The (Possible) Effect of Plain Packaging on Smoking Prevalence in Australia: A Trend Analysis

Ashok Kaul and Michael Wolf

June 2014

The (Possible) Effect of Plain Packaging on Smoking Prevalence in Australia: A Trend Analysis*

Prof. Dr. Ashok Kaul†

Institute for Policy Evaluation (IPE), Saarland
and Department of Economics, Saarland University

Prof. Michael Wolf, Ph.D.‡

Department of Economics
University of Zurich

June 2014

Abstract

A stated objective of the Australian Plain Packaging Act 2011 is to reduce smoking prevalence. We use the Roy Morgan Single Source (Australia) data set over the time period January 2001 to December 2013 to analyze whether this goal has been achieved in the first year since the implementation. In particular, we carry out a statistical trend analysis to study the (possible) effect of plain packaging on smoking prevalence. Two informative analyses help to draw conclusions on the (actual) effect of plain packaging on smoking prevalence in Australia. First, we look at the year of data *before* plain packaging was introduced, which happened in December 2012. Second, we compute confidence intervals around the estimated treatment effects.

Our main results can be summarized as follows. First, if a statistical significance level of 5% is required, then there is no evidence at all for a plain packaging effect on smoking prevalence. Second, if one is willing to accept a relatively low level of statistical significance (that is, 10%), then there is evidence for a very short-lived plain packaging effect on smoking prevalence, namely in December 2012 only (after which smoking prevalence is statistically indistinguishable from its pre-existing trend).

A formal power analysis demonstrates that the power of our inference methods is remarkably high.

KEY WORDS: Plain packaging, smoking prevalence, treatment effect, trend analysis.

JEL CLASSIFICATION NOS: C13, C22, H43, I18.

1   Goals and Basic Setup

The Australian Tobacco Plain Packaging Act 2011 prescribes that from December 2012 on, cigarettes and other tobacco products have to be sold in plain packages in Australia, that is, in packs with a standardized design and shape. Australia is thereby the first country to introduce such a regulation. The key objective of the Plain Packaging Act 2011 is the improvement of public health by discouraging the taking up of smoking and by encouraging the giving up of smoking and the use of other tobacco products. So far, there is no empirical evidence that the measures prescribed by the Plain Packaging Act 2011 are effective in attaining the stated goals of the Australian government. In fact, there is hitherto only a single research paper that empirically studies the (possible) effect of plain packaging in Australia on changes in smoking prevalence: Kaul and Wolf (2014) provide a trend analysis similar to the one in this paper but focusing on minors (aged 14–17 years) only.[1] They fail to find any evidence for a plain packaging effect on Australians aged 14–17 years.

Plain packaging in Australia was implemented in December 2012 and thus had been in place for one year in December 2013. As a consequence, reliable data that cover both the pre-implementation period and a sufficiently long post-implementation period are now available for a first thorough empirical assessment of the effects of plain packaging. Given the unprecedented nature of the intervention, no one could predict for sure what the intervention would lead to. In a notable contribution, Pechey et al. (2013) run an elicitation survey on over 30 internationally-renowned experts on tobacco control policies, asking them about their expectations of the effect of plain packaging on smoking prevalence rates two years after its introduction. The experts were asked to provide estimates, *holding all other factors constant*. In the case of Australia, the introduction of plain packaging came together with an enlargement of graphical health warnings. Assuming both effects work in the same direction, the Australian case should therefore show a bigger reaction than what would be expected based on an isolated plain packaging experiment alone. The median estimate of the experts in Pechey et al. (2013) for the impact on adult smoking prevalence was a one percentage point decline. Taking the expected reaction for adults as a lower bound, we can therefore expect to find at least a drop in smoking prevalence of one percentage point two years after the introduction of plain packaging (if the expert opinions are correct predictors of what to expect). Since we have one year of post-implementation data, it is important to ensure that an actual plain packaging effect of less than one percentage point is picked up by the chosen statistical inference methods with reasonable power.

This paper addresses the question whether there is empirical evidence showing that the pre-implementation trend in smoking prevalence in Australia has been changed by plain pack-

---

[1]Since a major reason for the introduction of plain packaging was the objective of reducing smoking prevalence of minors in particular, there is considerable interest in analyzing the sub-population of minors separately.

aging. The research question guiding our statistical analysis is the following: Can we find any plain packaging effect on smoking prevalence *at all* over the 13 months from December 2012 to December 2013? In principle, a careful analysis requires the use of a multiple-testing adjustment to take the possibility of "cherry picking" into account (that is, the possibility of searching for a statistically significant effect over the entire period).[2] Note, however, that in most of the paper, we employ a statistical approach more favorable to finding a plain packaging effect, namely by asking whether there is a plain packaging effect in any *specific* month. This approach ignores "cherry picking" and does not require any multiple-testing adjustment. A formal power analysis demonstrates that our approach can identify even small reductions in smoking prevalence with reasonable power.

2 Data Description and Construction

We use the Roy Morgan Single Source (Australia) data set (RMSS subsequently) over the time period January 2001 to December 2013. The total sample size over this 13-year period is around 700,000; the average annual sample size is around 54,200.

Roy Morgan is a major Australian market research firm and the Single Source data set has been drawn from the so-called establishment survey. These are weekly surveys realized as computer-assisted personal interviews (CAPI) that are administered door-to-door; see Roy Morgan Research (2012).

