Causal Inference for Observational Time-Series with Encoder-Decoder Networks (Online Appendix)

January 4, 2018

Contents

1 Encoder-decoder architecture and training

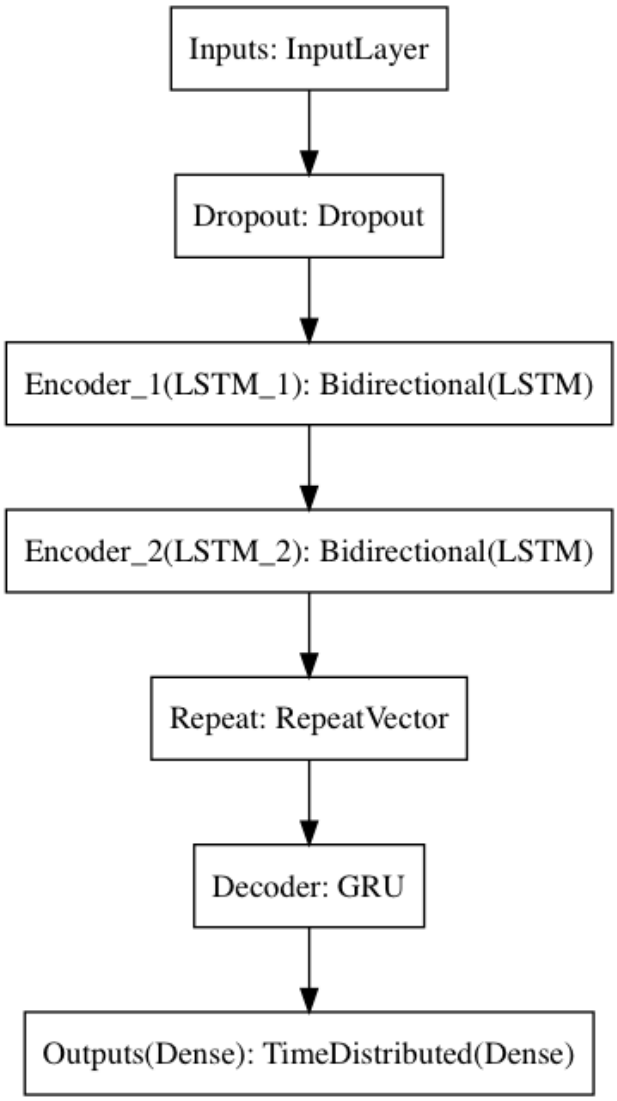

Figure 1: Encoder-decoder networks architecture. Dropout is applied to the visible input sequences, which are then fed to a two-layer bidirectional LSTM encoder. The encoder encodes the input sequences into a single vector that contains information about the entire sequence. The output of the encoder is repeated t' times and fed to the single-layer GRU decoder, which translates the encoded sequence into the predicted sequence. Finally, a dense layer is applied to the decoder output to generate predictions.

