Overcoming Bias in Predictive Regression

A Multi-purpose Methodology

Yakov Amihud, Clifford M. Hurvich and Yi Wang

Department of Finance and Department of IOMS
Stern School of Business, New York University
Outline

- Why predictive regression?
  1) Empirically important, e.g.
     (i) Stock return prediction using dividend yield, B/M, CAY, etc.
     (ii) Risk Premium Hypothesis: return on volatility
     (iii) Forward Rate Unbiasedness Hypothesis (FRUH) of exchange rates
     (iv) Inflation vs. Interest Rate
  2) Theoretically interesting

- Empirical Examples

- Basic Predictive Regression Model: stationary AR(1) predictor

- Generalized Models for Predictive Regression
  1) AR(p) predictor
  2) Long memory predictor
  3) Autoregressive response variable and Multiple predictors

- Analysis of Empirical Examples
Empirical Examples

Predictive Regression with an AR(1) Predictor

The model is

\[ y_t = \alpha + \beta x_{t-1} + u_t \]
\[ x_t = \theta + \rho x_{t-1} + v_t \]

where \(|\rho| < 1\) and

\[
\begin{pmatrix} u_t \\ v_t \end{pmatrix} \sim i.i.d. N(0, \Sigma), \quad \Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}
\]

- OLS estimate \( \hat{\beta} \) is \( \sqrt{n} \)-consistent and asymptotically normal.
- But Stambaugh (1999) shows that, \( \hat{\beta} \) is biased in small samples, if \( \sigma_{uv} \neq 0 \).
Bias Expression for $\hat{\beta}$:

$$E(\hat{\beta} - \beta) = \phi E(\hat{\rho} - \rho)$$

where $\phi = \sigma_{uv}/\sigma_v^2$.

The bias vanishes if $\sigma_{uv} = 0$.

Small-sample bias correction for OLS $\hat{\rho}$ by Kendall (1954)

$$E(\hat{\rho} - \rho) = -\frac{1 + 3\rho}{n} + O(\frac{1}{n^2}) \approx -\frac{1 + 3\rho}{n}$$

Feasible small-sample bias correction formula for $\hat{\beta}$

$$\hat{\beta}^s = \hat{\beta} + \hat{\phi}s \frac{1 + 3\hat{\rho}}{n},$$

where $\hat{\phi}^s = \frac{\sum \hat{u}_t \hat{v}_t}{\sum \hat{v}_t^2}$ and $\hat{u}_t, \hat{v}_t$ are OLS residuals.

No statistical inference method based on $\hat{\beta}^s$ is suggested.
It is easy to show that, for a bivariate normal \((u_t, v_t)'\), we can write

\[ u_t = \phi v_t + e_t, \]

where \(\phi = \sigma_{uv} / \sigma_v^2\), \(\{e_t\}\) are i.i.d. normal with zero mean and independent of both \(\{x_t\}\) and \(\{v_t\}\). Thus,

\[ y_t = \alpha + \beta x_{t-1} + u_t \quad \rightarrow \quad y_t = \alpha + \beta x_{t-1} + \phi v_t + e_t \]

OLS estimate \(\hat{\beta}\) obtained by regressing \(\{y_t\}_{t=2}^n\) on \(\{x_t\}_{t=1}^{n-1}\) augmented with \(\{v_t\}_{t=2}^n\), would be unbiased if we knew \(\{v_t\}\). \(\rightarrow\) Need a proxy for \(\{v_t\}_{t=2}^n\).

Can we use OLS residuals \(\hat{\sigma_t}\)? No! Since they are orthogonal to \(\{x_{t-1}\}\) and thus \(\hat{\beta}\) is the same.

