Predicting Debt Distress in Low-Income Countries

Clemens Graf von Luckner (Stanford University)
Sebastian Horn (University of Hamburg & Kiel Institute)
Aart Kraay (World Bank)
Rita Ramalho (World Bank)

November 15, 2024
25th Jacques Polak Annual Research Conference

Motivation: general

- Sovereign debt crises have large economic and social costs
  - Lower growth and productivity; higher poverty (Reinhart and Rogoff 2009, Aguiar and Amador 2021, Farah-Yacoub et al. 2024a)
- Early warning systems for “debt distress” can have large benefits if they enable preventative measures
  - Large literature on predicting debt distress (Moreno Badia et. al. (2022) survey)
  - Premise of WB/IMF LIC Debt Sustainability Framework

Motivation: specific WB/IMF policy application

- The WB/IMF DSF for Low-Income Countries
  - Developed in mid-2000s to
    - provide early warning of debt vulnerabilities
    - prevent debt re-accumulation post-HIPC/MDRI
  - Sets borrowing limits, mix of grants and loans from IDA, **debt relief envelopes**
  - Last reviewed in 2017, new review ongoing

- Core of LIC DSF is an empirical model to predict debt distress
  - Used to derive country-specific debt thresholds reflecting countries' debt carrying capacity

Our contributions to literature and policy

1. Refining debt distress outcome measurement to reflect the *onset* rather than *resolution* of distress

2. Systematic approach to predictive model selection
   - *Evaluate 559,872 possible models based on J-K-fold cross-validated out-of-sample predictive performance*

3. Evaluate simple versus sophisticated prediction algorithms
   - *Best simple models strongly dominate more sophisticated alternatives such as Random Forests*

4. Policy implications for LIC-DSF
   - *Scope to improve predictive performance*
   - *Scope to reduce overoptimism bias through k-year-ahead predictions*

1. Measuring debt distress – *What are we trying to predict?*

1.1 Measuring debt distress: signals

LIC DSF Review 2017 – reflects typical set of debt distress signals in the literature:

Defaults on private creditors:
Data from S&P and Catao & Milesi-Ferreti, whenever available

Arrears:
Arrears > 5% of ppg debt stock, for 3 years

IMF Programs:
Rapid disbursements > 30 % of quota, all program types

Debt restructurings
Default assumed to start 1 year prior:
- Private creditors (Cruces & Trebesch)
- Paris Club creditors (Das et al.)

1.1 Measuring debt distress: signals

LIC DSF Review 2017 – reflects typical set of debt distress signals in the literature:

Defaults on private creditors:
Data from S&P and Catao & Milesi-Ferreti, whenever available

Arrears:
Arrears > 5% of ppg debt stock, for 3 years

IMF Programs:
Rapid disbursements > 30 % of quota, all program types

Debt restructurings
Default assumed to start 1 year prior:
- Private creditors (Cruces & Trebesch)
- Paris Club creditors (Das et al.)

Our paper:

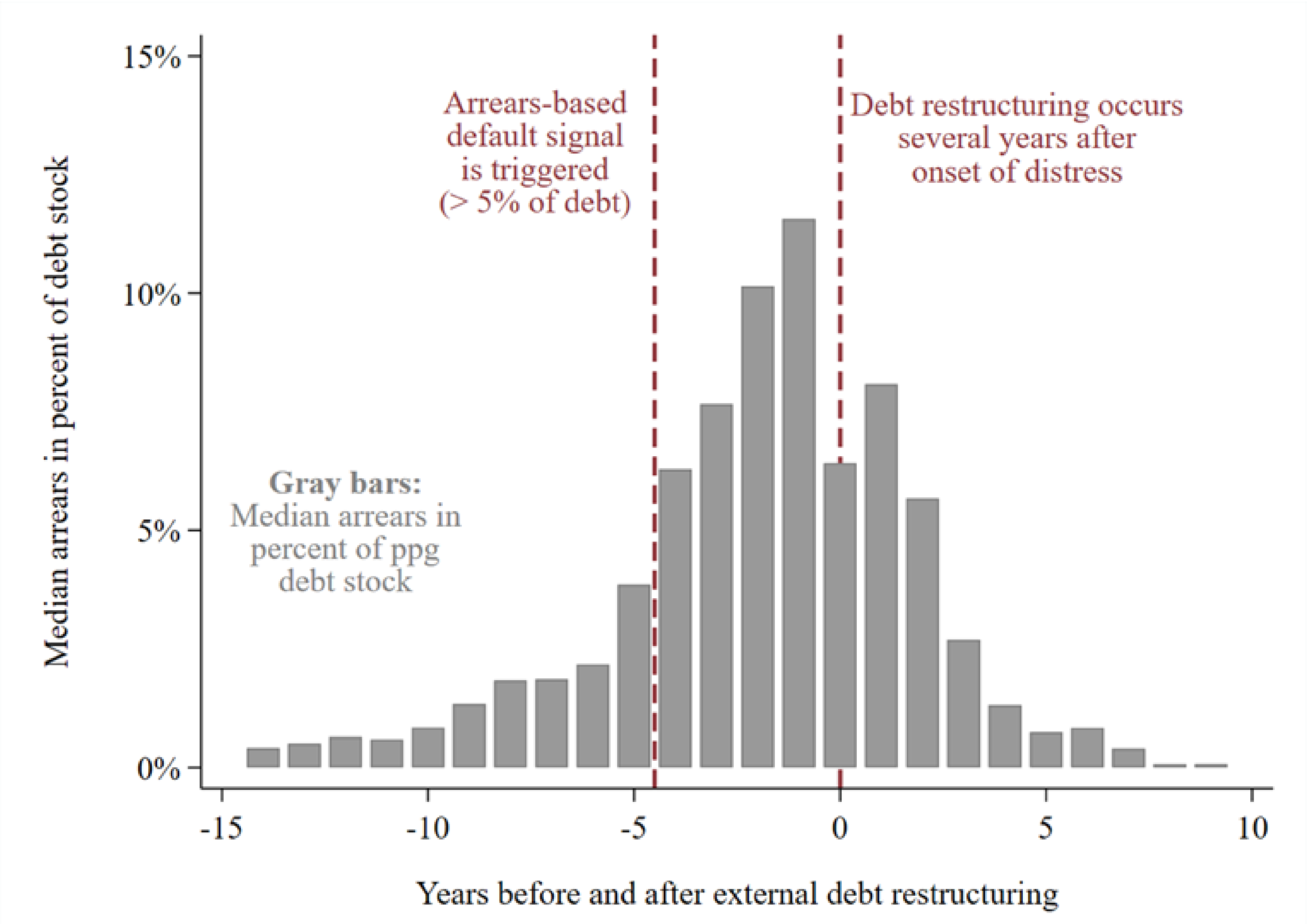
Defaults on private creditors:
New data for all LICs from Asonuma & Trebesch (2016) and Farah-Yacoub et al. (2024)

Arrears:
Arrears > 5% of ppg debt stock, for 3 years

IMF Programs:
Rapid disbursements > 30 % of quota, only non-concessional programs / no RFI

no restructuring signal – key timing point – see next slide

1.1 Measuring debt distress: timing of arrears and restructurings



- Restructurings mark conclusion rather than onset of distress (Asonuma & Trebesch 2016)
- Long and variable lags between defaults and restructurings (median of 4 years)