In each month, we compute (observed) smoking prevalence as the average of the 0-1 variable `smoker` in the RMSS data that indicates whether an individual in the sample smokes. Note that there is considerable variation in the sample size over time; see Figure 1. The sample sizes generally range between 3,500 and 5,000 and are thus quite large.[3] On the other hand, the composition of the sample changes from month to month; therefore, it is expected that monthly observed prevalence is unstable over time. This is indeed the case; see Figure 2.

3 Data Analysis

3.1 Fitting a Linear Time Trend

We start by modeling a simple linear time trend. This is achieved by estimating the regression model

$$p_t = \alpha + \beta \cdot t + \varepsilon_t \ . \tag{3.1}$$

---

[2]For example, Heckman et al. (2010) convincingly promote the use of multiple-testing adjustments to avoid the erroneous detection of treatment effects when "cherry picking" is possible.

[3]December 2013 is marked by a relatively low number because Roy Morgan decided not to interview in the week leading up to Christmas. Therefore, the sample size for December 2013 is 'only' 3,124. Future numbers are expected to be higher again.

Here, p_t denotes the observed prevalence in month t ($t = 1, \dots, 156$), α denotes the intercept of the linear time trend, β denotes the slope of the linear time trend, and ε_t denotes the error term in month t (that is, the deviation of the observed prevalence from the trend line).

We fit model (3.1) by weighted least squares, using the monthly sample sizes as the weights.[4] The fitted model is given by

$$\hat{p}_t = 24.61 - 0.040 \cdot t \; . \tag{3.2}$$

This model implies an average yearly decline of $12 \cdot 0.040 \approx 0.48$ percentage points in smoking prevalence over the period 2001 until 2013; see Figure 2 for a graphical display. We also include a local, nonparametric trend that does not make any assumptions on the parametric form of the trend (like linear or quadratic). Such a nonparametric trend provides a good local fit and avoids the problem of misspecification.[5] It can be seen that the (global) linear trend is not a very satisfactory fit to the observed data: it is somewhat too high early on and in the final years while somewhat too low in the middle.

Despite its flexible nature, the nonparametric fit resembles a straight line in the second two thirds of the observation period, which is the interval of main interest to us. For simplicity, and for ease of reproducibility of our results by other researchers, we match the nonparametric trend in the second two thirds of the data by fitting a linear time trend from 07/2004 on.[6] Furthermore, we exclude the data from 12/2012 until 12/2013 in fitting this linear time trend, thereby avoiding a possible contamination of the fitted trend line in case there should be a strong plain packaging effect. The fitted trend[7] based on the period 07/2004–11/2012 is given by

$$\hat{p}_t = 25.23 - 0.045 \cdot t \; . \tag{3.3}$$

(A more detailed regression output can be found in Table 1.) This model implies an average yearly decline of $12 \cdot 0.045 \approx 0.54$ percentage points in prevalence from 07/2004 on. The results are displayed in Figure 3. It can be seen that in the last two thirds of the period, the linear trend is, for all practical purposes, indistinguishable from the nonparametric trend.

¿From here on, we will therefore base the analyses on the fitted linear trend (3.3).

---

[4]Since the sample sizes vary considerably over time, as evidenced in Figure 1, weighted least squares (WLS) gives more accurate estimation results than ordinary least squares (OLS); for example see Hayashi (2000, Section 1.6).

[5]We use a standard off-the-shelf method, namely the function `loess` of the statistical package R (with default model parameters); see http://stat.ethz.ch/R-manual/R-patched/library/stats/html/loess.html.

[6]Other researchers who do not use the statistical package R might get slightly different results when fitting a nonparametric trend. But they will get the same results when fitting a linear time trend from 07/2004 on.

[7]Again obtained by weighted least squares.

3.2 Analyzing Deviations from the Linear Time Trend

3.2.1 A Naïve First Step

The deviations of the observed data from the fitted linear time trend from 12/2012 until 12/2013 are displayed in Figure 4. Of the 13 deviations, seven are negative and six are positive. The average deviation is −0.16 percentage points. A naïve (and incorrect) interpretation would be that, on average, plain packaging has resulted in a monthly reduction in prevalence of 0.16 percentage points.

However, one must take into account that the observed prevalence numbers are only estimates themselves. Therefore, one must not equate an estimated (treatment) effect of plain packaging in a given month — namely, the deviation of the observed prevalence from the fitted trend line — with the true effect.

3.2.2 A More Informative Analysis Based on Pre Plain Packaging Deviations

One robustness check is to also include *previous deviations* from the linear time trend in such a plot. If one starts the plot one year prior to the intervention, that is, in 12/2011 rather than in 12/2012, then the numbers post 12/2012 are not 'unusual' compared to the numbers pre 12/2012; see Figure 5.[8] In fact, given the generally larger deviations (in absolute value) pre 12/2012, the deviations post 12/2012, with the possible exception of 12/2012 itself, appear just like random noise. The largest negative deviation from 01/2013 on is −0.87 percentage points in 04/2013. But there are two larger negative deviations *before* 12/2012, namely −1.32 in 02/2012 and −1.56 in 04/2012. It is clear that a negative deviation from the fitted time trend alone cannot be equated with an actual plain packaging effect.

