Evaluating Conditional Forecasts from Vector Autoregressions

Todd Clark\textsuperscript{1}    Michael McCracken\textsuperscript{2}

\textsuperscript{1}Federal Reserve Bank of Cleveland

\textsuperscript{2}Federal Reserve Bank of St. Louis

March 2014
Introduction

<table>
<thead>
<tr>
<th>VARs are commonly used to produce conditional forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td>- conditional on policy paths (e.g., Doan, Litterman, and Sims 1984)</td>
</tr>
<tr>
<td>- conditional on oil prices (e.g., Giannone, et al. 2010)</td>
</tr>
<tr>
<td>- conditional on judgmental nowcasts (e.g., Schorfheide and Song 2013)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional forecasts are also very common in other forecast approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Judgment-driven macroeconomic forecasts of central banks</td>
</tr>
<tr>
<td>- Bank stress tests</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Challenge to evaluating conditional forecasts: familiar optimality properties do not apply</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Familiar tests for bias, efficiency, etc. apply to <strong>unconditional forecasts</strong></td>
</tr>
</tbody>
</table>
Example from a simple VAR(1):

\[
\begin{pmatrix}
y_t \\
x_t
\end{pmatrix} = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \begin{pmatrix} y_{t-1} \\
x_{t-1}
\end{pmatrix} + \begin{pmatrix} e_t \\ v_t
\end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}
\]

- **Condition**: \( \hat{x}_{t+1}^c = x_t \)
- **minimum-MSE conditional forecast (DLS)**:
  \[
  \hat{y}_{t+1}^c = \hat{y}_{t+1}^u + \rho(\hat{x}_{t+1}^c - \hat{x}_{t+1}^u)
  \]
  \[
  \hat{y}_{t+1}^c = \hat{y}_{t+1}^u + \rho(1 - c)x_t
  \]
- **Conditional forecast error**:
  \[
  \hat{u}_{t+1}^c = \hat{u}_{t+1}^u - \rho(1 - c)x_t
  \]

**Implications**: (MSE-optimal) conditional errors can be biased, serially correlated, and correlated with conditional forecasts.
Establish asymptotic normality of regression-based tests of bias and efficiency for conditional forecasts from VARs

Establish asymptotic normality of equal MSE tests applied to conditional and unconditional forecasts

- Consider both reduced form (DLS) and structural shock approaches to conditional forecasting
- Focus on large $R$, large $P$ asymptotics of West (1996)
- Discuss fixed $R$, large $P$ asymptotics of Giacomini and White (2006)

Precedents:

- “Modesty” statistic of DLS (1984)
- Efficiency test of Faust and Wright (2008) for judgmental forecasts
- Bayesian posterior predictive checks of Herbst and Schorfheide (2012)
Develop inference approaches
- OLS-based standard errors and $N(0,1)$ critical values
- Bootstrap provides an easier way to correct for effects of parameter estimation error

Provide Monte Carlo evidence on size and power
- Bootstrap often more reliable than Normal-based inference
  - Partly due to HAC challenges and partly due to PEE

Examine an example: U.S. macro forecasts conditioned on unchanged monetary policy
- Small and large BVARs
1 Theory
2 Monte Carlo Analysis
3 Application Results
West and McCracken (1998) establish asymptotic normality of regression-based tests of predictability

- $\hat{y}_{t+\tau} =$ forecast, $\hat{\epsilon}_{t+\tau} =$ forecast error
- Test regression and null:
  $$\hat{\epsilon}_{t+\tau} = \hat{Z}_t' \gamma + \xi_{t+\tau}, \quad H_0 : \gamma = 0$$
- bias: $\hat{Z}_t = 1$
- efficiency: $\hat{Z}_t = (1, \hat{y}_{t+\tau})'$

But the assumptions of W-M rule out direct application to conditional forecasts
- Forecast errors must be uncorrelated with period $t$ information
Our extension to bias and efficiency of conditional forecasts:

- **General form:** \( \hat{\varepsilon}_{t+\tau}^c = \hat{Z}_t'\gamma + \xi_{t+\tau} \)
- **Bias:** \( \hat{\varepsilon}_{t+\tau}^c = \alpha_1 + \xi_{t+\tau}, \quad H_0: \alpha_1 = 0 \)
- **Efficiency:** \( \hat{\varepsilon}_{t+\tau}^c = \alpha_0 + \alpha_1 \hat{y}_{t+\tau}^c + \delta(\hat{y}_{t+\tau}^u - \hat{y}_{t+\tau}^c) + \xi_{t+\tau}, \quad H_0: \alpha_1 = 0 \)