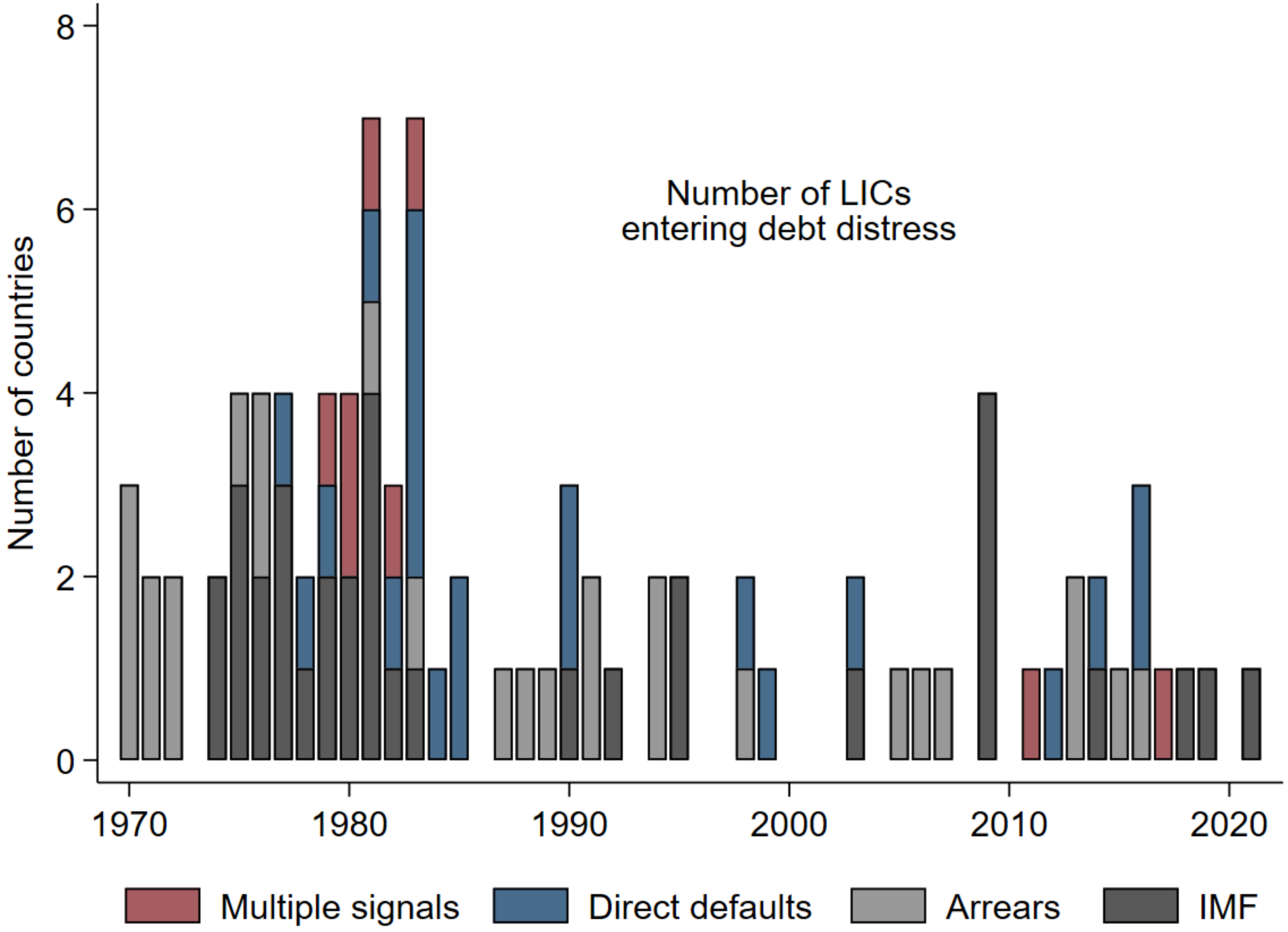
1.2 Measuring debt distress: episodes

- Define distress signal $S_{ct} = 1$ if any one of three distress signals is observed in country c and year t; $S_{ct} = 0$ otherwise
  - (1) defaults, (2) high arrears, (3) large and rapid IMF disbursement

- Define distress episode $Y_{ct+1} = 1$ if:
  - $S_{ct} = S_{ct-1} = S_{ct-2} = 0$: not currently/recently in distress, *and*
  - $S_{ct+1} = 1$: ***distress signal next year***

- Define non-distress episodes $Y_{ct+1} = 0$ if:
  - $S_{ct} = S_{ct-1} = S_{ct-2} = 0$: not currently/recently in distress, *and*
  - $S_{ct+1} = 0$: ***no distress signal next year***

1.4 Measuring debt distress: results



- Sample consists of 1,752 observations covering 80 LIC DSF-eligible countries 1970-2021

- 90 cases of debt distress represent 5.1 percent of sample

- Three signals of roughly equal importance in triggering distress episodes

1. Measuring debt distress

2. Predicting debt distress

2.1 Predicting debt distress: probit model

- Estimate predicted probability of distress using probit model:

$$P[Y_{ct+1} = 1] = \Phi(\beta' X_{ct}), \quad \hat{p}_{ct+1} = \Phi(\hat{\beta}' X_{ct})$$

- Cutoff probability p^* generates binary prediction $\hat{Y}_{ct+1} = 1$ when $\hat{p}_{ct+1} > p^*$
  - False positive rate: $FPR = \left(\sum_{ct}(1 - Y_{ct+1})\, \hat{Y}_{ct+1}\right) / \sum_{ct}(1 - Y_{ct+1})$
  - False negative rate: $FNR = \left(\sum_{ct} Y_{ct+1}\left(1 - \hat{Y}_{ct+1}\right)\right) / \sum_{ct} Y_{ct+1}\,)$

- Select p^* to minimize quadratic mean squared prediction loss function:

$$L(FNR, FPR) = \sqrt{wFNR^2 + (1-w)FPR^2}\ , \qquad w = 0.5$$

2.2 Predicting debt distress: standard covariates from literature

- Debt indicators
  - PPG/GDP, PPG/Exports, **NPV/GDP, NPV/Exports, TDS/Exports, TDS/Revenue**, domestic debt/GDP, Interest on Public Debt / Exports
- Policies and institutions
  - **Country Policy and Institutional Assessment (CPIA)**, years since last distress, decaying average of past distress
- Business cycle and level of development
  - **GDP growth**, inflation rate, depreciation rate, log GDP per capita
- Political cycle
  - Years in office, years until end of term
- External environment
  - Current account balance, FDI inflows, **remittances**, change in TOT, 10-year US Treasuries rate, **reserves/imports**, trade openness, **world growth**

