Forecasting Levels in Loglinear Unit Root Models

van Garderen, K.J.

DOI
10.1080/07474938.2023.2224175

Publication date
2023

Document Version
Final published version

Published in
Econometric Reviews

Taylor & Francis
Taylor & Francis Group

Econometric Reviews

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/lecr20

Forecasting Levels in Loglinear Unit Root Models

Kees Jan van Garderen

To cite this article: Kees Jan van Garderen (2023) Forecasting Levels in Loglinear Unit Root Models, Econometric Reviews, 42:9-10, 780-805, DOI: 10.1080/07474938.2023.2224175

To link to this article: https://doi.org/10.1080/07474938.2023.2224175

Published online: 12 Jul 2023.

ECONOMETRIC REVIEWS
2023, VOL. 42, NO. 9–10, 780–805
https://doi.org/10.1080/07474938.2023.2224175

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Forecasting Levels in Loglinear Unit Root Models

Kees Jan van Garderen

University of Amsterdam

ABSTRACT

This article considers unbiased prediction of levels when data series are modeled as a random walk with drift and other exogenous factors after taking natural logs. We derive the unique unbiased predictors for growth and its variance. Derivation of level forecasts is more involved because the last observation enters the conditional expectation and is highly correlated with the parameter estimates, even asymptotically. This leads to conceptual questions regarding conditioning on endogenous variables. We prove that no conditionally unbiased forecast exists. We derive forecasts that are unconditionally unbiased and take into account estimation uncertainty, non linearity of the transformations, and the correlation between the last observation and estimate, which is quantitatively more important than estimation uncertainty and future disturbances together. The exact unbiased forecasts are shown to have lower Mean Squared Forecast Error (MSFE) than usual forecasts. The results are applied to Bitcoin price levels and a disaggregated eight sector model of UK industrial production.

ARTICLE HISTORY
Received 18 November 2021
Accepted 3 June 2023

KEYWORDS
nonlinear transformation; stochastic growth; parameter uncertainty; unbiasedness

JEL CLASSIFICATION
C20; C22; C53

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 modeling strategy found in many empirical studies in economics is therefore to model the variables in log-differences if the hypothesis of a unit root in the log of the variables cannot be rejected. In this article, we will assume that taking log-differences is indeed correct and renders the series stationary, but that interest is nevertheless in predicting the level of the original, untransformed variables. This leads to a number of interesting issues not encountered in linear (non) stationary settings, even when abstracting from the complication of testing for a unit root or the log-transformation as has been addressed by e.g., McAleer and co-authors (Bera and McAleer, 1989; Franses and McAleer, 1998; Lim and McAleer, 2001). The aim of the article is to highlight the issues, to provide solutions for the problems encountered, and to investigate the quantitative importance of the results. We allow for regressors, but assume that they are strictly exogenous. This is trivial for constants, seasonal dummies, and deterministic trends, but a serious limitation for economic time series that are jointly determined or even cointegrated, which requires a higher dimensional, vector autoregressive generalization of the results here.

In order to set out the principal issues involved, we consider the following specific example of the model we have in mind, which is the Cobb-Douglas production function for Y_t, output in period t,

CONTACT Kees Jan van Garderen K.J.vanGarderen@uva.nl UvA-Econometrics, University of Amsterdam, Roetersstraat 11, 1018WB Amsterdam, The Netherlands.

given inputs labor, L_t, and capital K_t with time varying technology as employed by Rosanna (1995) in the context of optimizing firms and Binder and Pesaran (1999) in the growth literature:

$$\begin{aligned} Y_t &= Z_t L_t^{\beta_L} K_t^{\beta_K}, \\ \ln(Z_t) \equiv z_t &= \beta_0 + \delta\, t + u_t, \\ u_t &= u_{t-1} + \varepsilon_t, \end{aligned} \tag{1}$$

with ε_t *independent and identically Normal (i.i.N.)* distributed and $\ln(Z_t)$ the natural logarithm of Z_t. The term z_t represents technological progress which is assumed to grow deterministically over time, and further has a stochastic trend. The normality assumption allows explicit solutions for the problems we want to analyze. The model was also used in Garderen et al. (2000) in comparing various non linear aggregate and disaggregate models based on predictions of the output level Y.

The central issue is forecasting one or more steps ahead output level Y_{T+h} based on observations up to and including time T. We focus on unbiased forecasts since, in the absence of knowledge of the real cost of forecast errors, one often starts out with a squared error loss function and use the Mean Squared Forecasting Error (MSFE) criterion for assessing 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, as was pointed out as early as Goldberger (1968) in the context of Cobb-Douglas production functions with *independent and identically distributed (i.i.d.)* disturbances. The focus on unbiasedness is not only motivated by the squared error loss function, but also by its ubiquity and primary interest, even though other measures of central tendency are available.

After taking log-differences the model becomes:

$$\Delta y_t = \delta\ + \Delta l_t \beta_L + \Delta k_t \beta_K + \varepsilon_t,$$

where small letters indicate that natural logs have been taken. An example with only regressors that are clearly exogenous is a model with only a trend and seasonal dummies to allow different seasons to contribute to growth differently.

The transformed model is easily estimated, forecasting is standard, and inference is straightforward. The inverse transformation (exponentiation) is non linear and the non stationary nature of the log variable gives rise to further complications, mainly in forecasting 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 unknown parameter values. The non stationarity 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 the 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 forecasts like $\check{Y}_{T+1} = Y_T \exp\{\hat{\delta}\ + \Delta l_{T+1}\hat{\beta}_L + \Delta k_{T+1}\hat{\beta}_K + \frac{1}{2}\hat{\sigma}^2\}$, which seem reasonable and are in common use (either with or without the last bias correction term $\frac{1}{2}\hat{\sigma}^2$), can be significantly biased and are not even consistent predictors of the conditional expectation.

Second, the non linearity 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 non linear 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 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 theoretical conditional mean) and no explanatory variables. Ariño and Franses (2000) consider forecasting levels from a VAR in log-transformed time series. They find that the VAR analog of $\check{Y}_{T+1}$ with the equivalent term $\frac{1}{2}\hat{\sigma}^2$ to correct for the non linear transformation of the white noise performs better than without it, but abstract from estimation uncertainty and correlation with conditioning variables Y_T. Bårdsen and Lütkepohl (2011), also in a VAR context, do a simulation study and take estimation and selection of cointegration order into account. See also Mayr and Ulbricht (2015).

The third issue 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. Clements and Hendry (1998, p.111 ff) also stress the importance 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.

One fundamental issue in this article concerns conditioning and unbiasedness. How to define unbiasedness when conditioning on past observations in the presence of high correlation between parameter estimates and conditioning variables.

For prediction purposes, 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, e.g., 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. Second, 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-prediction.

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, but he notes that the Edgeworth expansions are not accurate for large autoregressive parameters. Appendix A.2 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. For example, 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 important result is new and holds obviously when conditioning on the whole past but more interestingly, also when conditioning only on the last observation.

We derive two new exact unbiased estimators, one unconditionally unbiased estimator 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.

Finally, assuming it known that the model is loglinear with unit root and drift warrants a comment. It means that there is no (pre-) testing for unit roots or loglinearity. Proper inference procedures should take into account, but it would obscure the issues we wish to highlight here, but see for instance Ng and Vogelsang (2002), Stock (1996, 1997) who moreover, as Dufour (1984), considers construction of conditional and unconditional unbiased forecasts in nonstationary models. Second, the drift parameter is more accurately estimated in the first (log)-difference model where the Cramér-Rao lower bound is attained, than the trend coefficient in (log-)levels. Parameter uncertainty is an important factor in adjusting the predictions and using a less accurate estimator would lead to even larger effects. Suitability of the log-transformation could be tested using various methods, see e.g. McAleer and Pesaran (1986), Bera and McAleer (1989), but correctness is again assumed.

The remainder of the article is organized as follows. The next section discusses the basic model. Section 3 deals with forecasting future levels, its MSFE both theoretical and in a simulation exercise. Section 4 applies these results to Bitcoin price levels and Section 6 to eight sectors in the UK economy and compares our suggestions with common solutions. Section 5 investigates the sensitivity of these results to misspecifications and deviations from the assumptions including local-to-unity specification and skewness of the disturbances. Section 7 concludes. Proofs are relegated to Appendix A.

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:

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

$$u_t = u_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim i.i.N(0, \sigma^2), \tag{3}$$

where x_t is a $(k \times 1)$ vector of regressors including a linear trend with parameter δ which is one element in the $(k \times 1)$ vector of unknown parameters β. The normality of ε_t is a strong assumption, but facilitates the exact finite sample results derived here.

The model implies that for the log-differences

$$\Delta y_t = \Delta x_t'\beta + \varepsilon_t, \tag{4}$$

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

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

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

Assumption 3. X is exogenous.

Assumption 1 is West (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, as in all our simulations, 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 (4) is uniquely defined. The assumption does imply that no constant term is included in x_t, but that is related to the fact that the constant cannot be identified from the equation in first differences.

Assumption 3 allows the conditioning on explanatory variables. This is innocuous if it only includes dummies for seasons or known structural breaks and deterministic trends etc., but in general is quite restrictive. It abstracts from a full vector autoregressive framework with cointegration for instance, which would be quite natural in this context. This would require non trivial multivariate generalizations of our results however, that are not necessary for the main issues in the paper: (i) high information in the last observation Y_T (which would become a vector), (ii) its high correlation with the estimators, (iii) parameter uncertainty, and (iv) forecasting horizon.

