UvA-DARE (Digital Academic Repository)

Forecasting growth and levels in loglinear unit root models

van Garderen, K.J.

Publication date
2005

Document Version
Submitted manuscript

Link to publication

Citation for published version (APA):
van Garderen, K. J. (2005). *Forecasting growth and levels in loglinear unit root models*. (UvA Econometrics discussion paper; No. 2005/04). Universiteit van Amsterdam, Amsterdam School of Economics.
http://aimsrv1.fee.uva.nl/koen/web.nsf/view/6384D435B763155EC12570A0002B503D/$file/0504.pdf

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:16 May 2025

UvA ECONOMETRICS

Discussion Paper: 2005/04

Forecasting Growth and Levels in Loglinear Unit Root Models

Kees Jan van Garderen

(this replaces the earlier Discussion Paper: 2004/04)

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

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Forecasting Growth and Levels in Loglinear Unit Root Models

Kees Jan van Garderen*

University of Amsterdam

30 September 2005

Abstract

This paper considers unbiased prediction of growth and levels when data series are modelled as a random walk with drift and other exogenous factors after taking 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 and 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 MSFE than usual forecasts. We derive exact unbiased estimators of the MSFE and show that they can succesfully be used in the construction of forecast intervals. The results are applied to a disaggregated eight sector model of UK industrial production.

JEL *classification*: C20; C22; C53.

Keywords: loglinear unit root models, stochastic growth, unbiasedness, parameter uncertainty.

* Faculty of Economics and Econometrics, University of Amsterdam, Roetersstraat 11, 1018WB Amsterdam, The Netherlands. tel:+31-20-5254220, fax:+31-205254349; K.J.vanGarderen@uva.nl

1 Introduction

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

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 with time varying technology as employed by Rosanna (1995) in the context of optimizing firms and Binder and Pesaran (1999) in the growth literature:

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

with ε_t i.i.N. 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 in the problems we want to analyze. The model was also used in Garderen, Lee, and Pesaran (2000) in comparing various nonlinear aggregate and disaggregate models based on predictions of the output level Y.

The central issue is forecasting one or more step ahead output level Y_{T+h}, and predicting growth $G_{h,T} = 100(Y_{T+h} - Y_T)/Y_T$ based on observations up to and including time T. We focus on unbiased forecasts since, in the absence of knowledge of the real cost of forecast errors, one often starts out with a square 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 by Goldberger (1968) in the context of Cobb-Douglas production functions with *i.i.d.* data.

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 logs have been taken. The transformed model is easily estimated, forecasting is standard, and inference is straightforward. The inverse transformation (exponentiation) is nonlinear and the nonstationary nature of the log variable gives rise to further complications, mainly in 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 nonstationarity causes the estimators associated with any variable trending in the levels to be highly correlated with the current level of the dependent variable. In a weakly dependent situation the influence of y_T on the estimator would be of order $1/T$ and dropping the last observation in a strict *i.i.d.* setting would actually make estimator independent of y_T. In the presence of a unit root this is no longer the case. Current y_T is correlated with all y_t's in the past and the covariance does not go to zero due to the stochastic trend. The covariance between y_T and the estimator is of order 1 and the correlation does not disappear asymptotically. Hence forcasts 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 term $\frac{1}{2}\hat{\sigma}^2$), are significantly biased and not even consistent predictors of the conditional expectation.

Second, the nonlinearity of the inverse transformation (taking exponentials) causes the expectation to differ from the exponential of the mean. In linear settings unbiasedness of predictors can be proved by symmetry arguments in certain cases, but these arguments do not apply in the presence of nonlinear transformations. Furthermore, the variance of the random walk component of the log-series is increasing linearly over time and increasingly affects the expected value of the levels of the original series. See also Granger and Newbold (1976) who consider forecasting series that are non-linearly transformed, including exponential transformations and consider quadratic transformations for non-stationary variables. They also compare different predictors and the loss involved, but do not consider estimation uncertainty (the optimal forecast is the 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 analogue of $\check{Y}_{T+1}$ with the equivalent term $\frac{1}{2}\hat{\sigma}^2$ to correct for the nonlinear transformation of the white noise performs better than without it, but abstract from estimation uncertainty and correlation with conditioning variables Y_T.

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 (1999, p111 ff.) also stress the importantance of estimation uncertainty, but show that when T and h are both allowed to increase proportionally, as in Sampson (1991), that the forecast variance of the difference stationary model outgrows that of the trend stationary model, the ratio of the two going to infinity. The exponential transformation only exacerbates the situation and the variance increases even faster. Phillips (1979) also analyzes the role of parameter estimation in forecasting from stable AR(1) models and discusses conditioning on the last observation explicitly.

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

One possible solution is to condition only on those variables that enter the conditional mean. In the leading example those are the exogenous variables and the endogenous variable Y_T. This was also discussed and analyzed by Phillips (1979) for the AR(1) case with stable parameter values, 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. E.g. if only a trend is included then $\hat{\delta} = y_T/T$ and the difficulty is the same as when conditioning on the whole past, namely that $\hat{\delta}$ is fixed. We show theoretically that when a linear trend is included no conditionally unbiased estimator exists at all. This important result holds obviously when conditioning on the whole past but more interestingly, also when conditioning only on the last observation.

We derive two exact unbiased estimators, one unconditionally unbiased estimator i.e. by directly solving the unconditional unbiasedness condition, and the second based on a conditional expression using the last observation Y_T, but taking into account the correlation between the estimator and the last observation and requiring unbiasedness unconditionally. It is interesting that, although both predictors are derived from very different perspectives, they are actually identical.

We also derive an unbiased estimator of the MSFE. Given the skewness in the distributions of future levels of the series Y_{T+h} and its unbiased forecasts, it is not obvious that this is a useful measure of uncertainty. We show however, that it can succesfully be used in the construction of forecast intervals. These forecast intervals are constructed in a standard fashion, using basic normal quantiles, yet have coverage probabilities close to their nominal values, and this seems an interesting and practical contribution.

Finally, a minor comment about assuming it is known that the model is a unit root with drift. This means that there is no (pre-) testing for unit roots which 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). Secondly, the drift parameter is more accurately estimated in the first (log)-difference model where the Cramer-Rao lowerbound 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 larger effects.

The remainder of the paper is organized as follows. The next section discusses the basic model. Section 3 deals with predicting growth in this model and Section 4 deals with forecasting future levels, its MSFE, and constructing forecast intervals. Section 5 applies these results to eight sectors in the UK economy and compares our suggestions with common solutions. Section 6 concludes. Proofs are relegated to the appendix.

2 The Model

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

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

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

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

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

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 $\Delta X = (\Delta x_1, \Delta x_2, ..., \Delta x_T)'$ has full column rank. ■

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

Assumption 2 ensures that the OLS estimator in (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.

With these assumptions the parameters β and σ^2 are simply 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, see Lemma 6 in the appendix, 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 trend or is $I(1)$, such that $(\Delta X'\Delta X) = O_p(T)$ and $x_T = O_p(T)$, then*

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

If x_t consists of a linear trend only then for all T:

$$Corr\left(y_T, \widehat{\delta}\right) = 1. \tag{8}$$

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.

3 Predicting Growth

Growth in the model does not depend on the level Y. It is a function only of the disturbance 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 :

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

which is easily shown using the Moment Generating Function (MGF) of a normal distribution.[1] The term $\frac{h}{2}\sigma^2$ is therefore due to the expectation of a nonlinear function of $\sum_{i=1}^{h} \varepsilon_{T+i}$. The expected growth can be estimated unbiasedly using a result in Van Garderen (2001) who also provides a method for deriving an unbiased estimate of the variance. This leads to:

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 $a_{T+h} = \Delta_h x'_{T+h}(\Delta X'\Delta X)^{-1}\Delta_h x_{T+h}$.

