Online Appendix for "What Drives Bank Performance?"

Luca Guerrieri and James Harkrader

# 1 Extended Dataset

This appendix includes results based on a panel of all 34 firms that participated in the 2020 stress tests conducted by the Federal Reserve, without dropping the 6 firms with fewer than 40 observations. Table 1 shows the full list of firms and the start and end date of the available data for each firm. Figure 1 shows the cumulative R-squares from the regressions of macro factors and banking factors on charge-offs and PPNR for each bank in our panel. This figure is analogous to Figure 2 in the main paper. For the firms that included in both analyses, a comparison of the two figures shows only small quantitative differences. Moreover, there are no qualitative differences. This robustness strengthens our results.

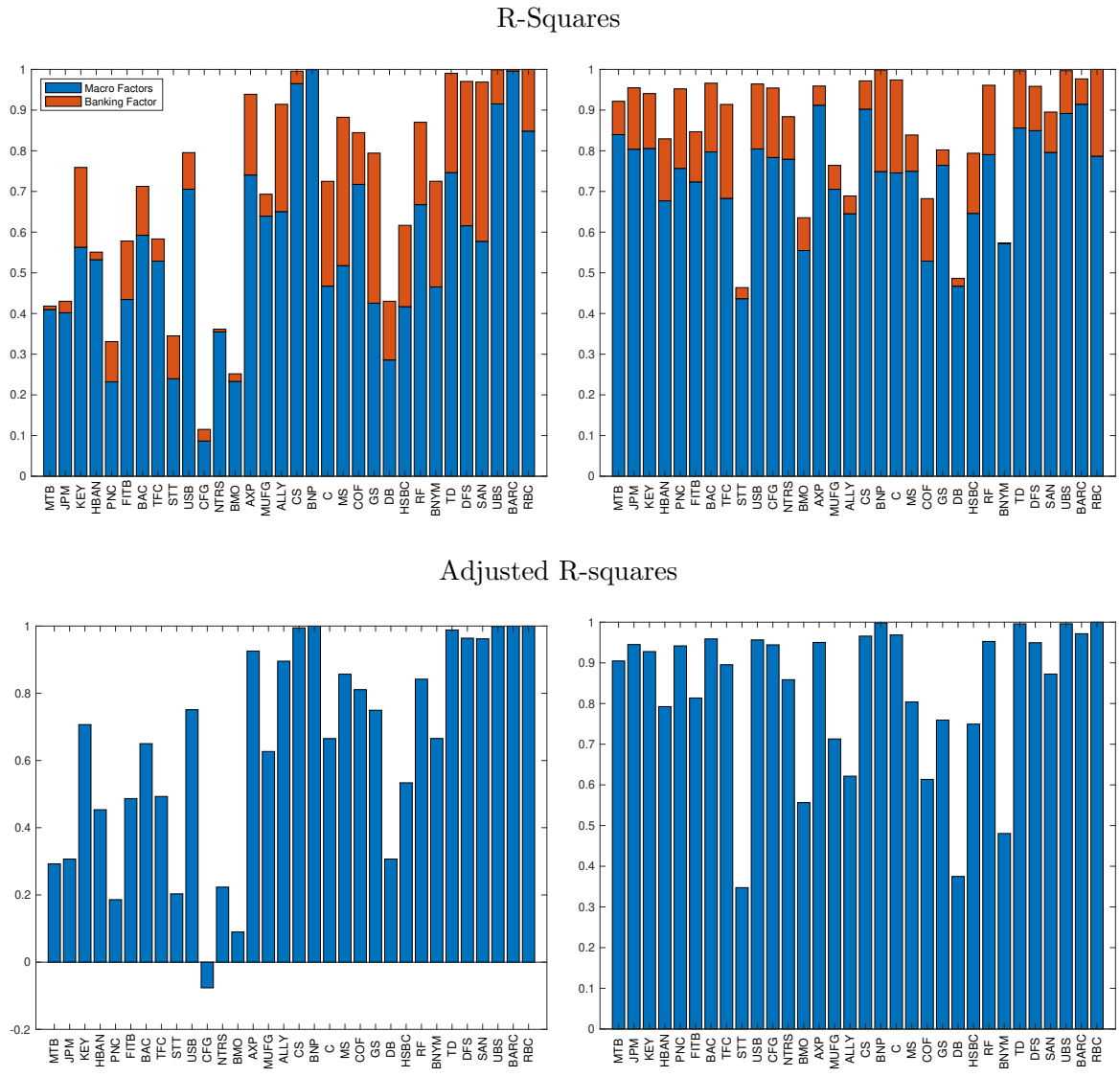
## 1.1 A Look at the Backcasts

After evaluating the relative importance of different sets of factors, we turn to our ability to use these factors to impute missing data. We rely on a pseudo-out-of-sample exercise for which, as a representative example, we backcast the charge-off rates for one bankc, JPMorgan Chase based on a shortened sample that drops the first 50 quarterly observations, so that the observed sample for this pseudo-out-of-sample exercise runs from the third quarter of 2014 through the third quarter of 2020. The backast for our baseline model, shown by the dashed red line, hugs the observed data, the solid back line. To help evaluate the importance of the macro factors the figure also shows an alternative backcast based on a model that simply

Table 1: Data Range

| Bank | Ticker | Start Date | End Date |
|---|---|---|---|
