Replication Paper: Droning On: Explaining the Proliferation of Unmanned Aerial Vehicles

Ino Terzi*

December 12, 2025

Abstract

This paper replicates Fuhrmann and Horowitz's "Droning On: Explaining the Proliferation of Unmanned Aerial Vehicles" and evaluates whether its substantive conclusions depend on the use of a probit model. Focusing on armed UAV program adoption, I replicate the original probit specification and estimate a linear probability model (LPM) using the same predictors. The two models are compared in terms of average marginal effects, predictive fit, and predicted probabilities across political regime types and levels of terrorism, with uncertainty assessed via bootstrap. The results show that while the probit model better captures nonlinearities at the extremes of the probability distribution, both specifications produce similar qualitative conclusions. Democracies and autocracies are more likely than mixed regimes to adopt armed UAV programmes, and higher levels of terrorism are associated with greater adoption probabilities. Overall, the findings suggest that the core results of the original study are robust to the choice of functional form.

*Graduate Student, Department of Political Science. Northwestern University. Email: inoterzi2030@u.northwestern.edu

Introduction

The purpose of this paper is to extend and replicate the findings of Fuhrmann and Horowitz's "Droning On: Explaining the Proliferation of Unmanned Aerial Vehicles" Fuhrmann and Horowitz (2017). In the original study, the authors develop a model to explain state-level adoption of armed and advanced drone programmes using an original dataset and a probit specification. Their results show that states facing higher levels of terrorism are more likely to obtain armed UAVs, and that both democracies and autocracies are more likely to pursue UAV programme development compared to hybrid regimes.

Building on these findings, this replication paper evaluates whether the probit model used in the original analysis can be substituted with a linear probability model (LPM) without altering the paper's substantive conclusions. Although nonlinear models such as probit are commonly used for dichotomous outcomes, a substantial methodological literature argues that linear regression models can perform comparably in many applied settings. Hellevik Hellevik (2009), as well as Chatla and Shmueli Chatla and Shmueli (2016), for example, show that LPMs often produce coefficient signs, statistical significance, and marginal effects that closely resemble those from probit or logit models. Angrist Angrist (2001) further argues that linear models may be preferable for many causal questions involving binary outcomes due to their transparency and ease of interpretation.

The contribution of this paper is therefore not to challenge the original theory, but to assess the robustness of its empirical results to an alternative modeling strategy. I first replicate the probit specification used by Fuhrmann and Horowitz for armed UAV programme adoption and then estimate an LPM using the same set of predictors. I compare the two models along three dimensions: overall predictive fit, predicted probabilities across political regime types, and predicted probabilities across varying levels of terrorism.

To evaluate model fit, I use calibration plots that compare predicted probabilities to observed outcome rates across deciles of the predicted distribution. I then examine how the probit and LPM differ in their substantive implications for regime type and terrorism by generating model-based predictions while holding other covariates at their means. Differences in predicted probabilities are assessed using bootstrap, allowing for a direct comparison of the two specifications' predictions and associated uncertainty. Overall, this paper asks whether replacing the probit model with an LPM meaningfully changes how we understand the determinants of armed UAV programme adoption. The results suggest that while the two models differ in their behavior at the extremes of the probability distribution, the central findings of the original paper remain intact.

Fitting the Probit and LPM models

The original paper specifies two models, one for armed drone programme development and one for advanced drone programme development. This paper focuses on armed drone programmes.

Fuhrmann and Horowitz's probit model is specified using five parameters: 1) number of territorial disputes that the state is engaged in 2) terrorism index 3) autocracy 4) democracy 5) GDP per capita 6) defence alliance with UAV manufacturing country. Table 1 shows the average marginal effects (AME) for each independent variable in the model specification, the corresponding p-values, and standard errors in parentheses.

Table 1: Probit AME with Clustered Standard Errors

term	coef_se	statistic	p.value
(Intercept)	-7.568 (1.282)	-5.90	0.000
Territorial Disputes	0.092 (0.064)	1.43	0.152
Terrorism	0.36 (0.081)	4.42	0.000
Autocracy	1.746 (0.614)	2.85	0.004
Democracy	1.498 (0.543)	2.76	0.006
GDP per Capita	0.485 (0.126)	3.86	0.000
Alliance with UAV Provider	-0.429 (0.414)	-1.04	0.300

All variables, except for alliance with a UAV provider are positively correlated with the development of an armed UAV programme. The highest AME is recorded for the autocracy variable followed by democracy, GDP per capita, alliance with UAV provider (absolute value), terrorism and finally territorial disputes, which have a much lower AME than the rest of the predictors. P-vlaues for the territorial disputes and alliance with a UAV provider variables are relatively high.