[1] If $z \sim N(\mu, \Sigma)$, then $MGF_z(r) = E\left[e^{r'z}\right] = e^{r'\mu + r'\Sigma\, r/2}$.

*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 the Appendix as an infinite sum and can be thought of as a generalization of the exponential function. See also Abadir (2001) 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 Cassella, 1998) leads to uniqueness.

Using these results we can attribute the term $\frac{1}{2}\hat{\sigma}^2(h - a_{T+h})$ in $\hat{G}_{h,T}$ to: (a) the uncertainty in the disturbances ε_{T+i}, $i = 1, ..., h$, which leads to the term $\frac{h}{2}\sigma^2$ in the expected growth equation (9), and (b) the uncertainty in the estimate $\Delta_h x'_{T+h}\hat{\beta}$, since $E[\exp\{\Delta_h x'_{T+h}\hat{\beta}\}] = \exp\{\Delta_h x'_{T+h}\beta + \frac{1}{2}\sigma^2 a_{T+h}\}$.

Definition 1 *Alternative Growth Predictors*

*(A) **Approximate Unbiased Predictor** of Growth and its Variance:*

$$\tilde{G}^{app}_{h,T} = 100\ (\exp\{\Delta_h x'_{T+h}\hat{\beta} + \tfrac{1}{2}\hat{\sigma}^2(h - a_{T+h})\} - 1), \tag{12}$$

$$\widetilde{var}(\tilde{G}_{h,T}) = 100^2 \exp\{2\Delta_h x'_{T+h}\hat{\beta}\} \times \left[\exp\{\hat{\sigma}^2(h - a_{T+h})\} - \exp\{\hat{\sigma}^2(h - 2a_{T+h})\}\right]. \tag{13}$$

*(B) **Naive Predictor***

$$\check{G}^{naiv}_{h,T} = 100\ (\exp\{\Delta_h x'_{T+h}\hat{\beta}\} - 1). \tag{14}$$

*(C) **Consistent Predictor**:*

$$\check{G}^{cons}_{h,T} = 100\ (\exp\{\Delta_h x'_{T+h}\hat{\beta} + \tfrac{h}{2}\hat{\sigma}^2\} - 1). \tag{15}$$

The approximate unbiased predictor is based on an approximation of the hypergeometric function by an exponential function. It therefore only involves exponential functions and appears much easier to calculate than the exact one, but it should be noted that many packages, such as *Mathematica*, include the hypergeometric function as standard, and a *Gauss* version is available from the author.

A second alternative, referred to as the naive predictor, ignores the uncertainty in both future ε_{T+i}'s and the estimation of β, and as a result is a biased and inconsistent estimator of expected growth. The bias effects work in opposite directions, however: ignoring the uncertainty in leads to under-estimation and ignoring uncertainty in $\hat{\beta}$ leads to over-estimation. The naive predictor can in certain circumstances be better in terms of MSE than the other predictors.

The consistent predictor takes into account the uncertainty in future ε_{T+i}'s, but ignores the uncertainty in the estimation of β and σ^2. This parameter uncertainty goes to zero as the sample size increases, and the predictor is a consistent estimator of expected growth.

Exact expressions and estimators are available for the variance of the consistent and naive predictors, but when growth itself is not estimated unbiasedly, then there is little reason for using an unbiased estimator of the variance or MSFE.

4 Forecasting Levels

Predicting growth is essentially straightforward, and is only complicated by the nonlinear function involved as we have just seen. 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 leads to significant bias. For this reason the obvious estimator based on the unbiased growth predictor:

$$\hat{Y}_{T+h} = Y_T(1 + \hat{G}_{h,T}/100), \tag{16}$$

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

$$\check{Y}_{T+h} = Y_T \exp\{\Delta_h x'_{T+h}\hat{\beta} + \tfrac{h}{2}\hat{\sigma}^2\}, \tag{17}$$

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 nonlinear 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 based directly on the expectation of Y_{T+h}.

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

$$E[Y_{T+h}] = \exp\{x'_{T+h}\beta + \tfrac{T+h}{2}\sigma^2\}, \tag{18}$$

$$E[Y_{T+h}|Y_T] = Y_T \exp\{\Delta_h x'_{T+h}\beta + \tfrac{h}{2}\sigma^2\}. \tag{19}$$

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*

$$\begin{aligned} F_{T+h} &= \exp\{x'_{T+h}\hat{\beta}\}\, {}_0F_1(m; m\hat{\sigma}^2\, z_{T+h}), \quad \textit{with}, \\ z_{T+h} &= \tfrac{1}{2}(T+h - x'_{T+h}\,(\Delta X'\Delta X)^{-1}\, x_{T+h}), \end{aligned} \tag{20}$$

is unbiased.

Proposition 2 *The Conditional Level Forecast*

$$\begin{aligned} F_{T+h|T} &= Y_T \exp\{\Delta_h x'_{T+h}\hat{\beta}\}\, {}_0F_1(m; m\hat{\sigma}^2 z_{T+h|T}), \qquad \textit{with}, \\ z_{T+h|T} &= \tfrac{1}{2}(h - 2x'_T\,(\Delta X'\Delta X)^{-1}\,\Delta_h x'_{T+h} - \Delta_h x'_{T+h}(\Delta X'\Delta X)^{-1}\Delta_h x_{T+h}), \end{aligned} \tag{21}$$

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 proved in Appendix A:

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

$$F_{T+h} = F_{T+h|T} \ . \tag{22}$$

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 (when varying h) 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 $\{Y_T\}$.*

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

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 (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 $h = 0$.

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, $E[(F_{T+h} - Y_{T+h})^2]$:*

$$\begin{aligned} MSFE(F_{T+h}) \quad &= \quad \exp\{2x'_{T+h}\beta + 2(T+h)\sigma^2\} \\ &\quad \times \left[\exp\{\sigma^2(h + a_{T+h})\}_0F_1(m; \sigma^4\, z^2_{T+h}) - 1\right], \end{aligned} \tag{23}$$

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:

$$\widehat{MSFE}(F_{T+h}) = F^2_{T+h} - \exp\{2x'_{T+h}\hat{\beta}\}_0F_1(m; -2m\hat{\sigma}^2(h + a_{T+h})). \tag{24}$$

The MSFE here is unconditional. In principal 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 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 $F_{T+h|T}$ (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 $\tilde{f}_{T+1} = a_T(y_T/T)$, 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.