| ALLY FINANCIAL INC. | ALLY | 2009:2 | 2020:3 |
| AMERICAN EXPRESS COMPANY | AXP | 2009:1 | 2020:3 |
| BANK OF AMERICA CORPORATION | BAC | 2002:1 | 2020:3 |
| BANK OF NEW YORK MELLON CORPORATION, THE | BNYM | 2007:3 | 2020:3 |
| BARCLAYS US LLC | BARC | 2016:3 | 2020:3 |
| BMO FINANCIAL CORP. | BMO | 2002:1 | 2020:3 |
| BNP PARIBAS USA, INC. | BNP | 2016:3 | 2020:3 |
| CAPITAL ONE FINANCIAL CORPORATION | COF | 2004:4 | 2020:3 |
| CITIGROUP INC. | C | 2002:1 | 2020:3 |
| CITIZENS FINANCIAL GROUP, INC. | CFG | 2002:1 | 2020:3 |
| CREDIT SUISSE HOLDINGS, INC. | CS | 2016:3 | 2020:3 |
| DB USA CORPORATION | DB | 2002:1 | 2020:3 |
| DISCOVER FINANCIAL SERVICES | DFS | 2009:2 | 2020:3 |
| FIFTH THIRD BANCORP | FITB | 2002:1 | 2020:3 |
| GOLDMAN SACHS GROUP, INC., THE | GS | 2009:1 | 2020:3 |
| HSBC NORTH AMERICA HOLDINGS INC. | HSBC | 2004:1 | 2020:3 |
| HUNTINGTON BANCSHARES INCORPORATED | HBAN | 2002:1 | 2020:3 |
| JPMORGAN CHASE & CO. | JPM | 2002:1 | 2020:3 |
| KEYCORP | KEY | 2002:1 | 2020:3 |
| MORGAN STANLEY | MS | 2009:1 | 2020:3 |
| MUFG AMERICAS HOLDINGS CORPORATION | MUFG | 2002:1 | 2020:3 |
| M&T BANK CORPORATION | MTB | 2002:1 | 2020:3 |
| NORTHERN TRUST CORPORATION | NTRS | 2002:1 | 2020:3 |
| PNC FINANCIAL SERVICES GROUP, INC., THE | PNC | 2002:1 | 2020:3 |
| RBC US GROUP HOLDINGS LLC | RBC | 2018:2 | 2020:3 |
| REGIONS FINANCIAL CORPORATION | RF | 2004:3 | 2020:3 |
| SANTANDER HOLDINGS USA, INC. | SAN | 2012:1 | 2020:3 |
| STATE STREET CORPORATION | STT | 2002:1 | 2020:3 |
| TD GROUP US HOLDINGS LLC | TD | 2015:3 | 2020:3 |
| TRUIST FINANCIAL CORPORATION | TFC | 2002:1 | 2020:3 |
| U.S. BANCORP | USB | 2002:1 | 2020:3 |
| UBS AMERICAS HOLDING LLC | UBS | 2016:3 | 2020:3 |