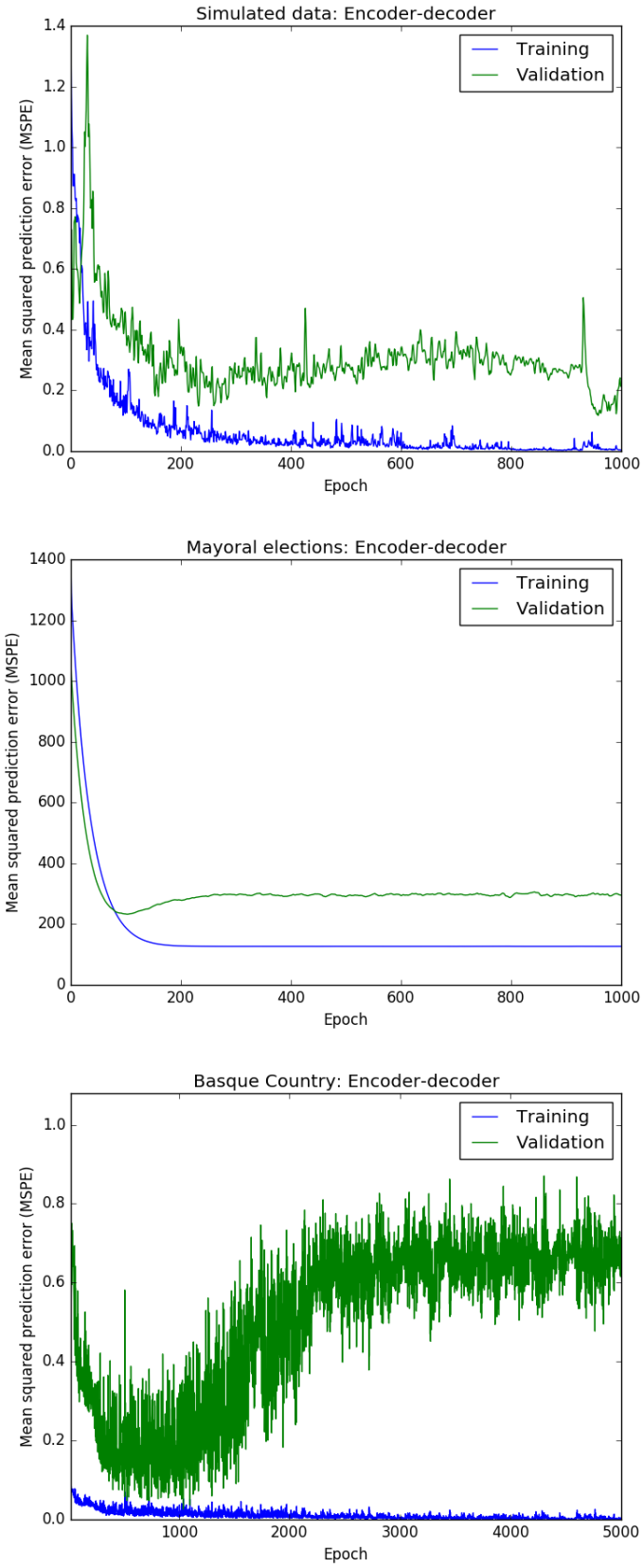

Figure 2: Evolution of encoder-decoder networks training and validation loss in terms of MSPE.

