Problem Set II Solution

Tobias Bodentien Philipp Grunenberg Alexander Haas Osama Warshagha

03-12-2025

Task 3

In Task 3, we use the **StoreData** panel dataset containing monthly county-level sales observed over 18 months, where months 1–6 form the pre-period and months 7–18 the post-period. Treatment counties are affected by a store closure, while control counties are unaffected. Our goal is (i) to develop time-series forecasting models for **total sales** (offline + online) for one treated and one control county, and (ii) to study the interaction between offline and online sales using a VAR model following Lecture Chapter 2.2. Before selecting counties, we enforce a clean panel structure by restricting the sample to counties with a constant treatment status over time and complete coverage of months 1–18, since missing months or changing treatment labels would distort time-series estimation and invalidate comparisons. From this eligible set, we then randomly draw one treated and one control county using a fixed seed to ensure reproducibility; in our run, the selected counties are **treatment: county_id = 39** and **control: county_id = 25**.

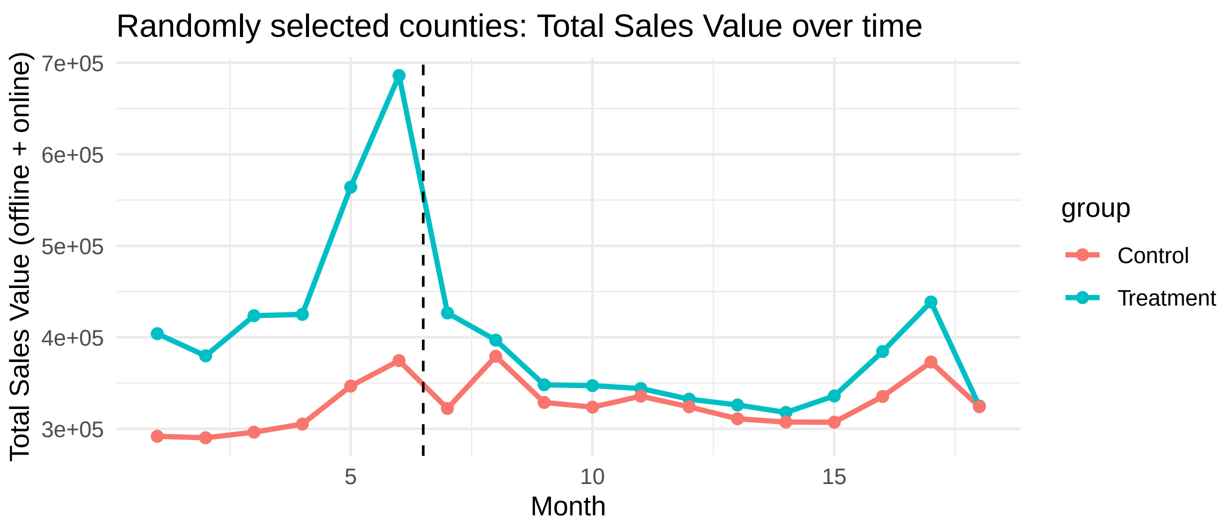

Figure 1: Total sales trajectories for the selected counties. Monthly total sales (offline + online) for the randomly selected treatment and control counties over months 1–18; the dashed vertical line indicates the start of the post-period (month 7).

The figure above plots the total sales value (offline + online) over time for both selected counties, with the dashed vertical line indicating the transition from the pre-period to the post-period (start at month 7). Visually, both counties show pronounced variation over time, especially around months 5–6 and later months, which motivates a careful time-series approach (stationarity checks, possible differencing, and lag selection) in the next step.

Time series setup (Total Sales)

We model monthly total sales value (offline + online) for each county as a univariate time series. The data are observed for 18 consecutive months; hence we treat the series as monthly data and focus on trend-stationarity / differencing, while seasonality (12-month cycle) cannot be identified reliably with such a short horizon.

Stationarity

We begin by assessing the time-series properties of total monthly sales in the treatment and control counties using the Augmented Dickey–Fuller (ADF) test. The ADF test evaluates whether a series contains a unit root. Formally, the null hypothesis is that the process has a unit root (and is therefore non-stationary), while the alternative hypothesis is that the series is stationary. In practice, a small p-value (e.g., below 0.05) provides evidence against the null, whereas a large p-value indicates that we cannot reject the presence of a unit root.