I proceed with specifying the LPM model, using the same predictors as the original probit model.

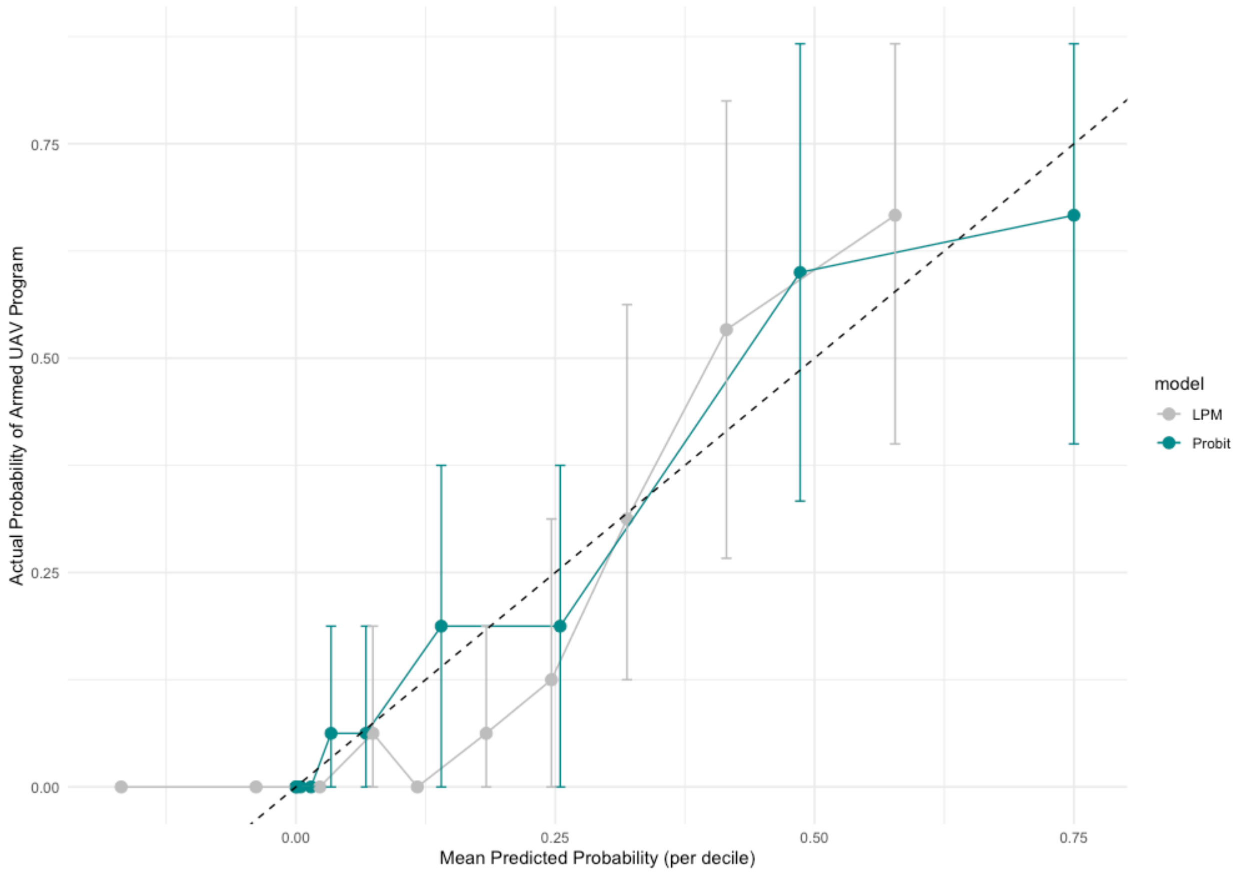
The LPM coefficients reflect the AME. The numerical values of AME are not comparable between the two models, however, we can see that overall the results are consistent in terms of direction and statistical significance, suggesting that the substantive conclusions are not sensitive to functional form. The probit model yields larger average marginal effects for binary predictors due to its nonlinear link function.