4.1 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 2 *Alternative forecasts:*

(A) ***Approximate unbiased forecast:***

$$
\begin{aligned}
F^{apu}_{T+h} &= Y_T \exp\{\Delta_h x'_{T+h}\hat{\beta} + \hat{\sigma}^2 z_{T+h|T}\}, \qquad with, \\
z_{T+h|T} &= \tfrac{1}{2}(h - 2x'_T\,(\Delta X'\Delta X)^{-1}\,\Delta_h x'_{T+h} - \Delta_h x'_{T+h}(\Delta X'\Delta X)^{-1}\Delta_h x_{T+h}),
\end{aligned} \tag{25}
$$

and approximate unbiased estimator of the mean squared forecasting error:

$$
\widetilde{MSFE}(F^{apu}_{T+h}) = \left(F^{apu}_{T+h}\right)^2 - \exp\{2x'_{T+h}\hat{\beta} - 2\hat{\sigma}^2(h + a_{T+h})\}. \tag{26}
$$

(B) ***Growth based forecast:***

$$
F^{grow}_{T+h} = Y_T \quad (1 + \frac{\hat{G}_{h,T}}{100}). \tag{27}
$$

(C) ***Naive forecast:***

$$
F^{naiv}_{T+h} = \exp\{x'_{T+h}\hat{\beta}\}. \tag{28}
$$

(D) ***Consistent forecasts:***

$$
F^{cons1}_{T+h} = Y_T \exp\{\Delta_h x'_{T+h}\hat{\beta} + \frac{h}{2}\hat{\sigma}^2\}, \tag{29}
$$

$$
F^{cons2}_{T+h} = \exp\{x'_{T+h}\hat{\beta} + \frac{T+h}{2}\hat{\sigma}^2\}. \tag{30}
$$

The approximate unbiased forecast F^{apu}_{T+h} is constructed by approximating the hypergeometric functions in the exact unbiased forecast and its MSFE, by using the exponentiated $\hat{\sigma}^2 z_{T+h|T}$, which corrects for uncertainty in Y_{T+h} caused by disturbances ε_t, $\hat{\beta}$ and the correlation between Y_T and $\hat{\beta}$, and by 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 as will be shown in Section 5, but is identical to the naive 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. 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, Lee, and Pesaran (2000) in their prediction based criterion for deciding between aggregate and disaggregate nonlinear models. This choice turns out not to be optimal for forecasting as we will see below. Another alternative would be $\exp\{x'_{T+h}\hat{\beta} + \frac{h}{2}\hat{\sigma}^2\}$, but, similar again to Theorem 2, equals F^{cons1}_{T+h} identically if a time trend is included in x_t.

4.2 MSFE Estimation

Theorem 4 gives an unbiased estimate of the MSFE, but does not provide standard errors or an indication how useful the estimates are in constructing forecast intervals. We simulated a log-linear unit root model which includes a trend, a random walk with drift, and a stationary variable as explanatory variables with parameter values similar to estimates based on the data in Nelson and Plosser (1982) or Sampson (1991) to investigate the MSFE estimator. Table 6 in Appendix B illustrates the unbiasedness of the exact unbiased estimator and its purpose is to show that the approximate unbiased MSFE estimator of Definition 2(A) is very close to the exact unbiased MSFE estimator. There are cases where the difference between exact and unbiased are larger, but the results are typical for a range of parameter values and data generating mechanisms for X, including the sectoral application in Section 5.

Table 6 only reports the first moment of these estimates and the standard deviations on these estimates are in fact very large. They are of a similar order of magnitude as the estimates themselves. There is, however, a large correlation between the MSFE estimate and the squared deviation $(F_{T+h} - Y_{T+h})^2$, ranging in the present model with $T = 50$ from 0.4 for $h = 5$ to 0.6 for $h = 20$. It may therefore still be possible to use the estimated MSFE to studentize the forecast for use in inference. This studentization results in a distribution that, apart form some skewness to the left, is not far from a standard normal distribution. Figure 1 shows the density of this studentized forecast estimated nonparametrically based on 100.000 replications and for three different sample sizes. The densities for the three different cases are remarkably close, and close to the standard normal distribution.

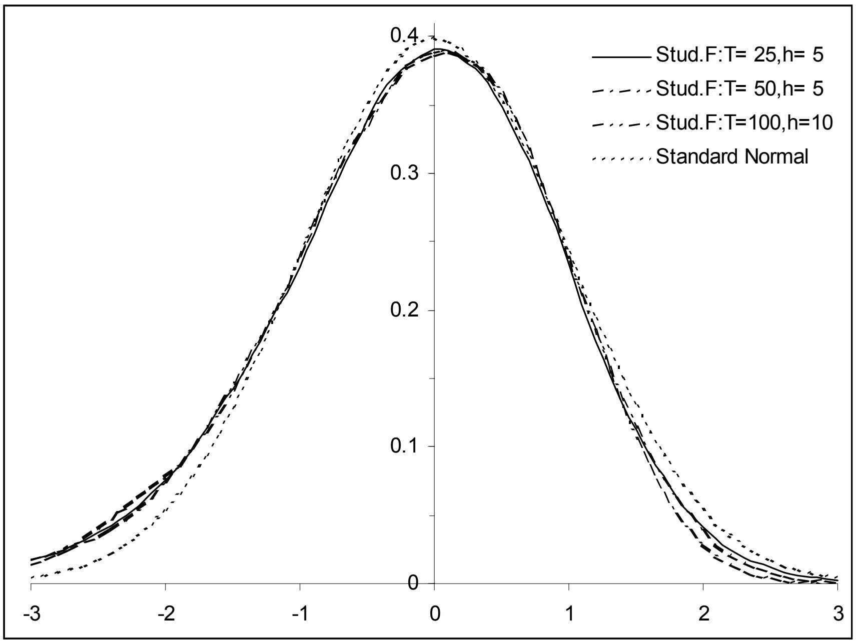

Figure 1: Density of Studentized Forecast. $\sigma = 0.05267$, $\beta' = (0.05, 0.2, 0.4, 0.5)$. Based on 100.000 replications.

4.3 Forecast Intervals

Given the closeness to the normal distribution of the standardized statistic, it is natural to check if confidence intervals based on the studentized statistic and the normal approximation provide a simple way of constructing confidence intervals with a reasonable coverage probability. Table 1 shows the percentage of Y_{T+h}'s that are included in the forecast intervals constructed in the standard fashion using the estimated MSFE and the $(1-\alpha/2)$-quantile of the standard normal distribution $z_{1-\alpha/2} = \Phi(1-\alpha/2)$, with confidence limits:

$$F_{T+h} \pm z_{1-\alpha/2}\widehat{MSFE}^{1/2}. \tag{31}$$

Table 1: Coverage Probabilities of Forecast Interval in %

	$T=25$		$T=50$		$T=100$		$T=500$	
$Nominal$	$h=1$	$h=5$	$h=1$	$h=5$	$h=1$	$h=10$	$h=1$	$h=25$
90%	88	88	89	89	90	89	90	88
95%	94	93	94	94	95	94	95	92
99%	98	98	99	98	98	99	99	96

$\sigma = 0.05267, \beta' = (0.05, 0.2, 0.4, 0.5)$. Based on 100.000 replications.

These coverage probabilities are surprisingly good, given the skewness of the distribution of Y_{T+h}. It was not obvious *a priori* that the MSFE is a good measure for use in inference. These results suggest, however, that in the context of this model it is reasonable to report the estimated MSFE as a measure of forecasting uncertainty for F_{T+h} since standard confidence intervals constructed with the estimated MSFE have coverage probabilities that are close to their nominal values. It should be noted though, that the rejection probability for larger h is predominantly due to rejection on the right hand side, i.e. the actual Y_{T+h} being larger than the upper bound. One might therefore be able to shorten, and hence improve the confidence intervals, by increasing both the upper and lower limits, but this is not pursued here.

4.4 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 forcasting accuracy. First note that in general the level of the series Y_{T+h} tends to increase over time.[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 when divided by the expected level: $\text{MSFE}(F_{T+h})/E[Y_{T+h}]^2$

In order to illustrate this we can use the basic loglinear unit root model with drift to obtain:

$$\frac{MSFE(F_{T+h})}{E[Y_{T+h}]^2} = exp\{(T+h)\sigma^2\}\left[\exp\{\sigma^2 h(1+h/T)\}{}_0F_1(m; \frac{1}{4}\sigma^4\left(h(1+h/T)\right)^2) - 1\right]. \qquad (32)$$

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 2.