Before running the tests, we consider two representations of total sales for each county: levels and (when admissible) log-levels. Because the logarithm is only defined for strictly positive values, we first check whether each county's total sales series is positive throughout the sample. Both selected series are strictly positive, so applying the natural log transformation is valid. Using log-levels is useful in this context because it often stabilizes the variance for revenue-like data and allows changes to be interpreted approximately in percentage terms.

We then apply the ADF test to total sales in levels and log-levels for the treatment county (ID 39) and the control county (ID 25). The resulting p-values are relatively large (around **0.39–0.50** in levels and **0.38–0.48** in log-levels), implying that we cannot reject the null hypothesis of a unit root for any of the four series. In other words, both the level and log-level total sales series exhibit behavior that is statistically consistent with non-stationarity over the observed 18-month period.

Table 1: ADF test p-values

Series	ADF_p_value
Treatment 39 (level)	0.3873
Control 25 (level)	0.5011
Treatment 39 (log level)	0.3760
Control 25 (log level)	0.4839

To move the data closer to stationarity, we subsequently work with the first difference of the log-transformed series, $\Delta\log(\text{Total Sales})$, for both treatment and control counties. These differenced log series can be interpreted as approximate monthly growth rates in total sales. Differencing is a standard approach to remove deterministic trends and unit-root components, thereby yielding series that fluctuate around a more stable mean. Moreover, focusing on growth rates instead of levels facilitates comparison between counties that may differ substantially in their absolute sales levels.

We visualize these growth-rate series over time and mark the beginning of the post-treatment period again with a vertical dashed line, while a horizontal dotted line at zero indicates no growth. The plot shows how growth rates fluctuate around a roughly stable mean in both the pre- and post-periods, without obvious deterministic trends, which is consistent with the objective of obtaining a more stationary series.

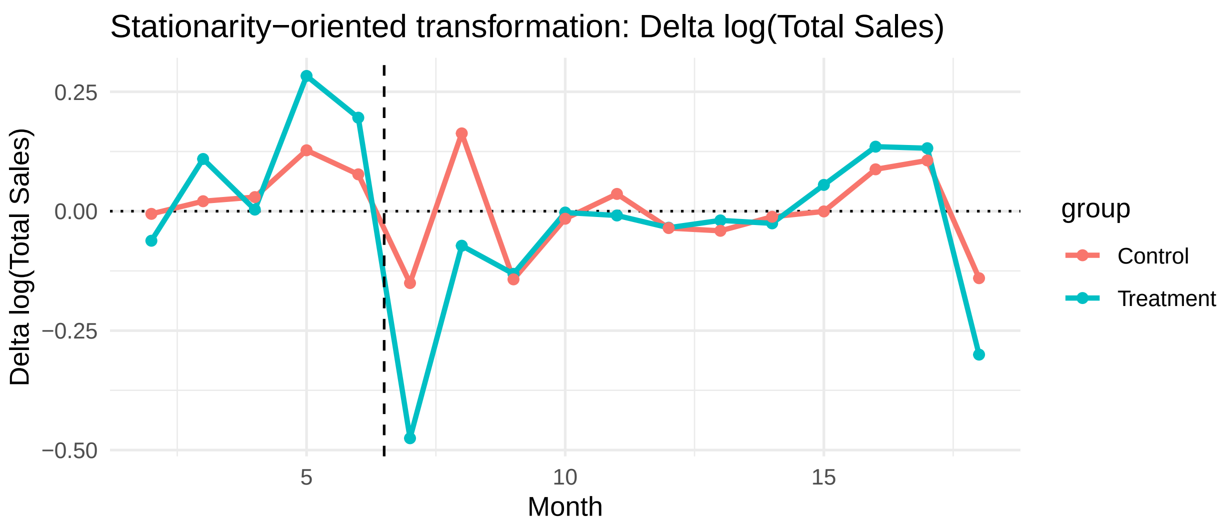

Figure 2: Growth-rate transformation of total sales (Treatment vs. Control). First differences of log total sales, $\Delta \log(\text{Total Sales})$, for the treatment and control counties over months 2–18; the dashed vertical line marks the start of the post-period (month 7) and the dotted horizontal line indicates zero growth.