Comparing overall fit of Probit and LPM

To assess each model's fit with the observed data, I use a calibration plot (Figure 1). The plot compares how well each model translates predicted probabilities into actual observed outcomes Austin and Steyerberg (2014). On the x-axis, the plot displays the mean predicted probability within each of ten equally sized quantile bins (deciles) of the model's predicted values. The deciles are formed by sorting the sample according to predicted probability and then splitting it into ten groups from the lowest to the highest predicted probability.

The y-axis shows the actual empirical probability of having an armed UAV programme within each bin, that is the proportion of observations where the outcome is 1. On the x-axis are plotted the predictions of each model. The horizontal position of each point therefore is what the model predicts on average for that decile, while its vertical position shows the outcome of the observed data for that decile. Since the 45 degree dashed line is where y=x, converging towards that line suggests good fit with the observed sample. The vertical error bars reflect the 95% binomial confidence intervals around the observed outcome rate.

Figure 1: Calibration Plot: Probit vs LPM (with 95% CIs)



The calibration plots show that the LPM systematically compresses predicted probabilities toward the center of the distribution. As expected for a linear specification, the LPM underestimates probabilities in bins where the true probability is relatively high and overestimates them where the true probability is low. In contrast, the probit model more accurately captures curvature in the probability space, particularly near the extremes (close to 0 or 1), where the binary nature of the outcome introduces natural nonlinearity.

Despite these differences, both models exhibit good calibration over most of the probability range. The predicted and observed probabilities for the probit and LPM specifications track the 45-degree line closely in the mid-range deciles, indicating that both models provide a broadly accurate mapping from covariates to predicted probabilities. The probit model provides a somewhat closer fit in the tails, but the overall difference in calibration between the two approaches is modest.

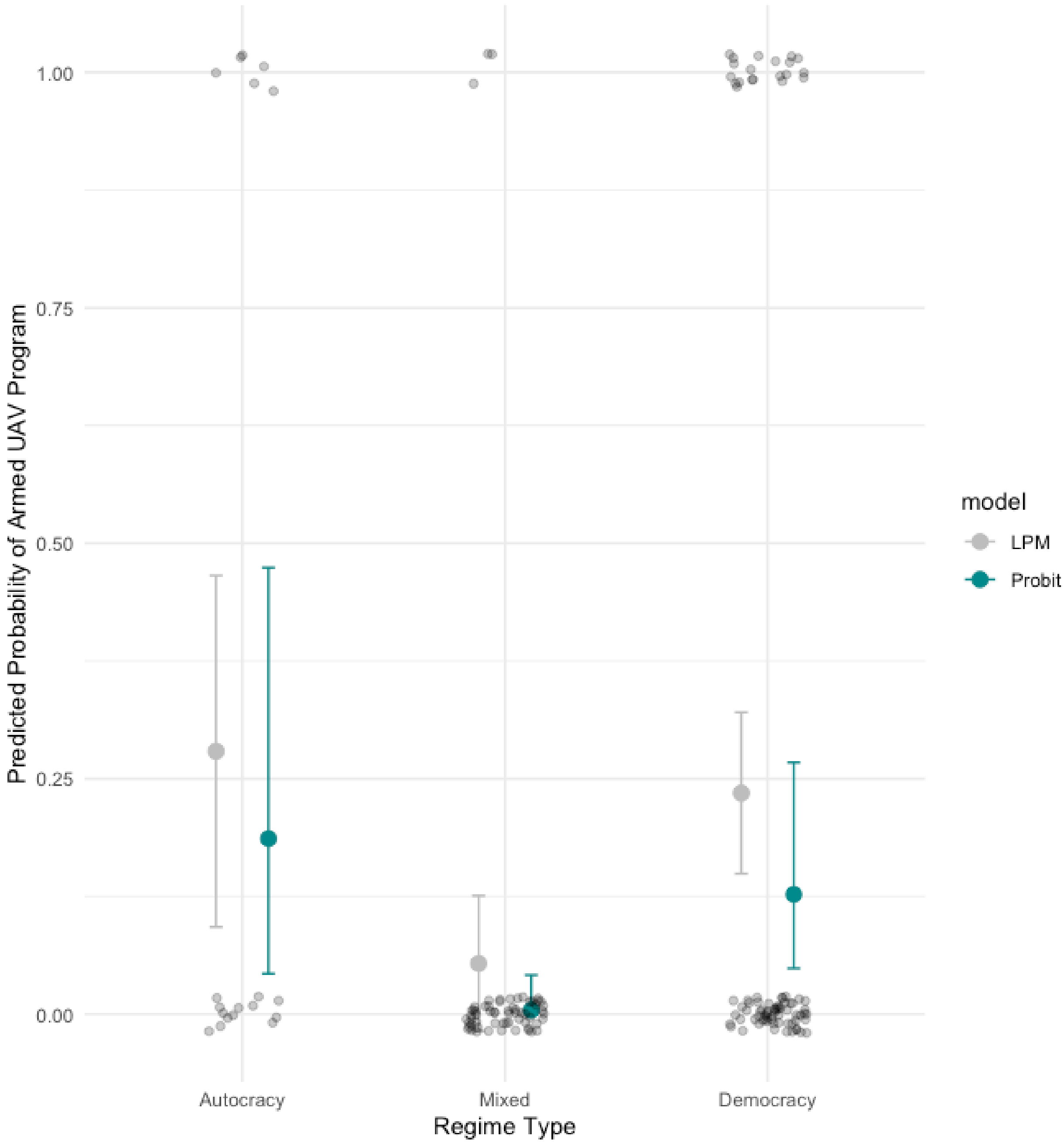
The following sections specifically examine how much the models' predictions diverge over specific variables of interest.