With these assumptions, the parameters β and σ^2 can simply be estimated using OLS in (4). Forecasting Δy_T is straightforward, as is forecasting the log-variable y_t. Using Goldberger (1962) it is easily shown that the optimal (minimum variance linear unbiased) predictor is given by:

$$\hat{y}^*_{T+1} = y_T + \Delta x'_{T+1}\hat{\beta}. \tag{5}$$

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

Lemma 1. *If $\widehat{\beta}$ is the OLS estimator of the model in log-differences, then:*

$$Cov\left(y_T, \widehat{\beta}\right) = \sigma^2 x'_T \left(\Delta X' \Delta X\right)^{-1}. \tag{6}$$

Corollary 1. *If x_t includes a linear and possibly a stochastic trend, such that $\left(\Delta X' \Delta X\right) = O_p(T)$ and $x_T = O_p(T)$, then conditional on $\{x_t\}_{t=1}^T$:*

$$Cov\left(y_T, \widehat{\beta}\right) = O(1), \tag{7}$$

If x_t consists of a linear trend only, s.t. $\beta_1 = \delta$, then for all T:

$$Corr\left(y_T, \hat{\beta}_1\right) = 1. \tag{8}$$

The problem is that y_T is very informative about the future levels y_{T+h}, while at the same time also contains much information about the parameter that needs to be estimated. 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 and bias results.

Lemma 1 is conditional on all regressors up to x_T in a finite sample. In the corollary the sample size is increased without bounds to show the problem persists when at least one of the regressors includes a deterministic trend.

Finally note that Assumptions 1–3 do not exclude higher order deterministic trends. It is the inclusion of the linear trend t in x_t that leads to the high correlation and results in the next section. In the Bitcoin application, we will exploit this feature and include a quadratic trend in the log-price for exuberant times.

3. Forecasting levels

There are two obvious methods for predicting the future level of the series.[1] One method is to predict y_{T+h}, exponentiate, and correct for the bias as desired. The second is to predict growth and apply this to the last observation Y_T, but taking care of the correlation between Y_T and the estimator.

Predicting growth is essentially straightforward, and is only complicated by the non linear function involved. Growth in the model does not depend on the level Y. It is a function only of the disturbance

[1] Smoothing methods could of course be considered like McAleer and coauthors do, e.g. Lim and McAleer (2001), but moving average or exponential smoothing are not based on the model under consideration.

term, parameters, and exogenous variables. At time T, growth over the next h periods equals: $G_{h,T} = 100(\exp\left\{\Delta_h x'_{T+h}\beta + \sum_{i=1}^{h} \varepsilon_{T+i}\right\} - 1)$, where $\Delta_h x_t = x_{t+h} - x_t$, and has expectation:

$$E\left[G_{h,T}\right] = 100(\exp\left\{\Delta_h x'_{T+h}\beta + \tfrac{h}{2}\sigma^2\right\} - 1), \tag{9}$$

A simple application of Van Garderen (2001) gives the following theorem

Theorem 1. *The* ***Exact Minimum Variance Unbiased Growth Predictor*** *is given by:*

$$\hat{G}_{h,T} = 100(\exp\{\Delta_h x'_{T+h}\hat{\beta}\}\, {}_0F_1(m, m\tfrac{1}{2}(h - a_{T+h})\hat{\sigma}^2) - 1), \tag{10}$$

where $\hat{\beta}$ and $\hat{\sigma}^2$ are the OLS estimators of the model in first differences, $m = (T-k)/2$ and $a_{T+h} = \Delta_h x'_{T+h}(\Delta X'\Delta X)^{-1}\Delta_h x_{T+h}$. The ***Exact Minimum Variance Unbiased Estimator of the Variance*** *is given by*

$$\widehat{var}(\hat{G}_{h,T}) = 100^2 \exp\{2\Delta_h x'_{T+h}\hat{\beta}\} \times [{}_0F_1(m; m\tfrac{1}{2}(h - a_{T+h})\hat{\sigma}^2)^2 - {}_0F_1(m; m(h - 2a_{T+h})\hat{\sigma}^2)]. \tag{11}$$

The ${}_0F_1$-confluent hypergeometric function is defined in Appendix A as an infinite sum and can be thought of as a generalization of the exponential function. See also Abadir (1999) who reviews the use of hypergeometric functions in economics. The proof in Appendix A essentially uses the result that $E[{}_0F_1(m, m\hat{\sigma}^2 z)] = \exp\{\sigma^2 z\}$, see Van Garderen (2001). The fact that the density is complete in a statistical sense (e.g., Lehmann and Casella, 2006) leads to uniqueness.

Forecasting future levels is more involved because future levels are not simply a function of parameters, but also depend on the current level of the series. This current level is highly correlated with the parameter estimates, and ignoring this dependence can lead to significant biases. For this reason the obvious estimator based on the unbiased growth predictor:

$$\hat{Y}_{T+h} = Y_T(1 + \hat{G}_{h,T}/100), \tag{12}$$

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

$$\check{Y}_{T+h} = Y_T \exp\{\Delta_h x'_{T+h}\hat{\beta} + \tfrac{h}{2}\hat{\sigma}^2\}, \tag{13}$$

which makes it clear that the correlation between current level Y_T and $\hat{\beta}$ will cause problems for inference in general, and for unbiasedness in particular. Level forecasts based on growth estimates are generally not unbiased for three reasons: (a) the non linear exponential transformation, (b) the parameter uncertainty, (c) the fact that Y_T and the estimator $\hat{\beta}$ are highly correlated. We will therefore have to bias-correct predictors for these factors, or consider estimators for the expectation of Y_{T+h} directly.

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:

$$E[Y_{T+h}] = \exp\{x'_{T+h}\beta + \tfrac{T+h}{2}\sigma^2\}, \tag{14}$$

$$E[Y_{T+h}|Y_T] = Y_T \exp\{\Delta_h x'_{T+h}\beta + \tfrac{h}{2}\sigma^2\}. \tag{15}$$

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 forecasts that are both unbiased:

Proposition 1. *The Unconditional Level Forecast*

$$F_{T+h} = \exp\{x'_{T+h}\hat{\beta}\}\, {}_0F_1(m; m\hat{\sigma}^2 z_{T+h}), \quad with,$$
$$z_{T+h} = \tfrac{1}{2}(T + h - x'_{T+h}\left(\Delta X'\Delta X\right)^{-1} x_{T+h}), \tag{16}$$

is unbiased.

Proposition 2. *The Conditional Level Forecast*

$$F_{T+h|T} = Y_T \exp\{\Delta_h x'_{T+h}\hat{\beta}\}\, {}_0F_1(m; m\hat{\sigma}^2 z_{T+h|T}), \qquad with,$$
$$z_{T+h|T} = \tfrac{1}{2}(h - 2x'_T\left(\Delta X'\Delta X\right)^{-1}\Delta_h x'_{T+h} - \Delta_h x'_{T+h}(\Delta X'\Delta X)^{-1}\Delta_h x_{T+h}), \tag{17}$$

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 $(T+h)$ originates from 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 originates from the h disturbance terms ε_t in the future between time T and $T+h$, the second term $-2x'_T\left(\Delta X'\Delta X\right)^{-1}\Delta_h x'_{T+h}$ derives from the covariance between y_T and $\Delta_h x'_{T+h}\hat{\beta}$, and the third term $x'_{T+h}\left(\Delta X'\Delta X\right)^{-1} x_{T+h}$ is correcting for parameter uncertainty in the estimate $\Delta_h x'_{T+h}\hat{\beta}$.

Although the predictors can be very different in practice and are derived from very different perspectives, they are in fact identical if a drift term is included as is proved in Appendix A:

Theorem 2. *If the model includes a deterministic trend, such that ΔX includes a constant term, then:*

$$F_{T+h} = F_{T+h|T}. \tag{18}$$

This remarkable equality between the conditional and unconditional predictors can be explained by noting that the unconditional predictor F_{T+h} equals Y_T for the limiting case $h = 0$ (but only if a constant is included).[2] So, although we are averaging over all possible sample paths for $\{Y_t\}$, each prediction based on any realized sample path still goes through the last observation Y_T. The difference is that the conditional predictor $F_{T+h|T}$ is an explicit function of Y_T, whereas the unconditional predictor only depends implicitly on Y_T, but behaves exactly the same.

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. 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 Y_T only. 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 a predictor that is conditionally unbiased given Y_T. This holds generally as stated in the following theorem:

Theorem 3. *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) only $\{Y_T\}$.*

The proof is given in Appendix A, but (a) is obvious since any forecast is constant given $\{Y_1, ..., Y_T\}$ and with probability 1 does not equal the conditional expectation of Y_{T+h}. The more interesting part

[2] With $h = 0$ we have $x'_{T+0}\hat{\beta} = \iota'\Delta X(\Delta X'\Delta X)^{-1}\Delta X'\Delta y = \iota'\Delta y = y_T$ and hence $exp\{x'_T\hat{\beta}\} = Y_T$. $z_T = \frac{1}{2}(T - \iota'\Delta X(\Delta X'\Delta X)^{-1}\Delta X'\iota)) = 0$, and ${}_0F_1(m, 0) = 1$ and hence $F_{T+0} = Y_T$. For the conditional expressions note that by definition $\Delta_0 x_t = x_{t+0} - x_t = 0$.

(b) is less obvious but also follows from a degeneracy in the conditional distribution of $\hat{\beta}$. The proof determines a condition that can only be satisfied when $h = 0$.

The results in theorems 2 and 3 depend on the assumption that the initial condition is fixed. As a referee pointed out, however, if the initial condition is assumed to be $o_p(T^{1/2})$, an assumption sometimes used in the unit root literature, the distribution of the estimated drift parameter conditional on Y_T, is no longer degenerate and the construction of conditional unbiased forecast might be possible.

The same referee pointed out that the distribution of the h-step ahead forecast could also be obtained by simulation methods and, given the complete structure of the model here, allow for a more general setup, including conditional heteroskedasticity. There are also bootstrap techniques available that capture the estimation uncertainty yet keep Y_T fixed. If we condition on the initial and last values of Y, no conditionally unbiased forecast exists but if we condition on the initial value only, an unconditional unbiased forecast can be constructed. But it is also possible to combine the estimates in each re-sample, not with the last value of the re-sampled data, but with the last value of the original sample, see, e.g., Pascual et al. (2004, 2006), Fresoli et al. (2015). This procedure would destroy the correlation between the estimator and the last value of Y but is expected to produce a MSFE that is lower than the unconditional MSFE.