Finally, we re-run the ADF test on the growth-rate series $\Delta\log(\text{Total Sales})$ for both counties. Since total sales are strictly positive throughout, the log transformation is well-defined. The resulting p-values remain relatively large (about 0.48–0.60), so we still cannot reject the unit-root null hypothesis at conventional significance levels. However, this outcome should be interpreted cautiously because the sample is extremely short (only 17 observations after differencing), which severely limits the power of unit-root tests. From a practical modeling perspective, working with $\Delta\log(\text{Total Sales})$ is still a sensible variance-stabilizing transformation that captures monthly growth rates. We therefore proceed with the differenced log series as the main outcome in the subsequent analysis, while acknowledging that formal stationarity evidence is weak in such a small sample.

Table 2: ADF tests on $\Delta\log(\text{Total Sales})$

Series	ADF_p_value
Treatment 39 (Δlog)	0.4803
Control 25 (Δlog)	0.6049

How many lags to include

To inform the choice of the dynamic structure in our subsequent time-series regressions, we examine the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the transformed outcome variable, i.e., the first difference of the log-transformed total sales, for both the treatment county (ID 39) and the control county (ID 25). The ACF summarizes the linear dependence of the series with its own past values at different lags, while the PACF isolates the incremental contribution of each lag after controlling for all shorter lags. Together, these diagnostics provide guidance on whether the series exhibits short-run persistence that would call for an autoregressive specification with one or more lags.

The Figure below displays the ACF and PACF of Delta log(Total Sales) for both counties. Overall, the plots do not suggest strong or long-lasting autocorrelation. At all positive lags, the sample autocorrelations are small and lie within the approximate 95% confidence bands, and the PACF does not show any pronounced spikes. Given the very short time dimension of our data (only 17 monthly observations after differencing), these diagnostics should be interpreted cautiously, but they nonetheless indicate that any remaining serial dependence is weak and, if present at all, likely to be of very low order.