2 Estimates on simulated data

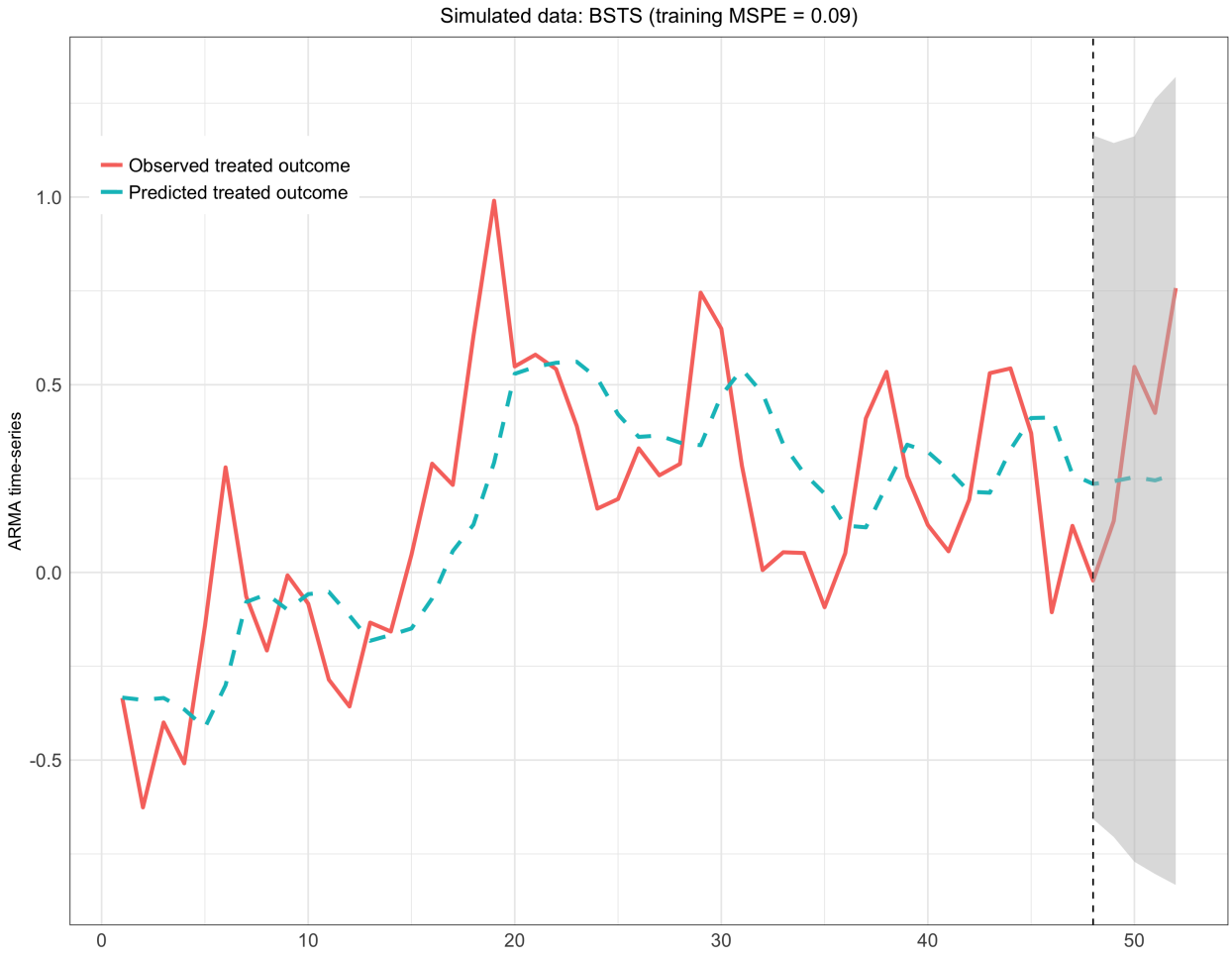

Figure 3: Observed ARMA simulated time-series and BSTS predictions. BSTS model with semilocal linear trend and seasonal components and spike-and-slab priors is trained with 10,000 MCMC samples with the first 1,000 samples discarded as burn-in. The reported training loss is the mean squared prediction error (MSPE) over time-steps 43 to 47. Predictions are obtained by averaging across MCMC draws and 95% credible intervals (shaded region) are obtained from the distribution of MCMC draws. The dashed vertical line indicates the start of the intervention.

Figure 4: Observed ARMA simulated time-series and synthetic control predictions. Synthetic control is constructed using input means over various periods during the pre-period as covariates. The reported training loss is the mean squared prediction error (MSPE) over time-steps 43 to 47. The shaded region represents 95% confidence intervals formed using the standard deviation of 1,000 block bootstrap replicates of the predicted time-series with optimal block lengths chosen following the procedure of Politis and White (2004). The dashed vertical line indicates the start of the intervention.