Construct a proxy for \(\{v_t\}\) through Kendall correction of \(\hat{\rho}\)

1) Correct \(\hat{\rho}\) by \(\hat{\rho}^c = \hat{\rho} + \frac{1+3\hat{\rho}}{n}\)

2) Construct \(v_t^c = x_t - \hat{\rho}^c x_{t-1}\) and use it as the proxy.
We propose the Augmented Regression Method (ARM): regressing \( \{y_t\}_{t=2}^n \) on \( \{x_{t-1}\}_{t=2}^n \) augmented with \( \{v_t^c\}_{t=2}^n \) constructed as above. The resulting equation is

\[
y_t = \hat{\alpha}^c + \hat{\beta}^c x_{t-1} + \hat{\phi}^c v_t^c + \hat{e}_t
\]

It is shown in Amihud and Hurvich (2005) that, using Kendall correction for \( \hat{\rho} \),

\[
\hat{\beta}^c = \hat{\beta}^s, \quad E(\hat{\phi}^c) = \phi
\]

i.e. the slope estimate generated by ARM is identical to the one obtained by using Stambaugh’s bias correction expression. But, ARM leads to a hypothesis test based on \( \hat{\beta}^c \).

How to estimate the standard error of \( \hat{\beta}^c \)?

\[
\widehat{Var}^c(\hat{\beta}^c) = \{\hat{\phi}^c\}^2 \widehat{Var}(\hat{\rho}^c) + \widehat{Var}(\hat{\beta}^c)
\]

which can be easily obtained using standard statistical package.
Predictive Regression with an AR(p) Predictor

- The AR(1) assumption of the \( \{x_t\} \) process is restrictive.
- Predictive regression model with a stationary AR(p) predictor

\[
\begin{align*}
  y_t &= \alpha + \beta_1 x_{t-1} + \ldots + \beta_p x_{t-p} + u_t \\
  x_t &= \theta + \rho_1 x_{t-1} + \ldots + \rho_p x_{t-p} + v_t
\end{align*}
\]

where the model assumptions are similar to those in the basic model.
- We decide to include \( x_{t-2}, \ldots, x_{t-p} \) in the predictive regression because for an AR(p) predictor, we would expect further predictive power from all lags.
- We use the similar Augmented Regression Method (ARM), which requires a proxy for \( \{v_t\} \).
Shaman and Stine (1988, 1989) provide the small-sample bias expression for the OLS estimates of the autoregressive parameters in an AR(p) process. For example, when \( \{x_t\} \) is AR(2) with unknown mean

\[
E(\hat{\rho}_1 - \rho_1) = -\frac{1 + \rho_1 + \rho_2}{n}, \quad E(\hat{\rho}_2 - \rho_2) = -\frac{2 + 4\rho_2}{n}
\]

Thus, we construct

\[
\hat{\rho}_1^c = \hat{\rho}_1 + \frac{1 + \hat{\rho}_1 + \hat{\rho}_2}{n}, \quad \hat{\rho}_2^c = \hat{\rho}_2 + \frac{2 + 4\hat{\rho}_2}{n}
\]

Consequently, we construct \( \{v_t^c\} \) as before by

\[
v_t^c = x_t - \hat{\theta} - \hat{\rho}_1^c x_{t-1} - \ldots - \hat{\rho}_p^c x_{t-p}
\]
The fitted equation is

\[ y_t = \hat{\alpha}^c + \hat{\beta}_1^c x_{t-1} + \ldots + \hat{\beta}_p^c x_{t-p} + \hat{\phi}^c v_t^c + \hat{e}_t \]

Some theoretical results

\[ E(\hat{\beta}_i^c - \beta_i) = \phi E(\hat{\rho}_i^c - \rho_i), \quad (i = 1, \ldots, p) \]

\[ E(\hat{\phi}^c) = \phi \]

\[ E[(\hat{\beta}_i^c - \beta_i)(\hat{\beta}_j^c - \beta_j)] = \phi^2 E[(\hat{\rho}_i^c - \rho_i)(\hat{\rho}_j^c - \rho_j)] + E[\text{cov}(\hat{\beta}_i^c, \hat{\beta}_j^c)] \]