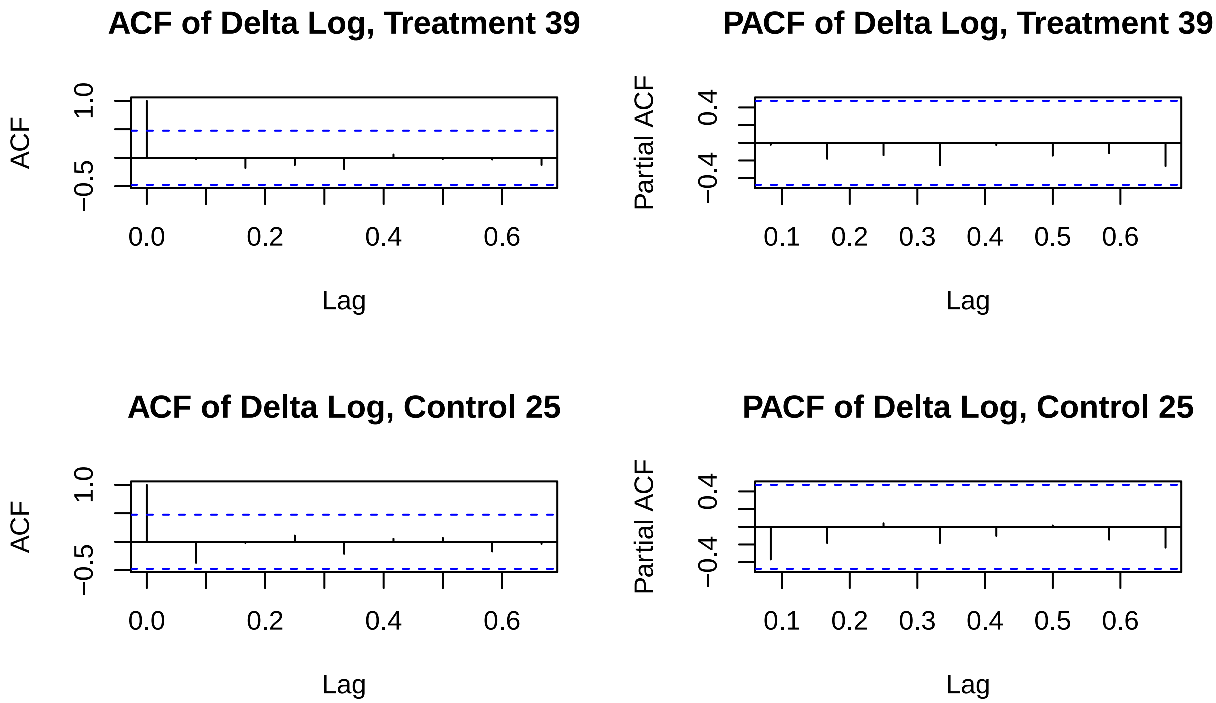

Figure 3: ACF/PACF of $\Delta \log($Total Sales$)$ (Treatment vs. Control). Autocorrelation (ACF) and partial autocorrelation (PACF) functions of the differenced log total sales series for the treatment county (top row) and the control county (bottom row); dashed lines indicate approximate 95% confidence bounds.

In light of these considerations and the **very short sample length** (18 months), we adopt a highly parsimonious specification and avoid heavily parameterized ARMA structures that would be poorly identified. Concretely, we model log(Total Sales) for each county using an ARIMA(0,1,0) with drift, i.e., a random walk with drift:

$$\log(y_t) = \log(y_{t-1}) + \mu + \varepsilon_t,$$

where μ denotes the average monthly growth rate in total sales (in log points) and ε_t is a white-noise innovation. This specification is equivalent to:

$$\Delta \log(y_t) = \mu + \varepsilon_t.$$

Thus, once we account for the integrated nature of log sales via first differencing (implicit in the ARIMA(0,1,0) structure), no additional autoregressive or moving-average terms are required to capture systematic short-run dynamics. This aligns with the diagnostic impression that any remaining serial dependence is weak and, in this small sample, unlikely to justify estimating additional lags. We therefore proceed with ARIMA(0,1,0) with drift for the treatment county (ID 39) and the control county (ID 25), and we interpret results cautiously given the limited time dimension.

Results (R-Output in rmd file)

The ARIMA(0,1,0) models with drift for log(Total Sales) confirm the very parsimonious dynamic structure suggested by the diagnostics. For the treatment county, the estimated drift is $\hat{\mu} = 0.0062$ (s.e. 0.0217), corresponding to an average monthly growth rate of about $e^{0.0062} - 1 \approx 0.62\%$. For the control county, the drift is $\hat{\mu} = -0.0128$ (s.e. 0.0420), i.e. approximately $e^{-0.0128} - 1 \approx -1.27\%$ per month. In both cases, the drift estimates are small relative to their standard errors, implying no statistically strong evidence of a systematic upward or downward trend in log sales over this short sample.

Consistent with the ARIMA(0,1,0) specification, no additional autoregressive or moving-average parameters are estimated. The innovation variance is $\hat{\sigma}^2 = 0.0085$ for the treatment county and $\hat{\sigma}^2 = 0.0319$ for the control county, indicating higher volatility of shocks to log sales in the control county. Overall, these results support using a random-walk-with-drift model as a defensible, low-parameter baseline given the limited time dimension.

Forecasting (Total Sales)

Based on the selected parsimonious specification, we forecast total sales (offline + online) for each county for the three months following the last observed month in the dataset (months **19–21**). Forecasts are generated on the log scale and then back-transformed to levels via exponentiation; the reported prediction intervals are obtained by applying the same transformation to the interval bounds and should be interpreted as an approximation.

The resulting point forecasts are very stable over the three-month horizon. For the treatment county, predicted total sales show a slight upward drift, increasing from approximately **326k** (month 19) to **330k** (month 21). For the control county, forecasts exhibit a small decline, from about **321k** (month 19) to **313k** (month 21). Overall, both counties are predicted to remain in a similar range in the immediate post-sample horizon.