3 Estimates on Basque Country data

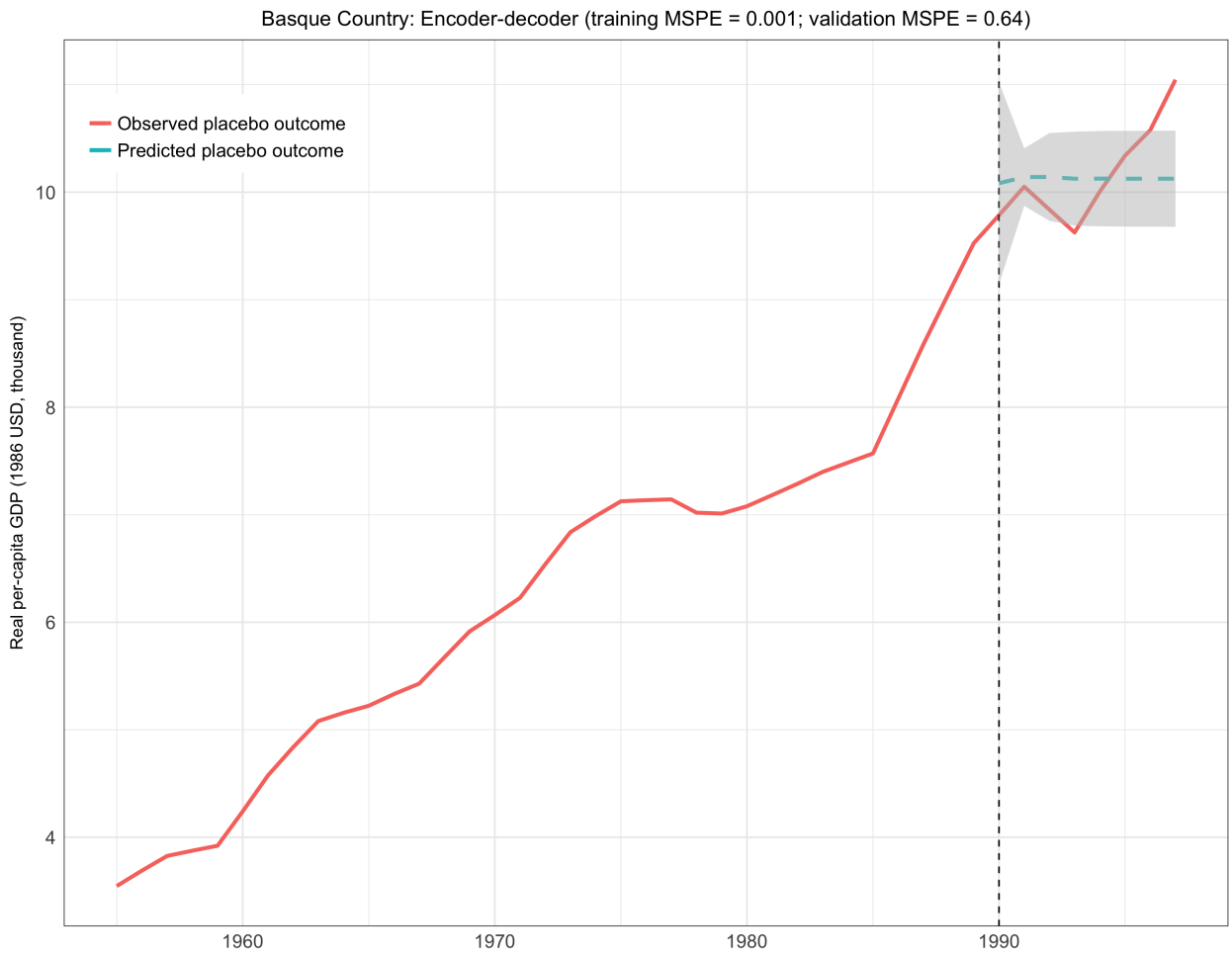

Figure 5: Observed placebo and encoder-decoder predicted outcomes in Basque Country data. Validation MSPE is calculated on a held-out sample. Shaded regions represent 95% prediction intervals obtained using the standard deviation of the prediction distribution (Eq. 7). The dashed vertical line indicates the start of the placebo intervention (1990).

