MIMOSA

Microfinance Index of Market Outreach and Saturation

Model Update: MIMOSA 2.1

September 2019

A special thanks to Jann Goedecke for lending his statistical skills during model development.

Model Update: MIMOSA 2.1

This is the first update to the MIMOSA 2.0 model, developed and published in 2015 (see "MIMOSA 2.0: Mapping the (micro)credit cycle"). In addition, as a companion to this model update, we are publishing an update to our full list of country scores, based on the 18 MIMOSA country and regional reports published to-date, as well as the 2017 survey data from Findex.

The MIMOSA 2.1 model is an incremental update. Its chief role is to refine how the model accounts for population density. Over the course of our work, we discovered that the 2.0 model, developed entirely based on country-level data, became increasingly inaccurate in assessing high-density areas that often occur at sub-national levels, such as states, districts, provinces and the like.

For this reason, reports published since 2015 have set an explicit cap on population density levels in the model's capacity calculations – a maximum of 600 persons/km^2 in all countries, except for India and Pakistan, where the maximum was raised to 1200 persons/km^2. The 600 limit was applied because high-density areas in these countries were due to the presence of large urban areas dominating a given administrative district (ex: Grand Casablanca in Morocco or Phnom Penh in Cambodia), whereas in the country-level dataset on which MIMOSA 2.0 was developed, all but one country (Bangladesh) had population densities below 600 persons/km^2. Thus, we felt it would be inappropriate to extrapolate the effects of high population density in urban environments based on a dataset that didn't have any comparable observations.

We made an exception for India and Pakistan because a large number of essentially rural districts there featured high population densities, including above 600 persons/km^2. Due to their similarity to Bangladesh (country population density of 1235 persons/km^2), in those cases we set the cutoff at 1200 – thus staying within the boundaries of the data used to construct MIMOSA 2.0, but not extrapolating beyond them.

Since that time, we have gathered extensive data at the sub-national level – a total of 403 observations from 9 countries, after excluding those where data was incomplete or unreliable. These observations have included very high density urban areas, such as Mumbai and Delhi districts in India (both above 20,000 persons/km^2), as well as a large number of high-density rural areas, especially in South Asia. These observations finally allow us to upgrade the capacity model, creating a more refined calculation for population density.

Number of observations, by population density

Country	<600	600-1200	>1200	Total
Bolivia	40			40
Cambodia	96		4	100
India	73	63	19	155
Jordan	11	2		13
Kyrgyzstan	22			22
Morocco	14		1	15
Nicaragua	6			6
Peru	50			50
Senegal	2			2
Grand Total	314	65	24	403

Building the MIMOSA 2.1 model

In order to refine the model, we used a two-step regression. In step one, we regressed penetration against HDI and Credit Bureau Score variables, but only at the national level (284 observations). This includes data from Findex 2011, 2014, and 2017 surveys and equivalent years for HDI and the Credit Bureau Score. In countries where we had done MIMOSA studies and had more reliable local data for penetration or credit bureau indicators, we replaced them with our verified data. In step two we add Population Density and run a second regression for the combined national and regional data (663 observations).

Step 1

The main reason for the two-step process is that regional HDI variables present significant problems of consistency. HDI itself is a combination of four indicators (per capita GNI, life expectancy at birth, mean years of schooling and expected years of schooling). While country-level data is reported by UNDP in a consistent manner, with UNDP's staff carrying out the necessary adjustments needed to ensure that the final HDI score is comparable across all countries, regional HDI data is typically far less consistent. In countries with good demographic and economic statistics at the regional level, we have been able to use the underlying statistical indicators to calculate regional HDIs ourselves that are comparable to those used by UNDP. However, in several of the 11 countries covered to-date, some of these component indicators are not available at the regional level, and we instead rely on human development reports that include regional HDI. These reports are typically published once a decade or so by the government statistics bureaus, sometimes in partnership with UNDP and sometimes independently. However, the HDI values in these reports are not always consistent with the UNDP methodology, and result in HDI figures that are really not comparable to those reported at the national level.