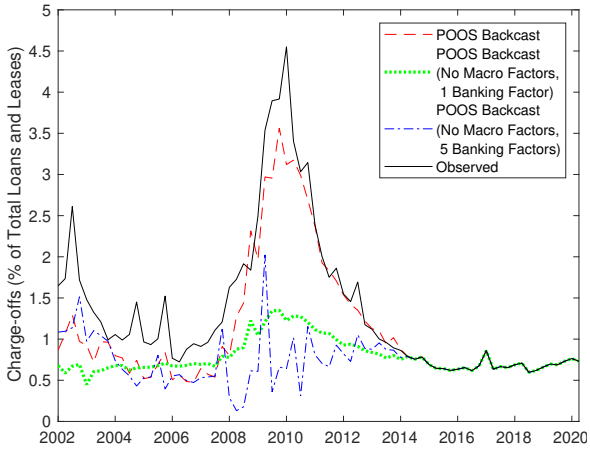
Figure 1: Macro Factors Explain a Large Portion of the Variation in Charge-off Rates as Opposed to PPNR. Left: PPNR, Right: Charge-offs.




drops the macro factors and retains one banking factor, as in the baseline model. The figures shows clearly a deterioration in performance in a pseudo-out-of-sample sense, as can be gauged from the greater gap between the dashed green line, for the alternative backcast, and the solid black line (the observed data) than the analogously gap for the dashed red line (for the baseline backcast). As a second alternative, we consider increasing the

number of banking factors to compensate for the exclusion of the macro factors. We use the test in Bai and Ng (2002) to determine the optimal number of banking factors, which calls for five factors. As can be seen from the figure, the dash-dotted blue line for this alternative backcast is closer to the observed data but not as close as our baseline backcast.

Figure 2: Comparing Pseudo-Out-of-Sample Backcasts of Charge-off Rates, JPMorgan Chase




Source: Authors' calculations and Federal Reserve Y-9C Release.

## 2 Matlab Implementation

The matlab routine backcast_ts.m takes as input an unbalanced dataset with an arbitrary pattern of missing observations. For contiguously missing observations at the beginning or at the end of each series, the routine relies on an extension of the method of Stock and Watson as described in the Guerrieri, L. and C. Harkrader "What Drives Bank Performance," which allows for use of external factors. Missing observations away from the begging or end sections of each series are filled in using cubic splines. The routine can be downloaded from a GitHub repository: https://github.com/lucashare/backcasting.

The dataset with missing observations needs to include at least as many series with non-missing observations at the beginning or at the end of the dataset as the number of factors to be extracted from the dataset.

The function has the following call syntax:

```
[ts_filled, ts_factors, ts_resid] = ...

backcast_ts(ts, n_factors, other_factors, max_iterations, tol0,
```

Inputs

- ts: array or timetable with series of different length.
- n_factors: integer indicating number of factors to extract from the dataset. The dataset with missing observations needs to include at least as many series with non-missing observations at the beginning or at the end of the dataset as the number of factors to be extracted from the dataset.
- other_factors: array or timetable of factors external to the first dataset to be used for backcasts. If omitted, we just add a constant.
- max_iterations: optional integer indicating maximum number of iterations before aborting. The default value is 100,000.
- tol0: optional parameter used to determine convergence, expressed as the maximum absolute difference in backcasts between iterations. The default value is 0.001.
- myfloor: optional theoretical lower bound for any observation to be imputed — if the imputed values fall below myfloor, myfloor is used instead. The default value is minus infinity.

Outputs

- ts_filled: array or timetable of the same shape ts but with missing values filled in.
- ts_factors: array or timetable with factors extracted from the time table ts.
- ts_resid: array or timetable of residuals for each series in ts_filled not explained by external factors or additional factors extracted from ts. NB: if ts is an array, ts_filled, ts_factors and ts_resid will be returned as arrays.

The replication code for Figures 2 and 3 shown in the main body of the paper provide examples on how to call the routine.

## References

Bai, J. and S. Ng. 2002. "Determining the Number of Factors in Approximate Factor Models." *Econometrica* pp. 191–221. 4