2.3 Predicting distress: model space

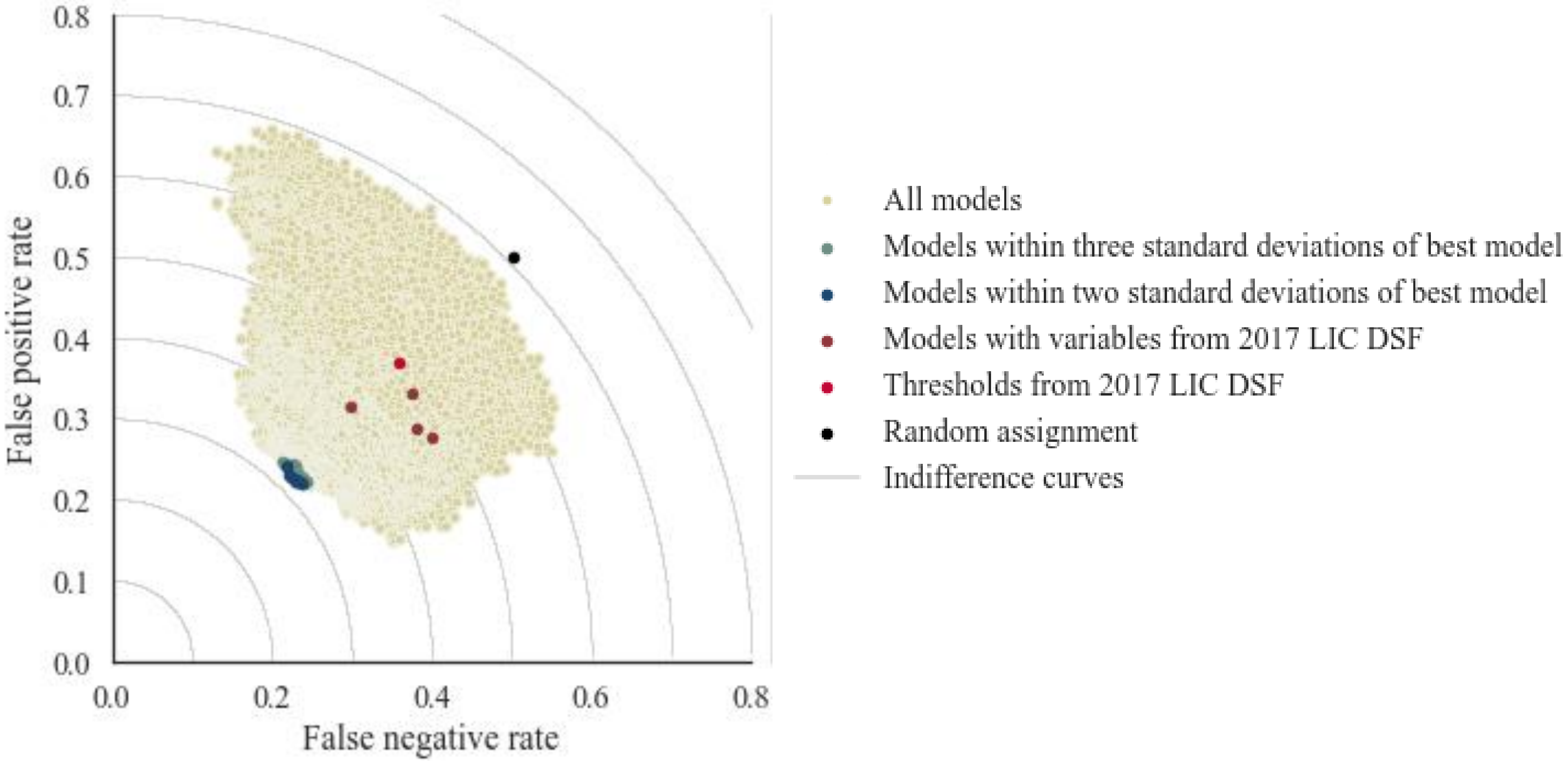
- "Brute force" approach to model selection – consider models defined by *all* relevant combinations of RHS variables
  - With 28 covariates we would have $2^{28} \approx 268$ million models to study
  - With $(J = 10) \times (K = 10)$ cross-validation, 26 billion probit regressions to estimate
- To limit scope of task to substantively interesting models, we impose a set of restrictions on the model space:
  - CPIA always included (*for LIC DSF policy application, not very binding constraint)*
  - *At least one* debt variable (for LIC DSF policy application)
  - *At most one* debt stock-, debt service-, credit history-, political cycle-, change in value of money-variable
  - With these restrictions, we consider 559,872 candidate prediction models

2.4 Predicting distress: cross-validation

- Evaluate models based on out-of-sample predictive performance using J-K-fold cross-validation

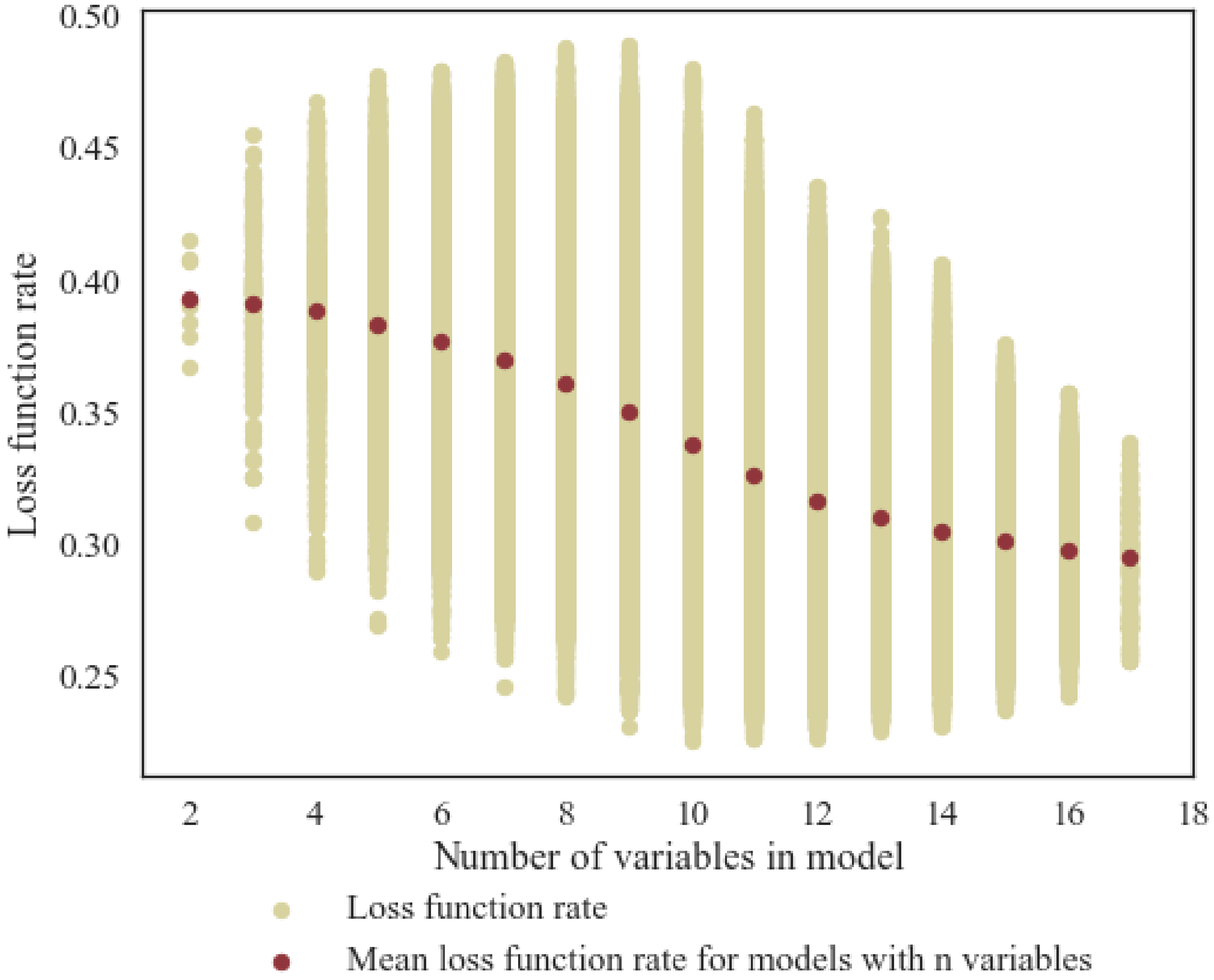
- For each combination of variables that defines a model:
  - Perform K-fold cross-validation for $K = 10$ exhaustive folds
    - *Estimate* probit model in training sample
    - *Select* p^* that minimizes prediction loss function in test sample
  - Repeat $J = 10$ times, retrieving minimized FPR, FNR, and $L(FNR, FPR)$
  - Calculate mean of FPR, FNR, and $L(FNR, FPR)$ across $J = 10$ replications
  - Construct confidence interval for $L(FNR, FPR)$