The average deviation post 12/2012 is −0.04 percentage points.[9] This is smaller than the average deviation pre 12/2012, which is 0.23 percentage points. However, this difference is not statistically significant: carrying out a two-sided t-test[10] yields a p-value of 0.38.

According to this analysis then, there is no evidence for a plain packaging effect beyond 12/2012 itself.

3.2.3 A More Informative Analysis Based on Confidence Intervals

Another robustness check is to add confidence intervals to the estimated effects of plain packaging in Figure 4. For a given month, this can be achieved as follows:

---

[8]The numbers pre 12/2012 are the numbers 12/2011–11/2012 and the numbers post 12/2012 are the numbers 01/2013–12/2013, so each set of numbers corresponds to twelve months (for reasons of symmetry).

[9]This number differs from the number −0.16 percentage points stated in Section 3.2.1, since 12/2012 itself is excluded now.

[10]Using a nonparametric inference method, such as a bootstrap test, does not change this conclusion. We report the outcome of the t-test, since this simple result can be easily reproduced by other researchers.

Algorithm 3.1 (Computation of Confidence Intervals for Plain Packaging Effects)

1. Compute a 90% prediction interval for the observed prevalence based on the fitted time trend (that is, assuming no plain packaging effect). This means if another random sample (with the same sample size) had been chosen instead for this month, then the resulting observed prevalence would have fallen in this interval with 90% confidence (assuming no plain packaging effect). Or, alternatively, 90% of all possible random samples (with the same sample size) would have resulted in observed prevalence numbers falling in this interval (assuming no plain packaging effect). By construction, this interval is centered at the linear time trend.

2. Subtract the observed prevalence based on the original data from the upper and the lower interval end points.

3. The thus shifted resulting interval can be interpreted as a 90% confidence interval for the actual (treatment) effect of plain packaging. By construction, this interval is centered at the deviation from the linear time trend. If the entire interval lies below zero, then there is evidence (at the 90% confidence level[11]) that plain packaging has lead to a reduction in prevalence.

The results are displayed in Figure 6. It can be seen that there is no statistical significance for a plain packaging effect beyond 12/2012 itself: for all other months, the number zero is contained in the confidence interval.

Several reasonable variations to the methodology used are possible and could in fact be called for, either because they are more standard than the method we use or because they are more appropriate (superior) given the properties of the data.

- We have computed the prediction intervals in step 1. of Algorithm 3.1 using standard textbook methodology based on an assumption of a normal distribution of the error terms ε_t in the linear model for the time trend. An analysis of the residuals[12] of the fitted model (3.2) indicates that this assumption is not violated in any noticeable way. It is possible in step 1. to use a more refined (and more computationally involved) bootstrap approach to compute prediction intervals that also incorporate potential non-normality of the error terms. The resulting changes would be minor, at most, and they would not change our conclusions.[13]

- The standard textbook methodology for the prediction intervals in step 1. of Algorithm 3.1 also assumes that the error terms ε_t around the linear time trend are independent and identically distributed (i.i.d.). This assumption might be violated in our

---

[11]Or, equivalently, at the 10% significance level.

[12]The residuals $\hat{\varepsilon}_t$ are computed as $\hat{\varepsilon}_t = p_t - \hat{p}_t$ (t = 43, ..., 143).

[13]Again, we opt for sticking with the simpler methodology, so that our findings can be more easily reproduced by other researchers.

application, since the data is collected over time and so the error terms might be auto-correlated. First of all, ignoring such a violation would only have a minor effect, since a (possible) autocorrelation of the error terms enters into the uncertainty of the estimated coefficients of the fitted model (3.3) (that is, the estimated trend line) but not the uncertainty due to a new observation (that is, the deviation from the trend line); the latter uncertainty far outweighs the former in determining the width of the interval. Second, ignoring a (possible) autocorrelation of the error terms generally makes the intervals smaller rather than wider, since error terms are generally positively autocorrelated rather than negatively autocorrelated, if autocorrelated at all. Third, an analysis of the residuals of the fitted model (3.3) does not show any autocorrelation whatsoever; see Figure 7.

- The confidence level could be changed from 90% to 95%. The latter is more standard in applied research and would result in wider confidence intervals. If the confidence level is changed to 95%, then there is no evidence for a plain packaging effect whatsoever, since even the confidence interval for 12/2012 contains zero. More precisely, the confidence interval for 12/2012 changes from $[-3.03, -0.25]$ to $[-3.30, 0.02]$; see Figure 8.

- We have computed *pointwise* confidence intervals. That is, the confidence of 90% holds for any given month. Doing so is appropriate if one is interested in whether there is a plain packaging effect in any *specific* month, say in December 2012. But if one is interested in whether there is any plain packaging effect *at all* over the 13 months under consideration, it is more appropriate to compute *uniform* confidence intervals, where the 90% confidence holds over all 13 months together.[14] Doing so results in wider intervals, and now even the interval for 12/2012 contains zero; see Figure 9.[15]

3.3 Power Analysis

As mentioned in Section 2, monthly observed prevalence is unstable over time and the deviations from the fitted trend line (3.3) are not small. This might raise the concern of whether our trend analysis has any reasonable power at all against a possible plain packaging effect beyond 12/2012 itself. We address this concern by carrying out a formal power analysis.

In particular, we consider the following inference methods to test for a plain packaging effect during the period 01/2013–12/2013 which is consistent with our previous analyses.