Theorem 4. *If the model includes a deterministic trend, such that ΔX includes a constant term, then the unbiased level forecast has mean squared forecasting error, $E[(F_{T+h} - Y_{T+h})^2]$:*

$$MSFE(F_{T+h}) = \exp\{2x'_{T+h}\beta + 2(T+h)\sigma^2\} \times \left[\exp\{\sigma^2(h + a_{T+h})\}_0F_1(m; \sigma^4 z^2_{T+h}) - 1\right], \tag{19}$$

with $a_{T+h} = \Delta_h x'_{T+h}(\Delta X'\Delta X)^{-1}\Delta_h x_{T+h}$. It can be estimated unbiasedly as:

$$\widehat{MSFE}(F_{T+h}) = F^2_{T+h} - \exp\{2x'_{T+h}\hat{\beta}\}_0F_1(m; -2m\hat{\sigma}^2(h + a_{T+h})). \tag{20}$$

The MSFE here is unconditional. In principle, one would like a conditional expression given $\{Y_1, ..., Y_T\}$ or $\{Y_T\}$, and we did derive such a theoretical expression, but it is not useful in practice since it cannot be used to indicate the accuracy of the forecast. The reason is that no conditionally unbiased estimates for the terms F^2_{T+h} and $-2F_{T+h}Y_{T+h}$ exist. This can be proved in the same fashion as Theorem 3.

Generally a constant would be included in the log linear regression. If no constant is included then F_{T+h} and $F_{T+h|T}$ are different and so are their MSFE's (and estimators thereof) which are not equal to Theorem 4. The proof in Appendix A.2 is readily adapted, however, but the results would include three terms for the MSFE and its estimate in both cases.

If no constant is included, the conditional $F_{T+h|T}$ should be used. It naturally satisfies $F_{T|T} = Y_T$ always, whereas $F_T \neq Y_T$ if no constant is included. The MSFE of unconditional F_{T+h} is much larger than that of the conditional predictor $F_{T+h|T}$ (unreported result show it to be comparable to the second consistent estimator discussed below). The situation can be compared to the simple random walk model $y_t = y_{t-1} + \varepsilon_t$ where the forecasts $f_{T+1} = 0$ and $\tilde{f}_{T+1} = a_T \cdot (y_T/T)$, with $a_T = T$ for conditional expression or $T + 1$ for forecasting from a model with a constant, are both unconditionally unbiased but $\tilde{f}_{T+1}$ has a much lower MSFE.

3.1. Alternative level forecasts

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

Definition 1. Alternative forecasts: (A) **Approximate unbiased forecast**:

$$F^{apu}_{T+h} = Y_T \exp\{\Delta_h x'_{T+h}\hat{\beta} + \hat{\sigma}^2 z_{T+h|T}\}, \qquad with,$$
$$z_{T+h|T} = \tfrac{1}{2}(h - 2x'_T \left(\Delta X' \Delta X\right)^{-1} \Delta_h x'_{T+h} - \Delta_h x'_{T+h}(\Delta X' \Delta X)^{-1} \Delta_h x_{T+h}), \tag{21}$$

and approximate unbiased estimator of the mean squared forecasting error:

$$\widetilde{MSFE}(F^{apu}_{T+h}) = \left(F^{apu}_{T+h}\right)^2 - \exp\{2x'_{T+h}\hat{\beta} - 2\hat{\sigma}^2(h + a_{T+h})\}. \tag{22}$$

(B) **Growth based forecast**:

$$F^{grow}_{T+h} = Y_T \ (1 + \frac{\hat{G}_{h,T}}{100}). \tag{23}$$

(C) **Naive forecast**:

$$F^{naiv}_{T+h} = \exp\{x'_{T+h}\hat{\beta}\}. \tag{24}$$

(D) **Consistent forecasts**:

$$F^{cons1}_{T+h} = Y_T \exp\{\Delta_h x'_{T+h}\hat{\beta} + \frac{h}{2}\hat{\sigma}^2\}, \tag{25}$$
$$F^{cons2}_{T+h} = \exp\{x'_{T+h}\hat{\beta} + \frac{T+h}{2}\hat{\sigma}^2\}. \tag{26}$$

The approximate unbiased forecast F^{apu}_{T+h} is constructed by approximating the hypergeometric functions in the exact unbiased forecast and its MSFE, using the exponentiated $\hat{\sigma}^2 z_{T+h|T}$, which corrects for uncertainty in Y_{T+h} caused by disturbances ε_t, uncertainty in $\hat{\beta}$, and the correlation between Y_T and $\hat{\beta}$, and using $(h + a_{T+h})$ from Theorem 4 for the MSFE estimator. We will show below that in practice these are very close to the exact level and MSFE estimators.

The growth-based forecast multiplies the current level with predicted growth, ignoring the correlation between Y_T and $\hat{G}$. Any of the growth predictors could be used, but in the applications below we have used the exact unbiased growth predictor to isolate the correlation effect between Y_T and $\hat{G}$. The naive growth predictor can also be used, especially since it can have the lowest MSFE, but is in fact identical to the naive level forecast.

The naive forecast simply substitutes the estimated parameters in the model function for Y_{T+h}. It ignores the uncertainty in future ε_{T+i}'s, $\hat{\beta}$, and the correlation between Y_T and $\hat{\beta}$. We could also think of a conditional version $Y_T \exp\{\Delta_h x'_{T+h}\hat{\beta}\}$, but this equals F^{naiv}_{T+h} identically if a time trend is included in x_t, similar to Theorem 2.

The consistent forecasts are based on the idea of substituting consistent estimators in the expressions for the conditional and unconditional mean of Y_{T+h}, respectively. They are not actually consistent since not even the bias goes to zero asymptotically. They take into account the increased expectation due to the disturbances, but ignore the increased bias due to estimation uncertainty. Moreover, F^{cons1}_{T+h} ignores the correlation between $\hat{\beta}$ and Y_T and F^{cons2}_{T+h} does not equal Y_T for $h = 0$. The conditional and unconditional versions are very different here, even if a constant is included in the regression. The forecast F^{cons2}_{T+h}, was used in Garderen et al. (2000) in their prediction based criterion for deciding between aggregate and disaggregate non linear models. This choice turns out not to be optimal for forecasting as we will see below. Another alternative would be $\exp\{x'_{T+h}\hat{\beta} + \frac{h}{2}\hat{\sigma}^2\}$, which Bårdsen and Lütkepohl (2011) and Lütkepohl and Xu (2012, Eq. (5)), Lütkepohl and Xu (2012, Eq.(5)), (Lütkepohl and Xu, 2012, Eq. (5)) refer to as the optimal predictor following Granger and Newbold (1976), but optimality is only if the parameters are known and it equals F^{cons1}_{T+h} identically if a time trend is included in x_t, similar again to Theorem 2.

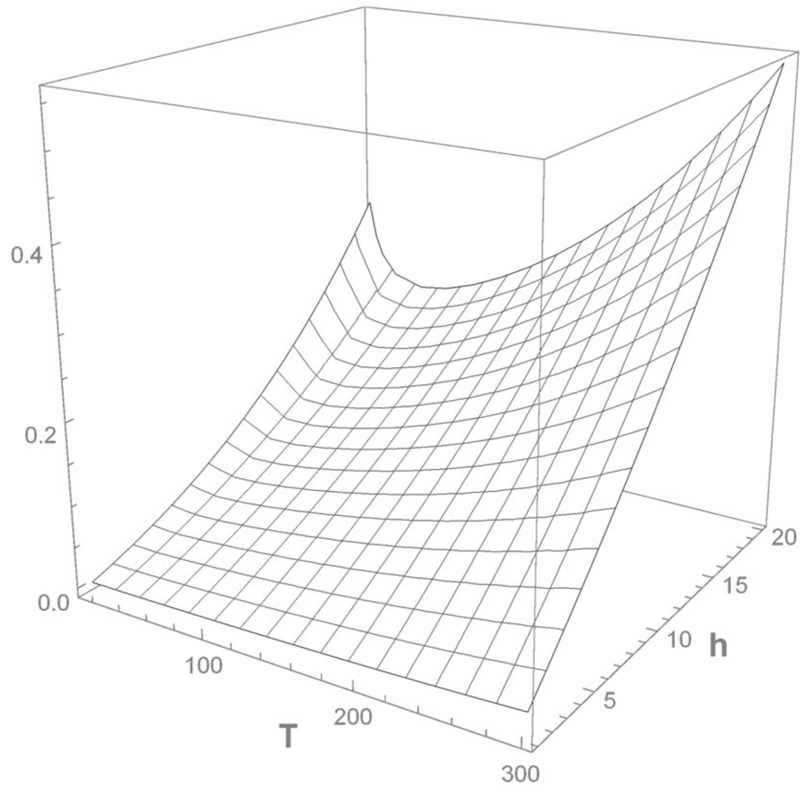

Figure 1. MSFE/$E[Y_{T+h}]^2$ in loglinear unit root model with drift, T : 15-200, $h : 1 - 20.\sigma = 0.1$.

3.2. Forecast horizon and sample size

We can use Theorem 4 to show the effects of increasing forecasting horizon h in relation to the available number of observations T on the forecasting accuracy. First note that in general the level of the series Y_{T+h} tends to increase over time when the drift parameter is large enough.[3] So the first effect of increasing T or h is an increase in the squared expected level for both Y and F in period $T + h$: $E[Y_{T+h}]^2 = E[F_{T+h}]^2 = \exp\{2x'_{T+h}\beta + (T + h)\sigma^2\}$. We would like to distinguish this level effect from the second effect which relates to the squared difference in Y and F after dividing by the expected level: $\text{MSFE}(F_{T+h})/E[Y_{T+h}]^2$.

In order to illustrate this, consider the basic loglinear unit root model with drift to obtain:

$$\frac{MSFE(F_{T+h})}{E[Y_{T+h}]^2} = exp\{(T+h)\sigma^2\}\left[\exp\{\sigma^2 h(1+h/T)\}{}_0F_1(m; \frac{1}{4}\sigma^4\,(h(1+h/T))^2) - 1\right]. \qquad (27)$$

The term in square brackets is governed by the forecasting horizon h when T is large, but T still plays an important role in the first term. Both h and T will increase the MSFE, even if divided by the expected level squared. This is shown in Figure 1.

