Multi-Step Forecasting and Predictive Regressions

Guillaume Chevillon

ESSEC Business School & CREST-INSEE, Paris

August, 4, 2009

FESAMES
University of Tokyo
Overview

1. Motivations
2. Local asymptotic framework in an AR(1)
3. Multi-step Forecasting
4. Predictive Regressions
I/ Motivations

- Multi-step forecasting: aim is to forecast some $y_{T+h}$ from a forecast origin at $T$ for $h \geq 1$ using
  
  $$y_t = \rho y_{t-1} + \epsilon_t$$

- Two methods:
  1. Iterated Multi-step (IMS), first estimate $\rho$ and then forecast
     
     $$\hat{y}_{T+1|T} = \hat{\rho} y_T$$
     $$\hat{y}_{T+h|T} = \hat{\rho} \hat{y}_{T+h-1|T} = \cdots = \hat{\rho}^h y_T$$
  2. Combine estimation and forecasting in Direct Multi-step (DMS):
     estimate $\rho_h$ in
     
     $$y_t = \rho_h y_{t-h} + u_{t,h}$$
     
     and forecast
     
     $$\tilde{y}_{T+h|T} = \tilde{\rho}_h y_T$$
Direct Multi-step forecasting has a long pedigree

**Drawbacks**

1. Loss of information (does not use all the information in estimation, equivalent sample of $T - h + 1$)
2. Inefficient for estimating the parameters in a well-specified model

Yet, it has been proved superior for (see Chevillon, 2007, for a survey)

3. Non-stationary and/or misspecified models (Clements & Hendry, 1996, Chevillon & Hendry, 2005)

Here we build upon 3.
Predictive Regressions

- **Aim:** establish the predictive power of $x_t$ on $z_t$ over some horizon $h$ when the DGP is

\[
\begin{align*}
Z_{t+1} & = \alpha + \beta x_t + \epsilon_{1,t+1} \\
(1 - \rho L) b(L) x_{t+1} & = \mu + \epsilon_{2,t+1}
\end{align*}
\]


- Often, evidence is mixed; hence the standard is to regress $Z^h_{t+1} = \sum_{i=1}^{h} z_{t+i}$ on $x_t$ or $X_t, X_{t-k}$.

- $\rho$ is often **close to unity**, hence distortions result.

- Link with **DMS** forecasting:

\[
\begin{align*}
Z_{t+h} & = \alpha + \frac{\beta \mu}{b(1)} \rho^{\{h-1\}} + \beta \rho^{h-1} x_t + \epsilon_{1,t+h} \\
& + \beta b(L)^{-1} \left( \sum_{i=1}^{h-1} \rho^{h-1-i} L^{h-i} \right) \epsilon_{2,t+h}
\end{align*}
\]
Aims of the present paper

The two literatures share features

1. the estimated model is not a priori that which would be most efficiently chosen

the multi-step technique therefore works because

2. the model is misspecified with autocorrelated errors
3. the variables are non-stationary or nearly so

In this paper we

- Obtain a finite sample distribution of estimators and forecasts
- where models present deterministic & stochastic trends
- and we allow for potential misspecification

We show that the major benefit from long horizon regressions is that they are mostly unaffected by the error autocorrelation
II/ Framework for analysis: local asymptotic AR(1)

- We use the following model

\[ y_t = \tau + \rho y_{t-1} + \epsilon_t \]

1. \( \tau \) is close to zero,
2. \( \rho \) is close to unity and
3. \( \epsilon_t \) is autocorrelated.

- We use a *local asymptotic unit root* (Phillips, 1988), a *local drift* (Haldrup & Hylleberg, 1995):

\[ y_t = \psi \frac{\phi}{\sqrt{T}} + e^{\phi/T} y_{t-1} + \epsilon_t \]

\[ h = \lceil cT \rceil \text{ and} \]

and consider short-horizon \((h \text{ fixed}, \text{ see Hjalmarsson, 2006})\) or long-horizon \(h = \lceil cT \rceil\) (Kemp, 1999, Valkanov, 2003)
Why is it relevant?

