**Appendix A:**

In this Appendix, we illustrate the impact of structural break on forecasting accuracy using a simulation example. We construct a price variable with its values being 2.99 for most of the observations (say, weeks) but occasionally reduced to 2.29 or 1.99[[1]](#footnote-1). We assume the following unobserved true product sales[[2]](#footnote-2):

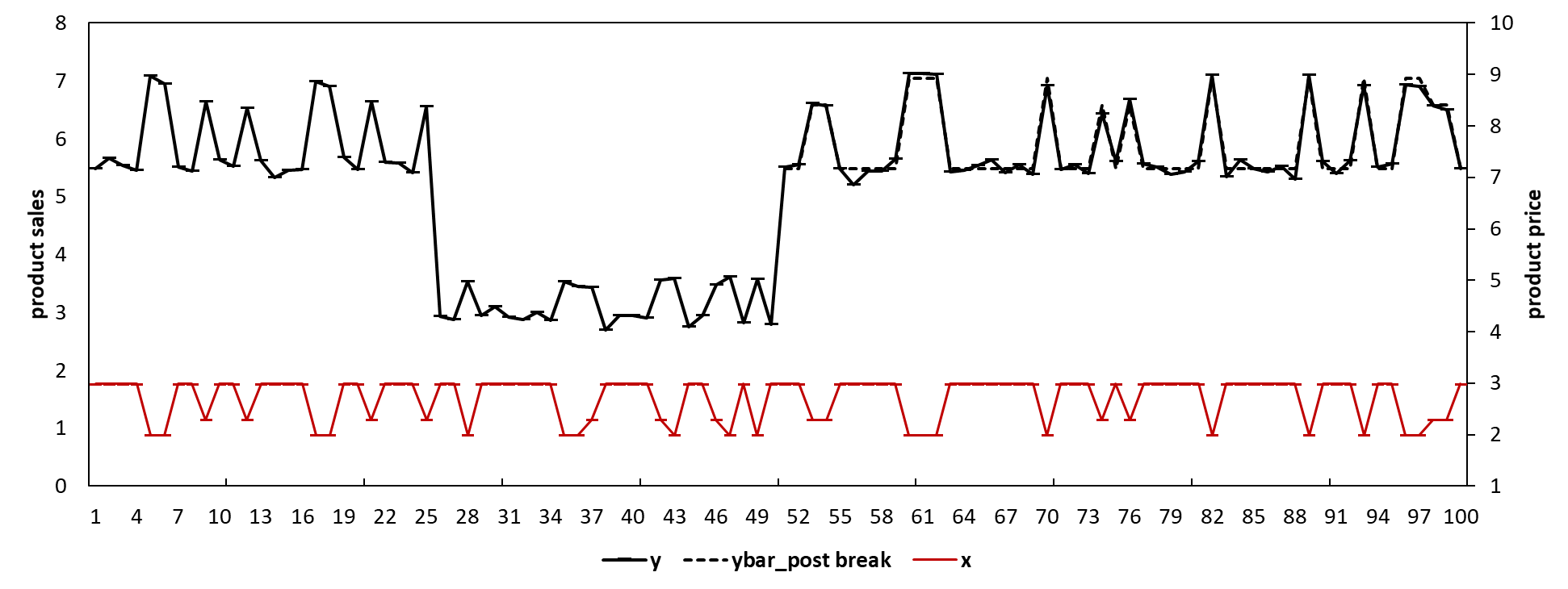
, , when

, , when

, , when

where and represent the product sales and the price at week *t*, and is the error term. There are two structural breaks for the model: the intercept and the parameter of the price changes at week 25 and then at week 50. In practice, this may be due to a new product introduction (which reduces the price elasticity of the focal product) or a credit crunch (e.g., customers become more price sensitive). The sales and price are represented in Figure A1 by the solid black line and the solid red line respectively.

Figure A1. Simulated sales with a structural break: model with post-break data[[3]](#footnote-3)



Suppose that we have the data from week 1 to week 75 and we forecast the product sales for the period from week 76 to week 100. If we know that there are changes for the effect of the product price, we may develop a congruent model (i.e., ) exclusively based on the post-break data (i.e., data from week 51 to week 75) and generate unbiased forecasts. Figure A1 represents the predictions/forecasts using the black dashed line (i.e., *ybar\_post breaks*). Table A1 shows the forecasting performance of the model with post break data (e.g., with MAE= 0.09, MSE= 0.01, MAPE= 1.5%, and SMAPE= 1.5%).

