Executive Summary
Columbia University SIPA Capstone Project
Election Data: Analyzing the Effectiveness and Accuracy of Ipsos Forecasting Models

The Columbia University SIPA Capstone Group prepared an election forecasting project for Ipsos, a leading global market research firm. The objective of this capstone project is to review the capabilities and limitations of the original Ipsos election prediction model, expand the dataset by including new variables, and enhance the robustness of the estimates through rigorous analysis and diagnostics of the model's constraints. With the addition of a new data set and added variables, a new model was created to predict the outcome of an election.

The election model estimates the likelihood that the incumbent party wins the election. The forecasting model is a logistic regression estimated with maximum likelihood. In addition to the data provided by Ipsos, the team expanded the data set to include additional independent variables, such as government approval ratings, GDP growth rate, inflation rate, and employment growth. The group also included a war variable that measures whether the country in which the election takes place is involved in a military conflict. The scope of the assembled election data was also increased to include elections in 87 countries from 1980 until the end of 2015. This additional data provides for more robust analysis by the Ipsos prediction model.

Seven different models were created and tested using the foundation of the original Ipsos forecasting model. The Columbia Capstone Group was able to successfully enhance the predictive capabilities of the original Ipsos model by addition of the new variables and collection of a more expansive election data set. The predictive power of the original model increased from 74% accuracy to 76%.

This report includes:

- A review of academic literature needed for an in-depth study of predicting elections;
- Methodology, data acquisition, and summarized codebook;
- The original model from Ipsos and its results using the updated data set;
- Estimates from new models using the additional variables; and
- Appendices, including an in-depth literature review, Stata code for each model, and a complete codebook.