Algorithm 3.2 (Inference Methods)

---

[14]Doing so prevents *data mining* or *cherry picking* by searching for *any* effect over the 13 months under consideration.

[15]Since there is no evidence for any autocorrelation in the error terms ε_t, uniform confidence intervals can by computed in the same fashion as pointwise confidence intervals, except that the confidence level is changed from 90% to 99.2%. Note here that $0.9^{1/13} = 0.992$.

1. Fit a linear time trend (using weighted least squares) based on the observation period 07/2004–11/2012, that is, based on $t = 43, \ldots, 143$.

2. Compare the average deviation pre 12/2012 to the average deviation post 12/2012, as done in Section 3.2.2. If the average deviation post 12/2012 is smaller than the average deviation pre 12/2012, carry out a formal two-sample t-test for the null hypothesis of zero difference in population (that is, for the null hypothesis of no treatment effect).[16] If the t-test rejects the null hypothesis, this is considered evidence for a plain packaging effect. We call this approach inference method 1 (IM-1).

3. Compute individual 90% confidence intervals for plain packaging effects from 01/2013 until 12/2013, as detailed in Section 3.2.3. If at least one of the resulting 12 confidence intervals is entirely negative, this is considered evidence for a plain packaging effect. We call this approach inference method 2 (IM-2).

4. Overall, evidence for a plain packaging effect is established if at least one of these two approaches, IM-1 or IM-2, finds evidence. We call this 'combined' approach inference method 3 (IM-3).

The next step is to generate pseudo data that are qualitatively similar to the observed data, but where a specified plain packaging effect is 'enforced'. Here some care must be taken, since the monthly samples sizes are not constant, which implies that the error terms ε_t around the trend line do not have the same variance. Denote the sample size in month t by n_t ($t = 43, \ldots, 156$). Then we may assume

$$\mathrm{Var}(\varepsilon_t) = \frac{\sigma^2}{n_t} \qquad \text{for some } \sigma^2 > 0 \ .$$

The fitted model (3.3) yields the estimator $\hat{\sigma}^2 = 2589.7$.

We next detail how we generate pseudo prevalence data according to a model that is in agreement with the observed data but has a specified plain packaging effect $\Delta > 0$ 'enforced' from 12/2012 on, that is, from $t = 144$ on.[17]

Algorithm 3.3 (Generation of Pseudo Data with Specified Plain Packaging Effect)

1. Generate $\gamma_{43}^*, \ldots, \gamma_{156}^*$ independent and identically distributed as $N(0, 2589.7)$, where the notation $N(0, \sigma^2)$ denotes a normal distribution with mean zero and variance σ^2.

2. For $t = 43, \ldots, 156$, let

$$p_t^* = 25.23 - 0.045 \cdot t + \varepsilon_t^* \qquad \text{where} \qquad \varepsilon_t^* = \frac{\gamma_t^*}{\sqrt{n_t}} \ .$$

---

[16]There was no need to carry out such a t-test in Section 3.2.2, since the average deviation post 12/2012 was larger than the average deviation pre 12/2012.

[17]So Δ is the (fraction of) percentage points by which plain packaging has lowered prevalence beyond the time trend. It makes no difference for the purposes of this power analysis whether we enforce the effect from 12/2012 or from 01/2013 on.

3. For $t = 144, \ldots, 156$, let

$$p_t^* = p_t^* - \Delta .\text{[18]}$$

We finally detail how we 'compute' power against a specific plain packaging effect $\Delta > 0$ via Monte Carlo simulation.

Algorithm 3.4 (Computation of Power against Specific Plain Packaging Effect)

1. Generate pseudo data with a plain packaging effect Δ according to Algorithm 3.3.
2. Analyze the pseudo data according to Algorithm 3.2.
3. If evidence is claimed, record a one; otherwise, record a zero.
4. Repeat this process a large number B of times.
5. The 'computed' power is the fraction of ones over the B repetitions.

The resulting numbers are presented in Table 2. One can see that power is actually high in general. For example, power of the inference method 3 (IM-3) against a plain packaging effect of 0.5 percentage points is 0.85 and power against a plain packaging effect of 1.0 percentage point is 0.99. Power of 0.8 is a commonly accepted industry standard[19], so even the power against a plain packaging effect of only 0.5 percentage points is already very high.

4 Conclusion

We carried out a trend analysis to study the (possible) effect of plain packaging on smoking prevalence in Australia. More specifically, we fitted a linear time trend that explains well the fact that observed prevalence has declined steadily from mid 2004 on at an annual rate of about 0.54 percentage points.[20]

It is of particular interest to see how observed prevalence behaves relative to the fitted trend line from December 2012 on (that is, from the point when plain packaging was implemented). It was seen that observed prevalence lies sometimes above and sometimes below the fitted trend line.

Two informative analyses help to draw conclusions on the (actual) effect of plain packaging on smoking prevalence in Australia. First, we looked at the year of data *before* December 2012. Second, we computed confidence intervals around the estimated plain packaging effects (that is, around the deviations from the fitted trend line) from December 2012 on. Both analyses fail to find any evidence for an actual plain packaging effect on smoking prevalence in Australia after December 2012.

Our results can be summarized as follows. First, if one is willing to accept a relatively low level of statistical significance (10%), then there is evidence for a very short-lived plain

---

[18]This slight abuse of notation means that the final value of p_t^* equals the value of p_t^* after step 2. minus Δ.