However, the changes of the effect of the price are usually unknown. If we overlook the two structural breaks and estimate the model using all the available data (i.e., from week 1 to week 75), we would obtain estimates of the parameters as the weighted average of the true parameters before and after the breaks and generate biased forecasts. In this example, we tend to under-predict the sales from week 1 to week 25, over-predict the sales from week 26 to week 50, then again under-predict the sales from week 51 to week 70 and finally generate downwards-biased out-of-sample forecasts from week 76 to week 100. Figure A2 shows the biased predictions/forecasts with the black dashed line (i.e., *ybar\_1*). Table 1 shows the forecasting performance of the model with the full data (e.g., with MAE= 0. 949, MSE= 0. 961, MAPE= 15.8%, and SMAPE= 17.2%). The forecasts are substantially inferior compared to the model with post-break data.

Figure A2. Simulated sales with a structural break: model with full data

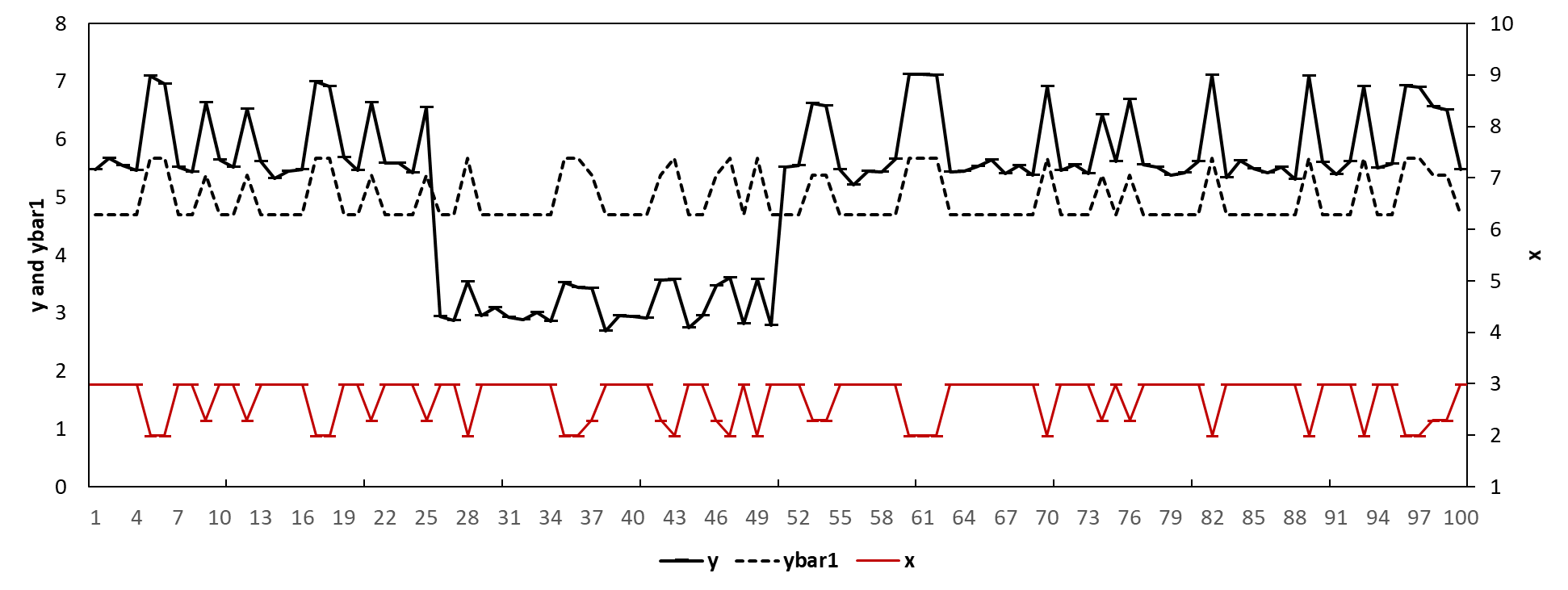


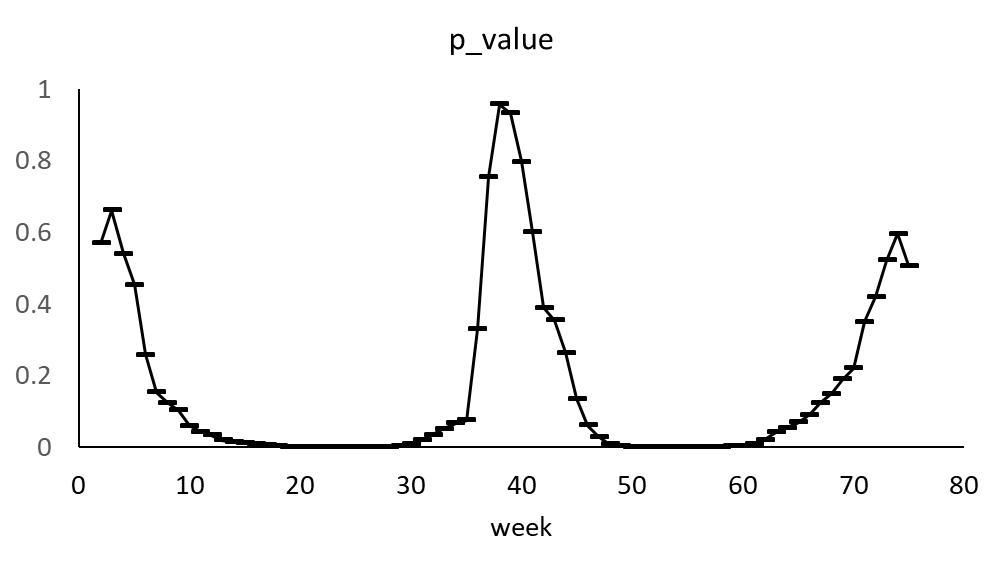
Table A1. The forecasting performance of different models in the simulation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | MAE | MSE | MAPE | SMAPE |
| Model with full estimation window | 0.95 | 0.96 | 15.8% | 17.2% |
| Model with post-break estimation window | 0.09 | 0.01 | 1.5% | 1.5% |
| Model with intercept correction | 0.20 | 0.07 | 3.2% | 3.2% |
| Model with estimation window combining | 0.86 | 0.80 | 14.2% | 15.4% |

