APPENDIX D
CVI PREDICTOR: EXCLUSION OF 2005 DATA POINT

LIST OF FIGURES

CVI Predictor: Exclusion of 2005 Data Point

The 2005 data point for the CVI predictor (see Figure II-1) was identified as an "outlier". Its *x*-value is nearly 50% greater than the next highest *x*-value, while its *y*-value is near the mid-range of the rest of the *y*-values. This suggests that the 2005 data point may have considerable influence on the linear predictor, and if so, would significantly influence the predictor not only at the high end of *x*, but at the low end of *x* as well. This is a particular concern given the situation in 2008, where the jack return (*x*-value) being used to predict the CVI is the lowest on record (more than 50% lower than the previous low) and near the origin.

Figure D-I displays several regression diagnostics for the predictor including the 2005 data point. By all of these measures it is clear that the 2005 data point is not only an outlier, but has excessive leverage on the resulting predictor. Figure D-II displays the same diagnostics for the predictor excluding the 2005 data point. These measures are more consistent with a linear predictor model, although it is suggested that the variance of *y* increases with *x* and that a weighted regression model may be a more efficient alternative.

Based on this information, the STT has decided to exclude the 2005 data point from the 2008 CVI predictor, but not to alter the prediction methodology.

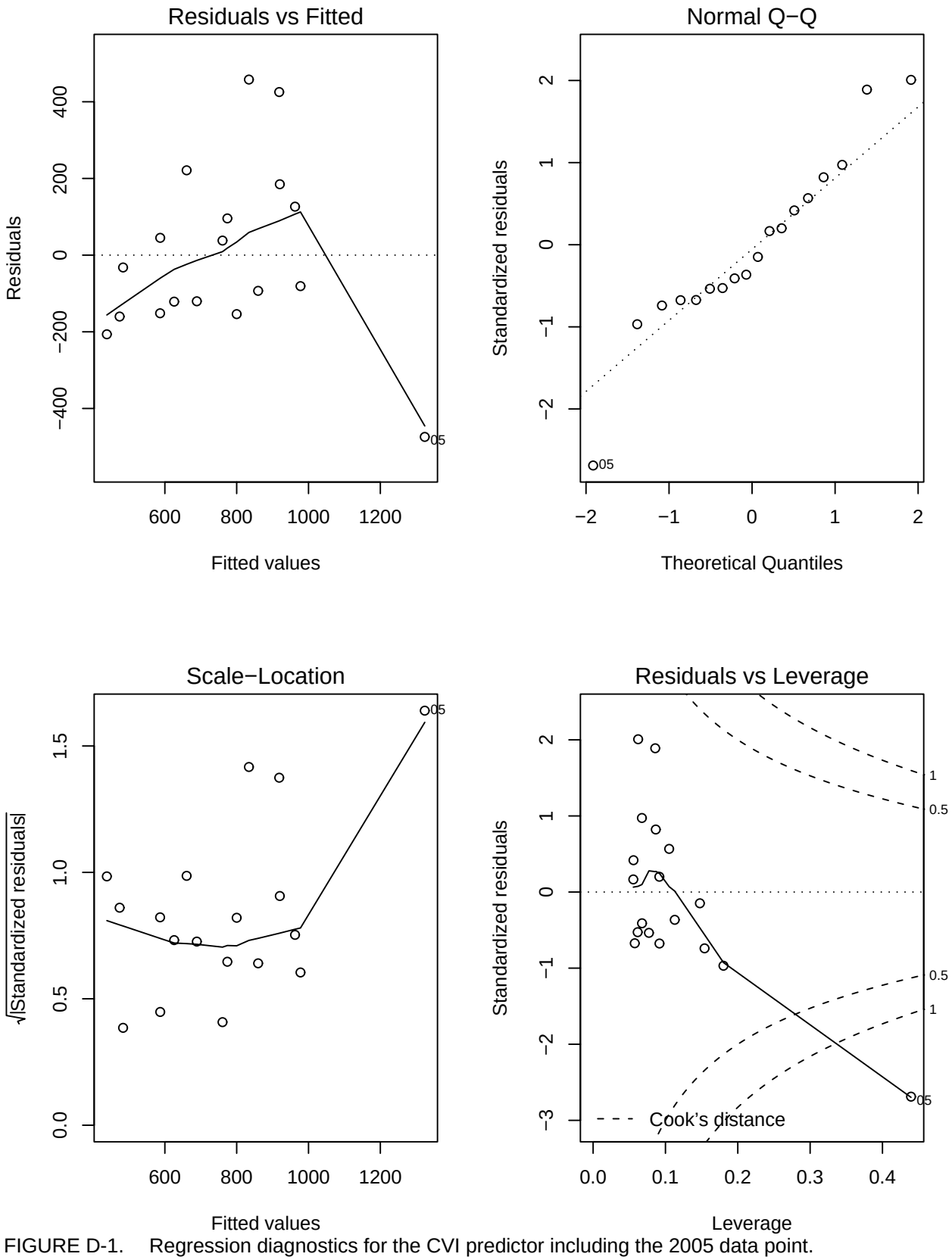

FIGURE D-1. Regression diagnostics for the CVI predictor including the 2005 data point.