When T is increasing, the ratio $\text{MSFE}(F_{T+h})/E[Y_{T+h}]^2$ is initially decreasing, but subsequently increases quickly when h is larger.[4] In order to explain the approximate linearity in h that can be seen in the graph, note that for small σ^2 and T fixed, the ratio in this loglinear unit root with drift model is

[3]Unless $x'_t\beta < -\frac{t}{2}\sigma^2$, which can happen with e.g. a negative drift term, and does happen for the mining sector in the application below.

[4]Not for $h = 0$, since $F_T = Y_T$ and the measure is identically 0.

approximately:

$$\frac{MSFE(F_{T+h})}{E[Y_{T+h}]^2} \overset{appr}{=} \sigma^2(h+\frac{h^2}{T})+\sigma^4(hT+\frac{2}{5}h^2+\frac{1}{2}\frac{h^2}{T}). \tag{28}$$

The figure further shows that, by letting h and T grow proportionally, as in Sampson (1991) or Clements and Hendry (1998), the increased uncertainty about the future dominates the increased accuracy in the estimation of unknown parameters, even when correcting for the fact that the level is exponentially increasing by dividing by $E[Y_{T+h}]^2$. Related is also West (1996), whose Assumption 4 allows $lim_{T+h\to\infty}h/T=0$, and parameters are estimated by regression functions, but interestingly, his Theorem 4.1 does not hold.[5]

The MSFE tends to infinity as T and/or h go to infinity (unless growth is sufficiently negative). In order to investigate the effects of the forecast horizon, we compare the unbiased forecast to alternative forecasting methods. We use a simple model for the price of Bitcoins to illustrate the point, but do not comment on the nature of the possibly rational bubbles.

4. Bitcoin price level

Crypto currencies are typically reported using price levels. Public interest is certainly in the extreme hikes in the price e.g., of a Bitcoin that rose above \$67000 in November 2021 from around \$7000 in May 2020. In December 2017 the price rose above \$19000 from less than \$400 in 2015. Modeling on the other hand, using log-differences is more appropriate than levels in periods leading up to these peaks. The first difference in log price appears trend stationary in the 24 months leading up to the peaks a the end of 2017 and 2020. Only a constant and trend are used in the log-difference equation, as a technical analyst might do, such that ΔX is obviously exogenous. Note that inclusion of the trend implies that the model includes a higher order trend in the log price and the price level itself. Recall that assumptions 1–3 do not exclude higher order trends and the exact results in Section 3 remain valid. Using the 25 observations from December 2015 - December 2017, we obtain the following results:

$$\begin{array}{lcccccc} \Delta\log(P_t) & = & -0.0155 & + & 0.00621 & t & + \; e_t \\ & & (0.0278) & & (0.00186) & & \end{array}$$

$T=25,\;\; se=0.06738,$

Jarque-Bera: $JB=0.478:p\text{-value}=0.787$

Breusch-Godfrey: $F=0.755:p\text{-value}=0.482$; $T\cdot R^2=1.679:p\text{-value}=0.432$

The null of normally distributed and uncorrelated disturbances cannot be rejected. Also for various other epochs that do not include events like the 2018 cryptocurrency- or other crashes, show similar results. So, although the model is unsophisticated, it does include a quadratic trend in log prices and passes specification tests that examine two assumptions that are essential for the theoretical results derived in this paper. It may serve as a building block in a more elaborate model that does, e.g., include crashes. More importantly, it may also be used, as a technical analyst might, to describe behavior and forecast during an exuberant period and it is interesting to investigate the short term predictive performance.

[5]Taking as moment functional $f(.)$ the one step ahead MSFE, (West 1996 uses R for T and his π becomes $lim_{T+h\to\infty}h/T=0$) we have for the theoretical expectation $E[f_{T+1}]=E(Y_{T+1}-E\left[Y_{T+1}\right])^2=E(Y_{T+1}-E\left[F_{T+1}\right])^2=\exp\{2x'_{T+1}\beta+\sigma^2(T+1)\}\left(\exp\{\sigma^2(T+1)\}-1\right)$. The sample MSFE $\bar{f}\equiv\frac{1}{P}\sum_{t=T+1}^{T+1}(Y_t-F_t)^2=(Y_{T+1}-F_{T+1})^2$ has expectation $E\left[\bar{f}\right]=\exp\{2x'_{T+1}\beta+2(T+1)\sigma^2\}\left\{\exp\{\sigma^2(1+a_{T+1})\}{}_0F_1(m;\sigma^4 z^2_{T+1})-1\right\}$. Hence the expectation of the difference equals: $E[\bar{f}-Ef_{T+1}]=\exp\{2x'_{T+1}\beta+2\sigma^2(T+1)\}\left[1-\exp\{-\sigma^2(T+1)\}-\exp\{\sigma^2(1+a_{T+1})\}{}_0F_1(m;\sigma^4 z^2_{T+1})+1\right]$. The second term in square brackets goes to 0 and the third goes to 1 and the third term does not tend to 2. So unless $(x'_{T+1}\beta+\sigma^2(T+1))$ goes to $-\infty$, $\bar{f}-Ef_{T+1}$ cannot be centered around 0. When the number of terms P is increasing, as in West's Assumption 4, this problem becomes worse as we increase h and set $P=h$.

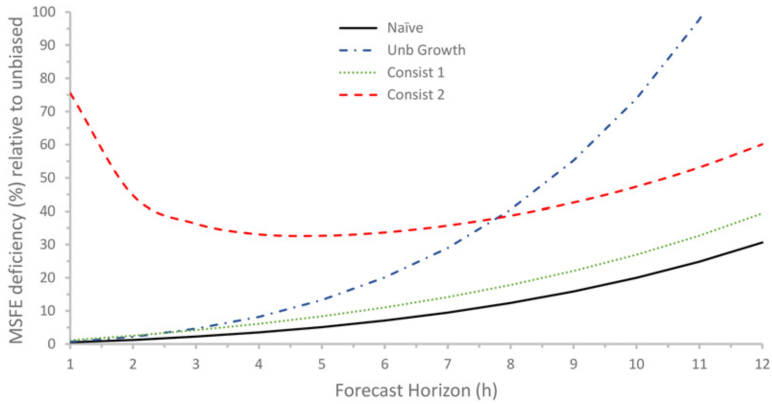

Figure 2. MSFE deficiency for horizons $h = 1 - 12$. Based on 100.000 replications using Bitcoin parameter estimates: $\sigma = 0.06738, \beta' = (-0.0155, 0.00621)$.

Using these parameter values, we simulate the process and the different forecasts up to $h = 12$ periods ahead. Figure 2 graphs how much worse the MSFE are of the alternative forecasts, growth based, naive, and consistent, relative to the exact unbiased forecast in terms of MSFE.

The figure shows that as we forecast further into the future the alternative forecasts become progressively worse. The extremely poor performance of the second consistent forecast is due to the fact that for the limiting case with $h = 0$, the predictor does not equal Y_T as the other forecasts, but the factor $e^{\hat{\sigma}^2 T/2}$ is used, which obviously worsens as T increases, e.g., with $T = 100$ and $h = 2$, it is more than 450% worse. Remarkable are also the poor performance of the forecast based on the optimal growth forecast which ignores the correlation and the relative good performance of the naive forecast that ignores both estimation uncertainty and correlation with the last observation.

5. Sensitivity

The exact unbiased forecast was derived under restrictive assumptions including normality and a unit root. Performance is expected to deteriorate when the conditions are not met and the data generating process (DGP) diverges from the model defined by (2) and (3). We investigate robustness to (i) deviations from the unit root and (ii) deviations from normality. We consider (i) both local to unity *stationary* processes, as well as mildly *explosive* processes for u by using $\rho_T = 1 + c/T$ in $u_t = \rho_T u_{t-1} + \varepsilon_t$ and (ii) both left- and a right skewness by simulating the innovations ε_t from a standardized χ^2_3 distribution. In all models we included a constant such that $\Delta x_{t,1} = 1$, specify a first order autoregressive (AR(1)) process for $x_{t,2} = \tilde{\rho}_T x_{t-1,2} + \epsilon_t$, and initialize all series at 0.

Table 1 reports results for sample size $T = 25, 100$, forecast horizons $h = 1, 4, 12$, $\rho_T = 1 + c/T$ with $c = -4$ (stationary), $c = 0$ (unit root), and $c = 1/4$ (mildly explosive) and also $\tilde{\rho}_T = 1 + \tilde{c}/T$. The middle three columns with $c = 0 = \tilde{c}$ are for the correctly specified model that satisfies (2), (3) and further assumptions of the model. The new forecast, although no longer unbiased, still outperforms the other forecasts for mild deviations from the unit root assumption. Table 1 reports a limited number of cases, but the "unbiased" forecast is also better for values $c \in \{-4, -2, 0, 0.25, 0.5\}$, $\tilde{c} \in \{-25, 0, 0.25, 0.5\}$, $T \in \{25, 100\}$ that are not reported here. The values in the other columns are rather typical. For one step ahead forecasts $h = 1$ the difference is small, since all forecasts take Y_T as point of departure, except the second consistent forecast which is so appalling it is not included in the tables. The approximate unbiased estimator is always less than 1% worse than the "exact unbiased" and the zero columns are dropped from the table. MSFE deficiencies for $h = 12$ vary between 7% and 23% for the naive forecast, which is always better in the cases considered than the growth based forecast which has deficiency varying between 10% and 76%. The first consistent forecast deficiency is generally worse and varies between 22% and 48%.

