



**University of
Zurich** ^{UZH}

University of Zurich
Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 149

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

Ashok Kaul and Michael Wolf

Revised version, May 2014

The (Possible) Effect of Plain Packaging on the Smoking Prevalence of Minors 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

First version: March 2014

This version: May 2014

Abstract

A key stated objective of the Australian Plain Packaging Act 2011 is to influence smoking prevalence, in particular of minors. We use the Roy Morgan Single Source (Australia) data set on minors, (that is, Australians aged 14 to 17 years) over the time period January 2001 to December 2013 to analyze whether there is evidence that this goal has been achieved. We carry out a statistical trend analysis to study the (possible) effect of plain packaging on smoking prevalence of minors in Australia. More specifically, we fit a linear time trend that explains well the fact that observed smoking prevalence has declined steadily over the last 13 years. Two informative analyses help to draw conclusions on the (actual) effect of plain packaging on smoking prevalence of Australian minors. 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 (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 Australians aged 14 to 17 years. Several reasonable variations to our methodology are discussed. All of these would only result in findings even more indicative of an absence of any plain packaging effect.

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

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

*Philip Morris International provided the funding for this research. At no time did we provide Philip Morris International with access to the underlying data.

[†]Saarland University, Department of Economics, Campus Building C3.1, D-66123 Saarbrücken, Germany. Email: a.kaul@ipe-saarland.de.

[‡]University of Zurich, Department of Economics, Wilfriedstrasse 6, CH-8032 Zurich, Switzerland. Email: michael.wolf@econ.uzh.ch. Michael Wolf is corresponding author.

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, specifically with a focus on preventing youth smoking. 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 not a single research paper that empirically links the introduction of plain packaging in Australia to changes in smoking prevalence in Australia. 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 cigarette market 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 (possible) effects of plain packaging.

There is a vast literature on the *expected* effects of plain packaging. In laboratory experiments like those presented by [Bansal-Travers et al. \(2011\)](#) or [Gallopel-Morvan et al. \(2013\)](#), the authors argue that plain packages are less attractive to customers and more likely to induce customers to think about health risks linked to smoking. In a similar vein, [Beede and Lawson \(1992\)](#) report that adolescents are more likely to recall health warnings when these are presented on plain instead of branded packages.

Given the unprecedented nature of the intervention, no one could know for sure what the intervention would lead to. A thorough inspection of the statistical evidence is thus called for. As pointed out in the *Handbook of Cancer Prevention* published by the World Health Organization, “As tobacco control policies are formulated and implemented, it is important that they undergo rigorous evaluation.”¹ In the section on causal interpretation of findings, the authors write that “we need to not only consider the size and nature of effects, we also need to consider the possibility that there is no meaningful effect. [...] We recognize that science cannot prove the null hypothesis, but it can and should make statements about interventions where there is a consistent failure to find evidence of any meaningful effect.”² In line with these statements, a scientific study of the *actual* effects versus the expected effects of such a drastic measure is needed.

On the side of expectations, Professor David Hill, then director of Cancer Council Victoria, argued in 2011 that “Plain packaging will slash smoking rates—and cigarette makers know it.”³

¹[IARC Handbooks of Cancer Prevention \(2008, p. x\)](#).

²[IARC Handbooks of Cancer Prevention \(2008, p. 29\)](#).

³www.smh.com.au/federal-politics/political-opinion/tobacco-industry-has-much-to-fear-20110407-1d63x.html, accessed most recently on May 12, 2014..

In an important contribution, [Pechey et al. \(2013\)](#) run a survey on 34 international experts, asking them about their expectations of the effect of plain packaging on smoking prevalence rates two years after its implementation. 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. Given the data at hand, the two effects cannot be disentangled by statistical methods. Assuming both effects work in the same way, the Australian case should therefore show a *bigger* reaction than what is expected based on plain packaging alone. The experts provided estimates for adults and children trying out smoking, which is not the same as youth prevalence. The median estimated effect on adult smoking prevalence was -1.0 percentage point, and the median estimated effect on children trying smoking was -3.0 percentage points. Taking the expected reaction for adults as a lower bound, we should expect to find at least a drop of one percentage point — if the expert opinions are correct predictors of what to expect.