**Interpretation of efficiency null:**
- Backward-looking conditions: \( \alpha_1 = 0 \) if unconditional forecast is efficient and unconditional error is uncorrelated with condition
- Conditions including future information: \( \alpha_1 = 0 \) if unconditional forecast is efficient and \( (\hat{y}_{t+1}^c - \hat{y}_{t+1}^u) \) is uncorrelated with period \( t \) information

**Equal MSE of conditional and unconditional forecasts:** \( H_0: E(\varepsilon_{t+1}^c)^2 - (\varepsilon_{t+1}^u)^2 = 0 \)

- **Approach 1:** \( t \)-test of Diebold-Mariano-West
- **Approach 2:** Test \( \alpha_1 = -1/2 \) in \( \hat{\varepsilon}_{t+\tau}^u = \alpha_1(\hat{\varepsilon}_{t+\tau}^c - \hat{\varepsilon}_{t+\tau}^u) + \eta_{t+\tau} \)
- **Other approaches:** joint restrictions that help with interpretation of rejection
Asymptotic distributions obtained by mapping regression tests into West (1996)

- West: Limiting distribution for sample average $f_{t+\tau}(\hat{\beta}_t)$
- Our tests can be expressed as functions $g(f_{t+\tau}(\hat{\beta}_t))$
- General null: $H_0: \alpha_1 = g(Ef_{t+\tau}) = 0$ for all $t$
- Using West’s result and Delta method yields:

$$P^{1/2}\hat{\alpha}_1 \rightarrow^d N(0, \Omega)$$

$$\Omega = \nabla g'(Ef_{t+\tau}) V \nabla g(Ef_{t+\tau})$$

$$V = S_{ff} + 2(1 - \pi^{-1} \ln(1 + \pi))(FBS_{fh} + FBS_{hh}B'F')$$

- In general, standard errors have to be adjusted to account for parameter estimation error
- But there are some special cases in which they do not
  - Bias test, backward-looking condition, recursive scheme
Inference option 1: Correct standard errors for PEE
- With conditional forecasts, just not practical

Inference option 2: Ignore PEE
- Use conventional HAC standard errors and hope for the best
- Under asymptotics of Giacomini and White (2006) and rolling estimation scheme, use conventional HAC standard errors

Inference option 3: Parametric bootstrap
- Fit VAR to full sample of data; retain coefficients and residuals $\hat{\epsilon}_t$
- Generate artificial innovations $n_t\hat{\epsilon}_t$, where $n_t \sim N(0, 1)$
- Use innovations and VAR process to get artificial data
- Generate artificial forecasts and test statistics
- Tabulate bootstrap percentiles
DGPs: Three bivariate or trivariate VARs for size, one for power

DGP 2:

\[
\begin{pmatrix}
  y_t \\
  x_t
\end{pmatrix} = \begin{pmatrix}
  0.25 & -0.5 \\
  0 & 0.8
\end{pmatrix} \begin{pmatrix}
  y_{t-1} \\
  x_{t-1}
\end{pmatrix} + \begin{pmatrix}
  e_t \\
  v_t
\end{pmatrix},
\]
\[
\Sigma = \begin{pmatrix}
  1 & -0.3 \\
  -0.3 & 0.2
\end{pmatrix}
\]

- We forecast \( y_{t+1} \) subject to conditions on \( x_{t+1} \), using VAR estimated by OLS
- Two different conditions:
  - backward: \( x_{t+1}^c = x_t \)
  - forward: \( x_{t+1}^c = x_{t+1} + \text{noise} \)
- Range of sample sizes
- Power DGP has a break in coefficients
- Other experiment results cover:
  - Conditioning via structural shocks to “policy”
  - Longer forecast horizons
  - BVAR estimation
Table 3: Size, DGP 2, recursive, backward condition
(nominal size = 10%, forecast horizon = 1 period)