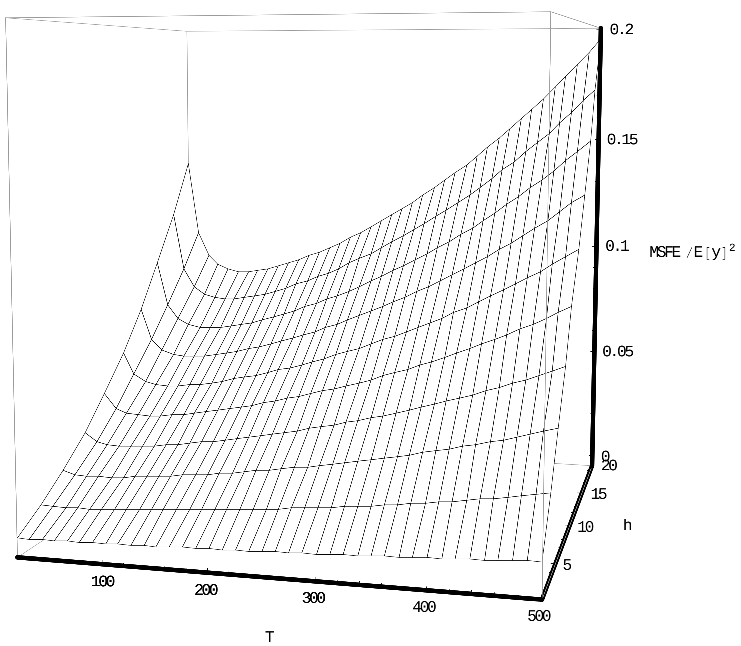

Figure 2: $\text{MSFE}/E[Y_{T+h}]^2$ in loglinear unit root model with drift, T : 25-500, h : 1-20

[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.

When T is increasing, the ratio 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 approximately:

$$\frac{MSFE(F_{T+h})}{E[Y_{T+h}]^2} \overset{appr}{=} \sigma^2(h+\frac{h^2}{T}) + \sigma^4(hT + \frac{2}{5}h^2 + \frac{h^2}{2(T-1)}). \tag{33}$$

The figure further shows that, by letting h and T grow proportionally, as in Sampson (1991) or Clements and Hendry (1999), 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 can compare the unbiased forecast to the alternative forecasting methods. We simulated a variety of slightly more involved log-linear unit root models and we report here the results for the model used in Section 4.2, which included a trend, a random walk with drift, and a stationary variable as explanatory variables with parameter values similar to estimates based on the data in Nelson and Plosser (1982) or Sampson (1991). Changing the way in which the X matrix is generated had some effect but did not qualitatively change the results, as long as a constant was included. We have calculated the MSFE for the alternative forecasts as a

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

[5] Taking as moment functional $f(.)$ the one step ahead MSFE, which in our notation is (West 1996 uses R for T and his π becomes $lim_{T+h\to\infty} h/T = 0$) we have for the theoretical expectation

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

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 $P = h$.

function of the forecastig horizon and for different sample sizes.

The exact unbiased forecast was uniformly best over all the experiments and all forecast horizons. It should be stressed that this general superiority should not have been anticipated in view of, for instance the superiority of the naive growth predictor.

Figure 3 graphs how much worse the alternative growth based, naive, and consistent estimators are relative to the exact unbiased forecast.

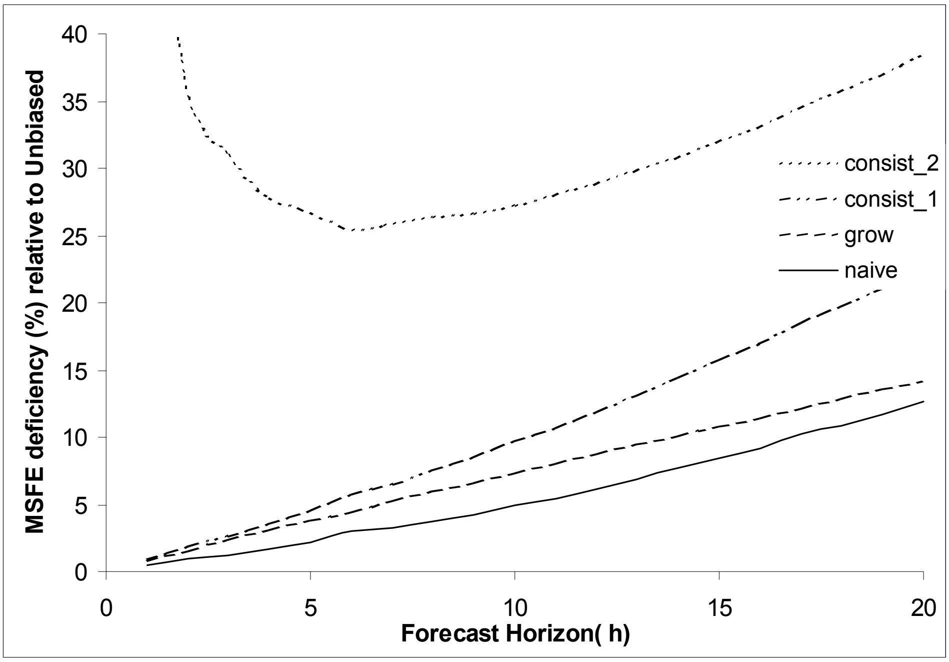

Figure 3: MSFE deficiency. $T = 25$, $\sigma = 0.05267, \beta' = (0.05, 0.2, 0.4, 0.5)$. Based on 100.000 replications.

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, whereas the other forecasts do, but instead is multiplied by the factor $e^{\hat{\sigma}^2 T/2}$. The problem therefore becomes worse as T increases. For example, in the same model with $T = 100$ and $h = 2$, it is 462% worse.

In Appendix C, we report the corresponding figure with $T = 50$ and 100, and a similar picture emerges. The relative MSFE deficiency for the growth based, naive, and first consistent forecasts appears nearly linear as h varies from 1 to 20, and ranges from less than 1% for $h = 1$ to over 16.2%, 8.4%, and 18.5% for $h = 20$.

5 Sectoral Production

In this section we make a comparison of the various predictors of growth and levels for eight industrial sectors. This serves two purposes. First, unbiasedness was motivated by the choice of MSFE as optimality criterion. Although the conditional mean minimizes the MSFE when the parameters are known, this is not necessarily true if the conditional mean is estimated, either unbiasedly or by substituting estimated parameters. Second, 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 case for growth prediction, but for level forecasts we do find relevant practical differences.

We have applied the different forecasts using the same model and set of data as Garderen, Lee, and Pesaran (2000). The sectoral models are Cobb-Douglas production functions with stochastic technology of the type given in Equation (1), but moreover, 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 only difference with the original application is that we report estimates of the model without imposing constant returns to scale restrictions. We simulated the model using the estimated parameters and keeping the exogenous regressors fixed. For observations after 1995 the regressors where dynamically generated using a VAR(1) estimated with data from 1955-1995. The dependent variables Y were then simulated according to the model. 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.

Results on Growth.

Table 2 reports the average of the actual values in the simulations and for the various forecasts the average of the predicted values, their bias (including the exact unbiased where we know the bias is 0 from theory), and their MSE deficiency, defined as the percentage worse their MSE is relative to the best. The individual sectoral results are reported in the appendix.