Focusing on actual effects, [Young et al. \(2014\)](#) present a time series analysis on the smoking cessation helpline “Quitline”. They show that calls to “Quitline” increased significantly after 1st of October 2012 (when first plain packages appeared on the market). Interestingly, the authors do not comment on a drop in calls which seems to start in November and continues in December 2012; see Figure 2 in [Young et al. \(2014\)](#).

As a reaction to an earlier version of this paper, [Laverty et al. \(2014\)](#) state that “in view of the short time span since the measure was introduced, the variability in the measure, and the small sample size” failing to find any evidence for a plain packaging effect “is neither an unexpected nor a meaningful conclusion.” Based on reasoning that is not explained in (sufficient) detail, they further claim that a reduction of 1.25 percentage points “would be required to be statistically significant using this analysis.” However, this claim is unjustified, since our approach actually allows to identify an effect much smaller than 1.25 percentage points with reasonable power already.

In this paper, we apply a straightforward trend analysis to monthly aggregated prevalence rates to study whether plain packaging has changed the pre-implementation trend in smoking prevalence of Australian minors. In particular, we address the criticism of [Laverty et al. \(2014\)](#) by adding a formal power analysis in Section [3.3](#).

2 Data Description and Construction

We use the Roy Morgan Single Source (Australia) data set on minors, that is, Australians aged 14 to 17 years over the time period January 2001 to December 2013. The total sample size over this 13-year period is 41,438. The average annual sample size is 3,187.

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 and sample about 54,000 Australians per year; see [Roy Morgan Research \(2012\)](#).

In each month, we compute (observed) smoking prevalence as the average of the 0-1 variable `smoker` that indicates whether an individual in the sample smokes. Note that there is considerable variation in the sample size over time; see [Figure 1](#). Since the monthly sample sizes are rather small, ranging mostly between 200 and 350, and since the minors included in the sample change from month to month, it is expected that the monthly observed prevalence is rather unstable over time. This is indeed the case; see [Figure 2](#). Nevertheless, as the overall and the annual sample sizes are big, a satisfactory trend analysis is possible.

3 Data Analysis

3.1 Fitting a Linear Time Trend

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

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

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

$$\hat{p}_t = 11.47 - 0.037 \cdot t . \tag{3.2}$$

(A more detailed regression output can be found in [Table 1](#).) This model implies an average yearly decline of $12 \cdot 0.037 \approx 0.44$ percentage points in smoking prevalence of Australian minors over the 2001 to 2013 period; see [Figure 2](#) for a graphical display. It can be seen that the linear trend generally fits the observed data well. The *individual* deviations of the observed data from the fitted line are typically quite large, which is unavoidable given the unstable development of observed smoking prevalence over time (for the reasons discussed above). *Globally*, however, the fit of such a simple linear time trend is surprisingly good, given the long time period and the numerous regulatory changes in tobacco control policies over this period.

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 December 2012 until December 2013 are displayed in [Figure 3](#). Of the 13 deviations, seven are negative and six

⁴Since 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).

are positive. The average deviation is -0.41 percentage points. A naïve interpretation would be that, on average, plain packaging has resulted in a monthly reduction in prevalence of 0.41 percentage points.

However, one must take into account that the observed prevalence numbers are only estimates themselves, based on small sample sizes and on constantly changing samples (in terms of which minors enter the sample) over time. 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 December 2011 rather than in December 2012, then the twelve numbers pre December 2012 are almost a mirror image of the twelve numbers post December 2012 (that is, the year following the intervention month); see Figure 4. Indeed, if the intervention month December 2012 were not indicated by a vertical **line**, it would seem impossible to infer it by looking at this plot. There is also very little difference in terms of the average deviations. More precisely, the average deviation in the year pre December 2012 is -0.60 percentage points while the average deviation in the year post December 2012 (excluding December 2012 itself) is -0.40 percentage points.⁵ So, if anything, the ‘effect’ of plain packaging was larger in the year pre December 2012 than in the year post December 2012.

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 3. For a given month, this can be achieved as follows:

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 of minors (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 of minors (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.

⁵Including or excluding December 2012 does not alter any of the conclusions.

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) that plain packaging has led to a reduction in prevalence.

