from $F(\cdot)$. Interference C and D resemble interference A, except that in C a successful treated trial only has a response drawn from $F^{\nu}(\cdot)$ if it follows 2 or more control trials, and in D if it follows 3 or more control trials.

Interference between units creates one type of dependence over successive trials, but there can also be other types of dependence that are present in the absence of interference, indeed present in the absence of any treatment effect. The upper half of Tables 5 and 6 is dependent over successive trials only due to interference. In the lower half of Tables 5 and 6, the responses above are added to stationary autoregressive errors with standard Normal marginal distributions and autocorrelation 0.5.

Each situation was simulated 5000 times, so the simulated power has a standard error of at most $\sqrt{.25/5000} = 0.007$.

5.2 Results of the simulation

Tables 5 and 6 contrast the size and power of four test statistics, namely the conventional pooled variance t-statistic and $T_{\mathbf{Z}}$ for three values of k, $k = 2$, $k = 5$, and $k = 10$. Recall that $k = 2$ corresponds with the Mann-Whitney-Wilcoxon statistic, and $k = 5$ is similar to the suggestion of Salzburg (1986) and Conover and Salzburg (1988).

The case of $\nu = 1$ in Tables 5 and 6 is the null hypothesis: it suggests that all four tests have size close to their nominal level of 0.05 in all sampling situations. This is expected for $T_{\mathbf{Z}}$ because it is a randomization test applied under the null hypothesis of no effect in a randomized experiment. For related results about the randomization distribution of statistics such as the t-statistic, see Welch (1937). Notice that, because this is a randomization test in a randomized experiment, it has the correct level even in the case of autocorrelated errors. In brief, because all four tests appear to be valid, falsely rejecting true hypotheses at the nominal rate of 5%, it is reasonable to contrast the tests in terms