**Appendix B:**

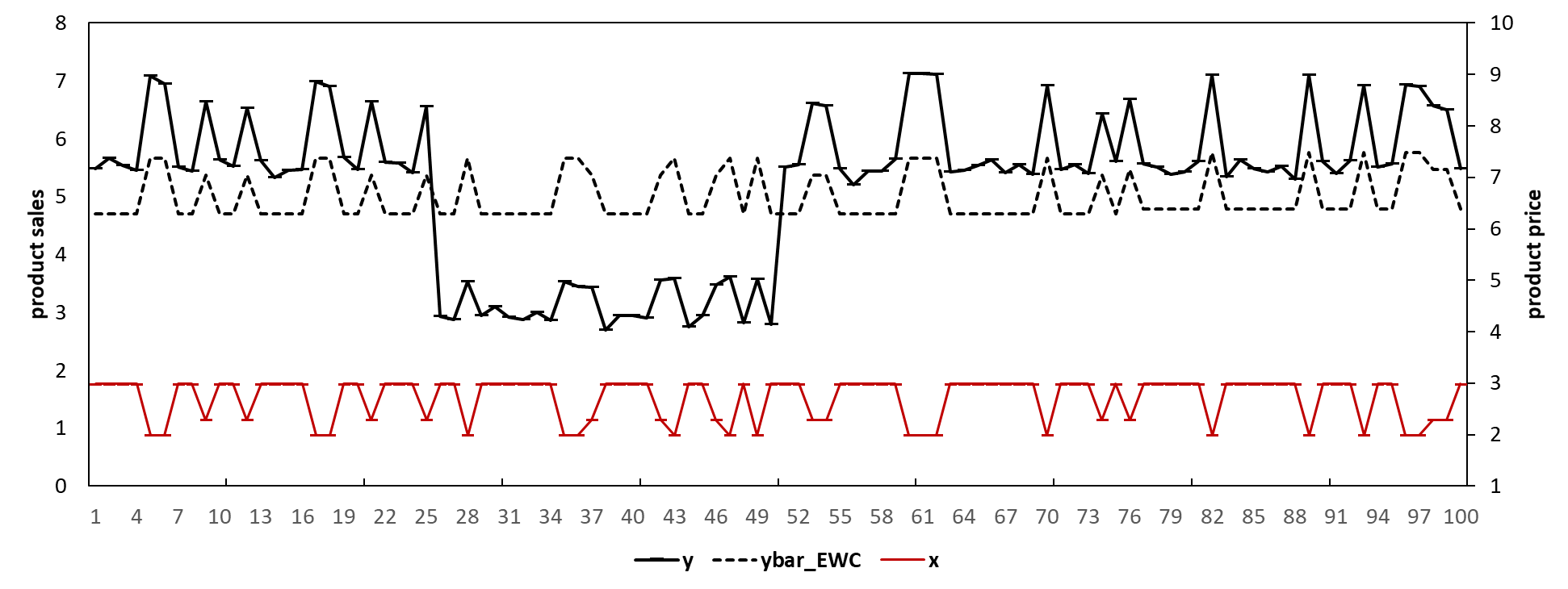
In this appendix, we illustrate how we may improve the accuracy of the forecasts using the estimation window combining (EWC) and the intercept correction (IC) method given there are structural breaks. We first conduct the test for structural break as we do not presume any priori knowledge of the structural break. We construct a congruent model as and we conduct a sequential [Chow (1960)](#_ENREF_16) test based on every observation during the estimation period[[4]](#footnote-4). The rejection of the null hypothesis of no structural break for any of the observations would suggest that the model is subject to structural break though without indicating how many structural breaks and their locations. Figure B1 plots the *p*-values of the sequential Chow test for the model in Appendix A. The results reject the null hypothesis of no structural break (especially for weeks closed to week 25 and week 50 where the p-values are close to zero) [[5]](#footnote-5). More advanced tests (e.g., considering multiple breaks, heteroskedasticity, and unit roots etc.) are available but require additional priori knowledge or assumptions such as the number of potential structural breaks ([Andrews, 1993](#_ENREF_4); [Andrews & Ploberger, 1994](#_ENREF_5); [Bai & Perron, 1998](#_ENREF_9), [2003](#_ENREF_10)).