There is very little to choose between the predictors in terms of bias or in terms of MSE. The naive predictor has a slight bias but is marginally better in terms of MSE. These results are not unexpected because the approximate relation between growth and logs. It can be seen that the average value of the drift term, and the average value of the standard deviation over the eight sectors are both very small. This means that the effect of the non-linear transformation on the expectation is also small.

Table 2: Sectoral Growth Predictions

Growth	Actual	*Exact Unbiased*	*Approx. Unbiased*	*Naive*	*Consistent*
1956-1980	$h = 5$				
Mean	11.38	11.38	11.38	11.38	11.61
Bias		0.00	0.00	0.00	0.22
MSFE deficiency		0.1%	0.1%	0.2%	0.6%
1956-2005	$h = 10$				
Mean	31.86	31.89	31.89	31.22	32.13
Bias		0.02	0.03	-0.64	0.27
MSFE deficiency		0.3%	0.3%	*	0.8%

* indicates best in all sectors. Based on 100.000 replications

Results on Levels

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 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

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, 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 appriciably 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 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.

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

The reason that the naive growth predictor performs well is that the nonlinear 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.

Results on Forecast Intervals

Table 5 reports the coverage probabilities of the forecast intervals for the eight industrial sectors using two time periods with 25 and 50 observations. We see that the coverage probabilities are again very close to their nominal values.

Table 5: Coverage Probabilities of Forecast Interval in %

	1956-1980			1956-2005		
nominal:	90%	95%	99%	90%	95%	99%
Agricult	88	93	98	89	94	98
Mining	88	93	98	88	93	97
Manufact	89	94	98	89	94	99
Energy	88	93	98	89	94	98
Construc	88	93	98	89	94	98
Transprt	88	93	98	89	94	99
Communic	89	94	98	89	94	99
Oth Serv	88	94	98	89	94	98
Average	88	93	98	89	94	98

[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. Hence, $-\iota' \Delta X \ \left(\Delta X' \Delta X\right)^{-1} \Delta_h x_{T+h}' = -x_T' \left(\Delta X' \Delta X\right)^{-1} \Delta_h x_{T+h}' = -h$

6 Conclusion

This paper has highlighted some of the issues involved when forecasting from log-linear unit root models with exogenous variables. Nonlinearity of the transformations of disturbances and estimators cause bias, which is well known. More important, however, 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. This was shown using the expression for $z_{T+h|T}$. The effect of parameter uncertainty disappears with increasing 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 also leads to deeper questions concerning conditioning on past information. We have shown that no conditionally unbiased estimator exists based on the regression estimates, which are statistically complete in this model, even when conditioning on only those terms that enter the conditional mean.

There are of course a various issues the paper has not dealt with, including structural breaks, serial correlation and heteroskedasticity in the transformed model, which is either trivial to deal with if the exact form is known, or analytically very complicated when it involves unknown parameters, or impossible to deal with exactly if its form is unspecified. Nor dynamic forecasts of exogenous variables, as e.g. Schmidt (1974), where the additional uncertainty would increase the bias correction term and would require the exact distribution of the vector autoregressive moving average estimators. We have concentrated on specific models that were helpful in highlighting the issues raised and allowed explicit finite sample solutions.

The paper is constructive in deriving the exact minimum variance unbiased growth predictor. More importantly, we derived two exact unbiased level forecasts. One is based on the unconditional expectation of the process at time $T + h$, and the other derived from the conditional expectation and 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. We showed that the unbiased level forecast was better than five alternative forecasts and provided a very accurate and convenient approximate unbiased forecast that only involved exponential functions. We derived its MSFE and an estimator thereof. The MSFE estimator proved useful in standardizing the estimates and for constructing reliable confidence intervals.

We investigated the effects of increasing sample size and forecast horizon and showed that the MSFE increases with the number of observations even if the forecast horizon is fixed. For increasing forecast horizon, the exact unbiased forecast gets better relative to the alternative forecasts.

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:

$$
\begin{aligned}
y_t &= x_t'\beta + u_t, \qquad t = 1, 2, ..., T \\
u_t &= u_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim IIN(0, \sigma^2)
\end{aligned} \tag{34}
$$

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

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

$y_{(t)} = (y_1, y_2, ..., y_t)'$ 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 $L\,y = LX\beta + Lu$, and hence:

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

Because of assumptions 1 and 2 we may write:

$$
\begin{aligned}
y_t &= \textstyle\sum_{s=1}^t \Delta x_s'\beta + \sum_{s=1}^t \varepsilon_s = x_t'\beta + S_t, \quad with: \\
S_t &= \textstyle\sum_{s=1}^t \varepsilon_s = \imath_{(t)}'\varepsilon_{(t)}.
\end{aligned} \tag{35}
$$

A.1 Proofs

Proof of Lemma 1.

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

Hence:

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

■

Proof of Propositions 1. Using Lemma 7 below, we have:

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

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

Proof of Propositions 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:

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

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

Note that this expectation is not equal $(a.s.)$ 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:

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

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

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 $x_T = \imath'_{(T)}\Delta X$. Now let $P_{\Delta X} = \Delta X\left(\Delta X'\Delta X\right)^{-1}\Delta X'$ be the projection matrix onto the column space of ΔX, then, $\imath'_{(T)}\Delta X\left(\Delta X'\Delta X\right)^{-1}\Delta X'\imath_{(T)} = \imath'_{(T)}P_{\Delta X}\imath_{(T)} = \imath'_{(T)}\imath_{(T)} = T$, since ΔX includes a column of ones (i.e. $\imath_{(T)}$) associated with the constant (the drift term).

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

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

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

since $\imath'_{(T)}\Delta X\left(\Delta X'\Delta X\right)^{-1}\Delta X' = \imath'_{(T)}$. 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 $\widehat{\beta} \mid Y_T \sim N(\mu,\Sigma)$ and for a fixed vector a we have $E[exp\{a\prime\widehat{\beta}\ \}|Y_T] = exp\{a'\beta + \frac{1}{2}a'\Sigma a\}$, we know that the function must be proportional to $exp\{a\prime\widehat{\beta}\ \}$ since the expectation function would otherwise not be log-linear in β. Hence, to find a conditional unbiased estimator we first need to solve:

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

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:

$$a'(I_k - (\Delta X'\Delta X)^{-1} x_T x_T' \frac{1}{T})\beta = \Delta_h x_{T+h}'\beta, \quad for\ all\ \beta \in \Re^k.$$

Let $B = (I_k - (\Delta X'\Delta X)^{-1} x_T x_T' \frac{1}{T})$ and let B^+ denote its Moore-Penrose generalized inverse (explicit expressions for B and B^+ are given in Lemma 2 below), then:

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

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:

$$B'B'^{+}\Delta_h x_{T+h} = \Delta_h x_{T+h}. \tag{36}$$

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 (36) 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 $a'\beta + a'(\Delta X'\Delta X)^{-1} x_T(-x_T')/T\ \beta = a\beta - a\beta = 0$ 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 $(\Delta X'\Delta X)^{-1} \Delta X'\Delta X = I_k$ and ι is the first column of ΔX we have $(\Delta X'\Delta X)^{-1} \Delta X'\iota = (1, 0, ..., 0)'$. 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:

$$E[(Y_{T+h} - F_{T+h})^2] = E(\underset{(1)}{Y_{T+h}^2} \underset{(2)}{-2Y_{T+h}F_{T+h}} \underset{(3)}{+F_{T+h}^2}). \tag{37}$$

The first term (1) in Equation 37 equals:

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

The second term (2): 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 $z_{T+h} = \frac{1}{2}(T + h - x'_{T+h}\left(\Delta X'\Delta X\right)^{-1} x_{T+h})$:

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

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:

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

The third term (3) in Equation 37 equals:

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

Combining the three terms gives

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

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:

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

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:

$$E\exp\{2x'_{T+h}\hat{\beta}\}{}_0F_1(m;-2m\hat{\sigma}^2(h+a_{T+h}))=$$

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

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 Equation (35) part (a) follows from:*

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

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 :*

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

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:

$$\begin{aligned} \Sigma_{12} \Sigma_{22}^{-1} (Y_2 - \mu_2) &= \begin{pmatrix} 1 \\ \vdots \\ T-1 \end{pmatrix} \frac{y_T - x'_T \beta}{T}; \\ \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} &= \frac{1}{T} \sigma^2 \begin{pmatrix} 1 \\ \vdots \\ T-1 \end{pmatrix} \begin{pmatrix} 1 & \cdots & T-1 \end{pmatrix}. \end{aligned}$$

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:*

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

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

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

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

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

$$\begin{aligned} L_{(T)}E[y \mid y_T] &= \Delta X\ \beta + \imath_{(T)}\frac{1}{T}\left(y_T - x_T'\beta\right), \\ L_{(T)}\begin{pmatrix} \Sigma_{(T-1)} & 0 \\ 0 & 0 \end{pmatrix}L_{(T)}' &= (I_T - \frac{1}{T}\imath_{(T)}\imath_{(T)}') = M_{\imath_{(T)}}. \end{aligned}$$

■

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

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) $\quad \widehat{\beta} \mid Y_T \sim N(\mu, \Sigma)$ *with:*

$$\begin{aligned} \mu &= E\left[\widehat{\beta}|Y_T\right] = \beta + \frac{y_T - x_T'\beta}{T}\left(\Delta X'\Delta X\right)^{-1}x_T, \\ \Sigma &= Var(\widehat{\beta}|Y_T) = \sigma^2\left(\left(\Delta X'\Delta X\right)^{-1} - \frac{1}{T}\left(\Delta X'\Delta X\right)^{-1}x_T x_T'\left(\Delta X'\Delta X\right)^{-1}\right), \end{aligned}$$

(B) $\quad (T-k)\dfrac{\widehat{\sigma}^2}{\sigma^2}| \, Y_T \ \sim \chi^2_{T-k}$,

(C) $\quad \widehat{\sigma}^2$ *and* $\hat{\beta}$ *are conditionally and unconditionally independent.*

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

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

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

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

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

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

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

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

(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: $M_{\imath_{(T)}} = (I_T - \frac{1}{T}\imath_{(T)}\imath'_{(T)}) = (I_T - \imath_{(T)}(\imath'_{(T)}\imath_{(T)})^{-1}\imath'_{(T)})$, is a projection matrix with $M_{\imath_{(T)}}\imath_{(T)} = 0$. This implies that if the model only includes a time trend, and therefore $\Delta X = \imath_{(T)}$, that the conditional variance of $\widehat{\beta}|Y_T$ is 0, which is as it should be since in that case $\widehat{\beta} = \frac{1}{T}y_T$, 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 $\widehat{\sigma}^2$.

Lemma 6 *Goldberger (1962). Let:*

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

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

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

Best in Goldberger (1962) means that the forecast minimizes $E\left[\left(\hat{y}^*_{T+1} - y_{T+1}\right)\left(\hat{y}^*_{T+1} - y_{T+1}\right)'\right]$, subject to (i) linearity $\hat{y}^*_{T+1} = C'y_{(T)}$ and (ii) unbiasedness $E\left[\hat{y}^*_{T+1} - y_{T+1}\right] = 0$. Since in the unit root case $\omega_{21}\Omega_{11}^{-1} = 1$ we have for Model (1): $\hat{y}^*_{T+1} = x'_{T+1}\hat{\beta} + 1.(y_T - x'_T\hat{\beta})$ and hence we have the following:

Corollary 3 *In Model (1), the Best Linear Unbiased Predictor of* y_{T+1} *is given by:*

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

A.3 Expectations involving exponentials and hypergeometric functions

Definition 3 *The hypergeometric function can be defined as the infinite sum:*

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

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

Lemma 7 *Let* $m\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_m$, *then for any real constant* z *we have,*

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

Proof See Van Garderen (2001). ■

References

Ariño, M.A. and P.H.Franses, 2000, Forecasting the levels of vector autoregressive log-transformed time series. International Journal of Forecasting, 16, 111-116.

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

Binder, M and M.H. Pesaran, 1999, Stochastic Growth Models and Their Econometric Implications, Journal of Economic Growth, **4**, 139-83.

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

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

Clements, M. P. and D.F. Hendry, 1998, Forecasting Economic Time Series, Cambridge University Press.

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

Goldberger, A., 1962, Best Linear Unbiased Prediction in the Generalized Linear Regression Model, Journal of the American Statistical Association **57**, 369—375.

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

Granger, C.W.J. and Newbold, P., 1976, Forecasting Transformed Series, Journal of the Royal Statistical Society (B), **38**, 189-203.

Lehman, E.L and G. Casella, 1998, Theory of Point Estimation.2nd Edition, Springer, New York.

Magnus, J.R. and H. Neudecker, 1988, Matrix Differential Calculus with Applications in Statistics and Econometrics. John Wiley, New York.

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

Ogasawara, T. and Takahashi, M., 1951, Independence of quadratic forms in normal system, J. Sci. Hiroshima University 15, 1-9.

Philips, P.C.B., 1979, The Sampling Distribution of Forecasts from a First-Order Autoregression, Journal of Econometrics, **9**, 241-261.

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

Ng, S. and T.J. Vogelsang, 2002, Forecasting autoregressive time series in the presence of deterministic components, Econometrics Journal, **5**, 196-224.

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

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

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

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

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

West, K.D., 1988, Asymptotic Nomality, When Regressors Have a Unit Root, Econometrica, **56**,1397-1417.

West, K. D., 1996, Asymptotic Inference about Predictive Ability, Econometrica, **64**, 1067-1084.

Yamamoto, T., 1976, Asymptotic Mean Square Prediction Error for an Autoregressive Model with Estimated Coefficients, Applied Statistics, **25**, 123-127.

B Additional Graphs and Tables

Table 6: MSFE and Estimates

h	*Theory*	*Simulation*	*Exact Unbiased Estimate*	*Appr. Unbiased Estimate*
1	5.5	5.5	5.5	5.5
2	11.2	11.2	11.2	11.2
3	12.1	12.2	12.1	12.1
4	11.9	12.0	12.0	12.0
5	49.3	49.9	49.5	49.5
6	45.2	45.8	45.4	45.4
7	72.5	73.6	72.9	72.7
8	69.4	70.2	69.7	69.5
9	97.9	99.5	98.3	98.1
10	555.7	563.6	558.3	556.9

$T = 25$, $\sigma = 0.05267, \beta' = (0.05, 0.2, 0.4, 0.5)$. 100.000 replications