- In finite sample, stochastic and deterministic trends are hard to distinguish (Sampson, 1991)
  - Bad inference results (Clements & Hendry, 2001)
  - Continuum between two types of trends
  - DMS forecasting seems more robust

- In predictive regressions, the data are often near-integrated (Rossi, 2005)
  - the largest autoregressive root is improperly estimated
  - the resulting model is misspecified
  - magnitude of $h/T$ matters (Richardson & Stock, 1989)
Local asymptotics

Why? both trends are asymptotically present,

\[ T^{-1/2} y_{Tr} \Rightarrow K_{\psi,\phi}(r) \]

\[ K_{\psi,\phi}(r) = \psi \frac{e^{\phi r} - 1}{\phi} + \sigma J_\phi(r) \]

where \( J_\phi \) is an Ornstein-Uhlenbeck process and

\[ T^{-1/2} \sum_{1}^{[Tr]} \epsilon_t \Rightarrow \sigma W(r). \]

And also define the forecast error

\[ \hat{e}_{h|T} = y_{T+h} - \hat{y}_{T+h|T} \]

Then for fixed \( h \), estimation and forecasting are a finite sample issue since

\[ \hat{e}_{h|T} \xrightarrow{T \to \infty} \sum_{j=0}^{h-1} \epsilon_{T+h-j}. \]

But with \( h = [cT] \) : in \( y_t = \tau_{T,h} + \rho_{T,h} y_{t-h} + w_{h,t} \), the error

\[ w_{h,t} = \sum_{i=0}^{h} \rho_{T}^i \epsilon_{t-i} \]

and \( \hat{e}_{h|T} \) exhibit a stochastic trend

We can therefore use an FCLT to derive all the finite sample distributions, the results follow from the CMT.
Asymptotic behaviour at short (fixed) horizons

Lemma

For fixed $h \in [1, T)$, as $T \to \infty$, in $y_t = \tau_{T,h} + \rho_{T,h}y_{t-h} + w_{h,t}$,

$$(a_h) \quad T^{-1/2} y_{[Tr]} \Rightarrow K_{\psi,\phi} (r)$$

$$(b_h) \quad T^{-1} \sum_{t=h}^{T} y_{t-h} w_{h,t} \Rightarrow h\sigma \int_0^1 K_{\psi,\phi} dW + \frac{1}{2} [h \sigma^2 - \sigma_{w_h}^2].$$

and

$$T^{-1} \left( \sum_{t=h}^{T} y_{t-h} w_{h,t} - h \sum_{t=1}^{T} \epsilon_t \right) \Rightarrow - \sum_{i=1}^{h-1} (h-i) \zeta_{\epsilon} (i),$$

$\zeta_{\epsilon} (i)$ is the autocovariance of $\epsilon_t$
Estimation error (fixed horizon)

- We apply this to OLS estimation of an AR(1) and $(\hat{\rho}^h = \sum_{i=0}^{h-1} \hat{\rho}_i)$

$$\hat{y}_t = \tau \hat{\rho}^h + \hat{\rho}^h y_{t-h}$$
$$\tilde{y}_t = \tilde{\tau}_h + \tilde{\rho}_h y_{t-h}$$

- Negative error autocorrelation benefits DMS

$$\begin{bmatrix} T^{1/2} (\tau \hat{\rho}^h - \tilde{\tau}_h) \\ T (\hat{\rho}^h - \tilde{\rho}_h) \end{bmatrix} \Rightarrow \frac{\sum_{i=1}^{h-1} (h - i) \xi_{\epsilon} (i)}{\int (K^u)^2} \begin{bmatrix} - \int_0^1 K_{\psi,\phi} \\ 1 \end{bmatrix}$$
Asymptotics at long (fractional) horizon