<table>
<thead>
<tr>
<th>test</th>
<th>source of critical values</th>
<th>R=50</th>
<th>R=50</th>
<th>R=100</th>
<th>R=100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>P=100</td>
<td>P=150</td>
<td>P=100</td>
<td>P=150</td>
</tr>
<tr>
<td>bias, uncond.</td>
<td>Normal</td>
<td>0.128</td>
<td>0.109</td>
<td>0.107</td>
<td>0.109</td>
</tr>
<tr>
<td>efficiency, uncond.</td>
<td>Normal</td>
<td>0.153</td>
<td>0.140</td>
<td>0.112</td>
<td>0.108</td>
</tr>
<tr>
<td>bias, conditional</td>
<td>Normal</td>
<td>0.066</td>
<td>0.053</td>
<td>0.066</td>
<td>0.051</td>
</tr>
<tr>
<td>efficiency, conditional</td>
<td>Normal</td>
<td>0.617</td>
<td>0.626</td>
<td>0.474</td>
<td>0.486</td>
</tr>
<tr>
<td>bias, conditional</td>
<td>bootstrap</td>
<td>0.101</td>
<td>0.094</td>
<td>0.092</td>
<td>0.088</td>
</tr>
<tr>
<td>efficiency, conditional</td>
<td>bootstrap</td>
<td>0.062</td>
<td>0.053</td>
<td>0.059</td>
<td>0.050</td>
</tr>
</tbody>
</table>

- Conditional efficiency test vs. N(0,1) critical values significantly oversized
- Bootstrap critical values generally reliable (HAC and PEE)
Monte Carlo Analysis

Size, DGP 2, recursive, forward condition  
*(nominal size = 10%, forecast horizon = 1 period)*

<table>
<thead>
<tr>
<th>test</th>
<th>source of critical values</th>
<th>$R=50$</th>
<th>$R=50$</th>
<th>$R=100$</th>
<th>$R=100$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$P=100$</td>
<td>$P=150$</td>
<td>$P=100$</td>
<td>$P=150$</td>
</tr>
<tr>
<td>bias, conditional</td>
<td>Normal</td>
<td>0.123</td>
<td>0.129</td>
<td>0.116</td>
<td>0.118</td>
</tr>
<tr>
<td>efficiency, conditional</td>
<td>Normal</td>
<td>0.185</td>
<td>0.169</td>
<td>0.179</td>
<td>0.135</td>
</tr>
<tr>
<td>bias, conditional</td>
<td>bootstrap</td>
<td>0.098</td>
<td>0.105</td>
<td>0.091</td>
<td>0.105</td>
</tr>
<tr>
<td>efficiency, conditional</td>
<td>bootstrap</td>
<td>0.084</td>
<td>0.089</td>
<td>0.091</td>
<td>0.081</td>
</tr>
</tbody>
</table>

- Conditional efficiency test vs. N(0,1) critical values still oversized
- Bootstrap critical values generally reliable
### Monte Carlo Analysis

#### Size, DGP 2, rolling, backward condition

*(nominal size = 10%, forecast horizon = 1 period)*

<table>
<thead>
<tr>
<th>test</th>
<th>source of critical values</th>
<th>R=50</th>
<th>R=50</th>
<th>R=100</th>
<th>R=100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>P=100</td>
<td>P=150</td>
<td>P=100</td>
<td>P=150</td>
</tr>
<tr>
<td>bias, uncond.</td>
<td>Normal</td>
<td>0.015</td>
<td>0.007</td>
<td>0.061</td>
<td>0.018</td>
</tr>
<tr>
<td>efficiency, uncond.</td>
<td>Normal</td>
<td>0.249</td>
<td>0.344</td>
<td>0.107</td>
<td>0.079</td>
</tr>
<tr>
<td>bias, conditional</td>
<td>Normal</td>
<td>0.008</td>
<td>0.003</td>
<td>0.026</td>
<td>0.007</td>
</tr>
<tr>
<td>efficiency, conditional</td>
<td>Normal</td>
<td>0.836</td>
<td>0.929</td>
<td>0.605</td>
<td>0.708</td>
</tr>
<tr>
<td>bias, conditional</td>
<td>bootstrap</td>
<td>0.096</td>
<td>0.090</td>
<td>0.093</td>
<td>0.098</td>
</tr>
<tr>
<td>efficiency, conditional</td>
<td>bootstrap</td>
<td>0.068</td>
<td>0.068</td>
<td>0.077</td>
<td>0.069</td>
</tr>
</tbody>
</table>

- Under the rolling scheme, PEE correction needed in all cases
- Failure to correct badly distorts inference
- Bootstrap reliable
Table 6: Power, DGP 4
(nominal size = 10%, forecast horizon = 1 period)