2.5 Results



0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
False positive rate
0.0
0.2
0.4
0.6
0.8
False negative rate
All models
Models within three standard deviations of best model
Models within two standard deviations of best model
Models with variables from 2017 LIC DSF
Thresholds from 2017 LIC DSF
Random assignment
Indifference curves

2.6 Predicting distress: parsimony vs. performance



- Some tradeoffs between model size and predictive performance
- Average predictive performance improves modestly with model size (red dots)
- Best model performance conditional on size is U-shaped in model size (lower envelope of yellow points)

2.7 Predicting debt distress: best models

- Our algorithms turn up many (many!) good models that outperform models in 2017 LIC-DSF
  - 431K models **(77%) outperform 2017 LIC-DSF** mechanical predictions
  - 395K models (71%) outperform best single probit with 2017 LIC-DSF variables

- To guide selection of "best models" we impose three further conditions:
  1. No perverse incentives: $\hat{\beta}_{CPIA} < 0, \hat{\beta}_{DEBT} > 0$
  2. Data availability: *data on all variables in model available for at least 90% of country-year observations since 2000.*
  3. Meaningful effect size: $\beta_x^{marginal}\, \sigma_x / \sigma_{\hat{p}} > 0.05$ *(top 20 percent)*

2.7 Predicting distress: selected best models

	Dependent variable: Incidence of external sovereign debt distress in t+1						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CPIA	-0.15**	-0.10*	-0.12**	-0.08*	-0.08**	-0.06	-0.11**
Ext. debt service / exports	0.22***	0.19***	0.18***	0.17***	0.17***	0.15***	
Reserves / imports		-0.24***		-0.21***	-0.17***	-0.15**	-0.17*
GDP p.c.			0.18***	0.14***	0.16***	0.13**	0.25***
Inflation			0.11**		0.09**		0.11
GDP growth				-0.09**		-0.09*	
Credit history					-0.07		-0.07
Commodities terms of trade						-0.08*	-0.09
US 10 year yield						0.08*	0.12*
Openness							-0.10
CA balance / GDP							-0.06
Ext. debt stock / exports							0.09
Number of variables	2	3	4	5	6	7	10
Loss function	0.37	0.31	0.29	0.27	0.26	0.27	0.29
False positive rate	0.37	0.32	0.33	0.21	0.25	0.19	0.30
False negative rate	0.36	0.30	0.24	0.32	0.28	0.34	0.27
Data coverage since 2000	0.96	0.94	0.91	0.93	0.91	0.92	0.92
Number of observations	1,002	1,002	1,002	1,002	1,002	1,002	1,002

- Model with only six regressors minimizes prediction loss function ($L = 0.26$)

2.7 Predicting distress: selected best models

	Dependent variable: Incidence of external sovereign debt distress in t+1						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CPIA	-0.15**	-0.10*	-0.12**	-0.08*	-0.08**	-0.06	-0.11**
Ext. debt service / exports	0.22***	0.19***	0.18***	0.17***	0.17***	0.15***	
Reserves / imports		-0.24***		-0.21***	-0.17***	-0.15**	-0.17*
GDP p.c.			0.18***	0.14***	0.16***	0.13**	0.25***
Inflation			0.11**		0.09**		0.11
GDP growth				-0.09**		-0.09*	
Credit history					-0.07		-0.07
Commodities terms of trade						-0.08*	-0.09
US 10 year yield						0.08*	0.12*
Openness							-0.10
CA balance / GDP							-0.06
Ext. debt stock / exports							0.09
Number of variables	2	3	4	5	6	7	10
Loss function	0.37	0.31	0.29	0.27	0.26	0.27	0.29
False positive rate	0.37	0.32	0.33	0.21	0.25	0.19	0.30
False negative rate	0.36	0.30	0.24	0.32	0.28	0.34	0.27
Data coverage since 2000	0.96	0.94	0.91	0.93	0.91	0.92	0.92
Number of observations	1,002	1,002	1,002	1,002	1,002	1,002	1,002