For the latter group of country reports (including India and Cambodia – countries with the largest number of data points), we have had to do a substantial amount of adjusting and re-scaling of regional HDIs to align them with the national figures reported by UNDP. This preserves the relative differences in HDI, such that the highest- and lowest-HDI regions in the country maintain their positions relative to the national UNDP-reported figure. While this is sufficient to provide the needed information to assess saturation levels across different regions of the same country, these adjustments don't give us confidence to use the adjusted regional HDI values for developing the model itself.

Thus, we have decided to use only national-level HDI figures in building the model, entailing the two-step process described above. The result of the first step (national-level) regression is as follows:

Dependent variable: Borrowing capacity

Variables	*Coefficient*	*Standard error*	*95% Confidence Interval*	
Intercept	-1.0282	1.8765	[-4.7219;	2.6656]
HDI	14.6434	3.3243	[8.0998;	21.1870]
CreditBureauScore	0.0333	0.0068	[.01990;	.0468]
Observations	284			
Adj. R^2	0.3201			
F value	67.612			

Note that both HDI and CreditBureauScore remain highly predictive factors for the model, yielding an adjusted R^2 of .3201 – nearly identical to the R^2 of the MIMOSA 2.0 model, whose predictive power (adj R^2 of .3167), despite having one fewer variable (no population density) and a significantly bigger dataset. Clearly, the addition of the 2017 Findex data and corrections from the MIMOSA reports have not altered the underlying logic of the original model.

Step 2

In step two, we use the residuals from the first regression (difference between observed penetration and the regressed output) and run a second regression using the combined country- and sub-country data (663 observations) against Population Density with the following adjustment for the variable:

$$PopulationDensity_{adjusted} = Log(Max(PopulationDensity, 20))$$

The reasons for this adjustment is that it better reflects the diminishing returns (in terms of credit capacity) of increasingly greater density, while also controlling for the impact on low-density areas, where differences become similarly reduced as density declines further. Indeed, the adjustment is very effective for high-density areas, but it remains somewhat weaker at the lower end. After all, even in low-density rural areas, population is rarely truly scattered, but instead resides in villages and towns. We believe that this remains a modest weakness of the model, and we will keep exploring ways to improve it.

The result of this second regression is as follows:

Dependent variable: Residuals from first regression

Variables	*Coefficient*	*Standard error*	*95% Confidence Interval*	
Intercept	-0.1824	1.0274	[-2.1997;	1.8348]
PopDensityAdj	0.0224	0.0049	[.0129;	.0032]
Observations	663			
Adj. R^2	0.0299			
F value	21.404			

Note again the high statistical significance of the adjusted population density variable, as well as the high adjusted R^2 (.0299), which is after all additional to the 1st step regression.

Finalizing the model

Combining the two models yields a strong adj R^2 (.350)[1], well above the value generated by MIMOSA 2.0. Clearly, the new model is better able to capture the effect of the three underlying variables and do so at both national and sub-national levels. However, before generating the final model, we make one more adjustment: we reduce the final predicted value by -3.725. This figure comes from comparing the output of the new model on all 663 observations against that generated by the original MIMOSA 2.0 model. The new model was on average 3.725 points higher. This is not because the earlier model underestimated capacity. Rather, it's because the input data on which the new model was built is unrepresentative – after all, most of the countries where we conducted MIMOSA studies were chosen because of concerns about their levels of penetration. Therefore, we expected that the average penetration rate in our expanded dataset would be higher and were not surprised by this outcome. This adjustment allows us to ensure that the new model remains valid even when applied to all markets, and moreover, it maintains continuity and comparability with MIMOSA 2.0.