The last result and the linearity of the Shaman-Stine expression is used to construct the estimated covariance matrix of \((\hat{\beta}_1^c, \ldots, \hat{\beta}_p^c)\). Thus corresponding hypothesis tests (both individual and joint) can be performed.
Predictive Regression with an ARFIMA(0,d,0) Predictor

- Can we generalize the \( \{x_t\} \) assumption even further? Yes, we consider a long memory process.
- Start with the simplest case: ARFIMA(0, d, 0) process. The model is

\[
\begin{align*}
y_t &= \alpha + \beta x_{t-1} + u_t \\
x_t &= \mu x + (1 - B)^{-d} u_t
\end{align*}
\]

where \( B \) is the backshift operator and \( d \) is the memory parameter.
- When \( d \) is not an integer, \( (1 - B)^{-d} \) is interpreted through the binomial series expansion

\[
(1 - B)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(j + d)}{\Gamma(j + 1)\Gamma(d)} B^j
\]

where \( \Gamma \) is the gamma function.
- So ARFIMA(0,d,0) can be regarded as a MA(\( \infty \)) or AR(\( \infty \)), but controlled by only one memory parameter: \( d \).
Simulated stationary ARFIMA (0,d,0) process.
When \( d \in (-0.5, 0.5) \) (stationary and invertible predictor), we obtain the \( \sqrt{n} \)-consistency and asymptotic normality for the OLS estimate \( \hat{\beta} \), generated by regressing \( \{y_t\}_{t=2}^n \) on \( \{x_{t-1}\}_{t=2}^n \),

\[
\sqrt{n}(\hat{\beta} - \beta) \Rightarrow_D N\left[0, \frac{\sigma_u^2}{\text{var}(x_t)}\right]
\]

where \( \text{var}(x_t) = \frac{\Gamma(1 - 2d)}{\Gamma^2(1 - d)} \). The convergence rate doesn’t depend on the memory parameter \( d \).

The \( t \)-statistic of \( \hat{\beta} \) is valid

\[
t(\hat{\beta}) \Rightarrow_D N(0, 1)
\]
When \( d \in (0.5, 1.5) \) (nonstationary predictor),

\[
n^d(\hat{\beta} - \beta) \Rightarrow D \frac{\sigma_u^2}{\sigma_v^2} \frac{\int_0^1 \tilde{W}_{d-1}(s)dW(s)}{\int_0^1 [\tilde{W}_{d-1}(s)]^2 ds}
\]

where

\[
W_d(s) = \frac{1}{\Gamma(d + 1)} \int_0^s (s - x)^d dW(x)
\]

is fractional Brownian motion for \( d \in (-0.5, 0.5) \) and \( W(s) \) is standard Brownian motion, \( \tilde{W}_{d-1}(s) = W_{d-1}(s) - \int_0^1 W_{d-1}(s)ds \) is the deviation of fractional Brownian motion from its Lebesgue integral from 0 to 1,

When \( d = 1 \) and the ARFIMA process has zero mean, this expression reduces to

\[
n(\hat{\beta} - \beta) \Rightarrow D \frac{\sigma_u^2}{2\sigma_v^2} \frac{[W(1)]^2 - 1}{\int_0^1 [W(s)]^2 ds}
\]

which is related to the results in Phillips (1987) studying the OLS estimate of random walk.
How to correct the small-sample bias for OLS $\hat{\beta}$?

First, what is the bias expression?

\[
E(\hat{\beta}) = \beta + \phi E\left[ \sum_{t=2}^{n} v_t (x_{t-1} - \bar{x}) \right] = \beta + \phi E\left( \frac{T_1}{T_2} \right).
\]

How to evaluate the bias expression?

1) Taylor expansion.
2) Augmented regression.
3) Bootstrapping-based method.
\[ y_t = \alpha + \beta x_{t-1} + \phi v_t + e_t \]
\[ x_t = \mu_x + (1 - B)^{-d} v_t \]

Augmented regression method: build a proxy for \( \{v_t\} \)

1) Estimate \( d \) using Whittle.
   (i) It is basically the frequency domain MLE approximation.
   (ii) It is computationally fast and asymptotically equivalent to MLE.

