Towards a Unified Statistical Framework to Evaluate Financial Crises Early Warning Systems

How to evaluate an EWS?

Bertrand Candelon†, Elena-Ivona Dumitrescu‡, Christophe Hurlin‡

†Maastricht University and ‡University of Orléans

2009
Introduction

→ From the subprime crisis to currency crises

→ Early Warning Systems (EWS) set up to ring before the occurrence of crises
Introduction

How can we specify an EWS model?
→ Rich literature (Kaminski et al. (1998), Kumar et al. (2003), Abiad (2003), etc.)

How can we evaluate the predictive abilities of an EWS?
→ Kaminski et al. (1998): signalling approach

➤ Threshold which minimizes the NSR criteria
➤ Type I and type II errors

→ Arbitrarily chosen cut-offs (Berg and Patillo (1999), Arias and Erlandsson (2005))
Our New EWS Evaluation Method

→ I. Optimal cut-off

→ II. Credit-scoring evaluation criteria
  QPS, LPS, AUC, Pietra Index, Bayesian Error, Kuiper’s score

→ III. Comparison tests
  ▶ Diebold-Mariano (1995) test for non-nested models
  ▶ Clark-West (2007) test for nested models
  ▶ Area under ROC comparison test
Contents

A New EWS Evaluation Method

EWS Specification and Estimation

Empirical Results

Conclusions
Step 1. A New EWS Evaluation Method
I. Optimal cut-off identification

\[ C^* = \text{Arg}\{C\}[Sensitivity(C) = Specificity(C)], \text{ where } C \in [0, 1] \]

**Definition 1.**  
*Sensitivity* is the number of crises correctly predicted for a cutoff \( C \) over the total number of crises in the sample

**Definition 2.**  
\( 1 – Specificity \) is the number of false alarms for a cutoff \( C \) over the total number of non-crises in the sample
Towards a Unified Statistical Framework to Evaluate Financial Crises Early Warning Systems
II. Performance assessment criteria

The Area Under the ROC Curve and the Quadratic Probability Score

What is the ROC curve? (Receiving Operating Characteristic)
II. Performance assessment criteria

The Area Under the ROC Curve

\[ A = \int_{0}^{1} Sensitivity(1 - Specificity)d(1 - Specificity) \]

- Measure of the model’s overall ability to discriminate between the cases correctly predicted and the false alarms
- For a perfect model AUC=1 while for a random one AUC=0.5
II. Performance assessment criteria

The Quadratic Probability Score

\[ QPS = \frac{1}{T} \sum_{t=1}^{T} 2(\hat{l}_t - l_t)^2 \]

- Comparison of forecasts \((\hat{l}_t)\) and realizations \((l_t)\)
- The closer QPS is to 0 the better the model is
III. Comparison tests

2. Clark-West (2007) test for nested models
3. Area under ROC comparison test (Delong et al. (1988))
III. Comparison tests

Proposition 1: Let us denote by $M_1$ and $M_2$ two EWS models, and by $\widehat{AUC}_1$ and $\widehat{AUC}_2$ the associated areas under the ROC curve.

$H_0 : \widehat{AUC}_1 = \widehat{AUC}_2$

$$\frac{(\widehat{AUC}_1 - \widehat{AUC}_2)^2}{\text{Var}(\widehat{AUC}_1 - \widehat{AUC}_2)} \xrightarrow{d} \chi^2(1)$$

Towards a Unified Statistical Framework to Evaluate Financial Crises Early Warning Systems
Step 2. EWS Specification and Estimation
To apply our evaluation methodology:

I. **Real crisis dating method** ($l_t$)
   - KLR modified pressure index - Lestano and Jacobs (2004)
   - The threshold equals two standard deviations above the mean

II. **Crisis probabilities** ($\hat{Pr}_t$)
    - Panel logit with fixed effects
    - Markov Switching Model with constant transition probabilities
I. Currency crisis dating method

KLR modified pressure index - Lestano and Jacobs (2004)

Definition 3. The 24 months crisis variable:

\[ l_t = C24_{n,t} = \begin{cases} 
1, & \text{if } \sum_{j=1}^{24} Crisis_{n,t+j} > 0 \\
0, & \text{otherwise}
\end{cases} \]
II. Empirical models

Model 1. Panel and time-series logit model

\[
\Pr(C24_{nt} = 1) = \frac{\exp(\beta' x + f_n)}{1 + \exp(\beta' x + f_n)} \quad \forall n \in \Omega_h,
\]

where

- \( f_n \) represents the fixed effects
- \( x \) is the matrix of economic variables
- \( n \) is the country identifier
- \( \Omega_h \) is the \( h^{th} \) cluster