<table>
<thead>
<tr>
<th>test</th>
<th>source of critical values</th>
<th>$R=50$</th>
<th>$P=100$</th>
<th>$R=50$</th>
<th>$P=150$</th>
<th>$R=100$</th>
<th>$P=100$</th>
<th>$R=100$</th>
<th>$P=150$</th>
</tr>
</thead>
<tbody>
<tr>
<td>bias, uncond.</td>
<td>Normal</td>
<td>0.532</td>
<td>0.551</td>
<td>0.779</td>
<td>0.806</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>efficiency, uncond.</td>
<td>Normal</td>
<td>0.497</td>
<td>0.604</td>
<td>0.430</td>
<td>0.498</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias, condit.-RF</td>
<td>Normal</td>
<td>0.811</td>
<td>0.844</td>
<td>0.964</td>
<td>0.978</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>efficiency, condit.-RF</td>
<td>Normal</td>
<td>0.649</td>
<td>0.732</td>
<td>0.546</td>
<td>0.622</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias, condit.-RF</td>
<td>bootstrap</td>
<td>0.748</td>
<td>0.788</td>
<td>0.953</td>
<td>0.969</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>efficiency, condit.-RF</td>
<td>bootstrap</td>
<td>0.329</td>
<td>0.444</td>
<td>0.286</td>
<td>0.389</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Tests have power
- As expected, bootstrap rejection rates are lower than N(0,1)-based rates
Forecasts of U.S. GDP growth, unemployment rate, and PCE inflation from 22-variable BVAR

- BVAR specification taken from Giannone, Lenza, and Primiceri (2012)
  - Model in levels or log levels, with random walk priors (Sims and Zha, 1998)
- Model used to produce unconditional and conditional forecasts
- Conditions: Federal funds rate constant over forecast horizon
  - Reduced form (minimum-MSE) approach of DLS (1984)
  - Policy shock approach
- Horizons of 4 and 8 quarters ahead
- Recursive forecasting scheme
Results: point forecasts

Unemployment rate

forecast horizon = 8 quarters

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual</th>
<th>Unconditional</th>
<th>Condit.: DLS</th>
<th>Condit.: Policy Shock</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992</td>
<td>3.5</td>
<td>4.0</td>
<td>4.5</td>
<td>5.0</td>
</tr>
<tr>
<td>1994</td>
<td>4.0</td>
<td>4.5</td>
<td>5.0</td>
<td>5.5</td>
</tr>
<tr>
<td>1996</td>
<td>4.5</td>
<td>5.0</td>
<td>5.5</td>
<td>6.0</td>
</tr>
<tr>
<td>1998</td>
<td>5.0</td>
<td>5.5</td>
<td>6.0</td>
<td>6.5</td>
</tr>
<tr>
<td>2000</td>
<td>5.5</td>
<td>6.0</td>
<td>6.5</td>
<td>7.0</td>
</tr>
<tr>
<td>2002</td>
<td>6.0</td>
<td>6.5</td>
<td>7.0</td>
<td>7.5</td>
</tr>
<tr>
<td>2004</td>
<td>6.5</td>
<td>7.0</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>7.0</td>
<td>7.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Results: point forecasts

PCE inflation

forecast horizon = 8 quarters


-1 0 1 2 3 4 5 6

actual unconditional condit.: DLS condit.: policy shock
Application results: Tests of bias (t-statistics)

<table>
<thead>
<tr>
<th></th>
<th>Unconditional forecasts</th>
<th>Conditional forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(h=4)</td>
<td>(h=8)</td>
</tr>
<tr>
<td>GDP growth</td>
<td>-1.320</td>
<td>-0.535</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>0.843</td>
<td>0.637</td>
</tr>
<tr>
<td>PCE inflation</td>
<td>-2.392</td>
<td>-2.794</td>
</tr>
</tbody>
</table>

- Using normal critical values would imply both unconditional and conditional forecasts to be biased.
- But with more reliable bootstrap critical values, unbiasedness cannot be rejected.
Application results: Tests of efficiency (t-statistics)

<table>
<thead>
<tr>
<th></th>
<th>Unconditional forecasts</th>
<th>Conditional forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$h=4$</td>
<td>$h=8$</td>
</tr>
<tr>
<td>GDP growth</td>
<td>-3.299</td>
<td>-5.647</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>0.180</td>
<td>-1.724</td>
</tr>
<tr>
<td>PCE inflation</td>
<td>-7.220</td>
<td>-7.673</td>
</tr>
</tbody>
</table>

Using normal critical values would imply both unconditional and conditional forecasts to be inefficient.

With more reliable bootstrap critical values, inefficiencies rest primarily with inflation forecasts.
Conditioning on an unchanged funds rate path does not consistently reduce forecast accuracy.
Conclusions

- Establish asymptotic normality of regression-based tests of bias and efficiency for conditional forecasts from VARs
- Establish asymptotic normality of equal MSE tests applied to conditional and unconditional forecasts
- Develop inference approaches
- Provide Monte Carlo evidence on size and power
- Examine an example: U.S. macro forecasts conditioned on unchanged monetary policy