

This paper was published in Science on November 27, and is protected by copyright.

However, a full-text (pdf) version of the paper can be accessed by visiting <http://www.jblumenstock.com/research> and clicking on the corresponding link. (Science provides the corresponding author with a special link that allows anyone to access the article free of charge).

Otherwise, the article can be found at <https://www.sciencemag.org/content/350/6264/1073.full>

Apologies for the hassle!

ECONOMICS

## Predicting poverty and wealth from mobile phone metadata

Joshua Blumenstock,<sup>1\*</sup> Gabriel Cadamuro,<sup>2</sup> Robert On<sup>3</sup>

Accurate and timely estimates of population characteristics are a critical input to social and economic research and policy. In industrialized economies, novel sources of data are enabling new approaches to demographic profiling, but in developing countries, fewer sources of big data exist. We show that an individual's past history of mobile phone use can be used to infer his or her socioeconomic status. Furthermore, we demonstrate that the predicted attributes of millions of individuals can, in turn, accurately reconstruct the distribution of wealth of an entire nation or to infer the asset distribution of microregions composed of just a few households. In resource-constrained environments where censuses and household surveys are rare, this approach creates an option for gathering localized and timely information at a fraction of the cost of traditional methods.

individual's socioeconomic characteristics. This distinction is a scientific one, which also has several important implications: First, it allows for the method to be used in contexts for which recent census or household survey data are unavailable. Second, when an authoritative source of data does exist, it can be used to more objectively validate or refute the model's predictions. This limits the likelihood that the model is overfit on data from a single source, which is otherwise difficult to control, even with careful cross-validation (20). Third, our approach allows for a broad class of potential applications that require inferences about specific individuals instead of census tracts. As we discuss in the supplementary materials (section 6), future iterations of this approach could help to improve the targeting of humanitarian aid and social welfare, disseminate information to vulnerable populations, and measure the effects of policy interventions.

For this study, we used an anonymized database containing records of billions of interactions