[19]For example, see Section V.G. of FDA (2008).

[20]Observed prevalence had declined before also, but at a slower rate.

packaging effect on smoking prevalence, namely in December 2012 only (after which smoking prevalence is statistically indistinguishable from its pre-existing trend). Second, if a stronger statistical significance level (5%) is required, then there is no evidence at all for a plain packaging effect on smoking prevalence. Third, if the guiding research question is whether there is a plain packaging effect at all, one must adjust the confidence intervals to take the possibility of "cherry picking" into account (that is, the possibility of searching for a statistically significant effect over the entire period). Such an adjustment requires the use of uniform confidence intervals, in which case there is again no evidence for a plain packaging effect on smoking prevalence.

References

FDA (2008). Diabetis mellitus: Developing drugs and therapeutic biologics for treatment and prevention (Guidance for Industry). Available online at the following link: http://www.fda.gov/downloads/Drugs/Guidances/ucm071624.pdf.

Hayashi, F. (2000). *Econometrics*. Princeton University Press, Princeton, New Jersey.

Heckman, J., Moon, S. H., Pinto, R., Savelyev, P., and Yavitz, A. (2010). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics*, 1(1):1–46.

Kaul, A. and Wolf, M. (2014). The (possible) effect of plain packaging on the smoking prevalence of Australian minors: a trend analysis. Working Paper ECON 149, Department of Economcis, University of Zurich.

Pechey, R., Spiegelhalter, D., and Marteau, T. M. (2013). Impact of plain packaging of tobacco products on smoking in adults and children: an elicitation of international experts' estimates. *BMC public health*, 13(1):18.

Roy Morgan Research (2012). How we collect and process Single Source data in Australia. Melbourne (Australia). Roy Morgan Research. Available online at the following link: http://www.roymorgan.com/products/single-source/single-source-fact-sheets.
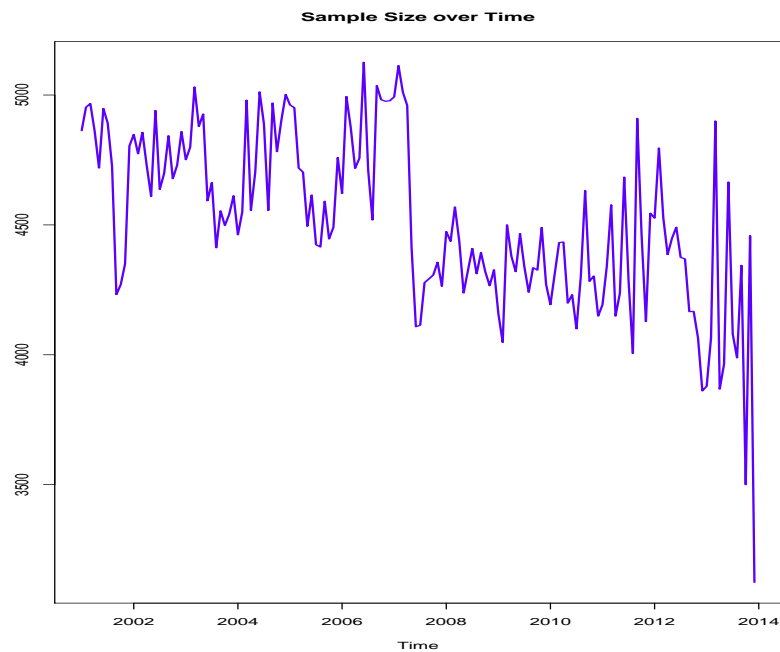
A   Figures and Tables

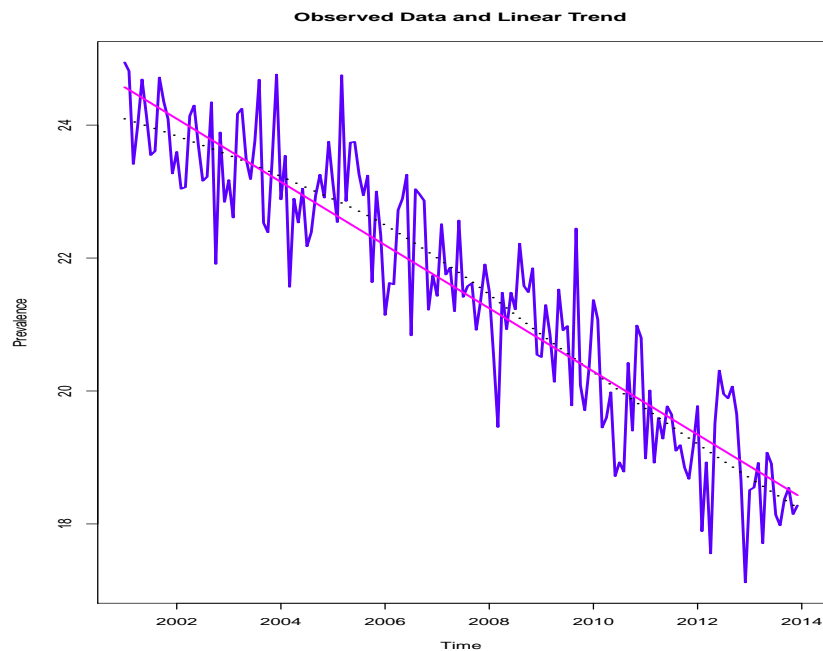Figure 1: Time series plot of the monthly sample sizes.