After these adjustments, the new MIMOSA 2.1 model is as follows:

$$Capacity = -4.936 + 14.643 * HDI + 0.03334 * CreditBureauScore + 0.02245 * Log\big(Max(PopulationDensity, 20)\big)$$

Comparing the models

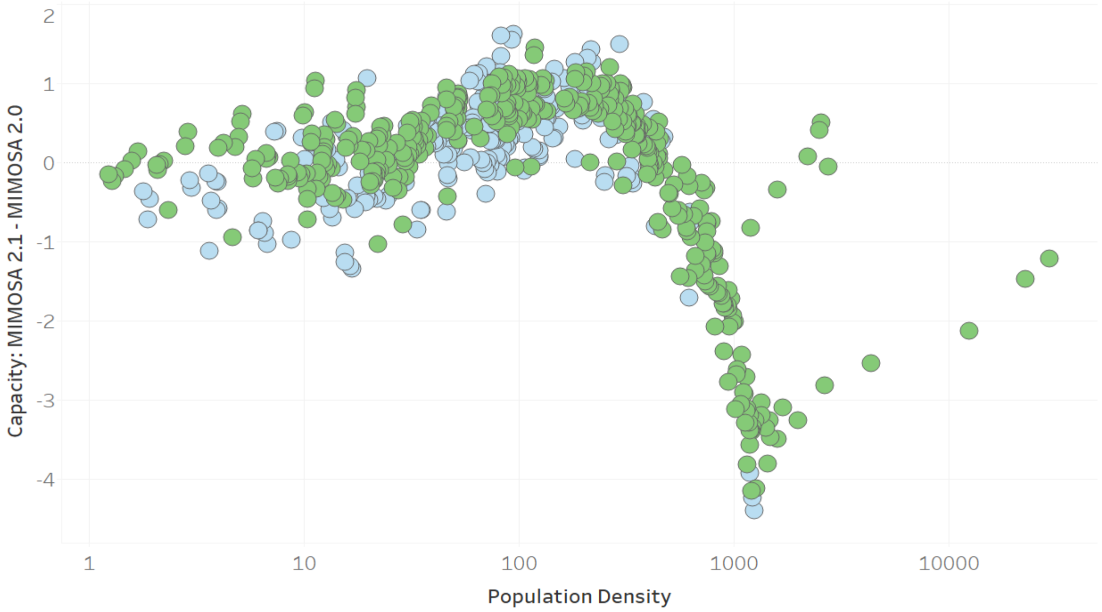
While the MIMOSA 2.0 and 2.1 models share a great deal of similarity, there are notable differences, especially with respect to how they treat population density. Most of the impact is for high-density areas, where capacity has been reduced, with the exception of regions above 1200 persons/km^2 (India & Pakistan) and 600 persons/km^2 (all others). The odd-looking outgrowths from the curve in the figure 1 are areas where these cutoffs were in effect under the old model, but have now been removed.

[1] Summing the adjusted R^2 from the two regressions can be somewhat problematic statistically, but it is sufficient for illustrative purposes.

Mimosa data
Findex data

Model Update: MIMOSA 2.1

Figure 1: Comparing MIMOSA models (Population Density)



In very low density regions, the effect has been a very small reduction in capacity (typically less than 1/2 point), and a modest increase of up to 1 point for moderate density regions (100-250 persons/km^2). Otherwise, the model update has largely maintained stability, and with the exception of very high density areas, no region has seen a change of more than 1.5 points in credit capacity – less than 1/2 of the value of a single MIMOSA score.

The result is also a tighter and more consistent distribution of MIMOSA scores across the HDI band (Figure 2). The capacity estimates at the upper edge of the old model have been brought closer to the norm.

Figure 2: Comparing MIMOSA models (HDI)

Model Update: MIMOSA 2.1

New MIMOSA Score Ranges

Updating the model also requires a minor update to the score ranges, which were based on the standard deviation (3.69) of the outputs of MIMOSA 2.0. In the new model, the standard deviation of outputs for national-level scores is 3.76 – a minor increase from earlier. We are thus updating the score ranges to reflect this new level:

MIMOSA Score	Penetration over/ under capacity		
	Percentage points	Standard Deviations	Market status
6	>11.3%	3+	Saturated
5	7.5-11.3%	2 to 3	
4	3.8-7.5%	1 to 2	
3	0-3.8%	0 to 1	Normal
2	-3.8 - 0%	-1 to 0	
1	< -3.8%	< -1	Underserved

The new model and score ranges will be applied to all reports and updates issued from September 2019 onwards, until the next update. We will not re-issue previously published reports using the new model, but we certainly do recommend revisiting the published scores of high density regions, where capacity may have been significantly reduced.