Model comparison for regime effects

To compare how the probit and linear probability model (LPM) specifications predict UAV programme adoption across political regime types, I generate predicted probabilities for three mutually exclusive regime categories: autocracy, democracy, and mixed regimes. The authors operate on the implicit (and intuitive) assumption that a regime cannot be both democratic and autocratic, but can be neither. I maintain the structure of the original model, which contains democracy and autocracy as separate binary predictors, but make the assumption explicit. I treat cases coded as 0 on both the democracy and autocracy variables as mixed regimes.

I construct a prediction grid that sets the regime variables to the three possible combinations, while holding all control variables at their sample means. I thus obtain model-based predicted probabilities that are directly comparable across regime types. Plotting the predicted probabilities (with clustered standard error intervals, as in Fuhrmann and Horowitz) for each regime category allows a clear visual comparison of how the probit and LPM differ in their substantive implications for regime effects. Figure 2 shows the results. For each regime type, the plot shows the model's point estimate along with a 95% confidence interval using clustered standard errors.

Figure 2: Predicted Probabilities by Regime Type



Across all three regime types, the probit model yields lower predicted probabilities than the LPM. However, the probit model also exhibits wider confidence intervals for autocracies and democracies, reflecting greater uncertainty in prediction at the extremes of the latent index where nonlinear models are more sensitive.

In contrast, the LPM produces slightly narrower intervals and more tightly clustered predictions, consistent with the model's linear structure, which tends to compress variation and underrepresent uncertainty near the boundaries of the probability scale. Despite these differences in magnitude and uncertainty, both models preserve the same ordering of regime types—democracies and autocracies display higher predicted probabilities than mixed regimes, indicating consistent substantive conclusions.

To assess whether these differences are statistically significant, I bootstrap the difference in their predicted probabilities for each regime category (autocracy, democracy, mixed). For each bootstrap resample, I re-estimate both models and compute the difference in the predicted probability of armed UAV adoption for each regime type, holding other covariates at their means. The resulting bootstrap intervals (Table 3) show that, while the probit and LPM agree on the ordering of regime types, the probit model predicts systematically lower probabilities than the LPM for democracies and autocracies. These differences are unlikely to be explained by sampling variation alone.