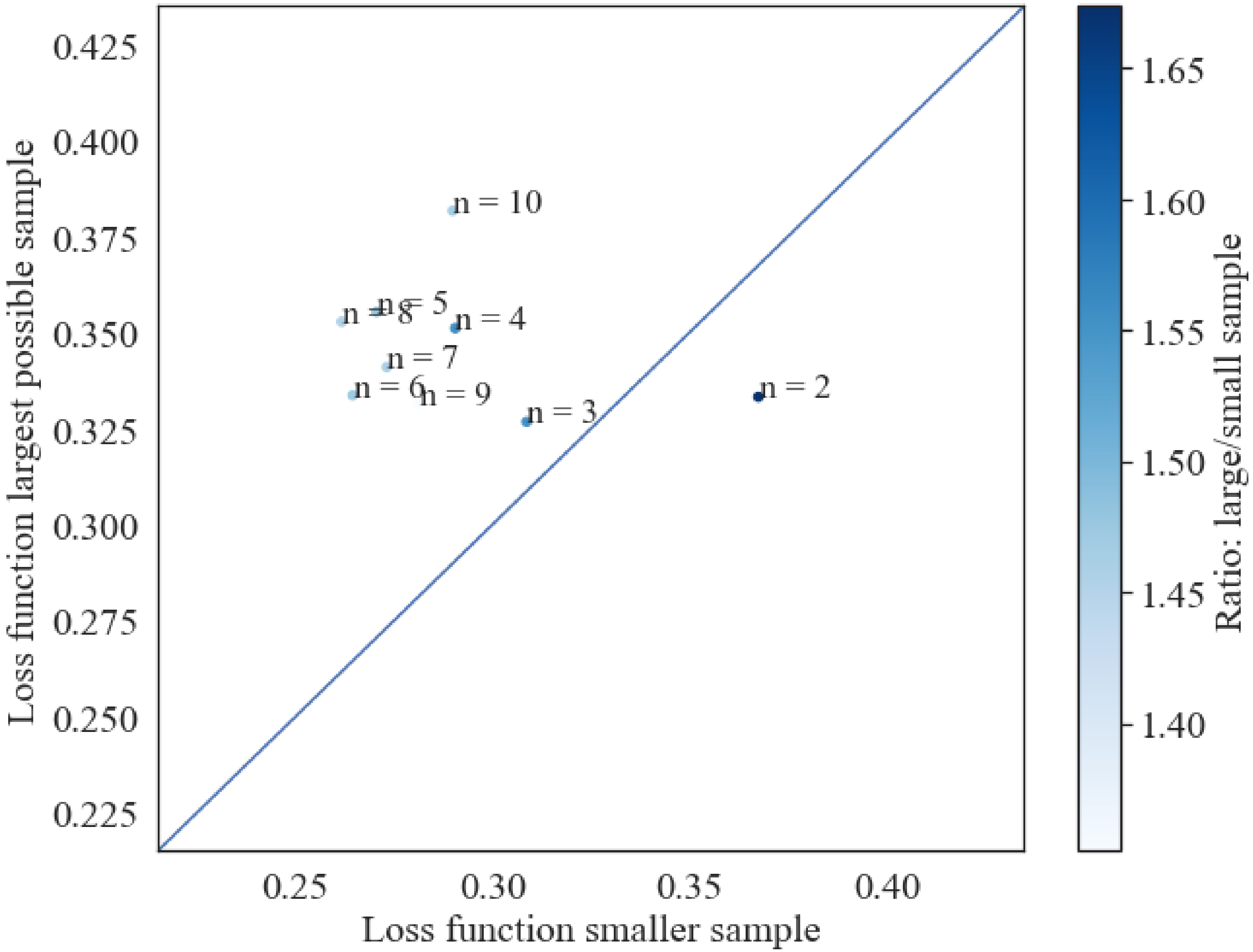
- Model with only six regressors minimizes prediction loss function ($L = 0.26$)
- Very parsimonious model with only three predictors does almost as well ($L = 0.31$) – "*Best Parsimonious Model" (BPM)*

2.7 Predicting distress: selected best models

	Dependent variable: Incidence of external sovereign debt distress in t+1						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CPIA	-0.15**	-0.10*	-0.12**	-0.08*	-0.08**	-0.06	-0.11**
Ext. debt service / exports	0.22***	0.19***	0.18***	0.17***	0.17***	0.15***	
Reserves / imports		-0.24***		-0.21***	-0.17***	-0.15**	-0.17*
GDP p.c.			0.18***	0.14***	0.16***	0.13**	0.25***
Inflation			0.11**		0.09**		0.11
GDP growth				-0.09**		-0.09*	
Credit history					-0.07		-0.07
Commodities terms of trade						-0.08*	-0.09
US 10 year yield						0.08*	0.12*
Openness							-0.10
CA balance / GDP							-0.06
Ext. debt stock / exports							0.09
Number of variables	2	3	4	5	6	7	10
Loss function	0.37	0.31	0.29	0.27	0.26	0.27	0.29
False positive rate	0.37	0.32	0.33	0.21	0.25	0.19	0.30
False negative rate	0.36	0.30	0.24	0.32	0.28	0.34	0.27
Data coverage since 2000	0.96	0.94	0.91	0.93	0.91	0.92	0.92
Number of observations	1,002	1,002	1,002	1,002	1,002	1,002	1,002

- Model with only six regressors minimizes prediction loss function ($L = 0.26$)
- Very parsimonious model with only three predictors does almost as well ($L = 0.31$) – *"Best Parsimonious Model" (BPM)*
- Total debt service on external debt is only debt indicator that features consistently in best models
- Fairly balanced FPR and FNR (due to choice of quadratic loss function)

2.8 Predicting distress: robustness



- Model selection algorithm uses common balanced sample with for so that all models are evaluated on the prediction of the *same set of episodes*.
- We re-estimate the top performing models in the largest available dataset
- More parsimonious models appear to be *more robust* to increases in sample size

1. Measuring debt distress

2. Predicting debt distress

3. More sophisticated models

3.1 More sophisticated models: RF

- Probit model is very simple – can more sophisticated prediction algorithms generate better out-of-sample predictions?

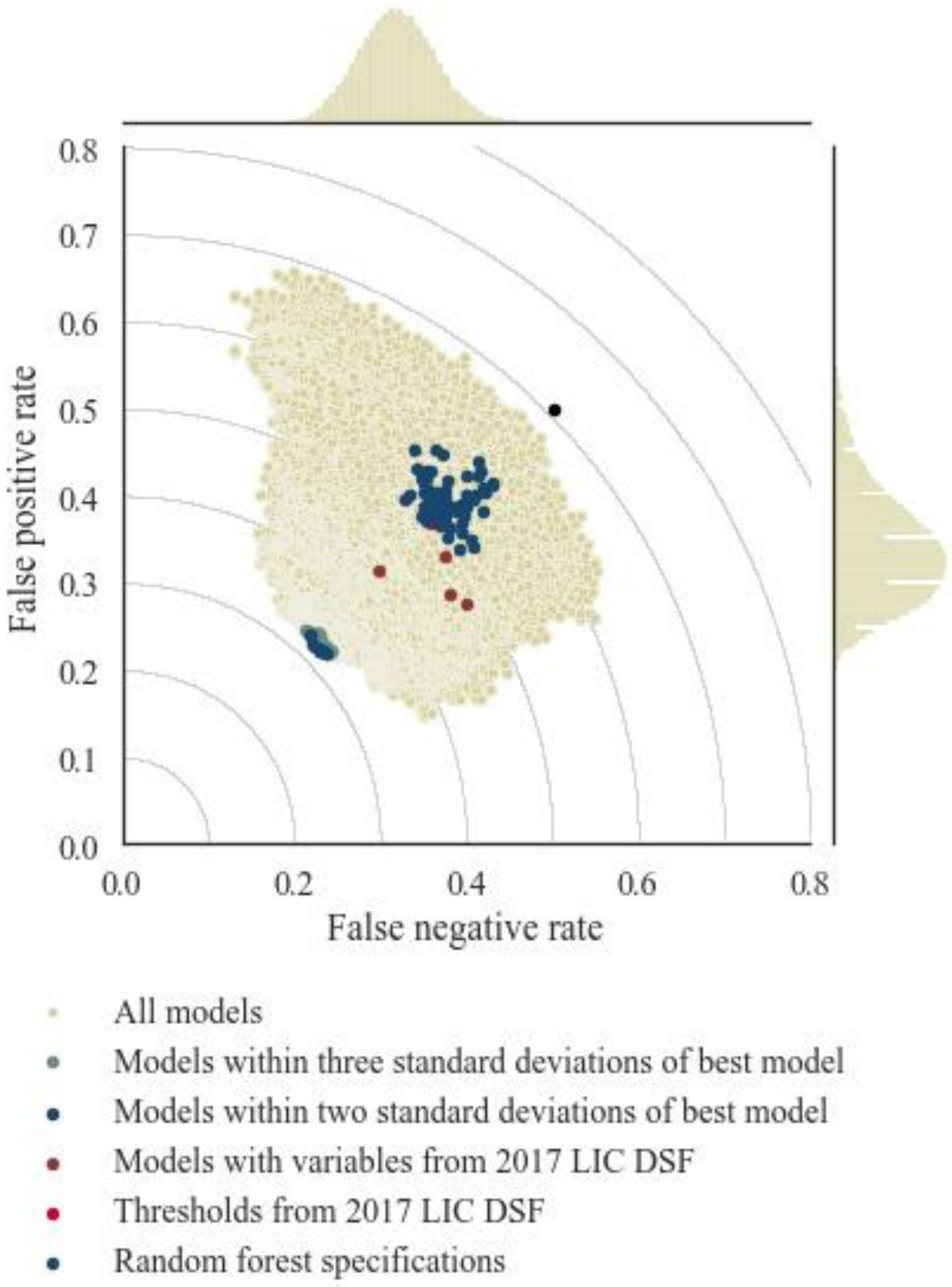
- Consider random forest (RF), apply in same sample, with same J-K-fold cross-validation

