Intervention analysis with time series data

FERSI Training Seminar on Evaluation
Bern, 19-21 May 2010

Jacques J.F. Commandeur

SWOV Leidschendam, Vrije Universiteit Amsterdam

Overview

- Part 1: the merits of *experimental designs* where we are allowed to *manipulate* circumstances
- Part 2: the opportunities and limitations presented in the much more realistic and common situation of *observational studies*, where we are afforded no such luxuries

Part 1: Experimental designs

- Even experimental designs have their advantages and drawbacks
- The relevance of these designs is that they should always be used as a *yardstick* against which to evaluate the merits of observational studies to be considered later in Part 2

The ideal experimental set-up

A randomized controlled trial (RCT):

		pre	inter-vention	post
treatment group	s1	17		8
	s2	22		2
	⋮	⋮		⋮
	s30	24		12
control group	s31	8		17
	s32	12		25
	⋮	⋮		⋮
	s60	23		26

where $17, 22, \ldots,$ are counts of accidents, casualties, etc.

The ideal experimental set-up

- the “subjects” or observation units are a *random sample* from the population (generalizability!)
- the “subjects” or observation units are *randomly assigned* to the treatment and control conditions
- the latter guarantees that the *only* difference between treatment and control group is the effect of the intervention (all other possible differences are identically distributed over the two groups)

The ideal experimental set-up

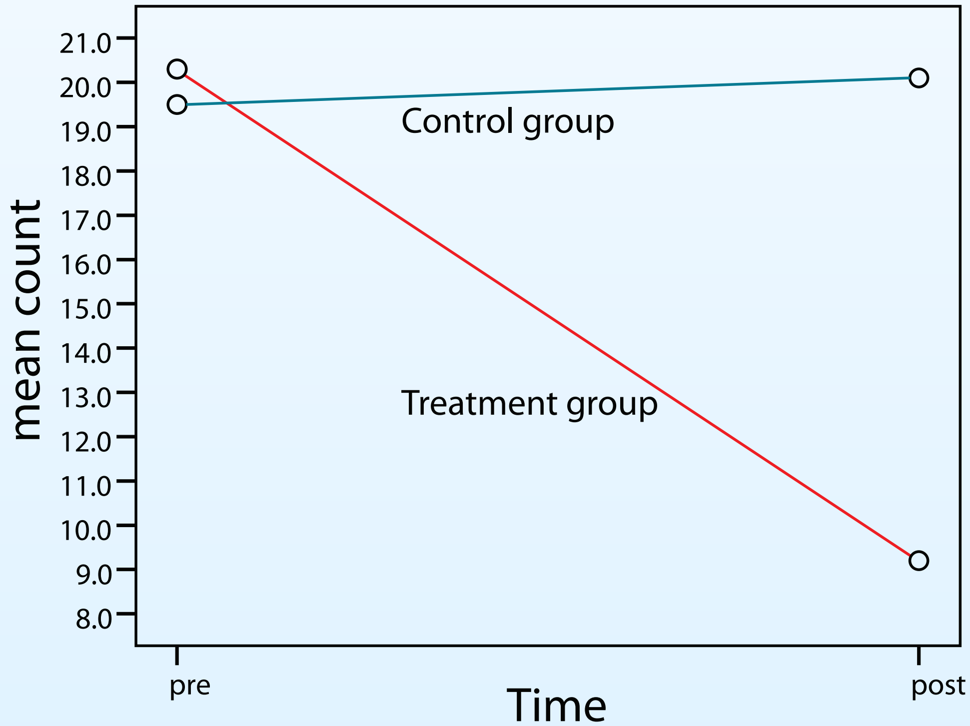
This ANOVA design allows for the evaluation of two main effects and one interaction effect:

- a main effect for Condition (treatment versus control group)
- a main effect for Time (pre- versus post-period)
- an interaction effect for Condition by Time.

We are of course only interested in the interaction effect!

The ideal experimental set-up

- The F-test for the interaction between Condition and Time is very significant ($p < 0.01$)
- Conclusion: the time change of the mean count in the treatment group is different from that in the control group
- Different in what sense?