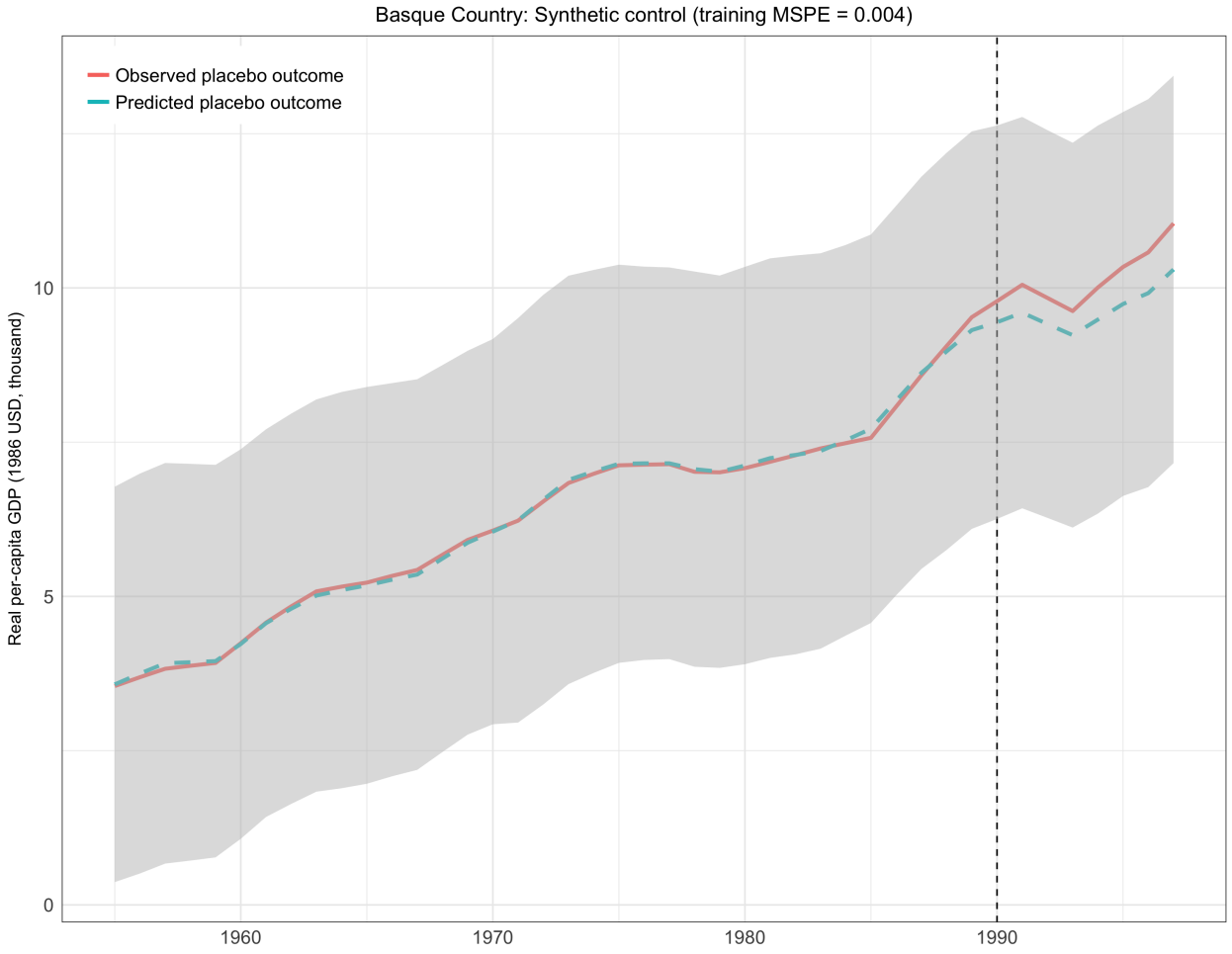

Figure 6: Observed placebo and synthetic control predicted outcomes in Basque Country data. Synthetic control is constructed using pretreatment covariates including population density and levels of investment and human capital. The reported training loss is the mean squared prediction error (MSPE) over the period of 1960 to 1989. The dashed vertical line indicates the start of the placebo intervention (1990). See notes to Fig. OA-4.