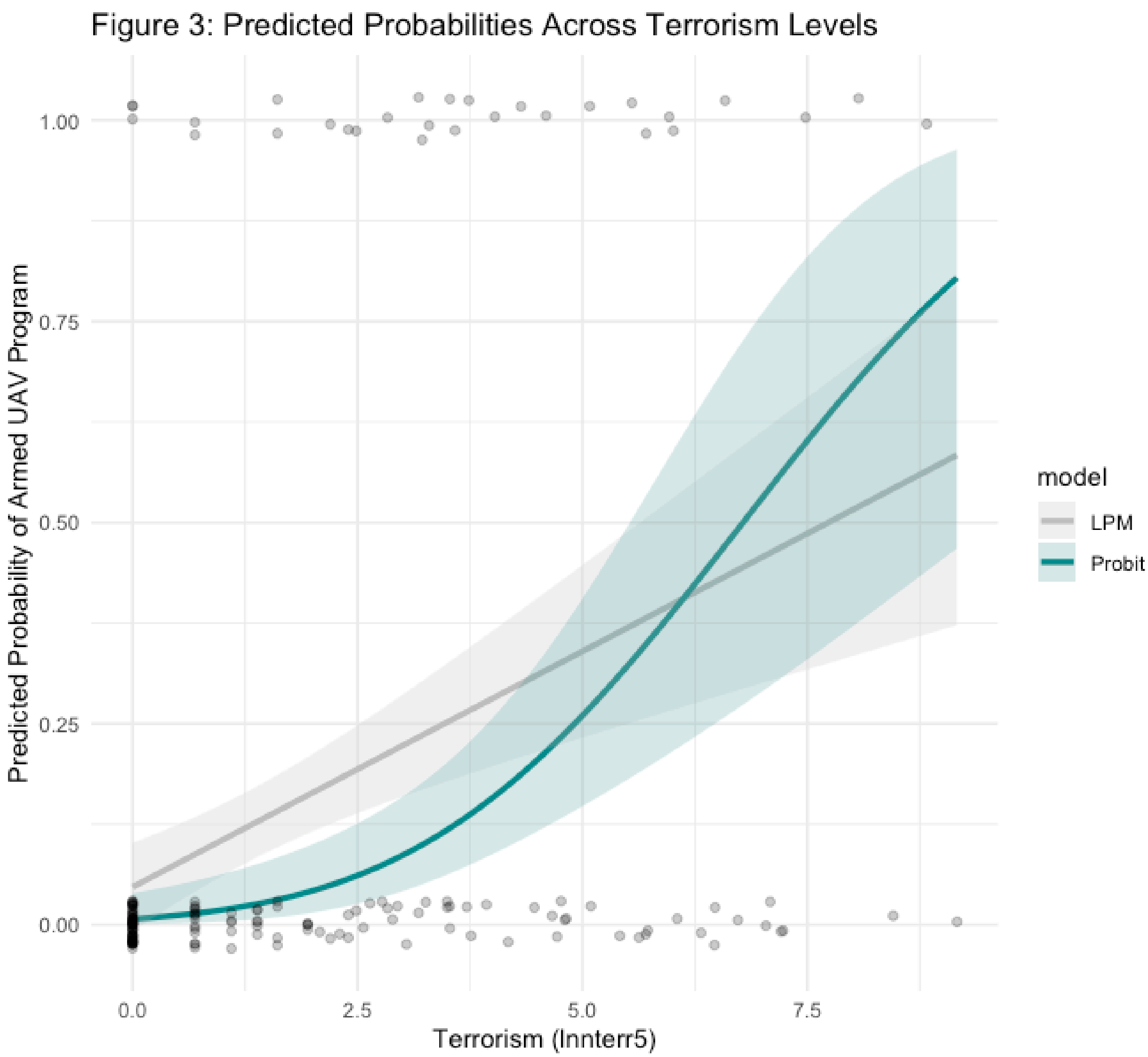
Table 3: Bootstrap Differences in Regime Predicted Probabilities

Regime	Low95	Median	High95
Autocracy	-0.2185359	-0.08916399	0.04461920
Democracy	-0.1908633	-0.11554035	-0.04942932
Mixed	-0.1198198	-0.05283337	0.01818662

For democracies, the bootstrap distribution lies entirely below zero, with a median difference of –0.12 and a 95% interval from –0.19 to –0.05, indicating that the LPM systematically overpredicts adoption relative to the probit model. For autocracies, the median difference is also negative, at –0.09, but the confidence interval includes zero (–0.22 to 0.04), suggesting that the two models' predictions are not statistically distinguishable. A similar pattern appears for mixed regimes, where the median difference is small (–0.05) and the bootstrap interval overlaps zero. Overall, the results indicate that while both models preserve the same ordering across regime types, the choice between a linear and a nonlinear specification has the greatest substantive impact for democracies, where the LPM appears to overstate the likelihood of UAV programme adoption.

Model comparison for terrorism

To evaluate whether the probit and linear probability model (LPM) differ in their substantive predictions for the terrorism variable, I bootstrap the difference in predicted probabilities between the two models across three "tiers" of terrorism: the 25th percentile, the median and the 75th percentile ("high terrorism"). For each of 1,000 bootstrap samples, I refit both models and compute the predicted probability of armed UAV adoption at each terrorism level, holding all other predictors at their sample means. For each terrorism value, I then calculate the difference between the probit and LPM predictions and extract percentile-based 95% confidence intervals.