Table 1. MSFE deficiency: Near unit root.

in %	Stationary $c=-4$, $\tilde{c}=-25$			Unit Root $c=0$, $\tilde{c}=0$			Explosive $c=1/2$, $\tilde{c}=1/4$		
$T=25$	$Y_T\hat{G}$	*Naiv*	$Cons_1$	$Y_T\hat{G}$	*Naiv*	$Cons_1$	$Y_T\hat{G}$	*Naiv*	$Cons_1$
$h=1$	1	1	2	1	1	3	1	1	2
4	4	2	8	8	6	13	6	4	10
12	27	7	27	46	23	48	29	13	30
$T=100$									
$h=1$	0	0	1	1	1	3	1	1	2
4	2	2	7	5	5	12	4	4	9
12	10	7	22	21	16	39	15	12	29

MSFE deficiency in %, relative to *"unbiased"* forecast. Based on 100.000 replications. $u_t=\rho_T u_{t-1}+\varepsilon_t$, $\rho_T=1+c/T$, $x_{t,1}=t$, $x_{t,2}=\tilde{\rho}_T x_{t-1,2}+\epsilon_t$, $\tilde{\rho}_T=1+\tilde{c}/T$. *y* Middle three columns: model correctly specified $\because$ $c=0$, $\beta=(0.2,1)'$, $\sigma_u=0.1$

Table 2. MSFE deficiency: Skewness.

in %	*Left*						*Right*					
$T=$	25			100			25			100		
	$Y_T\hat{G}$	*Naiv*	$Cons_1$	$Y_T\hat{G}$	*Naiv*	$Cons_1$	$Y_T\hat{G}$	*Naiv*	$Cons_1$	$Y_T\hat{G}$	*Naiv*	$Cons_1$
$h=1$	3	4	6	3	3	6	0	1	1	1	1	2
4	12	18	28	10	12	24	2	3	5	3	4	8
12	55	122	124	23	31	59	9	19	18	12	15	27

MSFE deficiency in %, relative to *"unbiased"* forecast. Right skewed: $u\sim\sigma_u\left(\chi_3^2-3\right)/\sqrt{6}$: *mean* $=0$, *standaard deviation* $=\sigma_u$. Left skewed: $-u$, same draws. Based on 100.000 replications. $\beta=(0.2,1)'$, $\sigma_u=0.1$, $x_{t,1}=t$, $x_{t,2}=0.6x_{t-1,2}+\varepsilon_t$

When $T=25$ then for values of $c\leq-10$ or $c\geq0.75$ it is no longer true that the *"unbiased"* estimator is generally better as shown in Appendix B which reports some further results. Tables B1 and B2 include cases where the naive forecast is better, and even the growth and consistent forecasts can be marginally better. For example, for stationary values $c=\tilde{c}=-10$ and $T=25$, such that $\rho_T=\tilde{\rho}_T=0.6$, the naive predictor is 6% better than the unbiased forecast when $h=12$. The growth and first consistent forecasts are still worse (20% and 22%, respectively, when $h=12$). For the mildly explosive $c=\tilde{c}=0.75$ the naive forecast is 3% better when $h=12$ and $T=25$, but the growth and first consistent forecast are not. Cases where all three alternatives are better for $h=12$ occur e.g. with $T=25$ and $c=1,\tilde{c}=-10$ or 0.75 : the first consistent and growth based forecasts are 17% better and the naive 10% better. Values $\rho_T=\tilde{\rho}_T=0.6$ are quite far removed from the assumed GDP as are $\rho_T=\tilde{\rho}_T=1.03$.

Skewed distributions are used to investigate deviations from normality as left and right skewness might be expected to affect performance differently because of the non linear transformation. We use a standardized χ_3^2 distribution with skewness $\sqrt{8/3}$ to illustrate that this is indeed the case. Table 2 shows that the "unbiased" forecast, although no longer unbiased, is still better than the other forecasts under left skewness, and much better under right skewness for the parameter values shown. Even the naive forecast is more than 50% worse in terms of MSFE than the "unbiased" forecast when $T=25$. A stationary AR(1) process was chosen for the second regressor, but further (unreported) experiments show that this holds whether the second regressor is stationary, a random walk, or mildly explosive.

Performance of the new unbiased forecast is reasonably robust relative to the other forecasts that seem reasonable alternatives under the loglinear unit root specification.

6. Sectoral production

In this section, we make a comparison of the various predictors for the levels in eight industrial sectors. This serves two purposes. First, although there might be important theoretical differences between the various predictors, it could be that the differences in practice are not important. This seems to be the

Table 3. Sectoral level forecasts.

Levels	*Actual Level*	*Exact Unbiased*	*Approx. Unbiased*	$Y_T\hat{G}$	*Naive*	*Cons*$_1$	*Cons*$_2$
1956-1980	$h = 5$						
Mean	2.629	2.628	2.628	2.640	2.640	2.646	2.676
Bias		0.000	0.000	0.011	0.012	0.017	0.048
MSFE deficiency		*	0.0%	1.0%	1.0%	1.6%	6.9%
1956-2005	$h = 10$						
Mean	7.583	7.585	7.585	7.649	7.688	7.701	7.968
Bias		0.002	0.002	0.067	0.105	0.119	0.385
MSFE deficiency		*	0.0%	2.3%	4.2%	4.9%	28.8%

*indicates best in all sectors. Based on 100.000 replications

case for growth predictions, but for level forecasts we do find relevant practical differences. Second, unbiasedness was motivated by the choice of MSFE as optimality criterion. Although the conditional mean theoretically minimizes the MSFE when the parameters are known, this is not necessarily true if the conditional mean is estimated, either unbiasedly or by substituting estimated parameters.

We have applied the different forecasting methods using the same model and set of data as Garderen et al. (2000). The sectoral models are Cobb-Douglas production functions with stochastic technology of the type given in Eq. (1), and 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. Tests for normality and functional form do not reject the model. The difference with the original application is that we report estimates of the model without imposing constant returns to scale restrictions and for prediction they used a consistent predictor, *Cons*$_1$ in the tables here, since the exact unbiased predictor was not yet available. Their focus was on the problem of cross-sectional aggregation of non linear micro behavioral relations. Our current contribution is that the much better accuracy of the exact unbiased predictor improves distinguishing aggregate- from disaggregate models.

We simulated the model using the estimated parameters and keeping the regressors fixed. For observations after 1995, we do not have actual observations, so the regressors where dynamically generated using a VAR(1) with parameter values equal to estimates based on 1955-1995 actual data. The dependent variables Y were then simulated according to the model and keeping the X variables fixed. All series were initialized to start at zero. Tables 1 and 2 report the unweighted average over all sectors to indicate the general tendencies. The full results for all sectors and two sample periods are given in Appendix B. The unweighted parameter estimates average over the eight sectors are for the period: 1956–1980: $\bar{\delta} = 0.014$, $\bar{\beta}_{\ln L} = 0.241$, $\bar{\beta}_{\ln K} = 0.430$, $\bar{\sigma} = 0.027$, and for the period :1956–2005: $\bar{\delta} = 0.024$, $\bar{\beta}_{\ln L} = 0.373$, $\bar{\beta}_{\ln K} = 0.258$, $\bar{\sigma} = 0.036$.

6.1. Results

Table 3 gives the results for the level forecasts of sectoral production based on the same model. The conditional unbiased forecast and unconditional unbiased forecast 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.

Table 4 shows the relative deficiencies in growth predictors and level forecasts for the period 1956-2005 for each sector. The naive growth predictor is best in every sector. This in itself is an interesting result but the differences with the other predictors are small and less than 0.7% for the exact unbiased predictor. When forecasting levels the naive estimator is no longer best and appreciably worse than the exact unbiased forecast and even worse than the growth-based predictor. The exact unbiased forecast is best in all sectors. The worst performance for the alternative forecasts is in the mining sector where the

Table 4. Relative Deficiency versus best (*) in %

	Growth(*MSE*)			Level (*MSFE*)				
1956-2005	*Exact Unb.*	*Naive*	*Cons*	*Exact Unb.*	$Y_T\hat{G}_{h,T}$	*Naive*	$Cons_1$	$Cons_2$
Agricult	0.2	*	0.7	*	2.7	5.1	5.8	34.2
Mining	0.7	*	2.0	*	6.4	12.0	13.9	86.5
Manufact	0.0	*	0.1	*	0.6	1.2	1.3	7.6
Energy	0.6	*	1.4	*	3.6	6.4	7.5	43.4
Construc	0.4	*	1.0	*	1.8	2.7	3.5	18.3
Transprt	0.4	*	1.0	*	1.8	2.8	3.5	18.1
Communic	0.1	*	0.3	*	1.0	1.9	2.5	12.1
Oth Serv	0.1	*	0.3	*	0.9	1.6	1.8	10.3
Average	0.3	*	0.8	*	2.3	4.2	4.9	28.8

* indicates best. h = 10. Approximate unbiased growth predictor and level forecast not reported since practically indistinguishable from exact unbiased

growth based forecast is 6.4% worse, the naive 12%, the first consistent 14%, and the second consistent 86% worse than the exact unbiased forecast.

The reason that the naive growth predictor performs well is that the non linear effects of future ε's and estimation uncertainty in $\Delta_h x'_{T+h}\beta$ partly cancel each other out. Following the comments below Theorem 1, we have $\hat{\sigma}^2$ times $\frac{1}{2}h = 5$ related to future ε's, and for parameter uncertainty $\hat{\sigma}^2$ times $\frac{1}{2}a_{T+h}$, which on average over the eight sectors is about -1.5. The naive estimator, does not use $\hat{\sigma}^2$ and its variance does therefore not contribute to the MSE.

When forecasting the level, we can use the three terms in $z_{T+h|T}$ associated with future ε's ($\frac{1}{2}h =$ 5 here), estimation uncertainty in $\Delta_h x'_{T+h}\beta$, (here $-\Delta_h x'_{T+h}(\Delta X'\Delta X)^{-1}\Delta_h x_{T+h}/2$: about -1.5 on average over the sectors, and the correction for the correlation between Y_T and $\Delta_h x'_{T+h}\hat{\beta}$ which is $-x'_T\left(\Delta X'\Delta X\right)^{-1}\Delta_h x'_{T+h} = -h = -10.$[6] This correlation term is therefore twice as large as the correction for future ε's and much larger than the estimation uncertainty term. The correlation between the level and the estimator is therefore quantitatively the most important factor.

The correlation term is $-h$ for all T and remains important even asymptotically. Only the estimation uncertainty term goes to zero as T increases.