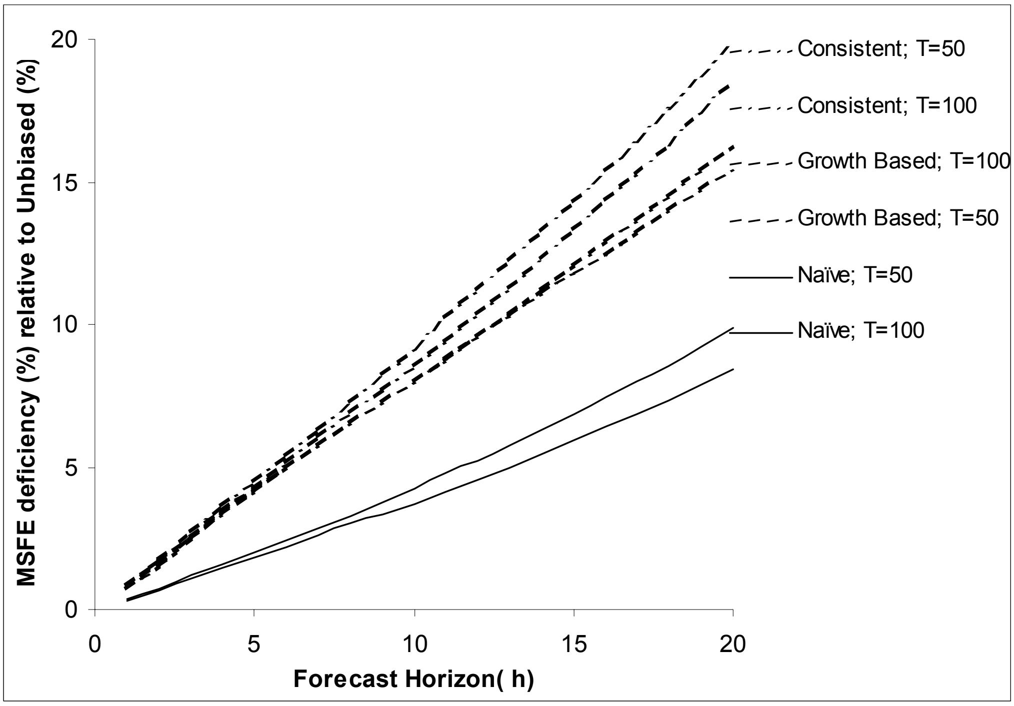

Figure 4: MSFE deficiency. $T = 25, \sigma = 0.05267, \beta' = (0.05, 0.2, 0.4, 0.5)$. 100.000 replications.

C Sectoral Results

1956-1980 : T = 25, Forecasting horizon h = 5 nr simulations = 100,000.

Sectoral Parameter Estimates										
1956-1980	incpt	dd72	d74	dd74	dd75	dd76	d80	dlli	dlki	$\hat{\sigma}$
Agricult	0.029				-0.063	-0.135		-0.025	-0.147	0.035
Mining	0.002	-0.127		-0.116				0.329	0.262	0.030
Manufact	-0.014							0.100	1.157	0.019
Energy	0.036							-0.319	0.715	0.040
Construc	0.006		-0.132				-0.061	0.518	0.387	0.027
Transprt	0.020						-0.066	0.009	0.429	0.028
Communic	0.039							0.707	0.055	0.021
Oth Serv	-0.004						-0.053	0.610	0.582	0.017
Average	0.014	-0.016	-0.016	-0.014	-0.008	-0.017	-0.023	0.241	0.430	0.027

Growth: Average Actual and predicted					
1956-1980	Actual	Unb	ApprUnb	Naive	Consist
Agricult	15.949	16.000	16.000	15.835	16.188
Mining	-5.413	-5.397	-5.397	-5.533	-5.323
Manufact	2.550	2.539	2.539	2.472	2.567
Energy	30.448	30.403	30.403	30.255	30.775
Construc	-0.276	-0.299	-0.299	-0.174	0.013
Transprt	5.031	5.038	5.038	5.530	5.734
Communic	22.033	22.051	22.051	21.989	22.127
Oth Serv	20.721	20.731	20.731	20.666	20.756
Average	11.380	11.383	11.383	11.380	11.605

Growth : Bias				
1956-1980	Unb	ApprUnb	Naive	Consist
Agricult	0.050	0.050	-0.114	0.238
Mining	0.016	0.016	-0.120	0.090
Manufact	-0.011	-0.011	-0.078	0.018
Energy	-0.045	-0.045	-0.194	0.327
Construc	-0.024	-0.024	0.102	0.288
Transprt	0.007	0.007	0.499	0.703
Communic	0.018	0.018	-0.044	0.094
Oth Serv	0.010	0.010	-0.055	0.036
Average	0.003	0.003	0.000	0.224

Growth: MSE				
1956-1980	Unb	ApprUnb	Naive	Consist
Agricult	43.739	43.739	43.633	43.935
Mining	13.857	13.857	13.834	13.887
Manufact	5.835	5.835	5.832	5.840
Energy	97.865	97.865	97.674	98.559
Construc	61.938	61.938	62.102	62.409
Transprt	145.387	145.387	146.964	147.774
Communic	18.551	18.551	18.533	18.586
Oth Serv	6.160	6.160	6.157	6.164
Average	49.167	49.167	49.341	49.644

Growth: MSE Deficiency vs best (*) in %				
1956-1980	Unb	ApprUnb	Naive	Consist
Agricult	0.245	0.245	*	0.692
Mining	0.172	0.172	*	0.388
Manufact	0.049	0.049	*	0.121
Energy	0.196	0.196	*	0.906
Construc	*	0.000	0.265	0.760
Transprt	*	0.000	1.085	1.642
Communic	0.102	0.102	*	0.288
Oth Serv	0.046	0.046	*	0.105
Average	0.101	0.101	0.169	0.613

Levels: Average Actual and predicted							
1956-1980	Actual	Unb	ApprUnb	$Y_T\hat{G}_{h,T}$	Naive	Consist1	Consist2
Agricult	2.284	2.285	2.285	2.295	2.299	2.302	2.338
Mining	0.821	0.821	0.821	0.824	0.825	0.826	0.835
Manufact	1.876	1.876	1.876	1.878	1.880	1.880	1.889
Energy	6.160	6.157	6.157	6.200	6.207	6.224	6.350
Construc	1.672	1.672	1.672	1.680	1.678	1.683	1.699
Transprt	1.719	1.719	1.719	1.734	1.726	1.737	1.754
Communic	3.984	3.984	3.984	3.991	3.993	3.996	4.018
Oth Serv	2.512	2.512	2.512	2.515	2.516	2.517	2.526
Average	2.629	2.628	2.628	2.640	2.640	2.646	2.676

Level : Bias						
1956-1980	Unb	ApprUnb	$Y_T\hat{G}_{h,T}$	Naive	Consist1	Consist2
Agricult	0.001	0.001	0.011	0.015	0.018	0.054
Mining	0.000	0.000	0.003	0.004	0.004	0.014
Manufact	0.000	0.000	0.002	0.003	0.004	0.013
Energy	-0.003	-0.003	0.039	0.046	0.064	0.189
Construc	0.000	0.000	0.008	0.006	0.011	0.027
Transprt	0.000	0.000	0.015	0.007	0.018	0.035
Communic	0.001	0.001	0.008	0.010	0.012	0.035
Oth Serv	0.000	0.000	0.003	0.004	0.005	0.014
Average	0.000	0.000	0.011	0.012	0.017	0.048

Level : MSFE						
1956-1980	Unb	ApprUnb	$Y_T \hat{G}_{h,T}$	Naive	Consist1	Consist2
Agricult	0.051	0.051	0.051	0.051	0.052	0.056
Mining	0.004	0.004	0.004	0.004	0.004	0.004
Manufact	0.009	0.009	0.009	0.009	0.009	0.009
Energy	0.548	0.548	0.557	0.559	0.564	0.622
Construc	0.029	0.029	0.029	0.029	0.029	0.030
Transprt	0.052	0.052	0.053	0.052	0.053	0.055
Communic	0.057	0.057	0.057	0.057	0.057	0.059
Oth Serv	0.012	0.012	0.012	0.012	0.012	0.013
Average	0.095	0.095	0.097	0.097	0.098	0.106

Level : relative deficiency MSFE vs best in perc						
	Unb	ApprUnb	$Y_T \hat{G}_{h,T}$	Naive	Consist1	Consist2
Agricult	*	0.000	1.170	1.631	2.217	10.902
Mining	*	0.000	0.744	1.221	1.513	8.187
Manufact	*	0.000	0.289	0.500	0.601	3.343
Energy	*	0.000	1.607	1.949	2.901	13.452
Construc	*	0.000	1.207	0.849	1.803	5.989
Transprt	*	0.001	2.122	0.838	2.738	6.590
Communic	*	0.000	0.454	0.623	0.850	4.041
Oth Serv	*	0.000	0.244	0.428	0.509	2.804
Average	*	0.000	0.980	1.005	1.641	6.913

1956 - 2005 : T = 50, Forecasting horizon h = 10, nr simulations = 100,000.