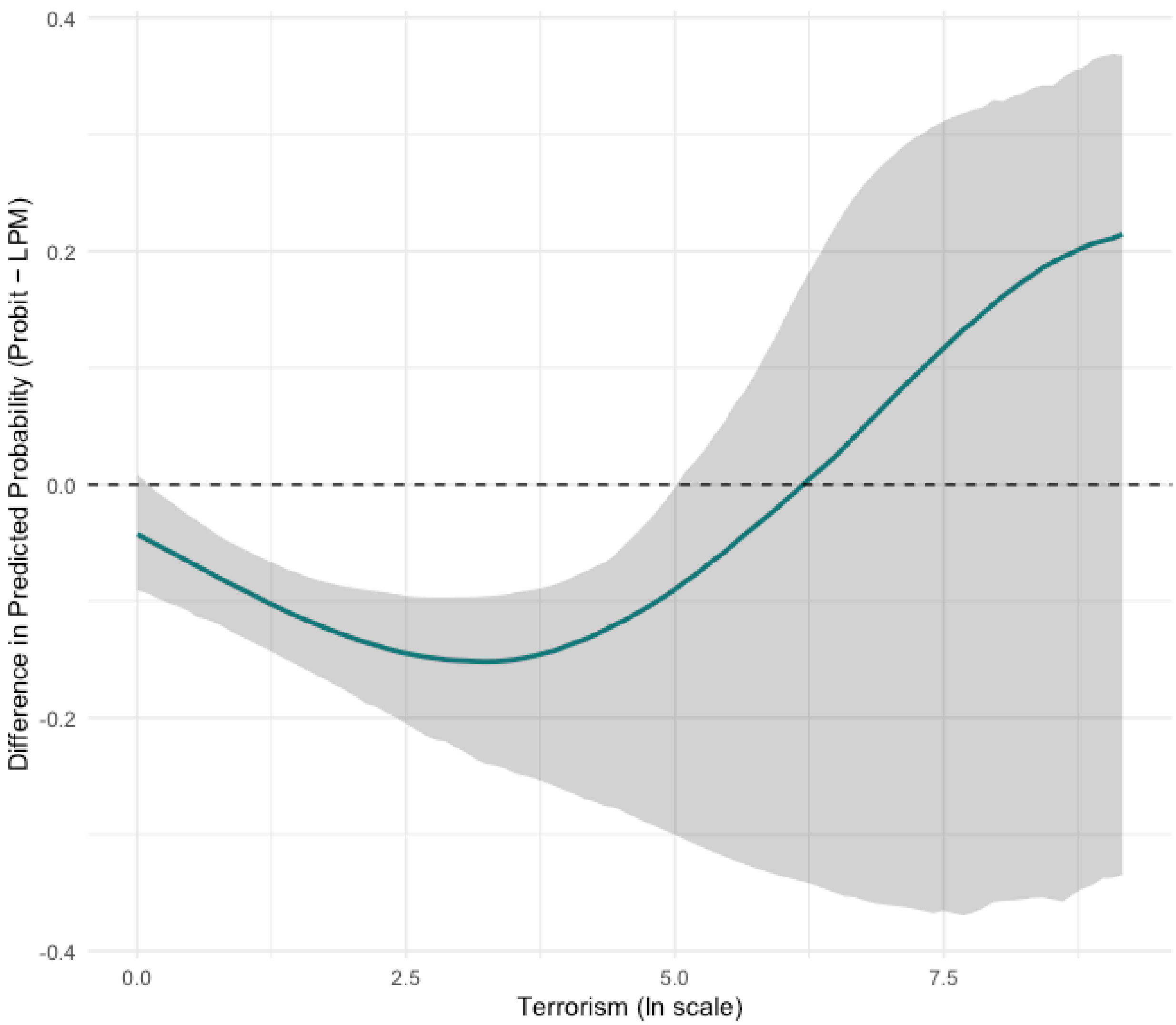


The results show that at low levels of terrorism the difference between the models is small even though the LPM systematically predicts higher probabilities than the , Probit, the latter being more sensitive to the high number of datapoints at the bottom left corner of the graph. At high terrorism levels, the probit model consistently predicts higher adoption probabilities than the LPM.