Lemma

The same analysis, in the case $h = [cT]$ leads to

\[
(f_\phi (r) = \frac{\exp\{\phi r\} - 1}{\phi})
\]

\[
(a_c) \quad T^{-1/2} y_{Tr,T} \Rightarrow K_{\psi,\phi} (r);
\]

\[
(b_c) \quad T^{-2} \sum y_{t-h,T} w_{h,t} \Rightarrow \int_c^1 K (r - c) \delta_c J_\phi (r) - \frac{c}{2} \psi^2 e^{-c\phi} f_\phi^2 (c)
\]

as now $T^{-1/2} w_{h,[rT]} \Rightarrow J_\phi (r) - e^{\phi c} J_\phi (r - c) = \delta_c J_\phi (r)$ is not degenerate.

- $\frac{c}{2} \psi^2 e^{-c\phi} f_\phi^2 (c)$ does no reflect misspecification, contrary to Hjalmarsson (2006) where $h = o(T)$.
- This correction is $O(c^3)$
Estimator comparison at long-horizon

As $c \to 0$

\[
\begin{bmatrix}
T^{-1/2} \left( \hat{\tau} \hat{\rho}^{\{h\}} - \tau \rho^{\{h\}} \right) \\
\hat{\rho}_T^h - \rho^h
\end{bmatrix}
- c
\begin{bmatrix}
T^{1/2} (\hat{\tau} - \tau) \\
T (\hat{\rho} - \rho)
\end{bmatrix}
= o_p (c)
\]

\[
\begin{bmatrix}
T^{-1/2} \left( \tilde{\tau} - \tau \rho^{\{h\}} \right) \\
\tilde{\rho}_h^h - \rho^h
\end{bmatrix}
- \sqrt{c}
\begin{bmatrix}
T^{1/2} (\hat{\tau}^\times - \tau) \\
T (\hat{\rho}^\times - \rho)
\end{bmatrix}
= o_p (c)
\]

$(\hat{\tau}^\times, \hat{\rho}^\times)$ are the estimators in the absence of autocorrelation of $\epsilon_t$. 
III/ Forecasting (1/3) \((\phi, \theta) = (0, 0)\)

The error is \(\epsilon_t = u_t + \theta u_{t-1}\)
Forecasting (2/3) \((\phi, \psi) = (0, 0.5)\)
Forecasting (3/3) ($\psi = 0$)
Monte Carlo setting: regression of $Z_{t+1}^h = \sum_{i=1}^h z_{t+i}$ on $(1, x_t)$ with DGP

\[
\begin{align*}
    z_{t+1} &= x_t + \epsilon_{1,t+1} \\
    x_{t+1} &= \frac{\psi}{\sqrt{T}} + e^{\phi/T} x_t + \epsilon_{2,t+1} \\
    \epsilon_{2,t+1} &= u_{t+1} + \theta u_t
\end{align*}
\]

with $(\epsilon_{1,t}, \epsilon_{2,t}) \sim \text{IN} (0, I_2)$

The estimated coefficient is $\sum_{i=0}^{h-1} \rho^i$ and we report the estimator of

\[
h^{-1} \sum_{i=0}^{h-1} \rho^i_T \rightarrow \left\{ \begin{array}{ll} 
1 & h \text{ fixed} \\
\frac{\exp\{c\phi\} - 1}{c\phi} & h = [cT]
\end{array} \right. 
\]
Predictive Regression (2/3) \((\psi = 0, \phi = -10, \theta = -0.2)\)

![Density plots](chart.png)

- Monte Carlo, \(T = 100\)
- Asymptotic, \(h\) fixed
- Asymptotic, \(h = [cT], \theta = 0\)
Long-horizon approximation works for smaller $h$ when $\theta \leq 0$. 
Conclusions

The paper

- derives an *approximation to the distributions* of estimators, test statistics and forecasts in a general setting, extending previous results (Valkanov, 2003, Hjalmarsson, 2006)
- shows that one source of gain in using multi-step estimation lies in its *robustness to misspecification* at long horizons.