**Optimal country clusters:** (Kapetanios procedure (2003))
II. Empirical models


\[ KLRm_t = \mu_t(S_t) + \beta(S_t)x_t + \epsilon_t(S_t), \]

where

- \( KLRm_t \) is the pressure index vector
- \( x_t \) represents the matrix of economic variables
- \( S_t \) follows a two states Markov chain

\[ S_t = \begin{cases} 
1, & \text{if there is a crisis at time } t \\
0, & \text{if not} 
\end{cases} \]
II. Empirical models

Definition 4. The 24 months ahead forecasts (Arias and Erlandson (2005)):

\[
\Pr(S_{t+1...t+24} = 1|\Omega_t) = 1 - \Pr(S_{t+1...t+24} = 0|\Omega_t)
\]

\[
= 1 - \left\{ [P_{10}P_{00}^{(23)}]\Pr(S_t = 1|\Omega_t)] + [P_{00}^{24}\Pr(S_t = 0|\Omega_t)] \right\},
\]

where $P_{10}$ and $P_{00}$ are elements of the transition probability matrix.
II. Empirical models

From crisis probabilities to crisis forecasts

\[ \hat{I}_t = \begin{cases} 1, & \text{if } \Pr(C_{24t} = 1) > C^* \\ 0, & \text{otherwise} \end{cases}, \]

where \( C^* \) is an **optimal cut-off** (see section 1)
Empirical Results

I. Dataset

II. Optimal country clusters

III. Comparison tests

IV. Optimal model: cut-off identification and performance assessment criteria
I. Dataset

→ Monthly data in US dollars for the period 1985-2005 (6 Latin-American and 6 South-Asian Countries)

→ Market expectation (m.e.) variables:
  ▶ Yield spread
  ▶ Growth of stock market price index

→ Macroeconomic variables: Jacobs et al. (2003)
II. Optimal country clusters

Kapetanios procedure (2003)

1. Argentina, Brazil, Mexico, Venezuela
2. Peru, Uruguay
3. Korea, Malaysia, Taiwan
4. Philippines, Thailand
5. Indonesia
III. Comparison tests

Testing strategy

1. Logit with market-expectation variables vs. simple logit

2. Markov with market expectation variables and spread switching vs. Markov with market expectation variables

3. Best logit vs. best Markov specification
### III.1. Logit with m.e. variables vs. simple logit

<table>
<thead>
<tr>
<th>Country</th>
<th>ROC</th>
<th>Clark-West</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>test statistic</td>
<td>p-value</td>
</tr>
<tr>
<td>Argentina</td>
<td>0.0301</td>
<td>0.8622</td>
</tr>
<tr>
<td>Brazil</td>
<td>5.7105</td>
<td><strong>0.0169</strong></td>
</tr>
<tr>
<td>Indonesia</td>
<td>7.9917</td>
<td><strong>0.0047</strong></td>
</tr>
<tr>
<td>Korea</td>
<td>4.5357</td>
<td><strong>0.0332</strong></td>
</tr>
<tr>
<td>Malaysia</td>
<td>0.3859</td>
<td>0.5345</td>
</tr>
<tr>
<td>Mexico</td>
<td>&lt;0.001</td>
<td>1.0000</td>
</tr>
<tr>
<td>Peru</td>
<td>0.0028</td>
<td>0.9577</td>
</tr>
<tr>
<td>Philippines</td>
<td>0.8738</td>
<td>0.3499</td>
</tr>
<tr>
<td>Taiwan</td>
<td>10.475</td>
<td><strong>0.0012</strong></td>
</tr>
<tr>
<td>Thailand</td>
<td>6.9801</td>
<td><strong>0.0082</strong></td>
</tr>
<tr>
<td>Uruguay</td>
<td>0.7443</td>
<td>0.3883</td>
</tr>
<tr>
<td>Venezuela</td>
<td>6.6647</td>
<td><strong>0.0098</strong></td>
</tr>
</tbody>
</table>

* The coefficients significant at a 5% level are in bold.
### III.2. Markov with m.e. variables and spread switching vs. Markov with m.e. variables