The results are displayed in Figure 5. It can be seen that there is no statistical significance for a plain packaging effect whatsoever: for all 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. All of these methods would only result in findings even more indicative of an absence of any plain packaging affect, as they would result in wider confidence intervals.⁶

- 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 (which will all contain zero as well then, of course).
- 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 (3.1). An analysis of the residuals⁷ of the fitted model (3.2) indicates that the error terms have a distribution that is (slightly) skewed to the right and heavy tailed.⁸ It is possible in step 1. to use a more refined (and more computationally involved) bootstrap approach to compute prediction intervals that also incorporate such skewness and excess kurtosis. Again, this would make the final confidence intervals wider.
- The standard textbook methodology for the prediction intervals in step 1. of Algorithm 3.1 also assumes that the error terms ε_t in the linear model (3.1) are independent and identically distributed (i.i.d.). This assumption might be violated in our application, since the data is collected over time and so the error terms might be autocorrelated. 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 (3.2) (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 (possi-

⁶These variations could therefore be called more *conservative* approaches, while our method could be called more *liberal* in the sense that it makes it more likely to ‘find’ a plain packaging effect in the data when in fact no such effect exists in the population.

⁷The residuals $\hat{\varepsilon}_t$ are computed as $\hat{\varepsilon}_t = p_t - \hat{p}_t$ ($t = 1, \dots, 156$).

⁸Both the sample skewness and the sample excess kurtosis are equal to 0.42. For a graphical display, see Figure 6.

ble) 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 at all. Third, an analysis of the residuals of the fitted model (3.2) does not show any autocorrelation whatsoever; see Figure 7.

- 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.⁹ Again, this would result in wider intervals.

3.3 Power Analysis

As mentioned in Section 2, the monthly sample sizes range mostly between 200 and 350 only. As a result, monthly observed prevalence is rather unstable over time and the deviations from the fitted trend line (3.2) are typically quite large. This might raise the concern of whether our trend analysis has any reasonable power at all against a possible plain packaging effect. For example, Laverly et al. (2014) state that “in view of the short time span since the measure was introduced, the variability in the measure, and the small sample size” failing to find any evidence for a plain packaging effect “is neither an unexpected nor a meaningful conclusion.”¹⁰

We address this concern by carrying out a formal power analysis. In particular, we consider the following inference method to test for a plain packaging effect which is consistent with our previous analyses.

Algorithm 3.2 (Inference Method)

1. Fit a linear time trend (using weighted least squares).
2. Compare the average deviation pre December 2012 to the average deviation post December 2012, as done in Section 3.2.2. If the average deviation post December 2012 is smaller than the average deviation pre December 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).¹¹ If the *t*-test rejects the null hypothesis, this is considered evidence for a plain packaging effect.
3. Compute individual 90% confidence intervals for plain packaging effects from December 2012 until December 2013, as detailed in Section 3.2.3. If at least one of the resulting 13

⁹Doing so prevents *data mining* or *cherry picking* by searching for *any* effect over the 13 months under consideration.

¹⁰This statement is based on a previous version of this paper that did not contain a power analysis.

¹¹There was no need to carry out such a *t*-test in Section 3.2.2, since the average deviation post December 2012 was larger than the average deviation pre December 2012.

confidence intervals is entirely negative, this is considered evidence for a plain packaging effect.

4. Overall, evidence for a plain packaging effect is established if at least of these two approaches, 1. or 2., finds evidence.

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 in the model (3.1) do not have the same variance. Denote the sample size in month t by n_t ($t = 1, \dots, 156$). Then we may assume

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

The fitted model (3.2) yields the estimator $\hat{\sigma}^2 = 862.01$.

We next detail how we generate pseudo prevalence data according a model that is in agreement with the observed data but has a specified plain packaging effect $\Delta > 0$ ‘enforced’ from December 2012 on.¹²

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

1. Generate $\gamma_1^*, \dots, \gamma_{156}^*$ independent and identically distributed as $N(0, 862.01)$, where the notation $N(0, \sigma^2)$ denotes a normal distribution with mean zero and variance σ^2 .
2. For $t = 1, \dots, 156$, let

$$p_t^* = 11.47 - 0.037 \cdot t + \varepsilon_t^* \quad \text{where} \quad \varepsilon_t^* = \frac{\gamma_t^*}{\sqrt{n_t}} .$$

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

$$p_t^* = p_t^* - \Delta .$$

(Note that $t = 144$ corresponds to December 2012 and $t = 156$ corresponds to December 2013.)

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.

¹²So Δ is the (fraction of) percentage points by which plain packaging has lowered prevalence beyond the time trend.