Figure B1 *P*-values of the sequential Chow test



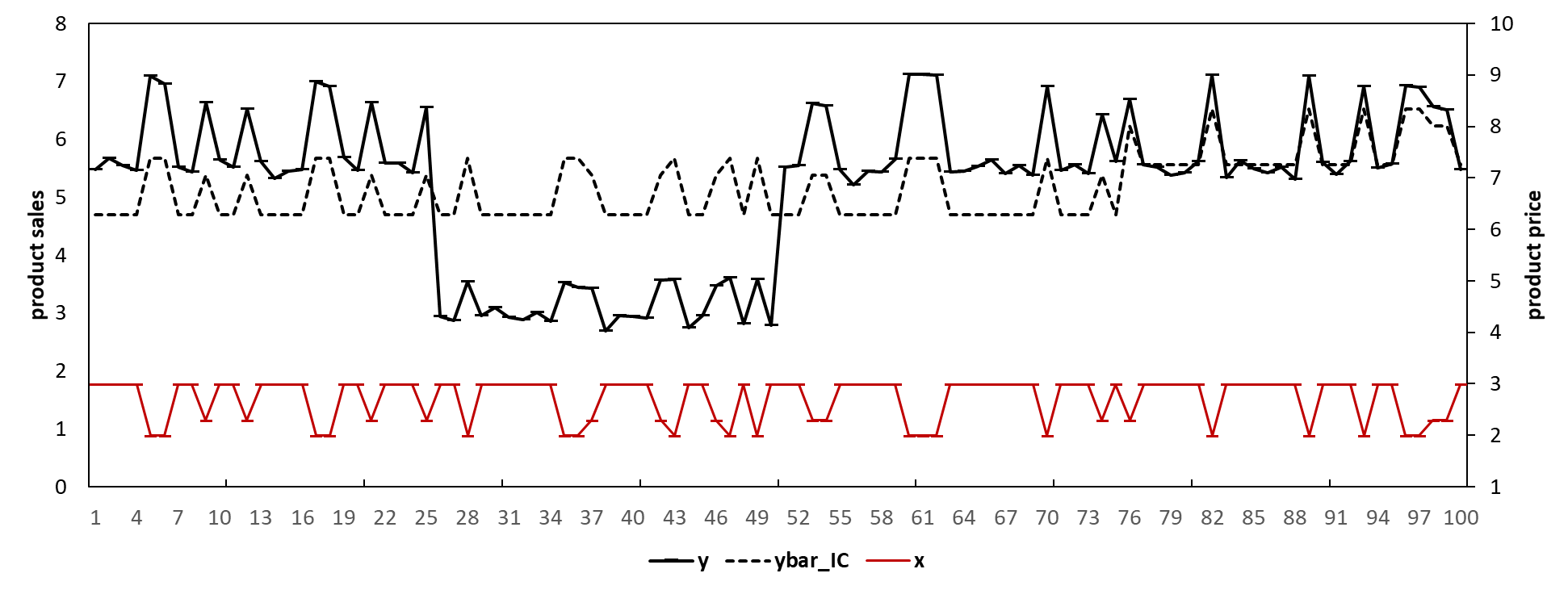
Therefore we confirm that the model is subject to structural break and we consider the forecasts as biased. We may improve the accuracy of the forecasts using the estimation window combining (EWC) method by combining the forecasts by the same model but with different estimation windows. Specifically, we may estimate the model using the data from week 1 to week 75, and generate the forecasts which are subject to the full bias (referred as ). We may then estimate the model with one less observation (e.g., from week 2 to week 75) and generate a second set of forecasts (referred as ), and so forth. The forecasts including are less biased compared to but associated with inflated forecasting error variance because they were generated by models with less pre-break information. We may arbitrarily choose to be 16, which gives us 60 sets of forecasts. Thus we calculate the final forecasts as the average of these 60 sets of forecasts. e.g.,. where are the final forecasts by the EWC model. Figure B2 represents the predictions/forecasts with the black dashed line (as *ybar\_EWC*). Table A1 shows the forecasting performance of the model with the full data (e.g., 0. 86for MAE, 0. 80 for MSE, 14.2% for MAPE, and 15.4% for SMAPE). The EWC method outperforms the conventional model with the full data.