- Perform grid search over three key tuning parameters to find best RF model:
  - Node purity criterion
  - Number of trees
  - Depth of trees

3.2: More sophisticated models: results

- Best RF does significantly worse in predicting debt distress than simple linear probit models
  - FPR=0.35 (vs. 0.32 in BPM)
  - FNR=0.37 (vs. 0.30 in BPM)

- In line with general principle that ML prediction algorithms adds little value in small datasets (Shmueli, 2010)



1. Measuring debt distress

2. Predicting debt distress

3. More sophisticated models

4. LIC DSF implications

4.1 LIC DSF implications: better predictions

- Apply old LIC-DSF model to our new sample of events through 2021
  - New model predicts much better than mechanical predictions from LIC-DSF model
  - *Not entirely fair comparison because LIC-DSF model was trained on different sample and a different definition of events*

	Predicted by Best Parsimonious Model		**Predicted by 2017 LIC DSF Model**	
Actual	No distress	Distress	No distress	Distress
No distress	903	394	814	483
Distress	15	44	19	40
False positive rate	0.30		0.37	
False negative rate	0.25		0.32	
Quadratic loss function	0.28		0.35	

4.2 LIC-DSF implications: better predictions

- Re-estimate Best Parsimonious Model in 2017 LIC-DSF sample, with old dependent variable and linear loss function from previous review
  - Pick cutoff probability to match in-sample predictive performance
  - *Not entirely fair comparison for BPM because its predictor list was selected in a different sample, yet BPM outperforms slightly.*

	Predicted by Best Parsimonious Model		**Predicted by 2017 LIC DSF Model**	
Actual	No distress	Distress	No distress	Distress
No distress	172	156	206	122
Distress	9	54	13	50
False positive rate	0.48		0.37	
False negative rate	0.14		0.21	
Linear loss function	0.25		0.26	

4.3 LIC-DSF implications: optimism bias

- LIC DSF predicts debt distress based on whether projected future debt ratios cross thresholds implied by probit regressions
  - Predicting debt ratios into the future is difficult (numerator and denominator)
  - Risk of optimism bias

- Instead of "*predicting the predictors*" of debt distress, how well can current values of predictors predict distress k periods into the future?

- Define new dependent variable $Y_{ct+k} = 1$ if :
  - $S_{ct} = S_{ct-1} = S_{ct-2} = 0$: not currently/recently in distress, *and*
  - $\max(S_{ct+1}, \ldots, S_{ct+k}) = 1$: distress signal **any time in next $\boldsymbol{k = 5}$ years**

4.3 LIC-DSF implications: 5-year predictions

- 5-year-ahead predictions are nearly as good as or even better than one-year-ahead predictions, e.g. for 3-variable model
  - FP=0.30 (compared with 0.32 for one-year-ahead)
  - FN=0.29 (compared with 0.30 for one-year-ahead)
- Suggests scope to improve LIC-DSF by reducing reliance on predicted future debt ratios

	Dep. variable: Incidence of external sovereign debt distress within next five years						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CPIA	-0.41***	-0.41***	-0.36***	-0.37***	-0.31***	-0.32***	-0.33***
Ext. debt service / exports	0.66***	0.72***	0.66***	0.64***	0.55***	0.51***	0.63***
GDP p.c.		0.59***	0.60***	0.64***	0.53***	0.60***	0.57***
Inflation			0.19**	0.18**			
Openness				-0.09		-0.15	-0.17
CA balance / GDP					-0.25***	-0.29***	-0.21***
Credit history					-0.30***	-0.28***	-0.28***
US 10 year yield					0.31***	0.31***	
Reserves / imports							-0.27***
Y.o.y. change in FX rate							-0.24***
Number of variables	2	3	4	5	6	7	8
Loss function	0.40	0.30	0.29	0.29	0.28	0.28	0.28
False positive rate	0.41	0.30	0.30	0.30	0.29	0.26	0.27
False negative rate	0.40	0.29	0.27	0.27	0.28	0.29	0.28
Data coverage since 2000	0.96	0.93	0.91	0.91	0.93	0.93	0.93
Number of observations	899	899	899	899	899	899	899

Conclusion

- Improved and simplified definition of debt distress
- Systematic approach to model selection generates better predictions
- Low return to prediction model complexity – probit dominates RF
- Five-year-ahead predictions almost as good as one-year-ahead predictions
- Scope to simplify prediction model to make LIC-DSF more transparent