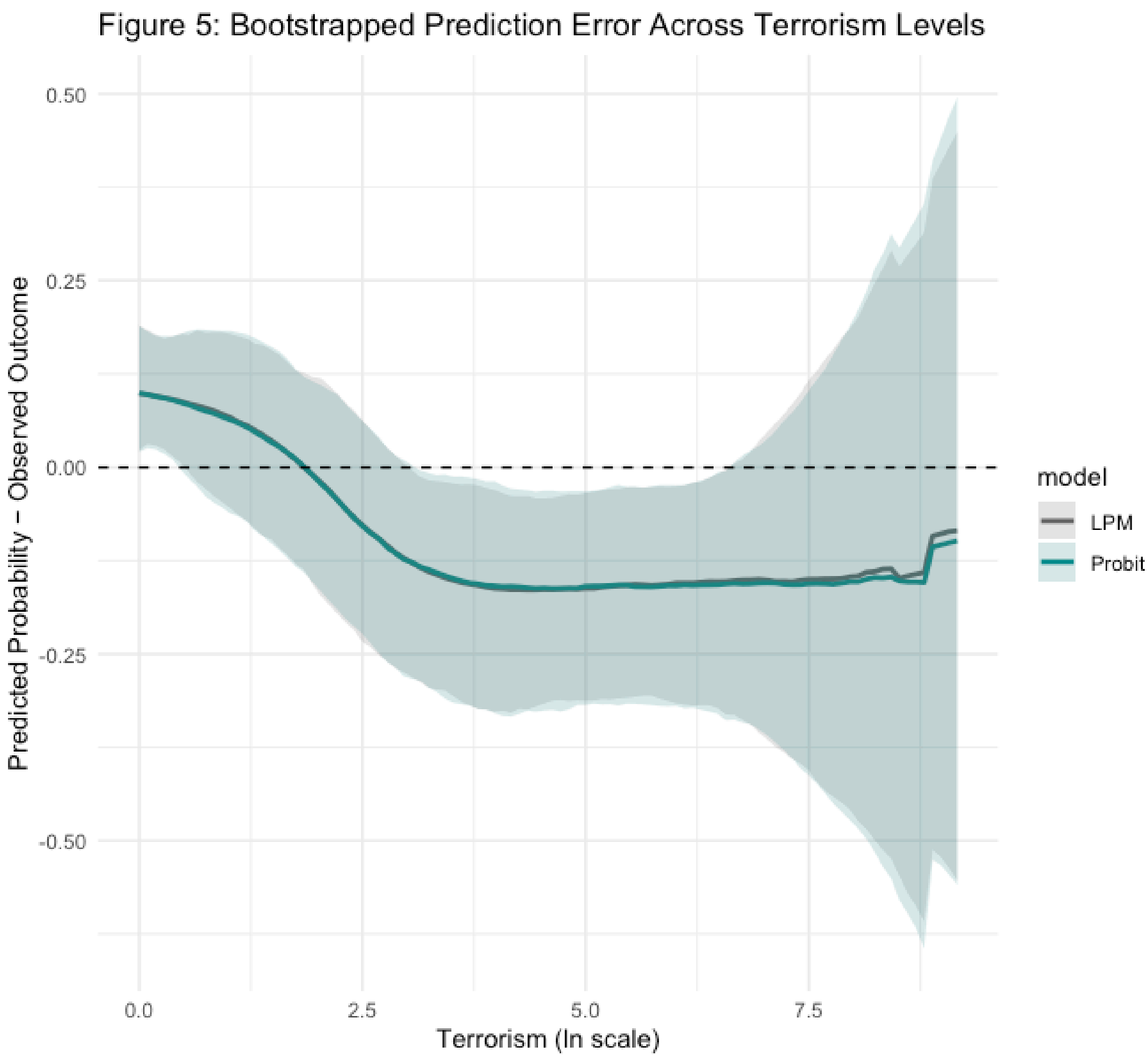
The difference between the predictions of the two models is illustrated in Figure 4, which shows the results of a 1000-trial bootstrap. The curve represents the 50th percentile of 1000 iterations, while the ribbon marks the 2.5th and 97.5th percentile. The difference at the

50th percentile ranges from around -0.18 for lower terrorism values to just over 0.2 for the highest terrorism values. The difference at lower values of terrorism is tightly clustered and consistently below 0 across the ribbon, which substantiates the claim that for lower values of terrorism, the LPM systematically overpredicts the probability of having an armed UAV programme compared to the Probit. However, at higher values of terrorism, the ribbon widens to a maximum range of 0.708 at 8.884 units of terrorism. This reflects both models' widening range of predictions at high levels of terrorism levels (see Figure 3). This is likely due to the scarcity of datapoints at high levels of terrorism, and the fact that they are divided in terms of the outcome variable.