Figure 2: Time series plot of observed prevalence with fitted linear trend based on all observations (solid line). In addition, a fitted nonparametric trend has been added (dotted line).
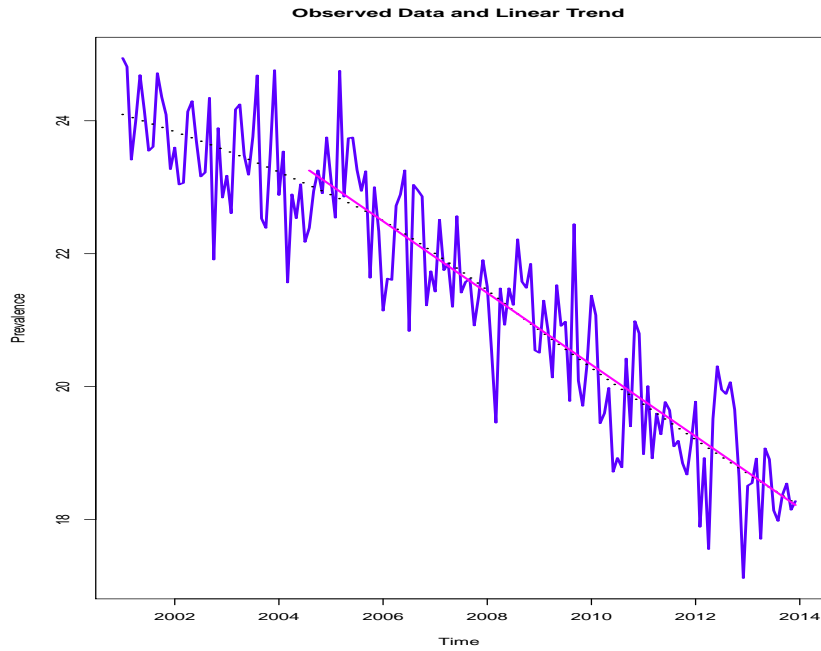
11

Figure 3: Time series plot of observed prevalence with fitted linear trend based on the observations from 07/2004 on (solid line). In addition, a fitted nonparametric trend has been added (dotted line).
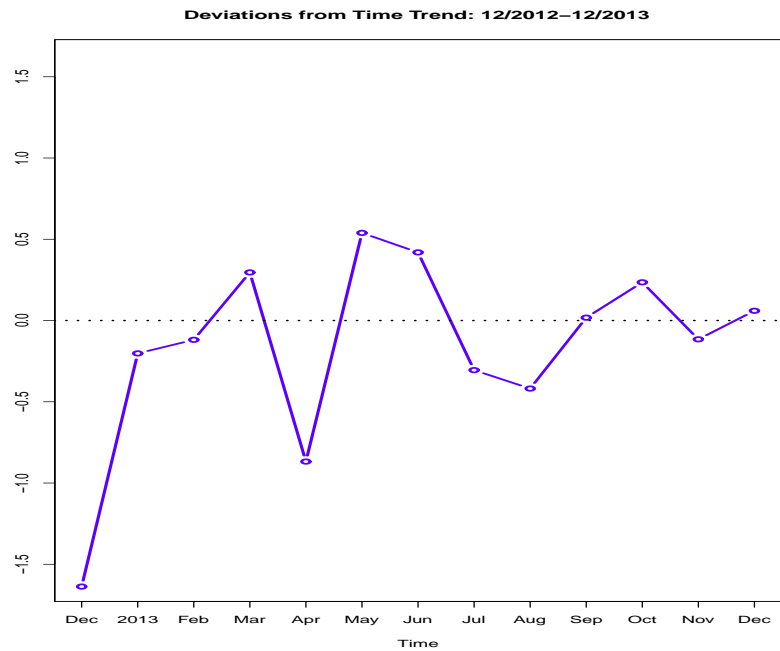

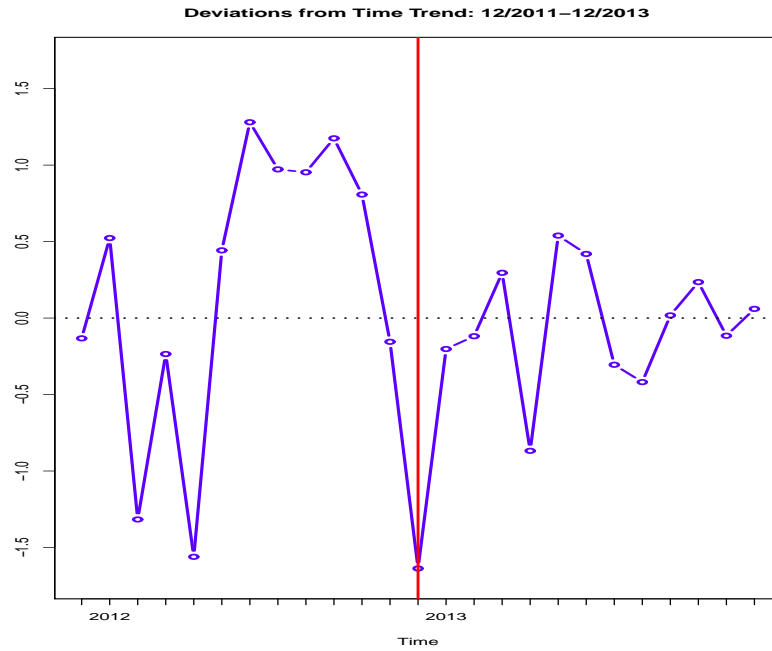Figure 4: Deviations of observed prevalence from fitted time trend.