Uncertainty, however, increases noticeably with the forecast horizon, as reflected in the widening 95% prediction intervals. Importantly, the prediction intervals for the control county are substantially wider than for the treatment county (e.g., by month 21 roughly **170k–573k** in the control county versus **242k–452k** in the treatment county), indicating higher volatility of shocks in the control county and therefore lower forecast precision. This difference is also clearly visible in the forecasting plots: while both panels show relatively flat point forecasts after month 18, the shaded 95% prediction band is markedly broader for the control county. In sum, while the point forecasts suggest only mild changes in expected sales over months 19–21, the wide intervals—especially for the control county—underline that inference is limited by the short sample and the inherent uncertainty of forecasting with only 18 monthly observations.

Table 3: 3-month forecasts for Total Sales (levels), with 95% PI.

month	treat_forecast	treat_lo95	treat_hi95	ctrl_forecast	ctrl_lo95	ctrl_hi95
19	320718.8	225939.4	455257.4	326160.4	272264.3	390725.5
20	316634.4	192934.6	519644.1	328185.7	254207.3	423692.9
21	312602.0	170408.2	573446.7	330223.5	241515.6	451513.6

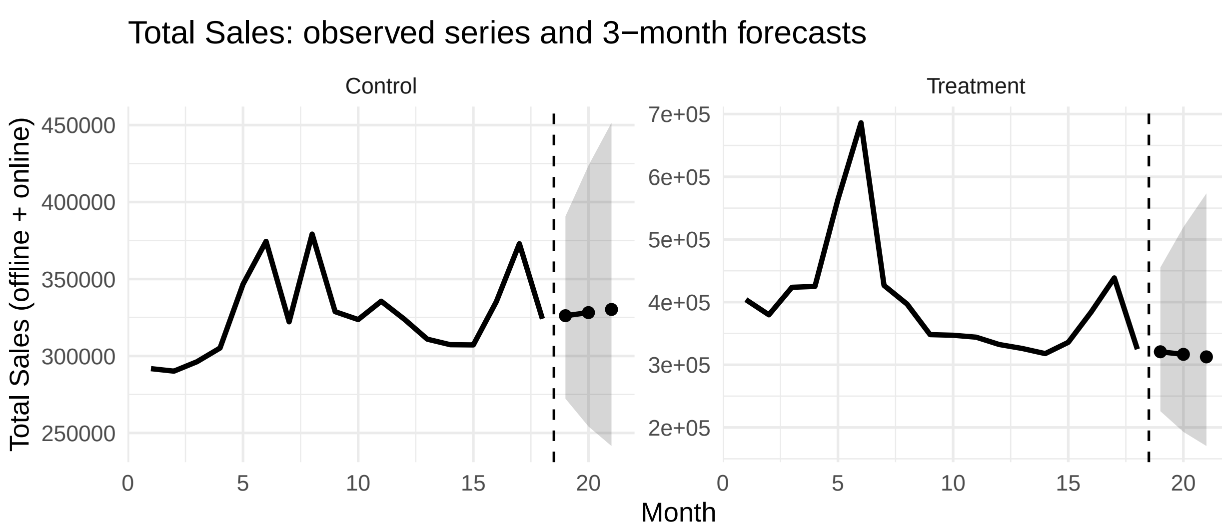

Figure 4: Total sales forecasts (Treatment vs. Control). Observed total sales (offline + online)

for months 1–18 and 3-month ahead forecasts for months 19–21 from the fitted ARIMA(0,1,0) with drift model (points/line); the dashed vertical line marks the forecast start and the shaded band shows the 95% prediction interval.

Residual ACF/ Seasonality

The residual ACF plots for both counties show no pronounced spikes outside the 95% confidence bands at positive lags, suggesting that the ARIMA(0,1,0) specification captures the main time-series structure reasonably well. In particular, there is no clear seasonal pattern (e.g., no prominent spike around the 12-month lag); given the very short sample (18 months), this also supports our decision **not** to model seasonality explicitly.