<table>
<thead>
<tr>
<th>Country</th>
<th>ROC test statistic</th>
<th>p-value</th>
<th>Clark-West test statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>10.930</td>
<td><strong>0.0009</strong></td>
<td>-6.7740</td>
<td>1.0000</td>
</tr>
<tr>
<td>Brazil</td>
<td>19.200</td>
<td>&lt;<strong>0.001</strong></td>
<td>8.0833</td>
<td>&lt;<strong>0.001</strong></td>
</tr>
<tr>
<td>Indonesia</td>
<td>36.319</td>
<td>&lt;<strong>0.001</strong></td>
<td>19.003</td>
<td>&lt;<strong>0.001</strong></td>
</tr>
<tr>
<td>Korea</td>
<td>4.8024</td>
<td><strong>0.0284</strong></td>
<td>-0.7131</td>
<td>0.7621</td>
</tr>
<tr>
<td>Malaysia</td>
<td>0.0064</td>
<td>0.9361</td>
<td>4.8475</td>
<td>&lt;<strong>0.001</strong></td>
</tr>
<tr>
<td>Mexico</td>
<td>0.0001</td>
<td>0.9930</td>
<td>-26.953</td>
<td>1.0000</td>
</tr>
<tr>
<td>Peru</td>
<td>6.9116</td>
<td><strong>0.0086</strong></td>
<td>9.7281</td>
<td>&lt;<strong>0.001</strong></td>
</tr>
<tr>
<td>Philippines</td>
<td>0.0906</td>
<td>0.7634</td>
<td>11.102</td>
<td>&lt;<strong>0.001</strong></td>
</tr>
<tr>
<td>Taiwan</td>
<td>0.5000</td>
<td>0.4795</td>
<td>1.4058</td>
<td>0.0799</td>
</tr>
<tr>
<td>Thailand</td>
<td>6.5530</td>
<td><strong>0.0105</strong></td>
<td>-7.7623</td>
<td>1.0000</td>
</tr>
<tr>
<td>Uruguay</td>
<td>111.15</td>
<td>&lt;<strong>0.001</strong></td>
<td>8.1857</td>
<td>&lt;<strong>0.001</strong></td>
</tr>
<tr>
<td>Venezuela</td>
<td>0.0691</td>
<td>0.7927</td>
<td>17.209</td>
<td>&lt;<strong>0.001</strong></td>
</tr>
</tbody>
</table>

* The coefficients significant at a 5% level are in bold
### III.3. Logit with m.e. variables vs. Markov with m.e. variables and spread switching

<table>
<thead>
<tr>
<th>Country</th>
<th>ROC test statistic</th>
<th>ROC p-value</th>
<th>Diebold-Mariano test statistic</th>
<th>Diebold-Mariano p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>62.678</td>
<td>&lt;0.001</td>
<td>12.965</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Brazil</td>
<td>9.7859</td>
<td>0.0018</td>
<td>8.783</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Indonesia</td>
<td>46.529</td>
<td>&lt;0.001</td>
<td>29.244</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Korea</td>
<td>9.8754</td>
<td>0.0017</td>
<td>12.207</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Malaysia</td>
<td>21.455</td>
<td>&lt;0.001</td>
<td>17.066</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mexico</td>
<td>17.829</td>
<td>&lt;0.001</td>
<td>50.850</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Peru</td>
<td>45.942</td>
<td>&lt;0.001</td>
<td>12.164</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Philippines</td>
<td>7.4266</td>
<td>0.0064</td>
<td>9.7129</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Taiwan</td>
<td>34.195</td>
<td>&lt;0.001</td>
<td>16.591</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Thailand</td>
<td>45.902</td>
<td>&lt;0.001</td>
<td>18.281</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Uruguay</td>
<td>125.00</td>
<td>&lt;0.001</td>
<td>12.877</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Venezuela</td>
<td>17.351</td>
<td>&lt;0.001</td>
<td>9.4665</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

* The coefficients significant at a 5% level are in bold
Comparison tests

Remarks

→ The panel logit model with market expectation variables works better than the Markov specifications.

→ The introduction of market expectation variables has a positive effect on the forecasting performance of an EWS.
## Best model - Optimal cut-off

<table>
<thead>
<tr>
<th>Country</th>
<th>Cut-off</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Cut-off</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>0.300</td>
<td>82.76</td>
<td>82.61</td>
<td>0.620</td>
<td>41.38</td>
<td>100.0</td>
</tr>
<tr>
<td>Brazil</td>
<td>0.160</td>
<td>100.0</td>
<td>69.47</td>
<td>0.880</td>
<td>7.69</td>
<td>100.0</td>
</tr>
<tr>
<td>Indonesia</td>
<td>0.200</td>
<td>96.97</td>
<td>96.20</td>
<td>0.930</td>
<td>72.73</td>
<td>100.0</td>
</tr>
<tr>
<td>Korea</td>
<td>0.206</td>
<td>85.71</td>
<td>90.96</td>
<td>0.930</td>
<td>14.29</td>
<td>100.0</td>
</tr>
<tr>
<td>Malaysia</td>
<td>0.380</td>
<td>93.10</td>
<td>93.97</td>
<td>0.730</td>
<td>65.52</td>
<td>100.0</td>
</tr>
<tr>
<td>Mexico</td>
<td>0.379</td>
<td>100.0</td>
<td>99.15</td>
<td>0.390</td>
<td>75.00</td>
<td>100.0</td>
</tr>
<tr>
<td>Peru</td>
<td>0.260</td>
<td>100.0</td>
<td>82.72</td>
<td>0.940</td>
<td>12.90</td>
<td>100.0</td>
</tr>
<tr>
<td>Philippines</td>
<td>0.346</td>
<td>67.95</td>
<td>68.35</td>
<td>0.730</td>
<td>20.51</td>
<td>100.0</td>
</tr>
<tr>
<td>Taiwan</td>
<td>0.160</td>
<td>94.12</td>
<td>65.17</td>
<td>0.670</td>
<td>17.65</td>
<td>98.31</td>
</tr>
<tr>
<td>Thailand</td>
<td>0.120</td>
<td>90.32</td>
<td>61.29</td>
<td>0.321</td>
<td>25.81</td>
<td>96.24</td>
</tr>
<tr>
<td>Uruguay</td>
<td>0.119</td>
<td>93.33</td>
<td>75.73</td>
<td>0.900</td>
<td>50.00</td>
<td>100.0</td>
</tr>
<tr>
<td>Venezuela</td>
<td>0.225</td>
<td>85.71</td>
<td>67.90</td>
<td>0.330</td>
<td>64.29</td>
<td>77.78</td>
</tr>
</tbody>
</table>