2) Construct a proxy for \( \{v_t\} \) by

\[
\hat{v}_t = \sum_{j=0}^{t-1} \pi_j(\hat{d})(x_{t-j} - \bar{x}), \quad (t = 2, \ldots, n)
\]

3) Regress \( \{y_t\}_{t=2}^n \) on \( \{x_{t-1}\}_{t=2}^n \) together with \( \{\hat{v}_t\}_{t=2}^n \)

4) The resulting fitted equation is

\[
y_t = \hat{\alpha}^{Aug} + \hat{\beta}^{Aug} x_{t-1} + \hat{\phi}^{Aug} \hat{v}_t + \hat{e}_t
\]
Bootstrapping-based bias correction
1) Again, memory parameter $d$ is estimated using Whittle estimate.
2) Remind that the bias expression is

$$E(\hat{\beta}) = \beta + \phi E\left[\frac{\sum_{t=2}^{n} v_t(x_{t-1} - \bar{x})}{\sum_{t=2}^{n} (x_{t-1} - \bar{x})^2}\right] = \beta + \phi E\left(\frac{T_1}{T_2}\right).$$

3) So we bootstrap $T_1$ and $T_2$ using estimated parameters.
4) Define $R_i^* = T_1^* / T_2^*$, take the average of $\{R_i^*\}_{i=1}^{B}$ as the estimate of $E(\frac{T_1}{T_2})$.

$$\hat{\beta}^{PBS} = \hat{\beta} - \hat{\phi}^A \text{ mean}(R_i^*).$$

Statistical inference for $\beta$ can be performed.
1) Using the bias-corrected estimates with their corresponding standard errors. For simplicity, OLS standard error is used.
2) Bootstrapping the empirical distribution of $\hat{\beta}$. (BS method)
The basic predictive regression model does not specifically consider the persistence of the response variable, nor does it allow feedback effect from $y_{t-1}$ to $x_t$.

Recall that, the basic model is

$$\begin{align*}
y_t &= \alpha + \beta x_{t-1} + u_t \\
x_t &= \theta + \rho x_{t-1} + v_t
\end{align*}$$

or, equivalently

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \alpha \\ \theta \end{pmatrix} + \begin{pmatrix} 0 & \beta \\ 0 & \rho \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix}$$

An obvious generalization is

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \alpha \\ \theta \end{pmatrix} + \begin{pmatrix} \delta & \beta \\ \pi & \rho \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ v_t \end{pmatrix}$$
Multi-predictor Predictive Regression

- It is also interesting to allow multiple predictors in the basic model.
- We assume \( \{x_t\} \) a \( p \)-dimensional vector time series, which follows a stationary Gaussian vector autoregressive \( \text{VAR}(1) \) model. The overall model is

\[
\begin{align*}
y_t & = \alpha + \beta' x_{t-1} + u_t, \\
x_t & = \Theta + \Phi x_{t-1} + v_t,
\end{align*}
\]

where \( x_t, \Theta, \beta \) and \( v_t \) are all \( (p \times 1) \) vectors, and \( \Phi \) is a \( (p \times p) \) matrix,

\[
\Phi = \begin{pmatrix}
\Phi_{11} & \cdots & \Phi_{1p} \\
\vdots & \ddots & \vdots \\
\Phi_{p1} & \cdots & \Phi_{pp}
\end{pmatrix},
\]

The quantities \( y_t, \alpha \) and \( u_t \) are scalars. The vectors \( (u_t, v'_t)' \) are i.i.d. multivariate normal with mean zero. We allow \( u_t \) and \( v'_t \) to be contemporaneously correlated.
**Selective Simulation Results**