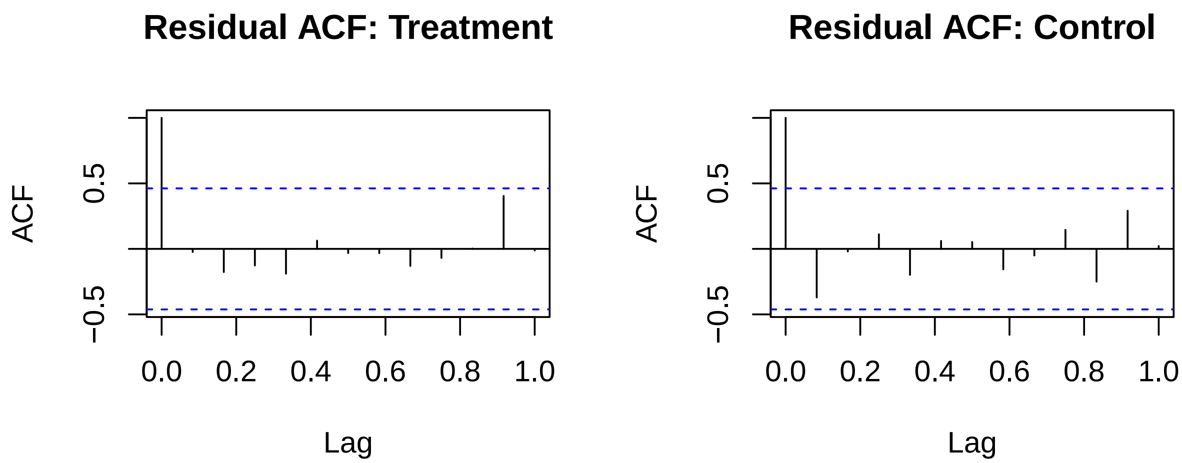

Figure 5: Residual ACF diagnostics (Treatment vs. Control). Residual ACFs of the fitted ARIMA(0,1,0) with drift models for the treatment (left) and control (right) county; dashed lines indicate approximate 95% confidence bounds.

VAR analysis (Online vs. Offline Sales)

We now turn to the second part of the task and analyze the interdependencies between offline and online sales using a Vector Autoregression (VAR) model for both the treatment and the control county. Given the very short sample (18 monthly observations), we make a few simplifying choices to keep the model identifiable and interpretable.

Transformation and simplifying assumptions.
Instead of modeling the sales levels, we work with log growth rates, i.e., the first differences of log sales, because the level series are likely non-stationary and a VAR in levels could lead to spurious dynamics in such a small sample. Working with $\Delta \log(\cdot)$ yields an approximately stationarity-oriented representation and has a natural economic interpretation as (approximate) monthly percentage changes. We further keep the specification parsimonious: we include a constant term and restrict the maximum lag order to a small number (due to limited degrees of freedom), and we do not attempt to estimate separate pre-/post-regimes.

Model setup.
For each county, we estimate a bivariate VAR on the transformed series

$$y_t \;=\; \begin{pmatrix} \Delta \log(\text{offline}_t) \\ \Delta \log(\text{online}_t) \end{pmatrix}, \qquad y_t \;=\; c + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

where c is a constant vector and u_t is a vector of innovations. The key quantities of interest are the cross-lag coefficients, i.e., how past online growth predicts offline growth (and vice versa).

Lag length selection.
To choose the lag order p, we use standard information criteria (AIC, HQ, SC, FPE) while restricting the search to very small lag lengths. For the treatment county, all criteria select $p = 2$ (VAR(2)), whereas for the control county, all criteria select $p = 1$ (VAR(1)). This yields a more flexible dynamic structure for the treatment county while remaining parsimonious.

Results (coefficients, R-Output in rmd file)

Treatment county (VAR(2)).
In the offline equation, we find a statistically significant cross-lag effect from online growth: $\Delta \log(\text{online})_{t-1}$ enters with a positive coefficient (estimate ≈ 0.62, $p \approx 0.009$). This suggests that increases in online sales growth tend to be followed by higher offline sales growth one month later in the treatment county. The second lag of offline growth is negative and marginally significant (estimate ≈ -0.55, $p \approx 0.074$), indicating some mean reversion in offline growth at the two-month horizon. Overall, the offline equation is jointly significant (F-test $p \approx 0.042$), consistent with meaningful short-run dynamics.

In the online equation, the lagged offline growth rate shows a positive but only weakly significant association ($\Delta \log(\text{offline})_{t-1}$ estimate ≈ 0.89, $p \approx 0.093$), while the remaining terms are not statistically strong. Hence, the most robust dynamic linkage in the treatment county runs from online (lag 1) to offline.