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 generally quite high instead of unreasonably low. For example, power against a plain packaging effect of 0.5 percentage points is 0.64 and power against a plain packaging effect of 1.0 percentage point is 0.79. Power of 0.8 is a commonly accepted industry standard¹³, so even the power against a plain packaging effect of only 0.5 percentage points is not unreasonably low.

3.4 Robustness Check

Imagine two different situations. First, a situation where there is no (noticeable) plain packaging effect. Second, a situation where there is a (strong) plain packaging effect. Then fitting a linear time trend would give a different result in the second situation, with a less steep linear trend. This might affect the analyses carried out previously.

Judging from Figure 2, we are not in such a scenario. Nevertheless we also fit model (3.1) using only the first 143 months (that is, the data pre December 2012).¹⁴ The thus newly fitted model is given by

$$\hat{p}_t = 11.41 - 0.036 \cdot t \quad (3.3)$$

and is very close to the fitted model (3.2) based on all the data. Both linear trends are displayed in Figure 8. One can see that, at the end of period, the fitted trend line using the pre December 2012 data only lies only slightly above the fitted trend line using all the data. However, the difference is very small and all the previous analyses qualitatively still go through if they are based on the alternative trend line (3.3) instead.¹⁵

Remark 3.1 Fitting the trend line using the data until November 2012 will be somewhat more powerful (in case of an existing plain packaging effect). If we repeat the power analysis of Section 3.3 based on this alternative approach, power generally increases by 0.02–0.03 percentage points; see Table 3. But as explained before, even this (slightly) more powerful method fails to detect any evidence for a plain packaging effect.

4 Conclusion

We carried out a trend analysis to study the (possible) effect of plain packaging on smoking prevalence of Australians aged 14 to 17 years. More specifically, we fitted a linear time trend that explains well the fact that observed prevalence has declined steadily over the last 13 years

¹³For example, see Section V.G. of FDA (2008).

¹⁴For the reasons mentioned before, we again use weighted least squares (WLS) instead of ordinary least squares (OLS) to estimate the model.

¹⁵Since these alternative results are virtually identical, there is no need to include them as well.

(that is, from January 2001 until December 2013) at an annual rate of about 0.44 percentage points.

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 of Australian minors. 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.

Altogether, we have applied quite liberal inference techniques, that is, our analysis, if anything, is slightly biased in favor of finding a statistically significant (negative) effect of plain packaging on smoking prevalence of Australians aged 14 to 17 years. Nevertheless, no such evidence has been discovered. More conservative statistical inference methods would only reinforce this conclusion.

References

- Bansal-Travers, M., Hammond, D., Smith, P., and Cummings, K. M. (2011). The impact of cigarette pack design, descriptors, and warning labels on risk perception in the us. *American journal of preventive medicine*, 40(6):674–682.
- Beede, P. and Lawson, R. (1992). The effect of plain packages on the perception of cigarette health warnings. *Public health*, 106(4):315–322.
- 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>.
- Gallopel-Morvan, K., Gabriel, P., Le Gall-Ely, M., Rieunier, S., and Urien, B. (2013). Plain packaging and public health: The case of tobacco. *Journal of Business Research*, 66(1):133–136.
- IARC Handbooks of Cancer Prevention (2008). *Methods for evaluating tobacco control policies*. International Agency for Research on Cancer. IARC Working Group on Methods for Evaluating Tobacco Control Policies and International Agency for Research on Cancer and others.
- Laverty, A. A., Watt, H. C., Arnott, D., and Hopkinson, N. S. (2014). Standardised packaging and tobacco-industry-funded research (Correspondence). *The Lancet*, 383(9926):1384.

- 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>.
- Young, J. M., Stacey, I., Dobbins, T. A., Dunlop, S., Dessaix, A. L., and Currow, D. C. (2014). Association between tobacco plain packaging and quitline calls: a population-based, interrupted time-series analysis. *The Medical Journal of Australia*, 200(1):29–32.