The ideal experimental set-up

- Wonder of wonders: with this ideal experimental set-up we are even allowed to conclude that the intervention has *caused* the decrease in the number of counts!
- This in contrast with every other study discussed in my presentation

First simplification: no pre-test

A completely randomized design:

		intervention	post
	s1		8
treatment	s2		2
group	⋮		⋮
	s30		12
	s31		17
control	s32		25
group	⋮		⋮
	s60		26

First simplification: no pre-test

- The average counts in the treatment and control group are 9.2 and 20.1; the main effect for the factor Condition is very significant ($p < 0.01$)
- If no random assignment of subjects to treatment and control groups was achieved, the observed difference between the experimental conditions may not necessarily (or only partly) have been caused by the intervention
- Other unobserved differences between the two groups may well explain the observed differences in average counts (e.g., confounding variables, regression to the mean)
- Possible solution if these confounding variables are known and have been observed: statistical control with the analysis of covariance (ANCOVA)

Second simplification: no control group

A randomized block design:

		pre	inter-vention	post
	s1	17		8
treatment	s2	22		2
group	⋮	⋮		⋮
	s30	24		12

The average counts in the pre- and post-period are 20.3 and 9.2
The main effect for the factor Time is very significant ($p < 0.01$)

- Can we conclude that the change was *caused* by the intervention?
- Most definitely not!

Second simplification: no control group

Alternative explanations of the effect: the observed change was not (or only partly) caused by the intervention, but is (also) the result of

- a general decreasing trend (maturation)
- regression to the mean
- changes in registration (instrumentation)
- other external events affecting the count changes (confounding variables)

or any combination of these.

The worst case scenario: one “subject” only

		pre	inter-vention	post
treatment “group”	s1	17		8

- We find a more than 50% reduction in counts, but what does it mean?
- Many alternative explanations, one of them being: pure coincidence!

Part 2: Observational studies

- In practice we usually can not manipulate variables when evaluating the effects of safety measures
- Does this mean we should give up evaluation altogether?
- The answer is: certainly not!

The central idea in observational studies is to emulate experimental designs as closely as possible by

- controlling for confounding variables
- trying to find a “natural” control or reference group
- and if possible: both