Control county (VAR(1)).
For the control county, the VAR(1) results are substantially weaker. Neither cross-lag coefficient is statistically significant in either equation, and the overall explanatory power is low—especially for the online equation. The only term that comes close to significance is the own-lag in the offline equation ($\Delta \log(\text{offline})_{t-1}$ estimate ≈ -0.58, $p \approx 0.097$), which is consistent with mild mean reversion in offline growth. Overall, these estimates suggest that short-run interactions between online and offline growth are much less pronounced in the control county.

Contemporaneous comovement.
In both counties, the residual correlation between the offline and online equations is relatively high (around 0.63–0.69), indicating that both channels are affected by common shocks within the same month, even when lagged dynamics are weak.

Taken together, the coefficient estimates point to stronger dynamic interdependence in the treatment county, in particular from online to offline sales growth, whereas the control county exhibits little evidence of systematic cross-channel lag structure. In the next step, we use impulse response functions (IRFs) to summarize and compare these dynamics more directly over time.

Impulse response functions (IRFs)

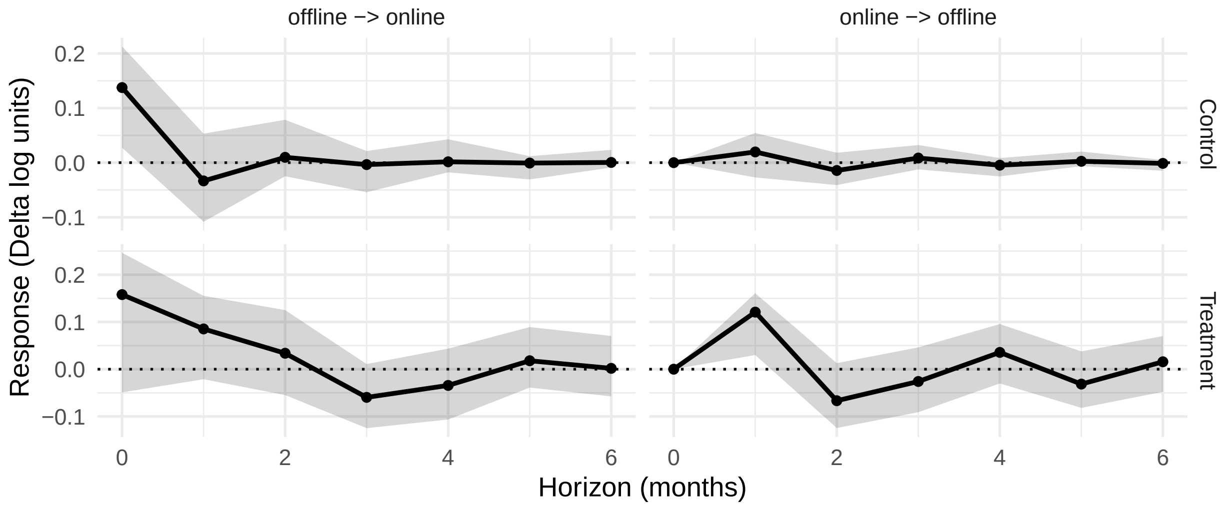

Figure 6: Impulse response functions (IRFs) from the VAR model (Treatment vs. Control). Estimated responses of $\Delta\log(\text{online})$ to a one-unit shock in $\Delta\log(\text{offline})$ (left column) and of $\Delta\log(\text{offline})$ to a one-unit shock in $\Delta\log(\text{online})$ (right column), shown separately for the control (top row) and treatment county (bottom row); the shaded areas denote bootstrap confidence bands and the dotted horizontal line marks zero response.

The IRFs broadly confirm the coefficient-based findings. In the control county, both the offline→online and online→offline responses remain close to zero across horizons, with confidence bands that typically include zero, indicating little evidence of systematic cross-channel dynamics. In contrast, the treatment county shows more pronounced short-run interactions: an offline shock is followed by a positive response of online growth at short horizons, and an online shock is associated with a noticeable response in offline growth (peaking early and then fading). Overall, the dynamic effects in the treatment county appear stronger but also imprecisely estimated, so results should be interpreted cautiously given the very short sample.