M. P., Hendry, D. F. (1999) "Forecasting Non-stationary Economic Time Series", MIT Press, Cambridge, Massachusetts.

Garderen, K.J.van, K.C.Lee, and M.H. Pesaran (2000) "Cross-sectional aggregation of non-linear models", Journal of Econometrics, **95**, 285-331.

Goldberger, A.S. (1968) The interpretation and estimation of Cobb-Douglas functions" Econometrica, **36**, 464-472.

Granger, C.W.J. and Newbold, P. (1976) "Forecasting Transformed Series", Journal of the Royal Statistical Society (B), 38(2), 1976.

Muirhead, R.J. (1982) "Aspects of Multivariate Statistical Theory", John Wiley, New York.

Nelson, C.R. and Plosser, C.I. (1982) “Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications”, Journal of Monetary Economics, **10**, 139-162.

Rosanna, R.J. (1995) Technology shocks and cointegration in quadratic model of the firm”, International Economic Review, **36**, 5-18.

Sampson, M. 1991 The Effect of Parameter Uncertainty on Forecast Variances and Confidence Intervals for Unit Root and Trend Stationary Time-series Models, Journal of Applied Econometrics **6**, 67-76,

Teekens, R., and Koerts, J. (1972) Some statistical implications of the log transformation of multiplicative models. Econometrica, **40**, 793-819.

van Garderen, K.J. (2001) “Optimal Prediction in Loglinear Models”, Journal of Econometrics, **104**,119-140.

Schmidt, P. (1974) "The Asymptotic Distribution of Forecasts in the Dynamic Simulation of an Econometric Model", Econometrica, 42, 303-310.

West, K.D. (1988) “Asymptotic Nomality, When Regressors Have a Unit Root”, Econometrica, 56,1397-1417

West, K. D. (1996) “Asymptotic Inference about Predictive Ability”, Econometrica, 64, 1067-1084.