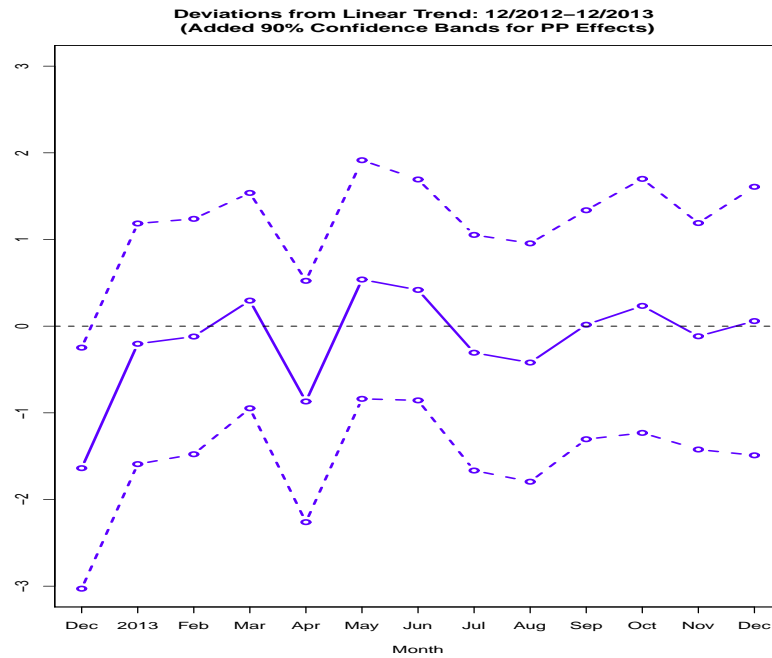Figure 5: Deviations of observed prevalence from fitted time trend.