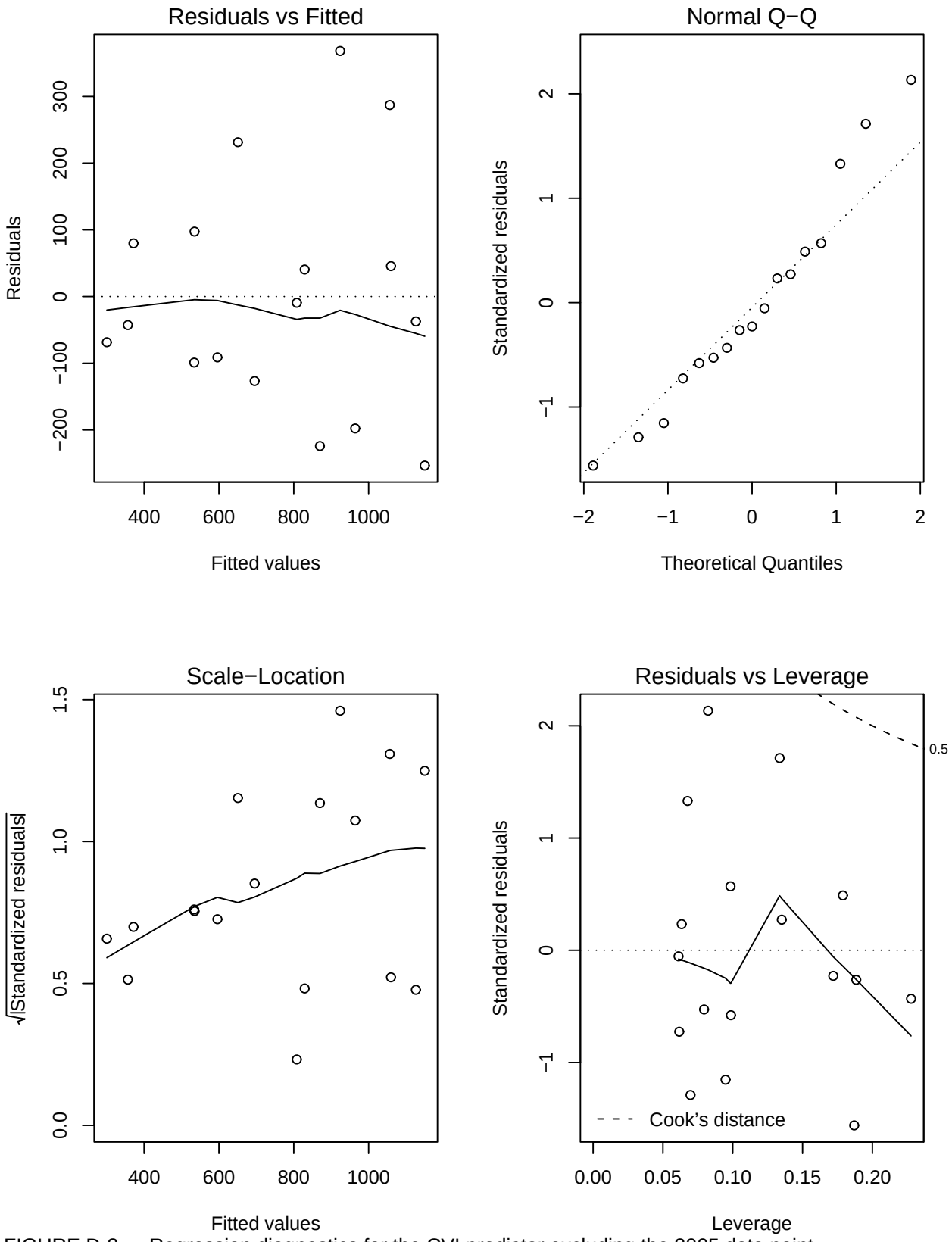

FIGURE D-2. Regression diagnostics for the CVI predictor excluding the 2005 data point.