7. Conclusion

This article has highlighted some of the issues involved when forecasting from log-linear unit root models with exogenous variables. Non linearity of the transformations of disturbances and estimators cause bias, which is well known, but more important is the high correlation between the last observation and parameter estimates. This correlation is often ignored, but is in fact larger than the effects of future disturbances and parameter uncertainty together as shown using $z_{T+h|T}$. The effect of parameter uncertainty decreases with sample size, but the correlation effect persists even asymptotically. This is a feature that is not limited to the stylized model employed in this paper but holds more generally. It leads to deeper questions concerning conditioning on past information. No conditionally unbiased estimator exists based on the regression estimates, even when conditioning on only those terms that enter the conditional mean.

There are of course various issues the article has not dealt with, including structural breaks, serial correlation, and heteroskedasticity, which is either trivial to deal with if the exact form is known, or analytically very complicated when it involves unknown parameters, or impossible to deal with exactly if its form is unspecified. The normality assumption is strong and we have not considered other distributions for which results may be derived using the Laplace inversion technique in Van Garderen

[6]The first column in ΔX is ι since x includes a deterministic trend. $\Delta X_i'\Delta X\ \left(\Delta X'\Delta X\right)^{-1} = (1, 0, ...,)$ and the first element of $\Delta_h x'_{T+h}$ is h. So $-\iota'\Delta X\ \left(\Delta X'\Delta X\right)^{-1}\Delta_h x'_{T+h} = -x'_T\left(\Delta X'\Delta X\right)^{-1}\Delta_h x'_{T+h} = -h$

(2001), but a distributional assumption is necessary to obtain exact finite sample results which will always be limited by this choice. We did not consider dynamic forecasts of exogenous variables, as e.g., Schmidt (1974). The additional uncertainty would increase the bias correction term but would require the exact distribution of the estimators in the VARMA. We have not (pre-)tested for unit roots or loglinearity. We have concentrated on specific models that were helpful in highlighting the issues raised and allowed explicit finite sample solutions.

The article is constructive in deriving the exact minimum variance unbiased growth predictor. More importantly, two exact unbiased level forecasts were derived based on the unconditional expectation of the process at time $T + h$, and the other based on the conditional expectation which is much better if no constant is included in the log-linear regression. A constant should be included in the regression however, in which case they are equal. The unbiased level forecast was better than five alternative forecasts and increasingly so as the forecast horizon increases. This superiority remains for models in a neighborhood of the normal log-linear unit root model considered here when the disturbances follow a local to unity process or mildly explosive process or when disturbances are skewed to the left or the right. The exact MSFE was used to investigate the effects of increasing forecast horizon and sample size. An exact unbiased estimator of this MSFE was also derived and can be reported as a measure of uncertainty.

Acknowledgments

I thank the Editor Esfandiar Maasoumi, Associate Editors, Robert Taylor, three anonymous referees, participants at the Econometric Study Group meeting in Bristol, Kevin Lee and Michael McAleer for hosting me when I was a young researcher.

References

Abadir, K. M. (1999). An introduction to hypergeometric functions for economists. *Econometric Reviews* 18(3):287–330. 10.1080/07474939908800447

Ariño, M. A., Franses, P. H. (2000). Forecasting the levels of vector autoregressive log-transformed time series. *International Journal of Forecasting* 16(1):111–116. 10.1016/S0169-2070(99)00025-4

Bårdsen, G., Lütkepohl, H. (2011). Forecasting levels of log variables in vector autoregressions. *International Journal of Forecasting* 27(4):1108–1115. 10.1016/j.ijforecast.2010.11.003

Bera, A. K., Mcaleer, M. (1989). Nested and non-nested procedures for testing linear and log-linear regression models. *Sankhyā: The Indian Journal of Statistics, Series B*, 51(2):212–224.

Binder, M., Pesaran, M. H. (1999). Stochastic growth models and their econometric implications. *Journal of Economic Growth* 4(2):139–183. 10.1023/A:1009802421114

Clements, M., Hendry, D. (1998). *Forecasting Economic Time Series*. Cambridge, UK: Cambridge University Press.

Dufour, J.-M. (1984). Unbiasedness of predictions from estimated autoregressions when the true order is unknown. *Econometrica* 52(1):209 10.2307/1911469

Franses, P. H., McAleer, M. (1998). Testing for unit roots and non-linear transformations. *Journal of Time Series Analysis* 19(2):147–164. 10.1111/1467-9892.00083

Fresoli, D., Ruiz, E., Pascual, L. (2015). Bootstrap multi-step forecasts of non-gaussian VAR models. *International Journal of Forecasting* 31(3):834–848. 10.1016/j.ijforecast.2014.04.001

Garderen, K. J. van, Lee, K., Pesaran, M. (2000). Cross-sectional aggregation of non-linear models. *Journal of Econometrics* 95(2):285–331. 10.1016/S0304-4076(99)00040-8

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association* 57(298):369–375. 10.1080/01621459.1962.10480665

Goldberger, A. S. (1968). The interpretation and estimation of cobb-douglas functions. *Econometrica* 36(3/4):464 10.2307/1909517

Granger, C. W. J., Newbold, P. (1976). Forecasting transformed series. *Journal of the Royal Statistical Society: Series B (Methodological)* 38(2):189–203. 10.1111/j.2517-6161.1976.tb01585.x

Lehmann, E. L., Casella, G.(2006). *Theory of Point Estimation*. Springer Science & Business Media. Berlin/Heidelberg, Germany.

Lim, C., McAleer, M. (2001). Forecasting tourist arrivals. *Annals of Tourism Research* 28(4):965–977. 10.1016/S0160-7383(01)00006-8

Lütkepohl, H., Xu, F. (2012). The role of the log transformation in forecasting economic variables. *Empirical Economics* 42(3):619–638. 10.1007/s00181-010-0440-1

Mayr, J., Ulbricht, D. (2015). Log versus level in VAR forecasting: 42 million empirical answers expect the unexpected. *Economics Letters* 126:40–42. 10.1016/j.econlet.2014.11.008
McAleer, M., Pesaran, M. (1986). Statistical inference in non-nested econometric models. *Applied Mathematics and Computation* 20(3-4):271–311. 10.1016/0096-3003(86)90008-1
Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Analysis*. New York: John Wiley and Sons.
Nelson, C. R., Plosser, C. R. (1982). Trends and random walks in macroeconmic time series. *Journal of Monetary Economics* 10(2):139–162. 10.1016/0304-3932(82)90012-5
Ng, S., Vogelsang, T. (2002). Forecasting autoregressive time series in the presence of deterministic components. *The Econometrics Journal* 5(1):196–224. 10.1111/1368-423X.t01-1-00081
Ogasawara, T., Takahashi, M. (1951). Independence of quadratic quantities in a normal system. *Journal of Science of the Hiroshima University, Series A (Mathematics, Physics, Chemistry)*. 15:1–9.
Pascual, L., Romo, J., Ruiz, E. (2004). Bootstrap predictive inference for ARIMA processes. *Journal of Time Series Analysis* 25(4):449–465. 10.1111/j.1467-9892.2004.01713.x
Pascual, L., Romo, J., Ruiz, E. (2006). Bootstrap prediction for returns and volatilities in GARCH models. *Computational Statistics & Data Analysis* 50(9):2293–2312. 10.1016/j.csda.2004.12.008
Phillips, P. C. (1979). The sampling distribution of forecasts from a first-order autoregression. *Journal of Econometrics* 9(3):241–261. 10.1016/0304-4076(79)90073-3
Rosanna, R. (1995). Technology shocks and cointegration in quadratic model of the firm. *International Economic Review*, 36(1):5–17.
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(1):67–76. 10.1002/jae.3950060106
Schmidt, P. (1974). The asymptotic distribution of forecasts in the dynamic simulation of an econometric model. *Econometrica* 42(2):303. 10.2307/1911980
Stock, J. H. (2009). Var, error correction and pretest forecasts at long horizons. *Oxford Bulletin of Economics and Statistics* 58(4):685–701. 10.1111/j.1468-0084.1996.mp58004006.x
Stock, J. H. (1997). Cointegration, long-run comovements, and long-horizon forecasting. *Econometric Society Monographs* 28:34–60.
Van Garderen, K. J. (2001). Optimal prediction in loglinear models. *Journal of Econometrics* 104(1):119–140. 10.1016/S0304-4076(01)00061-6
West, K. D. (1988). Asymptotic normality, when regressors have a unit root. *Econometrica* 56(6):1397. 10.2307/1913104
West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica* 64(5):1067. 10.2307/2171956

A. Appendix: Theory

The following notation will be used. Given initial values $y_0 = 0$ and $x_0 = 0$, the model can be written as:

$$\begin{aligned} y_t &= x_t'\beta + u_t, \qquad & t &= 1, 2, ..., T \\ u_t &= u_{t-1} + \varepsilon_t, \qquad & \varepsilon_t &\sim i.i.N(0, \sigma^2) \end{aligned} \tag{29}$$

$\iota_{(T)} = (1, ..., 1)'$, is a $T \times 1$ vector of ones, L is the first differencing matrix,

$$L = \begin{pmatrix} 1 & & & 0 \\ -1 & \ddots & & \\ & & & \\ & \ddots & \ddots & \\ 0 & & -1 & 1 \end{pmatrix}, \; with \;\; \left(L'L\right)^{-1} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & \cdots & T \end{pmatrix}.$$

$y_{(t)} = (y_1, y_2, ..., y_t)'$ is 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. We have $Ly = LX\beta + Lu$, and hence:

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

Because of assumptions 1 and 2 we may write:

$$\begin{aligned} y_t &= \sum\nolimits_{s=1}^{t} \Delta x_s'\beta + \sum\nolimits_{s=1}^{t} \varepsilon_s = x_t'\beta + S_t, \quad with: \\ S_t &= \sum\nolimits_{s=1}^{t} \varepsilon_s = \iota_{(t)}'\varepsilon_{(t)}. \end{aligned} \tag{30}$$

A.1. Proofs

Proof of Lemma 1.