A Figures and Tables

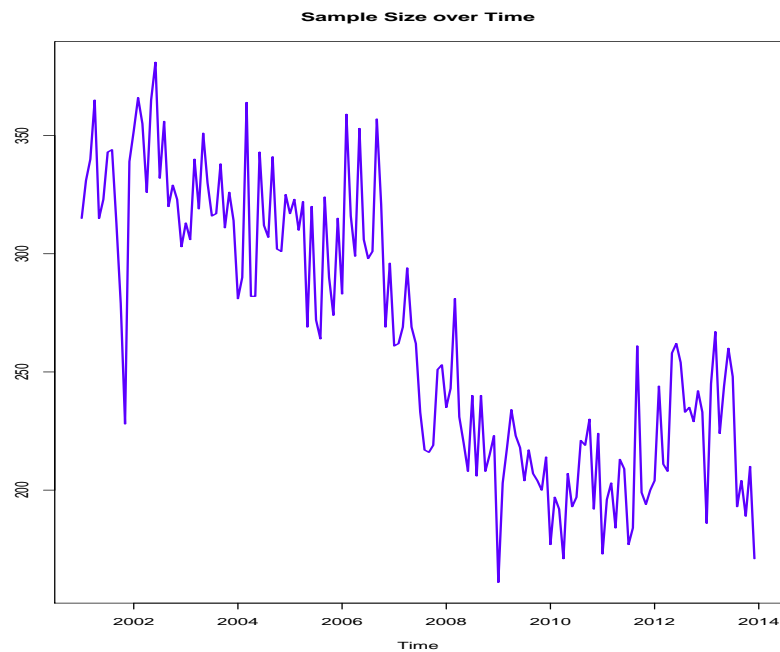


Figure 1: Time series plot of the monthly sample sizes.

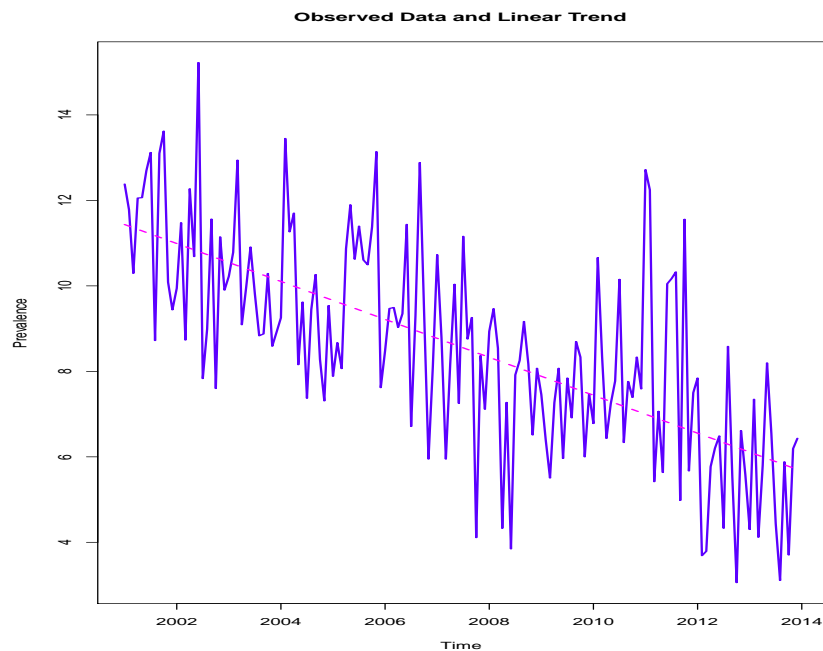


Figure 2: Time series plot of observed prevalence with fitted linear time trend.

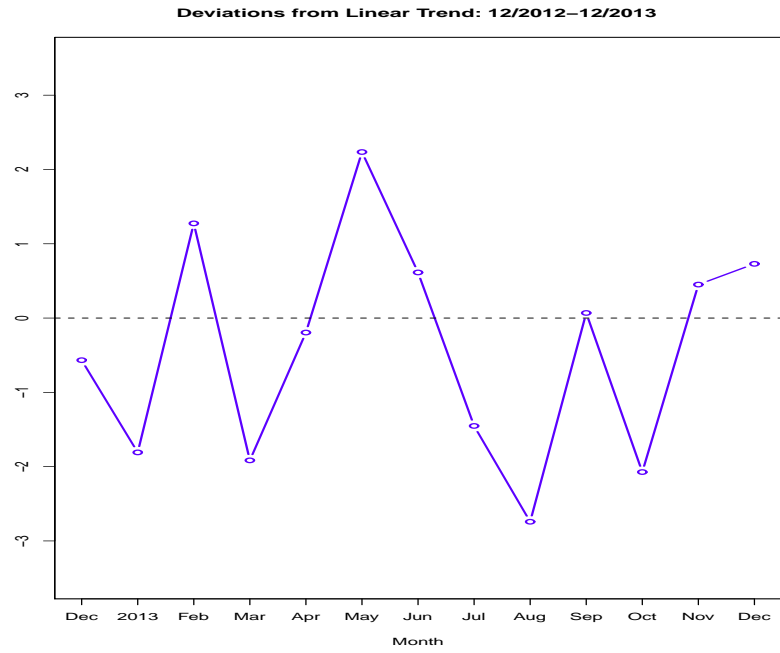


Figure 3: Deviations of observed prevalence from fitted linear time trend.