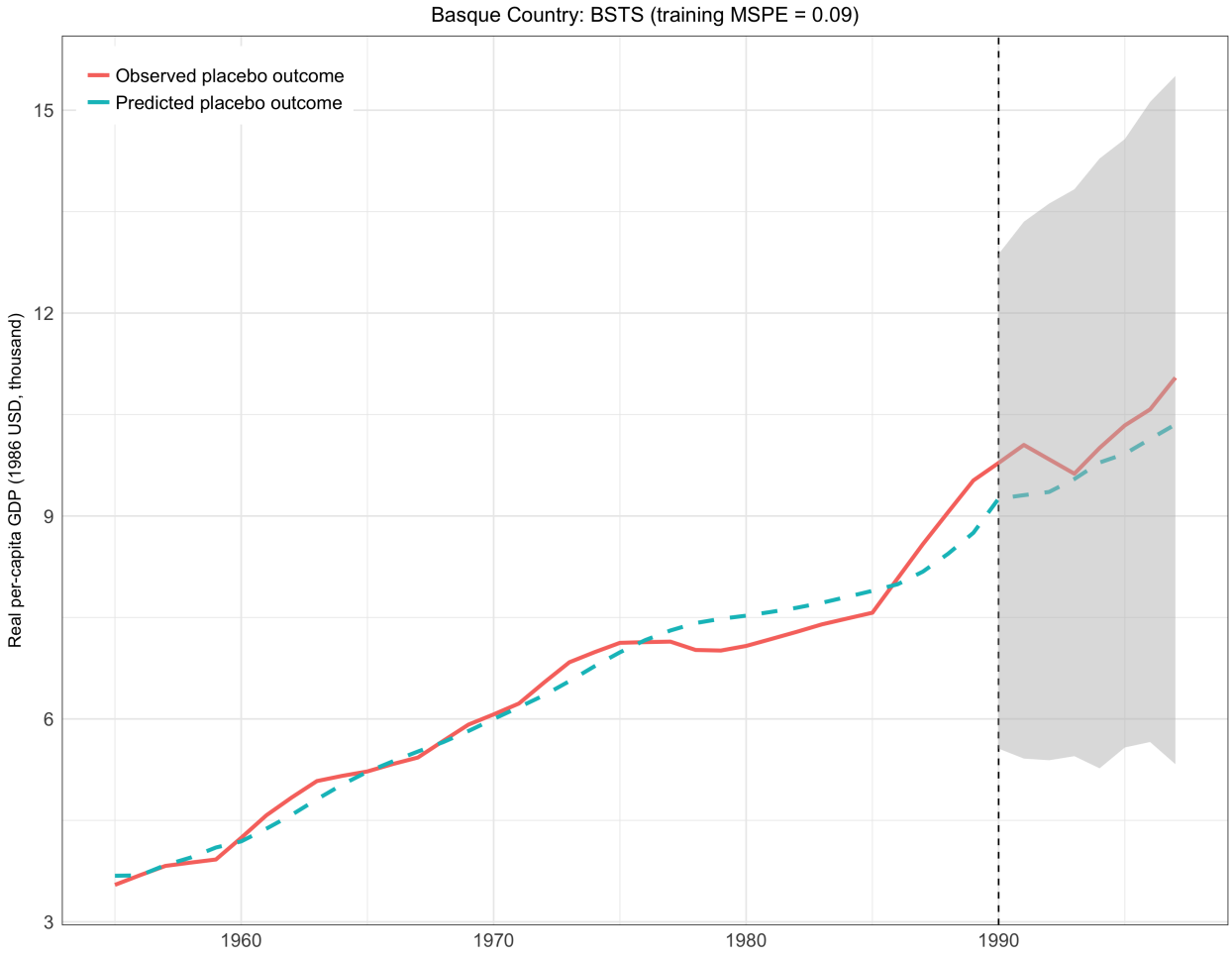

Figure 7: Observed placebo and BSTS predicted outcomes in Basque Country data. The reported training loss is the mean squared prediction error (MSPE) over the period of 1960 to 1989. The dashed vertical line indicates the start of the placebo intervention (1990). See notes to Fig. OA-3.

4 Estimates on mayoral elections data

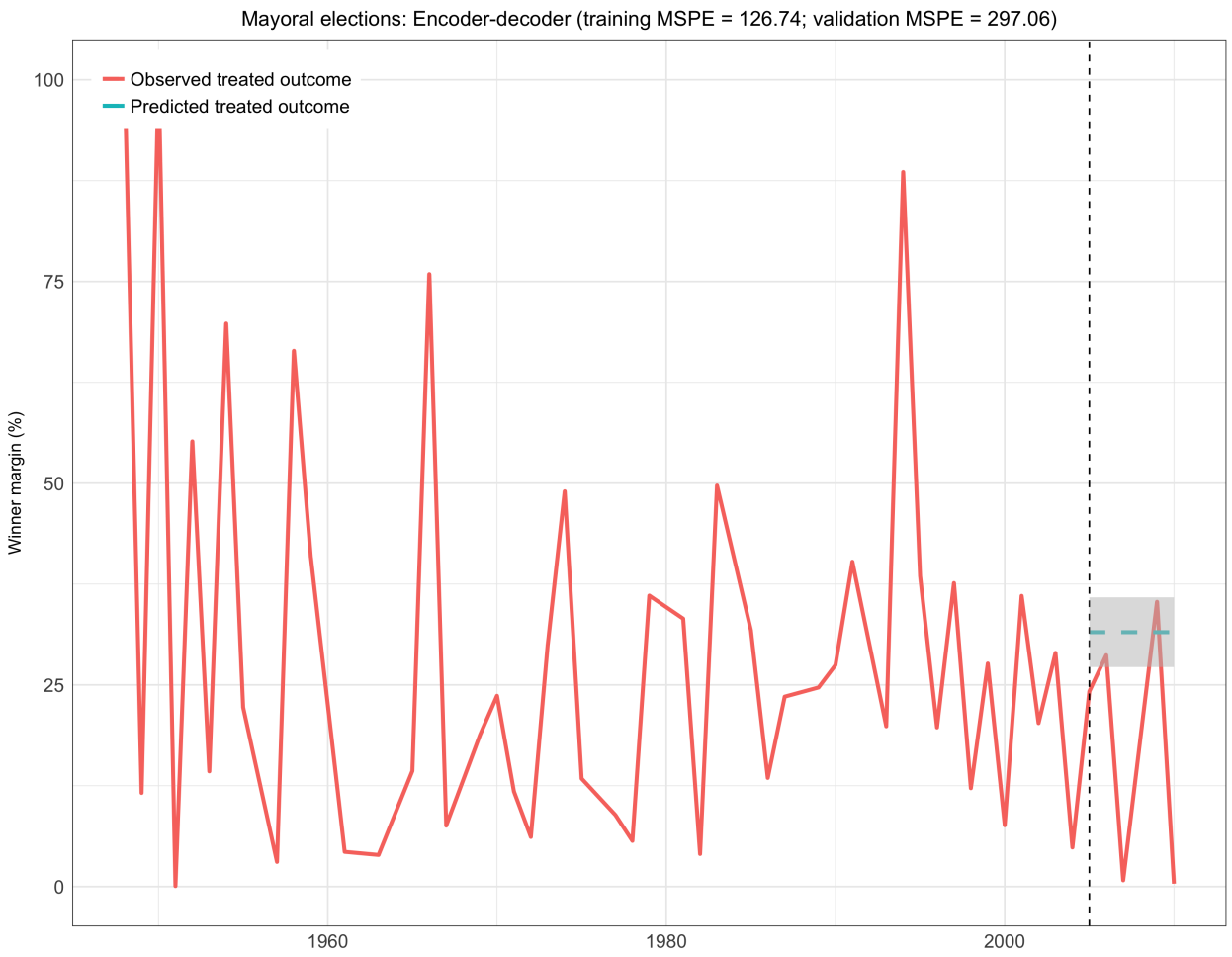

Figure 8: Observed and encoder-decoder predicted winner margins in Panagopoulos and Green treated cities. Validation MSPE is calculated on a held-out sample. Shaded regions represent 95% prediction intervals following the standard deviation of the prediction distribution of an ensemble of different checkpoints of a single model over time. The dashed vertical line is the start of the test set (2005) and the dotted vertical line represents the 2006 mayoral election.