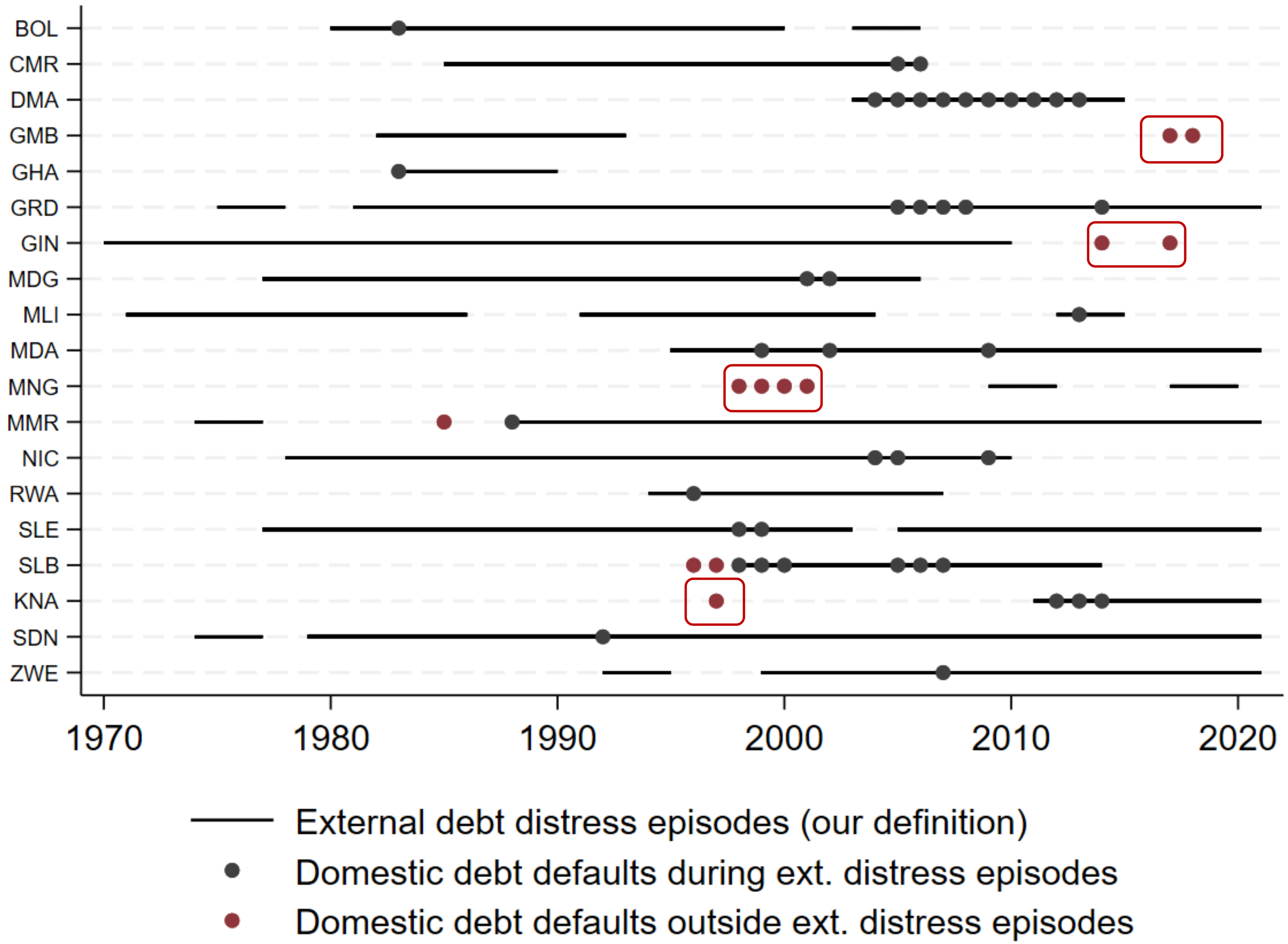
Supplementary Materials

External debt distress episodes by signal

External debt distress episodes in LICs, 1970 - 2015

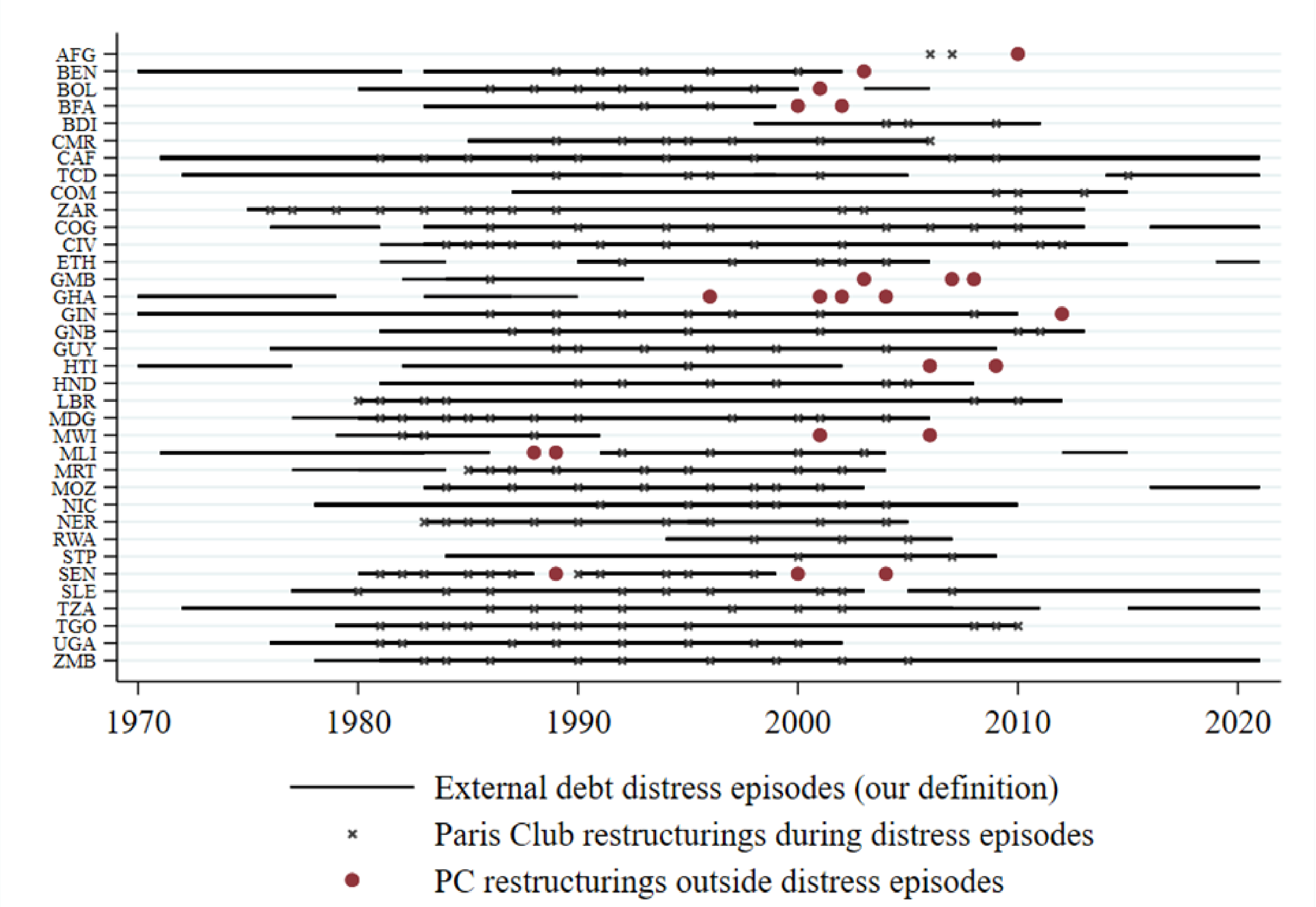
Distress signal	LIC DSF 2017	Our paper
Total number of episodes	**98**	**83**
Of which triggered by		
IMF Disbursements	35	31
Arrears	32	26
Defaults	1	19
Restructurings	22	-
Some combination of the above	8	7

1.5 Measuring debt distress: domestic debt



- Use data from IMF (2021) to capture 67 domestic debt restructurings in LICs (no data on default)
- Strongly correlated with external distress episodes (as expected)
- Yields only 4 new distress episodes

1.1 External debt distress episodes and Paris Club restructurings



- Only 25 out of 295 Paris Club restructurings occur outside our external debt distress episodes (9 percent of cases)
- Most of these 25 cases are related to the HIPC initiative and treat debts that had been contracted multiple decades ago in the 1970s and 1980s
- They “lag” rather than “lead” our external debt distress episodes.