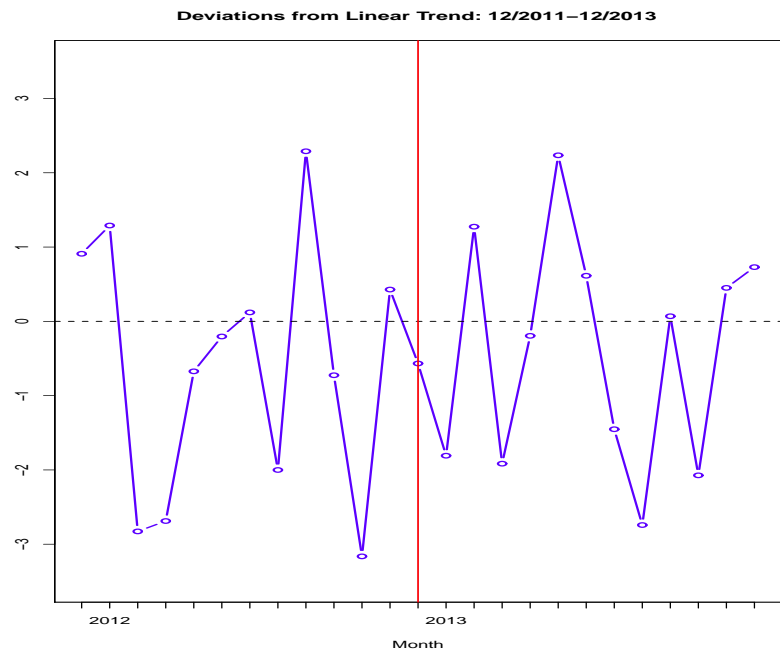


Figure 4: Deviations of observed prevalence from fitted linear time trend. The intervention month December 2012 is indicated by a vertical **line**.

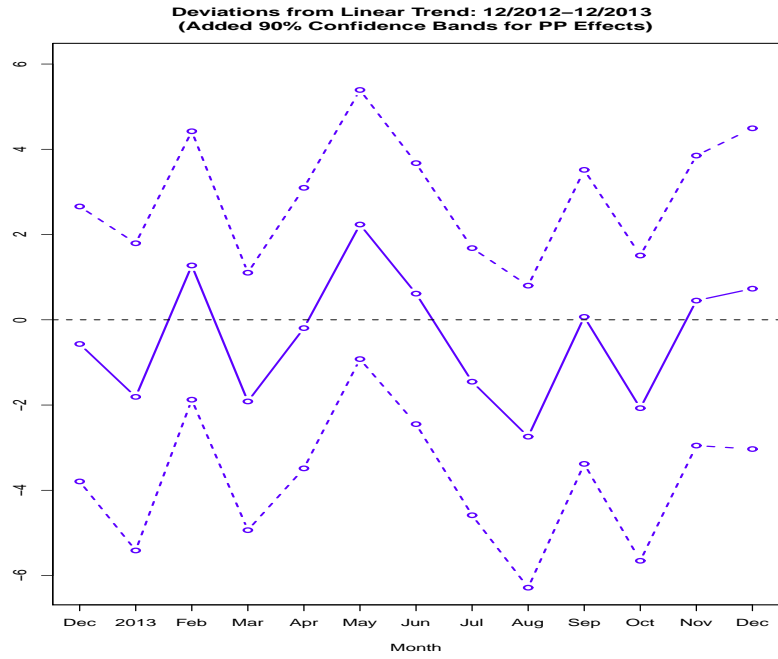


Figure 5: Deviations of observed prevalence from fitted linear time trend. Pointwise 90% confidence intervals for these estimated plain packaging effects have been added.