- AR(2) predictor case: simulation parameters: $\phi = -20, \rho_2 = 0.24$

<table>
<thead>
<tr>
<th>n</th>
<th>$\rho_1$</th>
<th>$\hat{\beta}_1$ [MSE]</th>
<th>$\hat{\beta}_2$ [MSE]</th>
<th>$\hat{\beta}_c^1$ [MSE]</th>
<th>$\hat{\beta}_c^2$ [MSE]</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.65</td>
<td>1.01 [9.82]</td>
<td>1.20 [8.82]</td>
<td>0.29 [9.04]</td>
<td>0.11 [8.63]</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.62 [9.37]</td>
<td>1.18 [9.09]</td>
<td>0.06 [9.27]</td>
<td>0.09 [8.97]</td>
</tr>
<tr>
<td>200</td>
<td>0.65</td>
<td>0.14 [1.87]</td>
<td>0.34 [1.87]</td>
<td>-0.05 [1.86]</td>
<td>0.05 [1.84]</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.09 [1.88]</td>
<td>0.32 [1.97]</td>
<td>-0.06 [1.89]</td>
<td>0.03 [1.94]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n</th>
<th>$\rho_1$</th>
<th>$t(\hat{\beta}_1)$</th>
<th>$t(\hat{\beta}_2)$</th>
<th>Wald</th>
<th>$t(\hat{\beta}_c^1)$</th>
<th>$t(\hat{\beta}_c^2)$</th>
<th>Wald$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.65</td>
<td>12.3%</td>
<td>10.6%</td>
<td>11.0%</td>
<td>7.6%</td>
<td>5.3%</td>
<td>8.3%</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>8.6%</td>
<td>9.9%</td>
<td>6.3%</td>
<td>6.0%</td>
<td>5.3%</td>
<td>7.1%</td>
</tr>
<tr>
<td>200</td>
<td>0.65</td>
<td>6.8%</td>
<td>6.6%</td>
<td>5.2%</td>
<td>4.9%</td>
<td>5.1%</td>
<td>5.3%</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>5.7%</td>
<td>7.2%</td>
<td>4.7%</td>
<td>4.9%</td>
<td>4.5%</td>
<td>4.6%</td>
</tr>
</tbody>
</table>
ARFIMA \((0, d, 0)\) predictor, bias-correction simulation: \(\phi = -20, \beta = 0, n = 50\).

<table>
<thead>
<tr>
<th>(d)</th>
<th>(\hat{\beta})</th>
<th>(\hat{\beta}_{Aug})</th>
<th>(\hat{\beta}_{PBS})</th>
<th>(\hat{d})</th>
<th>(\hat{\phi}_{Aug})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>mean</td>
<td>0.55</td>
<td>0.16</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>3.07</td>
<td>2.77</td>
<td>3.06</td>
<td>-</td>
</tr>
<tr>
<td>0.25</td>
<td>mean</td>
<td>1.15</td>
<td>0.30</td>
<td>0.03</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>3.96</td>
<td>3.07</td>
<td>3.08</td>
<td>-</td>
</tr>
<tr>
<td>0.45</td>
<td>mean</td>
<td>1.88</td>
<td>0.46</td>
<td>0.09</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>5.96</td>
<td>3.13</td>
<td>2.87</td>
<td>-</td>
</tr>
<tr>
<td>0.75</td>
<td>mean</td>
<td>2.58</td>
<td>0.68</td>
<td>0.27</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>8.73</td>
<td>2.75</td>
<td>2.26</td>
<td>-</td>
</tr>
<tr>
<td>0.95</td>
<td>mean</td>
<td>2.23</td>
<td>0.50</td>
<td>0.11</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>6.75</td>
<td>1.76</td>
<td>1.68</td>
<td>-</td>
</tr>
</tbody>
</table>
- ARFIMA(0, d, 0) predictor, hypothesis testing: $\phi = -20, \beta = 0, n = 50$.
- $H_0 : \beta = 0, H_a : \beta \neq 0$.