Figure 4: Model Disagreement Across Terrorism Levels



To understand whether the two models vary with regards to their predictive ability over terrorism vis-a-vis the observed data, I conduct a bootstrap analysis of model prediction error as a function of terrorism, shown in Figure 5. In each of 1,000 bootstrap resamples, I refit the Probit and LPM models, compute prediction errors (predicted probability minus the observed outcome), and smooth the conditional mean error over the continuous range of terrorism values. The solid lines in Figure 5 represent the pointwise median (50th percentile) of the bootstrapped error curves for each model, while the ribbons show pointwise 95% bootstrap intervals (2.5th–97.5th percentiles; $= 0.05$). The near-complete overlap of both the median curves and the uncertainty bands indicates that the two models exhibit very similar patterns of systematic over- and under-prediction across terrorism levels, suggesting little evidence that one specification yields meaningfully different prediction error than the other in this dimension. Both models' uncertainty bands widen at high terrorism values, which likely reflects limited data support in the upper tail of the terrorism distribution, making fitted relationships more sensitive to resampling and increasing the variability of predicted probabilities and associated errors at extreme terrorism levels.



Conclusion

This replication study assessed whether the substantive conclusions of Fuhrmann and Horowitz's analysis of armed UAV programme adoption depend on the choice of functional form, and more specifically, whether they are significantly altered by the selection of a LPM as opposed to the Probit model used by the authors.

Both the probit and LPM specifications yield consistent qualitative conclusions. States facing higher levels of terrorism are more likely to develop armed UAV programmes, and both democracies and autocracies exhibit higher adoption probabilities than mixed regimes. The direction and statistical significance of key predictors are preserved across models, suggesting that the substantive interpretations of the original paper do not hinge on the use of a nonlinear estimator. In this sense, the LPM functions effectively as a robustness check rather than a competing explanation.

At the same time, meaningful differences emerge in the models' predictive behavior. Calibration plots show that the Probit model better captures nonlinearities near the boundaries of the probability scale, particularly at very low and very high predicted probabilities. The LPM, by contrast, compresses predictions toward the center, leading to systematic overprediction in some regions of the covariate space. These differences are most pronounced when examining regime effects for democracies and at higher levels of terrorism, where bootstrap analyses indicate that the Probit tends to overstate adoption probabilities relative to the LPM.

However, prediction error analyses reveal that both models exhibit similar patterns of systematic over- and underprediction across terrorism levels, with overlapping uncertainty bands throughout most of the distribution. This suggests that, while the Probit model offers advantages in capturing curvature and respecting the bounded nature of probabilities, the practical gains in predictive accuracy relative to the LPM are modest in this application.

Overall, this replication reinforces the original study's conclusions, while highlighting the trade-offs inherent in modeling binary outcomes. The probit model provides a better fit at the extremes, whereas the LPM offers simplicity and direct interpretability of coefficients as average marginal effects. Importantly, the choice between these specifications does not alter the central substantive claims regarding the political and security determinants of armed UAV proliferation.

References

Angrist, Joshua D. 2001. "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice." *Journal of Business*

& Economic Statistics 19 (1): 2–28.

Austin, Peter C, and Ewout W Steyerberg. 2014. "Graphical Assessment of Internal and External Calibration of Logistic Regression Models by Using Loess Smoothers." *Statistics in Medicine* 33 (3): 517–35.

Chatla, Suneel, and Galit Shmueli. 2016. "Linear Probability Models (LPM) and Big Data: The Good, the Bad, and the Ugly." *Indian School of Business Research Paper Series*.

Fuhrmann, Matthew, and Michael C Horowitz. 2017. "Droning on: Explaining the Proliferation of Unmanned Aerial Vehicles." *International Organization* 71 (2): 397–418.

Hellevik, Ottar. 2009. "Linear Versus Logistic Regression When the Dependent Variable Is a Dichotomy." *Quality & Quantity* 43 (1): 59–74.