Sectoral Parameter Estimates											
1956-2005	incpt	dd72	d74	dd74	dd75	dd76	d80	dd84	dlli	dlki	$\hat{\sigma}$
Agricult	0.021				-0.068	-0.138		0.120	-0.031	0.177	0.042
Mining	0.014	-0.117		-0.111				-0.467	0.491	-0.019	0.064
Manufact	0.017							-0.030	0.490	0.515	0.020
Energy	0.036							-0.205	0.096	0.801	0.047
Construc	0.015		-0.125				-0.068		0.679	0.266	0.032
Transprt	0.023						-0.060		0.337	0.103	0.032
Communic	0.042								0.375	0.099	0.025
Oth Serv	0.023						-0.051		0.545	0.123	0.023
Average	0.024	-0.015	-0.016	-0.014	-0.008	-0.017	-0.022	-0.073	0.373	0.258	0.036

Growth: Average Actual and predicted					
1956-2005	Actual	Unb	ApprUnb	Naive	Consist
Agricult	24.573	24.606	24.607	23.759	24.835
Mining	-11.067	-11.101	-11.101	-12.497	-10.661
Manufact	27.918	27.937	27.937	27.735	27.994
Energy	76.283	76.318	76.318	74.890	76.843
Construc	8.933	8.940	8.940	8.613	9.164
Transprt	32.676	32.713	32.714	32.317	32.979
Communic	54.868	54.927	54.928	54.542	55.034
Oth Serv	40.697	40.736	40.736	40.434	40.821
Average	31.860	31.885	31.885	31.224	32.126

Growth : Bias				
1956-2005	Unb	ApprUnb	Naive	Consist
Agricult	0.033	0.033	-0.814	0.262
Mining	-0.034	-0.033	-1.429	0.406
Manufact	0.019	0.019	-0.184	0.076
Energy	0.035	0.035	-1.392	0.560
Construc	0.006	0.006	-0.321	0.231
Transprt	0.037	0.037	-0.359	0.303
Communic	0.060	0.060	-0.325	0.166
Oth Serv	0.039	0.039	-0.263	0.124
Average	0.024	0.025	-0.636	0.266

Growth: MSE				
	Unb	ApprUnb	Naive	Consist
Agricult	57.225	57.225	57.129	57.509
Mining	78.427	78.428	77.885	79.452
Manufact	14.521	14.521	14.519	14.537
Energy	185.455	185.456	184.374	186.929
Construc	48.750	48.750	48.545	49.021
Transprt	70.251	70.251	69.955	70.634
Communic	32.854	32.854	32.818	32.919
Oth Serv	24.006	24.006	23.984	24.047
Average	63.936	63.936	63.651	64.381

Growth: relative deficiency MSE vs best in perc				
1956-2005	Unb	ApprUnb	Naive	Consist
Agricult	0.167	0.168	*	0.665
Mining	0.695	0.696	*	2.012
Manufact	0.017	0.017	*	0.128
Energy	0.586	0.586	*	1.386
Construc	0.423	0.423	*	0.980
Transprt	0.423	0.423	*	0.970
Communic	0.112	0.112	*	0.308
Oth Serv	0.091	0.091	*	0.260
Average	0.314	0.314	*	0.839

Level : Average Actual and predicted							
1956-2005	Actual	Unb	ApprUnb	$Y_T\hat{G}_{h,T}$	Naive	Consist1	Consist2
Agricult	3.950	3.952	3.953	3.994	4.022	4.029	4.207
Mining	0.278	0.278	0.278	0.285	0.290	0.291	0.323
Manufact	4.295	4.296	4.296	4.307	4.313	4.315	4.359
Energy	23.549	23.553	23.553	23.886	24.081	24.153	25.533
Construc	1.979	1.979	1.979	1.993	1.999	2.003	2.054
Transprt	3.582	3.584	3.584	3.609	3.620	3.627	3.718
Communic	15.659	15.666	15.666	15.727	15.766	15.777	16.029
Oth Serv	7.368	7.370	7.370	7.395	7.411	7.415	7.518
Average	7.583	7.585	7.585	7.649	7.688	7.701	7.968

Level : Bias						
1956-2005	Unb	ApprUnb	$Y_T\hat{G}_{h,T}$	Naive	Consist1	Consist2
Agricult	0.002	0.002	0.044	0.071	0.079	0.257
Mining	0.000	0.000	0.007	0.012	0.013	0.045
Manufact	0.001	0.001	0.011	0.018	0.020	0.064
Energy	0.003	0.003	0.337	0.532	0.603	1.984
Construc	0.000	0.000	0.014	0.020	0.024	0.075
Transprt	0.001	0.001	0.027	0.037	0.045	0.136
Communic	0.007	0.007	0.068	0.107	0.118	0.370
Oth Serv	0.002	0.002	0.027	0.043	0.047	0.150
Average	0.002	0.002	0.067	0.105	0.119	0.385

Level : MSFE						
1956-2005	Unb	ApprUnb	$Y_T\hat{G}_{h,T}$	Naive	Consist1	Consist2
Agricult	0.368	0.368	0.378	0.386	0.389	0.493
Mining	0.005	0.005	0.006	0.006	0.006	0.010
Manufact	0.094	0.094	0.094	0.095	0.095	0.101
Energy	17.998	17.998	18.640	19.141	19.349	25.815
Construc	0.059	0.059	0.060	0.061	0.061	0.070
Transprt	0.193	0.193	0.197	0.199	0.200	0.228
Communic	1.983	1.983	2.003	2.020	2.026	2.223
Oth Serv	0.379	0.379	0.382	0.385	0.386	0.418
Average	2.635	2.635	2.720	2.787	2.814	3.670

Level : relative deficiency MSFE vs best in perc						
1956-2005	Unb	ApprUnb	$Y_T\hat{G}_{h,T}$	Naive	Consist1	Consist2
Agricult	*	0.001	2.711	5.100	5.827	34.170
Mining	*	0.003	6.414	11.895	13.856	86.535
Manufact	*	0.000	0.631	1.174	1.345	7.593
Energy	*	0.001	3.568	6.353	7.508	43.437
Construc	*	0.000	1.761	2.737	3.486	18.255
Transprt	*	0.000	1.779	2.770	3.510	18.162
Communic	*	0.000	1.012	1.881	2.151	12.102
Oth Serv	*	0.000	0.855	1.590	1.822	10.345
Average	*	0.001	2.341	4.187	4.938	28.825