<table>
<thead>
<tr>
<th>$d$</th>
<th>$t(\hat{\beta})$</th>
<th>$t(\hat{\beta}^{Aug})$</th>
<th>$t(\hat{\beta}^{PBS})$</th>
<th>BS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>right-tailed</td>
<td>7.2%</td>
<td>3.9%</td>
<td>6.6%</td>
</tr>
<tr>
<td></td>
<td>two-tailed</td>
<td>4.9%</td>
<td>4.3%</td>
<td>6.5%</td>
</tr>
<tr>
<td></td>
<td>left-tailed</td>
<td>2.6%</td>
<td>4.4%</td>
<td>6.4%</td>
</tr>
<tr>
<td>0.25</td>
<td>right-tailed</td>
<td>9.5%</td>
<td>6.8%</td>
<td>6.4%</td>
</tr>
<tr>
<td></td>
<td>two-tailed</td>
<td>5.6%</td>
<td>6.1%</td>
<td>8.1%</td>
</tr>
<tr>
<td></td>
<td>left-tailed</td>
<td>1.4%</td>
<td>5.9%</td>
<td>7.6%</td>
</tr>
<tr>
<td>0.45</td>
<td>right-tailed</td>
<td>12.5%</td>
<td>8.1%</td>
<td>6.9%</td>
</tr>
<tr>
<td></td>
<td>two-tailed</td>
<td>7.5%</td>
<td>8.2%</td>
<td>7.8%</td>
</tr>
<tr>
<td></td>
<td>left-tailed</td>
<td>0.3%</td>
<td>5.8%</td>
<td>7.8%</td>
</tr>
</tbody>
</table>
Empirical Illustrations: Stock Market Return Prediction

NYSE Value-Weighted Return and Dividend Yield: 1934:Q1 – 2001:Q4

Stern-Wharton Conference on Statistics in Business, Friday, April 28th 2006. – p. 24
Parameters estimation and hypothesis testing: $H_0 : \beta = 0$ vs. $H_a : \beta > 0$.

<table>
<thead>
<tr>
<th>Method</th>
<th>Estimation</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>$\hat{\beta}$</td>
<td>$t(\hat{\beta})$</td>
</tr>
<tr>
<td></td>
<td>0.8803</td>
<td>2.36</td>
</tr>
<tr>
<td>ARM-AR(1)</td>
<td>$\hat{\beta}^c$, $\hat{\rho}^c$, $\hat{\phi}^c$</td>
<td>$t(\hat{\beta}^c)$</td>
</tr>
<tr>
<td></td>
<td>0.6321, 0.9559, -17.2684</td>
<td>1.68</td>
</tr>
<tr>
<td>PBS-LM</td>
<td>$\hat{\beta}^{PBS}$, $\hat{d}$, $\hat{\phi}^{Aug}$</td>
<td>BS (p-value)</td>
</tr>
<tr>
<td></td>
<td>0.6298, 1.0830, -17.3847</td>
<td></td>
</tr>
</tbody>
</table>
Parameters estimation and hypothesis testing: $H_0 : \beta = 0$ vs. $H_a : \beta > 0$.

<table>
<thead>
<tr>
<th>Method</th>
<th>Estimation</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>$\hat{\beta}$</td>
<td>$t(\hat{\beta})$</td>
</tr>
<tr>
<td></td>
<td>0.8803</td>
<td>2.36</td>
</tr>
<tr>
<td>ARM-AR(1)</td>
<td>$\hat{\beta}^c$</td>
<td>$\hat{\rho}^c$</td>
</tr>
<tr>
<td></td>
<td>0.6321</td>
<td>0.9559</td>
</tr>
<tr>
<td>PBS-LM</td>
<td>$\hat{\beta}^{PBS}$</td>
<td>$\hat{d}$</td>
</tr>
<tr>
<td></td>
<td>0.6298</td>
<td>1.0830</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Method</th>
<th>Residual Sum of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.432</td>
</tr>
<tr>
<td>ARM-AR(1)</td>
<td>0.418</td>
</tr>
<tr>
<td>PBS-LM</td>
<td>0.417</td>
</tr>
</tbody>
</table>
The End

“Seven ‘provided’, thirteen ‘but’s’ and twenty seven ‘if’s’ - this is the best stockmarket forecast to clients I’ve ever read.”