Figure 6: Deviations of observed prevalence from fitted linear time trend. Pointwise 90% confidence intervals for these estimated plain packaging effects have been added.
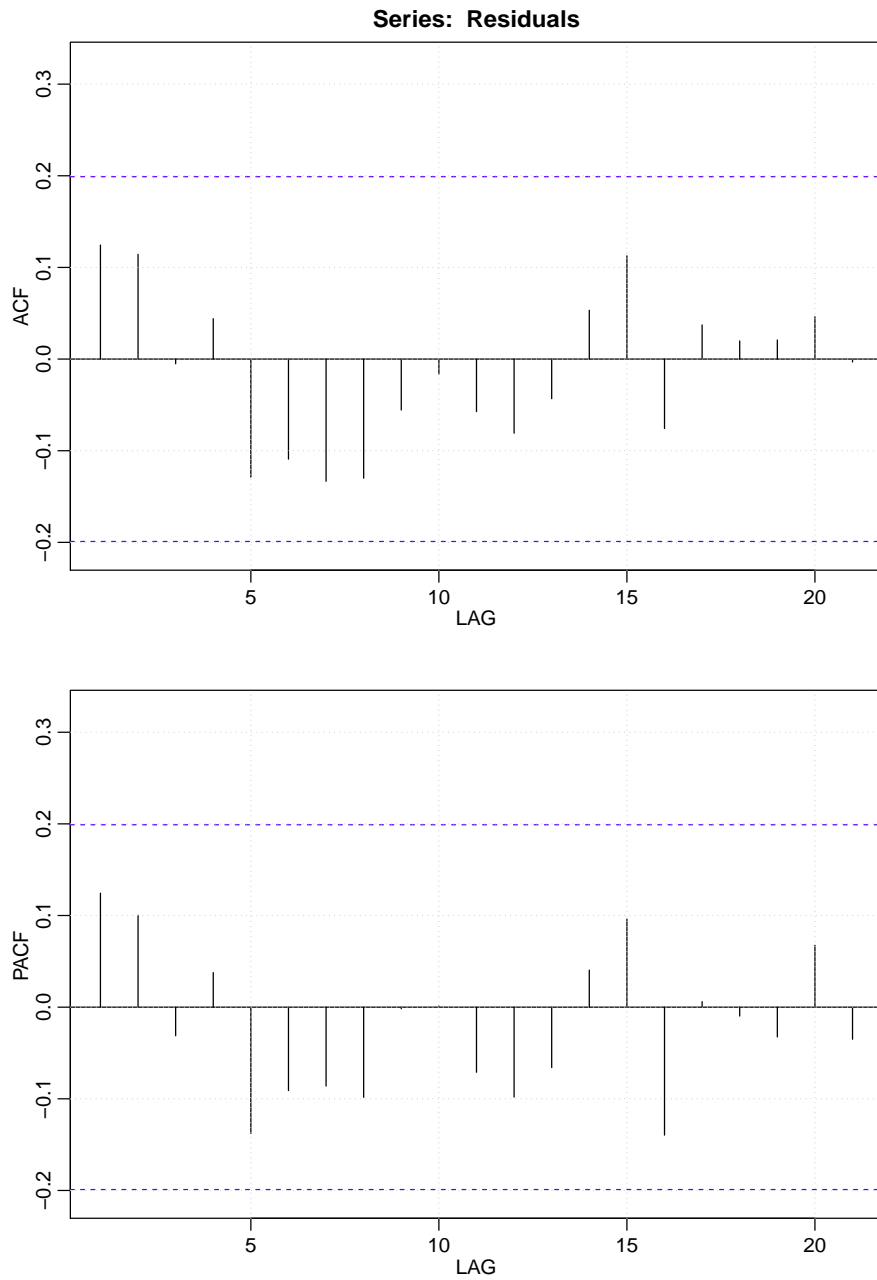
Figure 7: Autocorrelation function (ACF) and partial autocorrelation function (PACF) of the residuals of the fitted model (3.3). In each plot, bars outside the dotted bands would indicate the existence of autocorrelation.
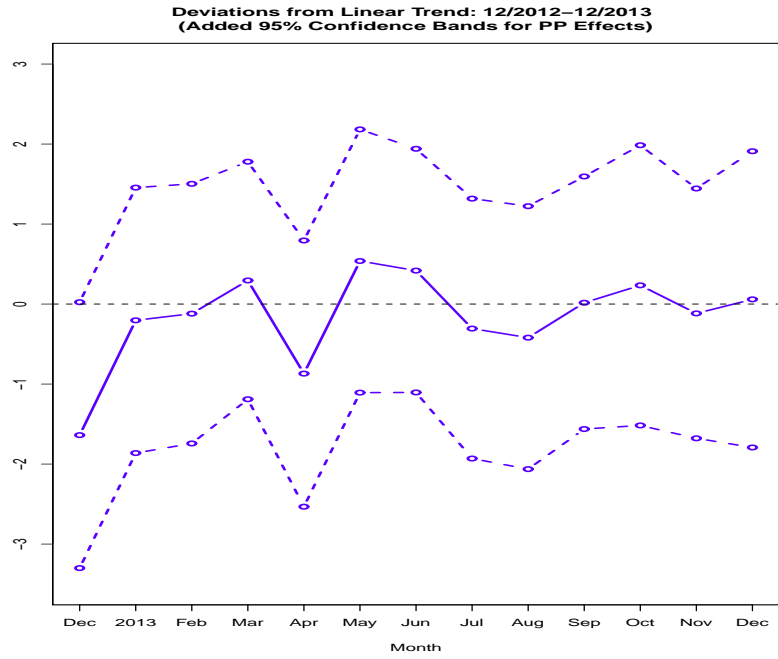
Figure 8: Deviations of observed prevalence from fitted linear time trend. Pointwise 95% confidence intervals for these estimated plain packaging effects have been added.
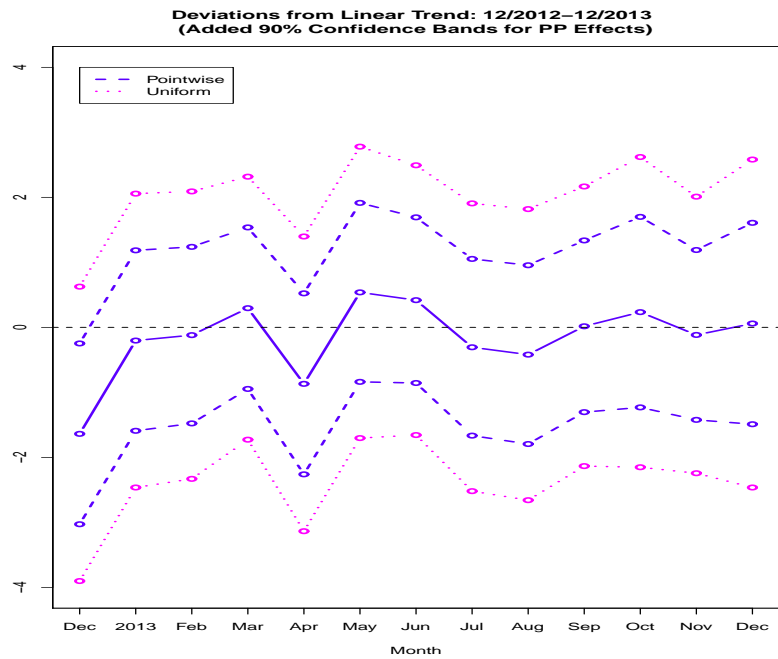

Figure 9: Deviations of observed prevalence from fitted linear time trend. Pointwise and uniform 90% confidence intervals for these estimated plain packaging effects have been added.

(Intercept)	25.230***
	(0.250)
t (Month)	-0.045***
	(0.003)
R^2	0.75
Adjusted R^2	0.75
Sample size	101
Degrees of freedom	99

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

Table 1: Regression output for the fitted model (3.3). The numbers in parentheses below the estimated coefficients are corresponding standard errors.

Effect Δ	IM-1	IM-2	IM-3
0.25	0.20	0.64	0.67
0.50	0.45	0.82	0.85
0.75	0.72	0.93	0.96
1.00	0.91	0.98	0.99

Table 2: Power against a permanent plain packaging effect Δ over the period 01/2013–12/2013. The inference methods IM-1, IM-2, and IM-3 are detailed in Algorithm 3.2. All numbers are based on $B = 50,000$ Monte Carlo repetitions in Algorithm 3.4.