- **Optimal cut-off:** $C \leq 0.38$
- **Crisis and calm periods:** correctly identified
## Best model - Evaluation criteria

<table>
<thead>
<tr>
<th>Country</th>
<th>AUC</th>
<th>Kuiper score</th>
<th>Pietra index</th>
<th>Bayesian error rate</th>
<th>QPS</th>
<th>LPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>0.898</td>
<td>65.37</td>
<td>0.235</td>
<td>0.132</td>
<td>0.215</td>
<td>-0.325</td>
</tr>
<tr>
<td>Brazil</td>
<td>0.907</td>
<td>69.47</td>
<td>0.249</td>
<td>0.132</td>
<td>0.202</td>
<td>-0.311</td>
</tr>
<tr>
<td>Indonesia</td>
<td>0.996</td>
<td>93.17</td>
<td>0.330</td>
<td>0.0138</td>
<td>0.034</td>
<td>-0.058</td>
</tr>
<tr>
<td>Korea</td>
<td>0.920</td>
<td>76.67</td>
<td>0.273</td>
<td>0.0780</td>
<td>0.135</td>
<td>-0.228</td>
</tr>
<tr>
<td>Malaysia</td>
<td>0.985</td>
<td>87.07</td>
<td>0.311</td>
<td>0.048</td>
<td>0.083</td>
<td>-0.131</td>
</tr>
<tr>
<td>Mexico</td>
<td>0.998</td>
<td>99.15</td>
<td>0.350</td>
<td>0.008</td>
<td>0.011</td>
<td>-0.023</td>
</tr>
<tr>
<td>Peru</td>
<td>0.947</td>
<td>82.72</td>
<td>0.292</td>
<td>0.107</td>
<td>0.166</td>
<td>-0.266</td>
</tr>
<tr>
<td>Philippines</td>
<td>0.739</td>
<td>36.30</td>
<td>0.163</td>
<td>0.235</td>
<td>0.368</td>
<td>-0.558</td>
</tr>
<tr>
<td>Taiwan</td>
<td>0.739</td>
<td>36.30</td>
<td>0.163</td>
<td>0.235</td>
<td>0.368</td>
<td>-0.558</td>
</tr>
<tr>
<td>Thailand</td>
<td>0.811</td>
<td>51.61</td>
<td>0.192</td>
<td>0.138</td>
<td>0.218</td>
<td>-0.348</td>
</tr>
<tr>
<td>Uruguay</td>
<td>0.939</td>
<td>69.06</td>
<td>0.257</td>
<td>0.105</td>
<td>0.165</td>
<td>-0.246</td>
</tr>
<tr>
<td>Venezuela</td>
<td>0.777</td>
<td>53.61</td>
<td>0.189</td>
<td>0.257</td>
<td>0.370</td>
<td>-0.530</td>
</tr>
</tbody>
</table>

- Performance assessment criteria: close to the optimal values
- Robustness of the model to sensitivity analysis
Towards a Unified Statistical Framework to Evaluate Financial Crises Early Warning Systems

**Fig. 1 – Predicted probability of crisis**

(a) Argentina  
(b) Brazil  
(c) Indonesia  
(d) Korea  
(e) Malaysia  
(f) Mexico
Towards a Unified Statistical Framework to Evaluate Financial Crises Early Warning Systems
Conclusion
Objective: Developing a new EWS evaluation framework based on optimal cut-offs, credit-scoring criteria and comparison tests

→ Substantial improvement of the predictive power of EWS

→ Markov models are not as efficient as panel logit model with market expectation variables
Conclusions

The optimal model

→ Predicts well most currency crises in the specified emerging markets

→ Robust to some sensitivity analysis

Extensions

→ Markov switching model with time varying probabilities

→ Other market expectation variables

→ A more consistent database (a longer period, more countries)

→ Out of sample validation