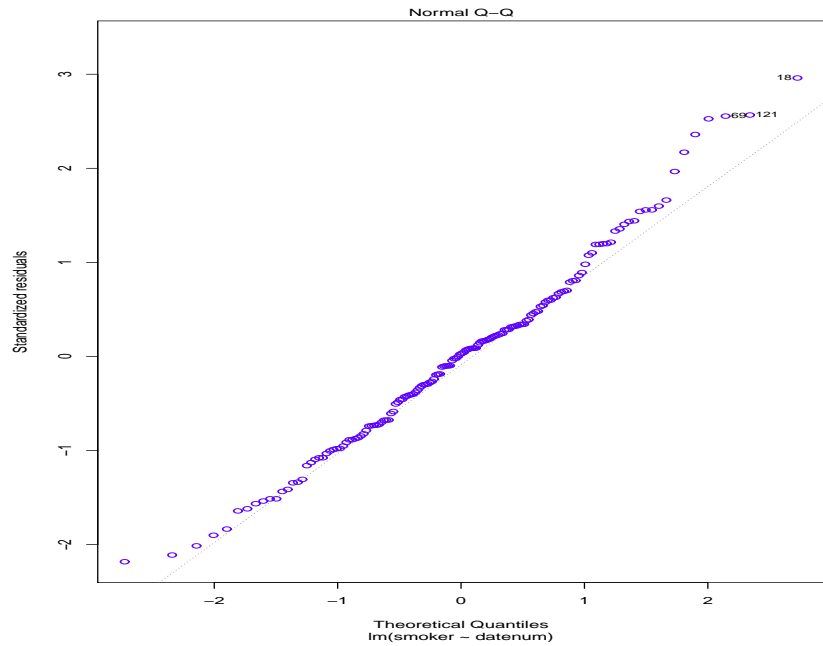


Figure 6: Normal Q-Q plot of the residuals of the fitted model (3.2). The pattern of the deviations of the circles from the straight line indicates a distribution that is (slightly) skewed to the right and heavy tailed.