$$\widehat{\beta} = \left(\Delta X'\Delta X\right)^{-1}\Delta X'\Delta y,$$
$$= \beta + \left(\Delta X'\Delta X\right)^{-1}\Delta X'\varepsilon \sim N\left(\beta, \sigma^2\left(\Delta X'\Delta X\right)^{-1}\right),$$
$$y_T = \sum\nolimits_{s=1}^{T}\Delta x_s'\beta + \sum\nolimits_{s=1}^{T}\varepsilon_s = \underbrace{x_T'\beta}_{E[y_T]} + \underbrace{S_T}_{\iota_{(T)}'\varepsilon}.$$

Hence:

$$Cov\left(y_T, \widehat{\beta}\right) = E\left[S_T\left(\left(\Delta X'\Delta X\right)^{-1}\Delta X'\varepsilon\right)'\right] = E\left[\iota_{(T)}'\varepsilon\varepsilon'\Delta X'\left(\Delta X'\Delta X\right)^{-1}\right],$$
$$= \sigma^2\iota_{(T)}'\Delta X'\left(\Delta X'\Delta X\right)^{-1} = \sigma^2 x_T'\left(\Delta X'\Delta X\right)^{-1}.$$

□

Proof of Proposition 1. Using Lemma 6 below, we have:

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

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

Proof of Proposition 2. The conditional expectation, given Y_T, of the conditional forecast can be derived using the results on conditional distributions given in the next subsection of this Appendix, which gives:

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

$$= Y_T\ \exp\{\ \Delta_h x_{T+h}'\beta + \sigma^2\tfrac{h}{2} + \Delta_h x_{T+h}'\left(\Delta X'\Delta X\right)^{-1}x_T\frac{(y_T - x_T'\beta)}{T} +$$
$$- \frac{1}{2T}\sigma^2\Delta_h x_{T+h}'\left(\Delta X'\Delta X\right)^{-1}x_T x_T'\left(\Delta X'\Delta X\right)^{-1}\Delta_h x_{T+h} - \sigma^2 x_T'\left(\Delta X'\Delta X\right)^{-1}\Delta_h x_{T+h}'\ \}.$$

Note that this expectation is not equal (*a.s.*) 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 $E[Y_T] = E[\exp\{y_T\}] = \exp\{x_T'\beta + \frac{T}{2}\sigma^2\}$. Hence:

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

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

and as a consequence:

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

showing that $F_{T+h|T}$ is unconditionally unbiased. □

Proof of Theorem 2. Using Assumption 1 we have $x_T = \iota_{(T)}'\Delta X$. 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,
$\iota_{(T)}'\Delta X\left(\Delta X'\Delta X\right)^{-1}\Delta X'\iota_{(T)} = \iota_{(T)}'P_{\Delta X}\iota_{(T)} = \iota_{(T)}'\iota_{(T)} = T$, since ΔX includes a column of ones (i.e. $\iota_{(T)}$) associated with the constant (the drift term).

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

Remains to be shown that $x'_{t+h}\hat{\beta} = \ln Y_T + \Delta_h x'_{T+h}\hat{\beta}$. By Assumption 1$x'_T = \iota'_{(T)}\Delta X$ and by definition of Δ_h we have $x'_{T+h} = x'_T + \Delta_h x'_{T+h}$ and hence:

$$\begin{aligned} x'_{T+h}\hat{\beta} &= x'_T\hat{\beta} + \Delta_h x'_{T+h}\hat{\beta}, \\ &= \iota'_{(T)}\Delta X \left(\Delta X'\Delta X\right)^{-1} \Delta X'\Delta y + \Delta_h x'_{T+h}\hat{\beta}, \\ &= y_T + \Delta_h x'_{T+h}\hat{\beta}, \end{aligned}$$

since $\iota'_{(T)}\Delta X \left(\Delta X'\Delta X\right)^{-1} \Delta X' = \iota'_{(T)}$. Hence $F_{T+h} = F_{T+h|h}$. □

Proof of Theorem 3. 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 $\widehat{\beta}|Y_T \sim N(\mu, \Sigma)$ 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:

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

Terms involving Y_T and σ^2 are immaterial 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:

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

Hence, we need to solve:

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

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 Lemma 2 below), then:

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

has general solution (as derived by Penrose, see e.g. Magnus and Neudecker 1988, 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:

$$B'B'^{+}\Delta_h x_{T+h} = \Delta_h x_{T+h}. \tag{31}$$

Lemma 2 below shows that $B'B'^{+} = \begin{pmatrix} 0 & 0 \\ 0 & I_{k-1} \end{pmatrix}$. The first element of $\Delta_h x_{T+h}$ equals h, and not 0, and the consistency condition (31) cannot be satisfied. No a exists such that $a'B = \Delta_h x'_{T+h}$. This implies that no conditional unbiased forecast based on the complete sufficient statistics exists. □

Comment. Increasing a increases $a'\hat{\beta}$ but the expectation associated with the drift term is reduced by an equal amount from the conditional expectation. This is most clearly seen when $x_t = t$, i.e. only a drift term is included in the regression. We then have $a'\beta + a'\left(\Delta X'\Delta X\right)^{-1} x_T\left(-x'_T\right)/T\,\beta = a\beta - a\beta = 0$ for any a.

Lemma 2. *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. Since $\left(\Delta X'\Delta X\right)^{-1}\Delta X'\Delta X = I_k$ and ι is the first column of ΔX we have $\left(\Delta X'\Delta X\right)^{-1}\Delta X'\iota = (1,0,...,0)'$. Hence $B = I_k - (1,0,...,0)'x'_T\frac{1}{T} = I_k - (1,0,...,0)'(1, \frac{1}{T}x'_{T,2:k})$. The second part of the Lemma is easily proved by verifying the four conditions of the Moore-Penrose inverse: (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) $B^+BB^+ = B^+$. □

Proof of Theorem 4. First write:

$$E[(Y_{T+h} - F_{T+h})^2] = E(\underset{(a)}{Y^2_{T+h}} \underset{(b)}{-2Y_{T+h}F_{T+h}} \underset{(c)}{+F^2_{T+h}}). \tag{32}$$

The first term (a) in Eq. (32) equals:

$$\begin{aligned} E[Y^2_{T+h}] &= E[\exp\{2x'_{T+h}\beta + 2\sum_{i=1}^{T+h}\varepsilon_i\}] = \exp\{2x'_{T+h}\beta + \frac{4}{2}(T+h)\sigma^2\}, \\ &= \exp\{2x'_{T+h}\beta + 2(T+h)\sigma^2\}. \end{aligned}$$

The second term (b): Since $\hat{\sigma}^2$ independent of the other terms present which involve ε, such as $\sum \varepsilon_t$ and $\hat{\beta}$, we have with the unconditional predictor $F_{T+h} = \exp\{x'_{T+h}\hat{\beta}\}\,{}_0F_1(m; m\hat{\sigma}^2 z_{T+h})$ with $z_{T+h} = \frac{1}{2}(T + h - x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_{T+h})$:

$$\begin{aligned} E[Y_{T+h}F_{T+h}] &= E[\exp\{x'_{T+h}\beta + \sum_{t=1}^{T}\varepsilon_t + \sum_{i=1}^{h}\varepsilon_{T+i} + x'_{T+h}\hat{\beta}\}{}_0F_1(m; m\hat{\sigma}^2 z_{T+h})], \\ &= E[\exp\{2x'_{T+h}\beta + \frac{h}{2}\sigma^2 + (\iota'\varepsilon_{(T)} + x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}\Delta X'\varepsilon_{(T)}) + \sigma^2 z_{T+h}\}, \\ &= \exp\{2x'_{T+h}\beta + \frac{h}{2}\sigma^2 + \\ &\quad + \frac{1}{2}(\iota' + x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}\Delta X')\sigma^2 I_T(\iota + \Delta X\left(\Delta X'\Delta X\right)^{-1}x_{T+h}) + \sigma^2 z_{T+h}\}, \\ &= \exp\{2x'_{T+h}\beta + \frac{h}{2}\sigma^2 + \frac{1}{2}\sigma^2(T + 2x'_T\left(\Delta X'\Delta X\right)^{-1}x_{T+h} + \\ &\quad + x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_{T+h}) + \sigma^2 z_{T+h}\}; \quad \text{since } \Delta X'\iota \text{ equals } x_T, \\ &= \exp\{2x'_{T+h}\beta + \frac{1}{2}\sigma^2(T + h + 2x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_T + \\ &\quad + x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_{T+h} + T + h - x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_{T+h})\}, \\ &= \exp\{2x'_{T+h}\beta + \frac{1}{2}\sigma^2(2T + 2h + 2x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_T)\}. \end{aligned}$$

So,

$$-2E[Y_{T+h}F_{T+h}] = -2\exp\{2x'_{T+h}\beta + 2(T+h)\sigma^2\} * exp\{-\sigma^2(T + h - x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_T)\}.$$

Now if a constant is included in the regression so that ι is the first column of ΔX then we have $\left(\Delta X'\Delta X\right)^{-1}\Delta X'\Delta X_i = e_i$ and

$$x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_T = x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}\Delta X'\,\iota = (T+h,...)(1,0,...,0)' \\ = T+h.$$

Hence:

$$-2E[Y_{T+h}F_{T+h}] = -2\exp\{2x'_{T+h}\beta + 2\sigma^2(T+h)\}.$$

For the third term (c) in Eq. (32) we use:

$$E\left[{}_0F_1(m;m\hat{\sigma}^2\, z_{T+h})^2\right] = \exp\{2\sigma^2 z_{T+h}\}{}_0F_1(m;\sigma^4\, z^2_{T+h}),$$

which was proved in Van Garderen (2001, proof Theorem 3). Hence, and by the independence of $\hat{\beta}$ and $\hat{\sigma}^2$, we obtain:

$$\begin{aligned}
E[F^2_{T+h}] &= E[\exp\{2x'_{T+h}\hat{\beta}\}\,{}_0F_1(m;m\hat{\sigma}^2\, z_{T+h})^2],\\
&= \exp\{2x'_{T+h}\beta + \frac{4}{2}\sigma^2 x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_{T+h} + 2\sigma^2 z_{T+h}\}{}_0F_1(m;\sigma^4\, z^2_{T+h}),\\
&= \exp\{2x'_{T+h}\beta + 2\sigma^2(x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_{T+h} + \tfrac{1}{2}(T+h-x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_{T+h}))\}\\
&\quad\times\ {}_0F_1(m;\sigma^4\, z^2_{T+h}),\\
&= \exp\{2x'_{T+h}\beta + \sigma^2(T+h+x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_{T+h}))\}{}_0F_1(m;\sigma^4\, z^2_{T+h}),\\
&= \exp\{2x'_{T+h}\beta + 2(T+h)\sigma^2\}\exp\{-\sigma^2(T+h-x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_{T+h})\}{}_0F_1(m;\sigma^4\, z^2_{T+h}).
\end{aligned}$$