2.2 Predicting debt distress: measurement challenges with domestic debt

- Domestic debt levels in LICs are on the rise, but systematic data remains scarce

- We construct series on *total public* (domestic plus external) debt to GDP by combining data from the IMF WEO, Abbas et al. (2010) and Reinhart and Rogoff (2009)
  - Near-complete coverage of country-year observations since 1970 in 2017 LIC DSF database

- Two main shortcomings:
  - Consistency of institutional coverage can not always be ascertained
  - Limited and noisy data on domestic debt *service* which matters most for debt distress in short run – longest available data covers only payments of *interest* not *principal*

External debt distress episodes: RFI

Including rapid disbursements under RFI as a distress signal creates 12 additional distress episodes in 2020

Country	Account	Arr. Type	Year	DSF Risk Rating
Albania	GRA	RFI	2020	-
Bangladesh	GRA	RFI	2020	Low
Benin	GRA	RFI	2020	Moderate
Comoros	GRA	RFI	2020	Moderate
Cote d'Ivoire	GRA	RFI	2020	Moderate
Kyrgyz Republic	GRA	RFI	2020	Moderate
Lesotho	GRA	RFI	2020	Moderate
Myanmar	GRA	RFI	2020	Low
Nicaragua	GRA	RFI	2020	Moderate
Nigeria	GRA	RFI	2020	-
Senegal	GRA	RFI	2020	Moderate
Solomon Islands	GRA	RFI	2020	Moderate

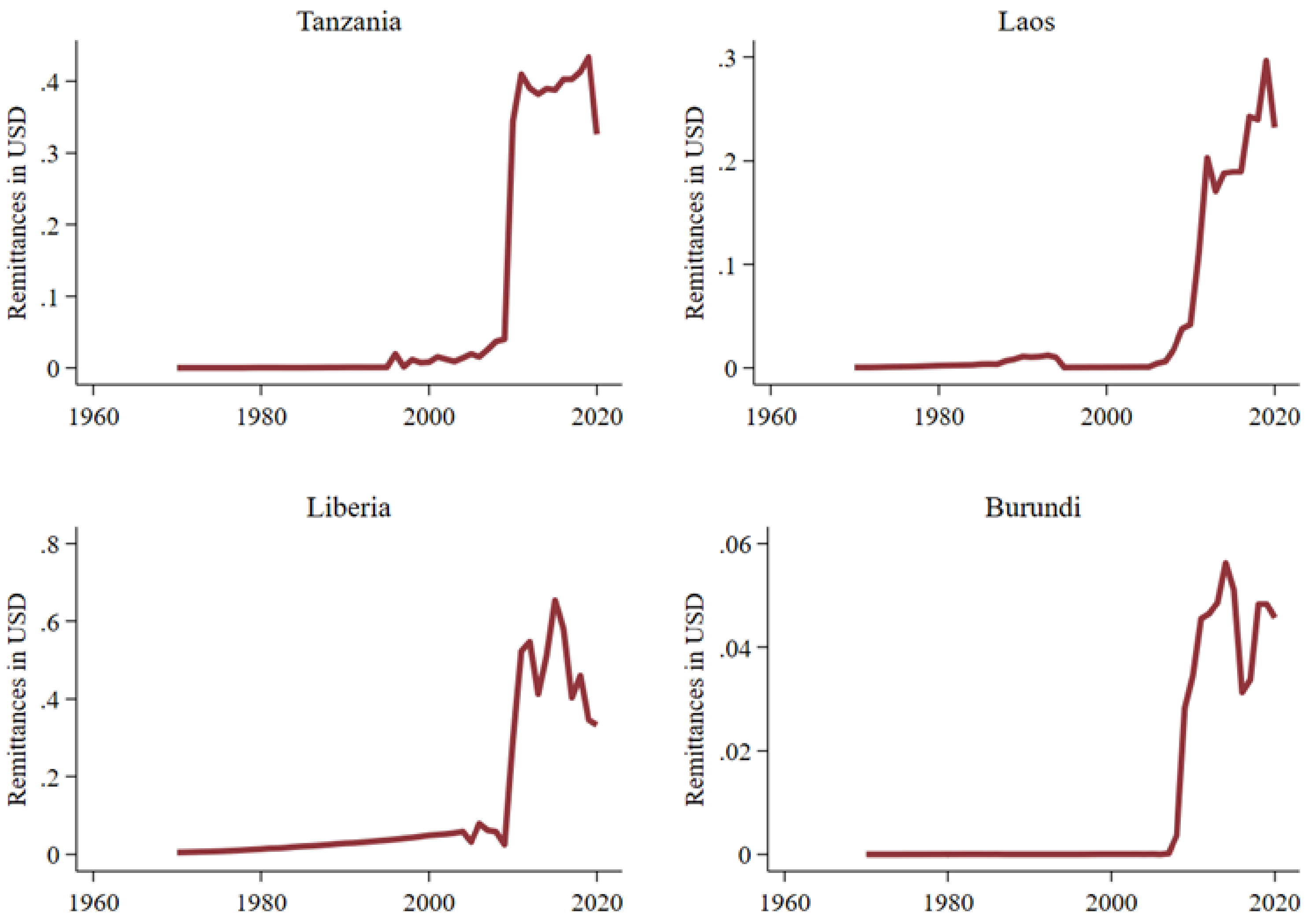
- None of these countries defaulted on private creditors and none accumulated significant payment arrears.
- They had comparatively low debt burdens in comparison to their debt servicing capacity.

Unconstrained top models

	Dependent variable: Incidence of external sovereign debt distress in t+1						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CPIA	-0.15**	-0.10*	-0.10*	-0.08*	-0.07*	-0.06*	-0.06**
Ext. debt service / exports	0.22***	0.19***	0.25***	0.22***	0.22***	0.12***	0.12***
Reserves / imports		-0.24***	-0.23***	-0.20***	-0.20***	-0.13**	-0.09**
Public debt / exports			-0.10	-0.11			
Inflation				0.08			
GDP p.c.					0.12***	0.12***	0.11***
NPV of ext. debt / exports					-0.08		
GDP growth					-0.10**		
Remittances / GDP						-1.74**	-1.64**
Post-2001 dummy						-0.07	0.01
Remittances / GDP x post-2001						1.62**	1.54**
US 10-year yield							0.08*
Ext. debt stock / GDP							-0.03
Years left in current term							-0.06
Number of variables	2	3	4	5	6	7	10
Loss function	0.37	0.31	0.29	0.27	0.26	0.25	0.23
False positive rate	0.37	0.32	0.30	0.25	0.20	0.25	0.22
False negative rate	0.36	0.30	0.28	0.28	0.31	0.24	0.23
Number of observations	1,002	1,002	1,002	1,002	1,002	1,002	1,002

- Loss function minimized by model with 10 predictor variables (LF = 0.23)
- Several top models not suitable for policy application:
  - “Wrong” coefficient signs lead to perverse policy incentives
  - Economically meaningless effect sizes
  - Predictors with low data coverage and large measurement error

Remittances: data peculiarities



- Many LIC remittance series exhibit structural breaks in early 2000s that cannot be explained by fundamentals
- Likely driven by improved recording of cross-border transactions, in particular by AML and CFT regulation implemented post 9/11 (Clemens & McKenzie 2018)
- We include remittance with post-2001 dummy and IA term to control for this pattern