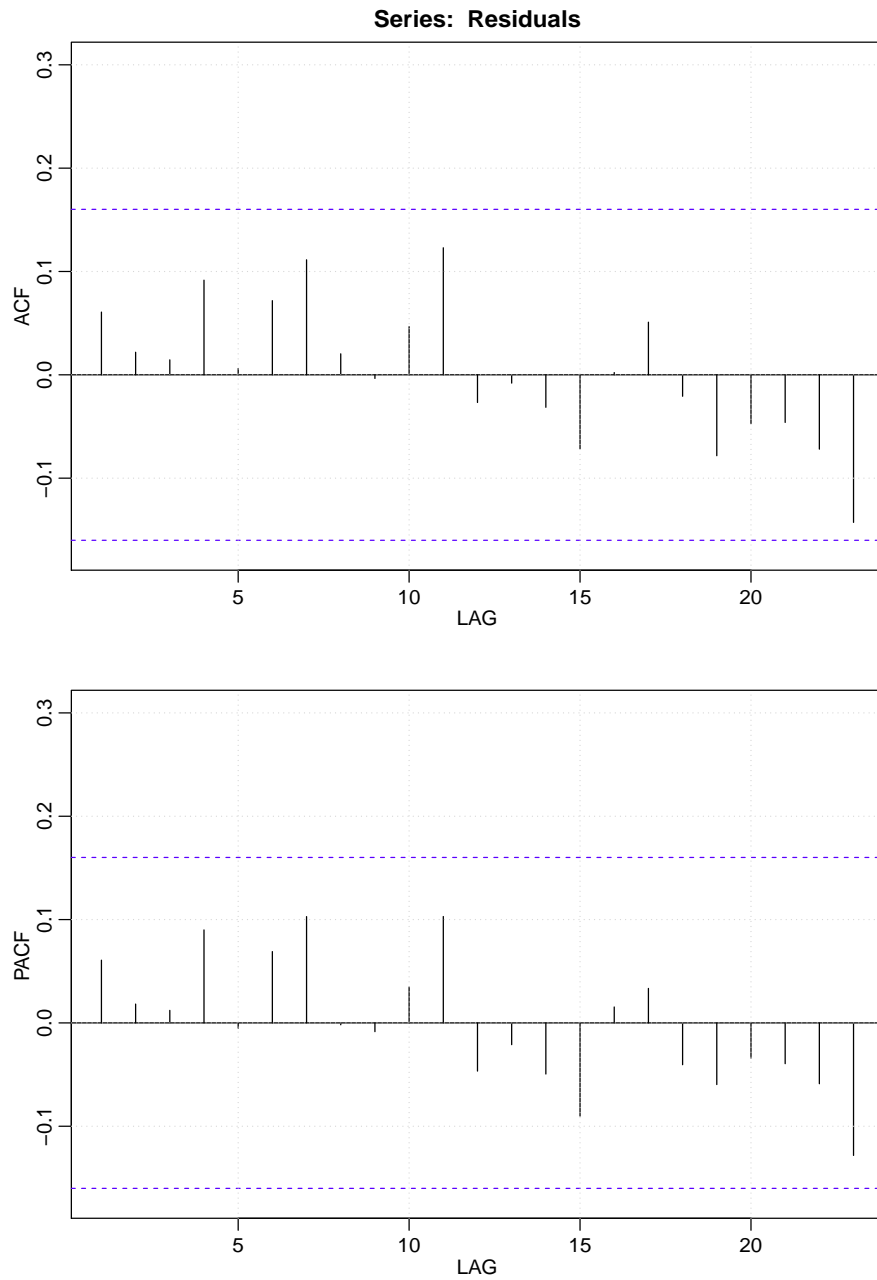


Figure 7: Autocorrelation function (ACF) and partial autocorrelation function (PACF) of the residuals of the fitted model (3.2). In each plot, bars outside the dotted bands would indicate the existence of autocorrelation.

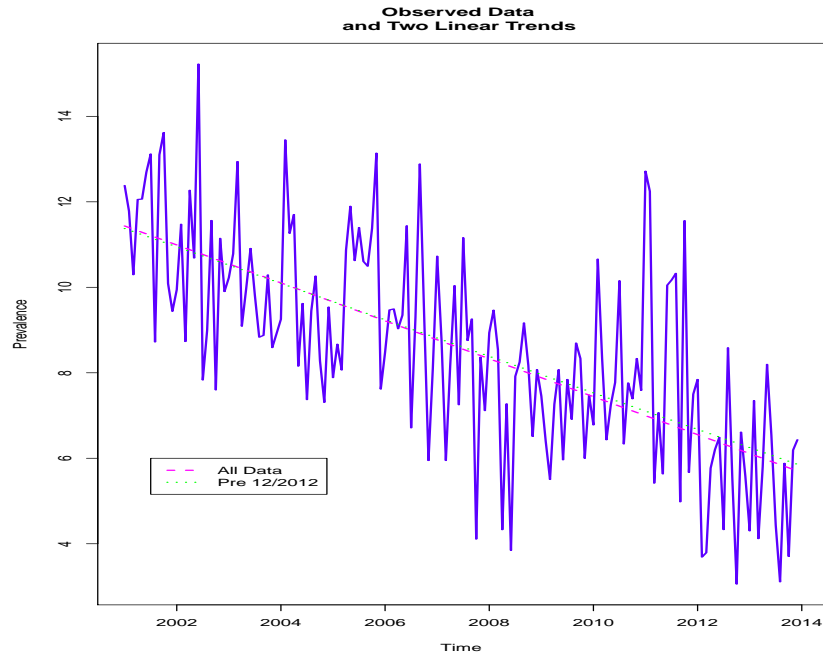


Figure 8: Time series plot of observed prevalence with two fitted linear trends.

(Intercept)	11.471*** (0.270)
t (Month)	-0.037*** (0.003)
R^2	0.46
Adjusted R^2	0.46
Sample size	156
Degrees of freedom	154

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

Table 1: Regression output for the fitted model (3.2). The numbers in parentheses below the estimated coefficients are corresponding standard errors.

Effect Δ	Power
0.25	0.56
0.50	0.64
0.75	0.72
1.00	0.79
1.25	0.85
1.50	0.90

Table 2: Power of the inference method detailed in Algorithm 3.2 against a plain packaging effect of size Δ . All numbers are based on $B = 50,000$ Monte Carlo repetitions in Algorithm 3.4.

Effect Δ	Power
0.25	0.58
0.50	0.67
0.75	0.75
1.00	0.83
1.25	0.88
1.50	0.93

Table 3: Power of the alternative inference method based on Section 3.4 against a plain packaging effect of size Δ ; see Remark 3.1. All numbers are based on $B = 50,000$ Monte Carlo repetitions.