Combining the three terms gives:

$$\begin{aligned}
E[(Y_{T+h}-F_{T+h})^2] = \exp\{2x'_{T+h}\beta + 2(T+h)\sigma^2\}*&\\
\Big(1-2\exp\{-\sigma^2(T+h-x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_T)\}+&\\
+\exp\{-\sigma^2(T+h-x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_{T+h})\}{}_0F_1(m;\sigma^4\, z^2_{T+h})\Big)&
\end{aligned}$$

Using $x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_T = T+h$ and $x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_{T+h} = T+2h+\Delta_h x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}\Delta_h x_{T+h}$, when a constant is included, we obtain:

$$\begin{aligned}
E[(Y_{T+h}-F_{T+h})^2] = \exp\{2x'_{T+h}\beta + 2(T+h)\sigma^2\}*&\\
\left\{\exp\{\sigma^2(h+\Delta_h x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}\Delta_h x_{T+h})\}{}_0F_1(m;\sigma^4\, z^2_{T+h})-1\right\}.&
\end{aligned}$$

The unbiased estimator of MSFE. First note that using the above we can write:

$$E[(Y_{T+h}-F_{T+h})^2] = E[F^2_{T+h}] - \exp\{2x'_{T+h}\beta + 2(T+h)\sigma^2\}.$$

The first term can be estimated unbiasedly by F^2_{T+h}, and the second term using Van Garderen (2001) in the familiar form $\exp\{2x'_{T+h}\hat{\beta}\}{}_0F_1(m;m\hat{\sigma}^2\,\tilde{z})$ with $\tilde{z} = -2(h+\Delta_h x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}\Delta_h x_{T+h})$ since:

$$\begin{aligned}
&E\exp\{2x'_{T+h}\hat{\beta}\}{}_0F_1(m;-2m\hat{\sigma}^2(h+a_{T+h})) =\\
&= \exp\{2x'_{T+h}\beta + \sigma^2\frac{4}{2}x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}x_{T+h} - \sigma^2 2\left(h+\Delta_h x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}\Delta_h x_{T+h}\right)\}\\
&= \exp\{2x'_{T+h}\beta + \sigma^2 2(T+2h+\Delta_h x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}\Delta_h x_{T+h} - \left(h+\Delta_h x'_{T+h}\left(\Delta X'\Delta X\right)^{-1}\Delta_h x_{T+h}\right)\}\\
&= \exp\{2x'_{T+h}\beta + 2(T+h)\sigma^2\}
\end{aligned}$$

which is equal to the second term in the expression for the MSFE. □

A.2. Conditional distributions given Y_T

Lemma 3. *(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. Using Eq. (30) part (a) follows from:

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

and calculating the mean and variance. Parts (b) and (c) follow from the moment generating function of a normal distribution: $E[e^{r'z}] = \exp\{r'\mu + \frac{1}{2}r'\Sigma r\}$ if $z \sim N(\mu, \Sigma)$. □

Lemma 4. *Conditional distribution of $y_{(T-1)}$ given y_T:* $y_{(T-1)}|y_T \sim N\left(X_{(T-1)}\beta + \begin{pmatrix} 1 \\ \vdots \\ T-1 \end{pmatrix} \frac{1}{T}\left(y_T - x'_T\beta\right) \quad \sigma^2\Sigma\right)$,
with

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

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

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

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:

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

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

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

Lemma 5. *The conditional distribution of Δy given y_T is a degenerate normal distribution:*

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

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

Proof. Using Lemma 4 the conditional distribution of $y_{(T)}$ given y_T is a degenerate normal distribution

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

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

$$L_{(T)}E[y|\, y_T] = \Delta X\,\beta + \iota_{(T)}\frac{1}{T}\left(y_T - x_T'\beta\right),$$

$$L_{(T)}\left(\begin{array}{cc} \Sigma_{(T-1)} & 0 \\ 0 & 0 \end{array}\right)L_{(T)}' = (I_T - \frac{1}{T}\iota_{(T)}\iota_{(T)}') = M_{\iota_{(T)}}.$$

□

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

Theorem 5. *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: (A)* $\widehat{\beta}|Y_T \sim N(\mu, \Sigma)$ *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)* $\widehat{\sigma}^2$ *and* $\hat{\beta}$ *are conditionally and unconditionally independent.*

Proof. (A) $\widehat{\beta} = \left(\Delta X'\Delta X\right)^{-1}\Delta X'\Delta y$. The distribution of Δy given y_T (or Y_T) is given in the previous lemma. The result follows by noting that $\Delta X'\iota_{(T)} = x_T$ and the mean and variance follow by basic matrix multiplication as follows:

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

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

since $x_T = \iota_{(T)}'\Delta X$ by Assumption 1.

(B) and (C). We will use a conditional version of a theorem by Ogasawara et al. (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 $M_{\Delta X}\Delta y_{(T)} = \sigma^2 M_{\Delta X} M_{\iota_{(T)}} M_{\Delta X} = \sigma^2 M_{\Delta X}$, since $M_{\Delta X} M_{\iota_{(T)}} = M_{\Delta X}$. 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 $rank(A\Sigma) = T - k$:

$$\frac{\Delta y' M_{\Delta X}\Delta y}{\sigma^2} | \, y_T \sim \chi^2_{T-k}.$$

(C) Follows since $\hat{\beta} = (\Delta X'\Delta X)^{-1}\Delta X'\Delta y$ and $M_{\Delta X}\Delta y$ are both linear functions of the conditionally and unconditionally normally distributed Δy and are 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. □

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

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

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

A.3. Expectations and hypergeometric functions

Definition 2. The hypergeometric function can be defined as the infinite sum:

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

See e.g. Abadir (1999) who reviews the use of hypergeometric functions in economics or Van Garderen (2001) who further proves:

Lemma 6. *Let $m\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_m$, then for any real constant z we have,*

$$E\left[{}_0F_1(\frac{m}{2}; z\frac{m}{2}\hat{\sigma}^2\right] = \exp\left\{\, z\;\sigma^2\right\},$$
$$E\left[\left\{{}_0F_1(\frac{m}{2}; z\frac{m}{2}\hat{\sigma}^2)\right\}^2\right] = \exp\left\{2\,z\;\sigma^2\right\}\;{}_0F_1(\frac{m}{2}; z^2\sigma^4).$$

Proof. See Van Garderen (2001). □

B. Additional simulation results

Some further forecast comparisons in misspecified models, related to Section 5.

Table B1. MSFE deficiency: Near unit root.

in %	Stationary $c = -10,\ \tilde{c} = -10$			Explosive $c = .75,\ \tilde{c} = -10$			Explosive $c = .75,\ \tilde{c} = .75$		
	$Y_T\hat{G}$	*Naiv*	$Cons_1$	$Y_T\hat{G}$	*Naiv*	$Cons_1$	$Y_T\hat{G}$	*Naiv*	$Cons_1$
h					$T = 25$				
1	1	0	2	1	0	1	1	1	2
4	2	1	7	4	2	5	2	1	5
12	19	**−7**	20	**−2**	**−4**	**−2**	0	**−3**	0
					$T = 100$				
1	0	0	2	**−1**	**−1**	**−1**	1	1	2
4	2	2	6	1	1	3	2	1	5
12	6	3	18	6	4	14	2	1	7

MSFE deficiency in %, relative to *"unbiased"* forecast. Bold negative values: alternative estimator better. $u_t = \rho_T u_{t-1} + \varepsilon_t$, $\rho_T = 1 + c/T$, $x_{t,1} = t$, $x_{t,2} = \tilde{\rho}_T x_{t-1,2} + \epsilon_t$, $\tilde{\rho}_T = 1 + \tilde{c}/T$. $\beta = (0.2, 1)'$, $\sigma_u = 0.1$. 100.000 replications.

Table B2. MSFE deficiency: Skewness and near unit root.

in %	Stationary: local to unity: $\rho_T = 1 - \frac{4}{T}$											
	Left skewness						*Right skewness*					
$T =$	25			100			25			100		
h	$Y_T\hat{G}$	*Naiv*	$Cons_1$	$Y_T\hat{G}$	*Naiv*	$Cons_1$	$Y_T\hat{G}$	*Naiv*	$Cons_1$	$Y_T\hat{G}$	*Naiv*	$Cons_1$
1	0	0	0	0	0	1	2	2	4	1	1	2
4	**−1**	**−1**	1	2	1	5	6	9	16	4	3	9
12	7	**−3**	8	6	4	15	21	57	59	11	15	31
					Mildly explosive: $\rho_T = 1 + \frac{1}{2}\frac{1}{T}$							
1	0	0	1	1	1	2	0	0	1	1	1	2
4	2	1	4	3	2	6	2	1	4	3	2	6
12	10	5	11	10	8	19	10	5	11	10	8	19

MSFE deficiency in %, relative to *"unbiased"* forecast. Bold negative values: alternative estimator better. Right skewed: $u \sim \sigma_u \left(\chi_3^2 - 3\right)/\sqrt{6}$ with std $\sigma_u = 0.1$. Left skewed: $-u$, same draws. $\beta = (0.2, 1)'$, $x_{t,1} = t$, $x_{t,2} = 0.6x_{t-1,2} + \varepsilon_t$. 100.000 replications.