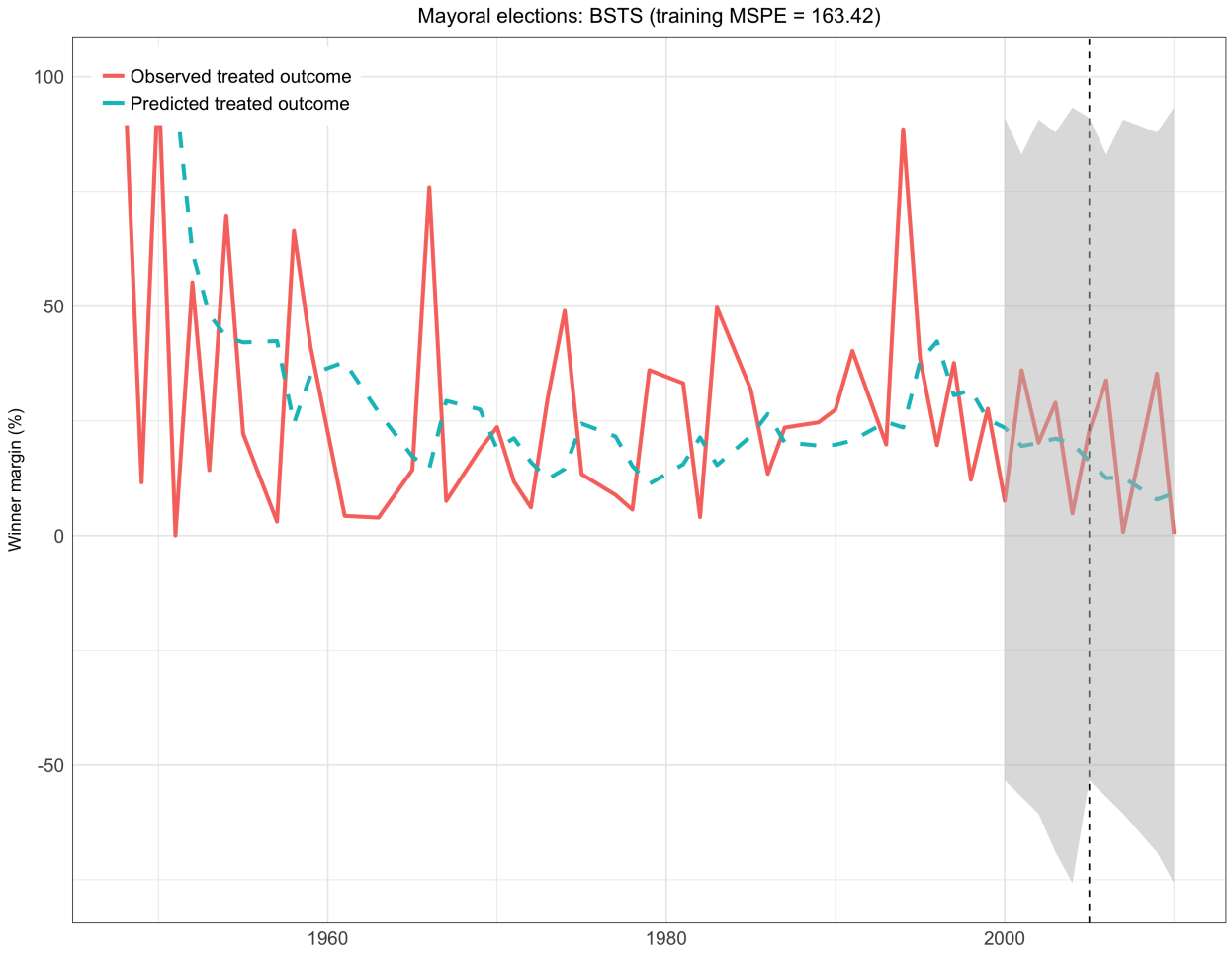

Figure 9: Observed and BSTS predicted winner margins in Panagopoulos and Green treated cities. The model predicts the time-series of the mean winner margin in Panagopoulos and Green treated cities using only winner margins in non-treated cities as predictors. See notes to Fig. OA-3.