Figure B2. Simulated sales with a structural break: model with EWC



We can also improve the accuracy of the forecasts using the intercept correction (IC) method. We may estimate the forecast bias as the average value of an ad hoc number (e.g., we choose four in this example) of the errors close to the forecast origin. e.g., where is the estimated forecast bias. We can obtain the final corrected forecasts by add the estimated bias back to the forecasts by the original model, e.g., , where are the final forecasts by the IC model. Figure B3 shows the predictions/forecasts with the black dashed line (as *ybar\_IC*). Table A1 shows the forecasting performance of the model with the intercept correction method (e.g., with MAE= 0.2, MSE= 0.07, MAPE= 3.2%, and SMAPE= 3.2%). The intercept corrected model substantially outperforms the model with the full data.

Figure B3 Simulated Sales with a structural break: model with intercept correction



1. This setting is typical in a retailer context. We arbitrarily make up the data series but keep the data series stationary. [↑](#footnote-ref-1)
2. For simplicity, we choose to illustrate the impact of structural breaks on forecasting accuracy using two structural breaks and also by holding the error variance to be constant before and after the breaks. Alternative settings (e.g., with different number of structural breaks and with changing error variance before and after the structural breaks) would provide the same indication. [↑](#footnote-ref-2)
3. In Figure A1, we use the blue area to represent the period before the first structural break (e.g., week [1,25]), use the yellow area to represent the period after the second structural break until the forecast origin (e.g., week [51, 75]), use the green area to represent the period between the two structural breaks (e.g., [26, 50]), and we use the red area to represent the forecast period (e.g., week [76, 100]). [↑](#footnote-ref-3)
4. The Chow test is a variant of F-test which compares the fitting of the model before and after the structural break. It assumes the locations of one structural break known a priori and also invariant error variations before and after the break. For a sequential Chow test, we conduct the Chow test assuming the break occurs at each individual week. For example, we may conduct the Chow test assuming there is a structural break at a specific week (e.g., week 10) and we obtain the p-value, and so forth. [↑](#footnote-ref-4)
5. We would consider the model not being subject to structural breaks only when all the p-values are above the threshold. To mitigate the multiple comparison problem, we adopt a very small threshold (e.g., 0.001) rather than the usual 0.05 for the p-values. [↑](#footnote-ref-5)