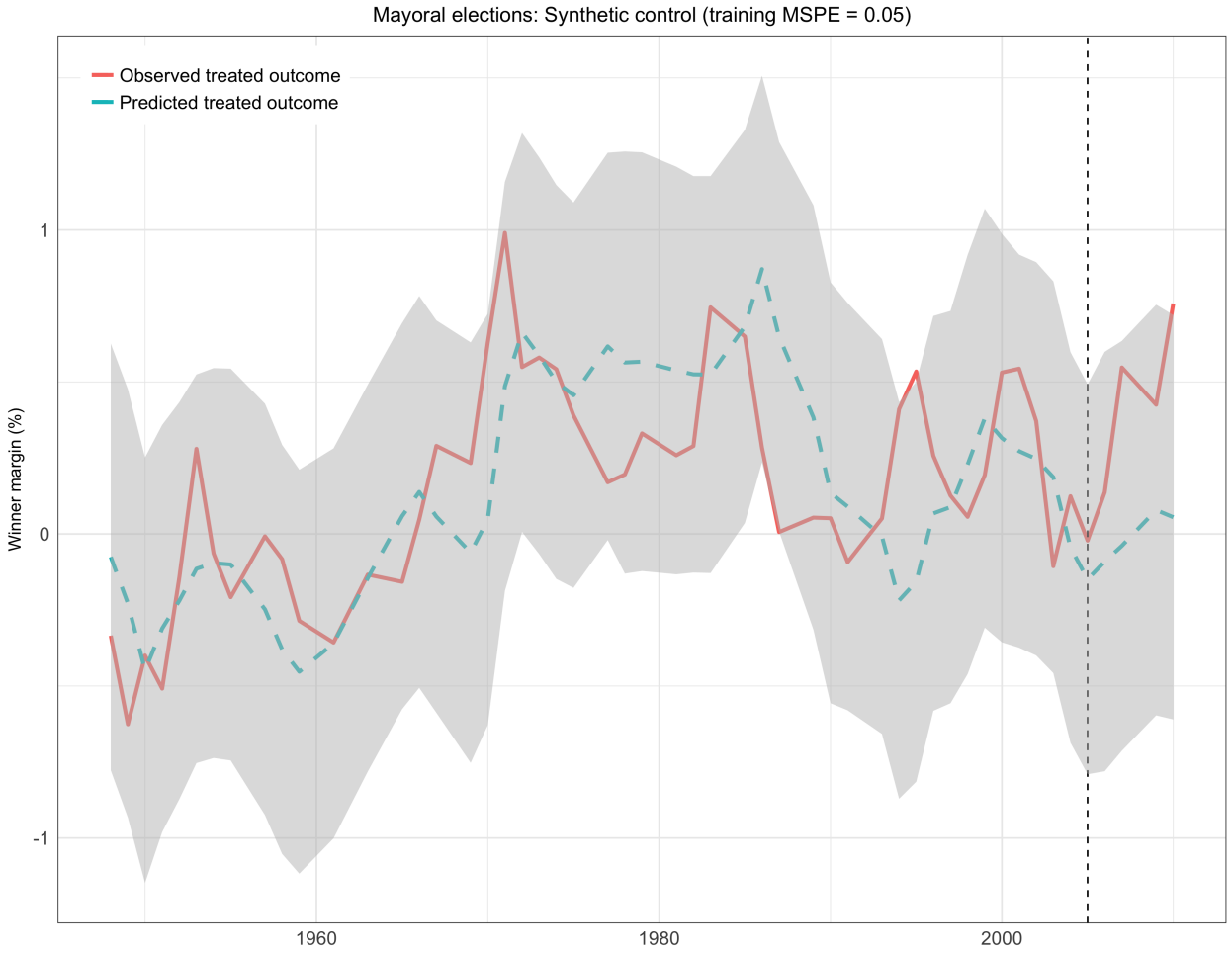

Figure 10: Observed and synthetic control predicted winner margins in Panagopoulos and Green treated cities. Synthetic control is constructed using pretreatment covariates including log vote total and winner margin means over the pre-period. The reported training loss is the mean squared prediction error (MSPE) over the period of 2000 to 2004. See notes to Fig. OA-4.

References

Politis, Dimitris N, and Halbert White. 2004. "Automatic Block-Length Selection for the Dependent Bootstrap." *Econometric Reviews* 23 (1): 53–70.