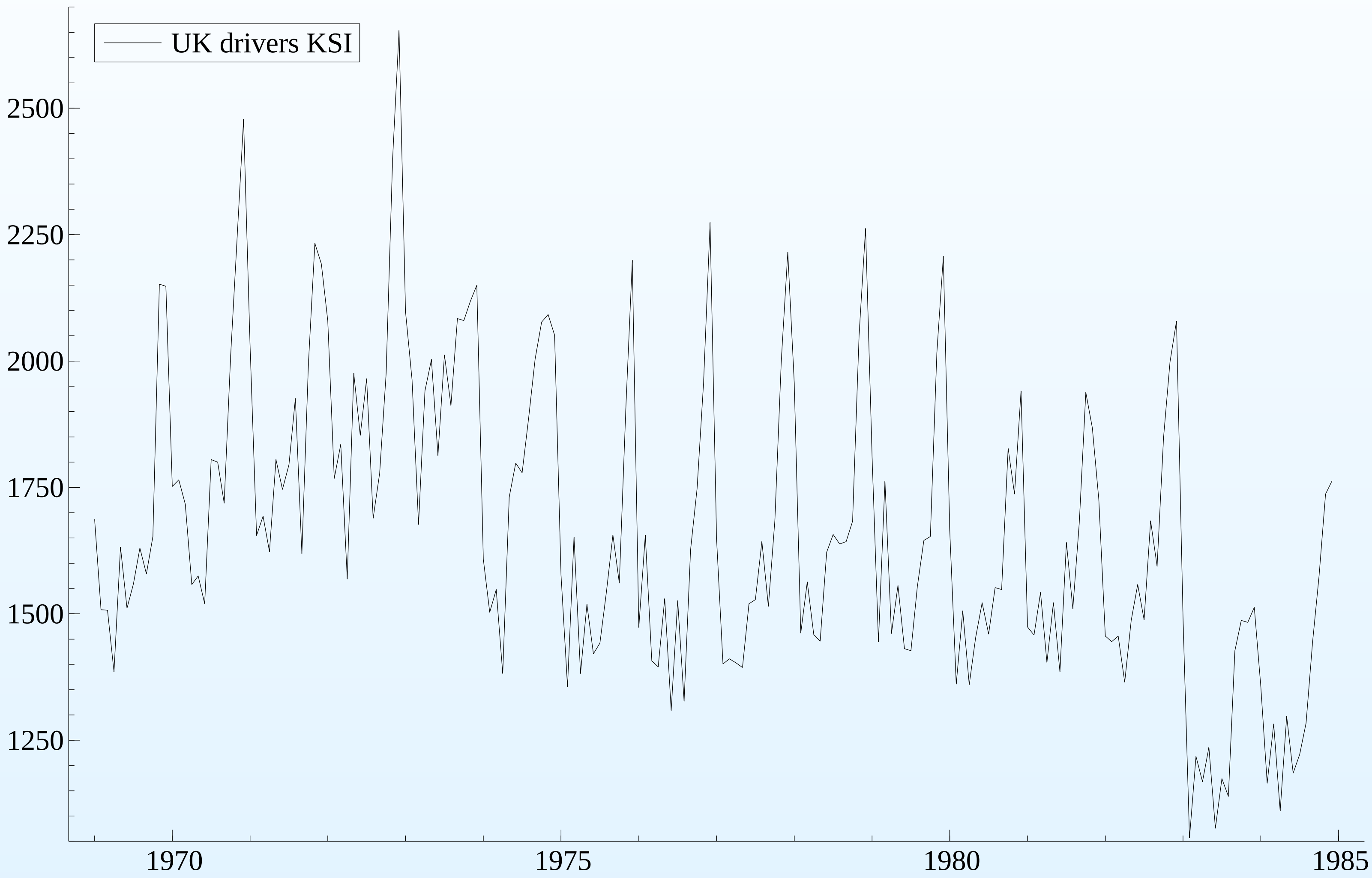
The effects of the seat belt law in the UK

- From 31 January 1983 onwards, the wearing of seat belts by car drivers and front seat passengers became mandatory
- By December 1982 the seat belt wearing rate was 40%
- By February 1983 the wearing rate had jumped to 90% and remained at approximately 95% from March 1983 onwards (Harvey $\&$ Durbin, 1986)

The effects of the seat belt

- The seat belt is supposed to affect the number of persons injured in car accidents, not the number of car accidents themselves
- Examples of confounding variables also affecting the injury severity of persons in car accidents are: speed, the weight of the vehicle, the type of car, the type of accident and the number of people in the car (Elvik & Vaa, 2004)

The UK car drivers KSI series



A very naive approach

UK drivers KSI:

	1982(2)-1983(1)	1983(2)-1984(1)	Total
treatment	19498	15335	34833
“control”	19498	19498	38996
Total	38996	34833	73829

For this contingency table we obtain $\chi^2 = 263.6$ ($p < 0.01$).
Conclusion: in the UK the seat belt law resulted in a $100(\frac{19498-15335}{19498}) = 21.35\%$ decrease in the number of car drivers KSI. Is this convincing?

Observational studies based on time series

The basic *descriptive* time series model is

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}(0, \sigma_\varepsilon^2)$$

for time points $t = 1, \ldots, n$, where μ_t is the trend in the series, γ_t captures the seasonal deviations from the trend, and ε_t are the residuals (unexplained part) of the model. The statement

$$\varepsilon_t \sim \mathrm{NID}(0, \sigma_\varepsilon^2)$$

is shorthand notation for the assumption that the residuals ε_t are normally and *independently* distributed with mean equal to zero and variance equal to σ_ε^2.

Deterministic approach to time series

- One way we could handle the trend μ_t is by regressing the dependent variable y_t on time $t = 1, \ldots, n$ itself:

$$\mu_t = a + bt,$$

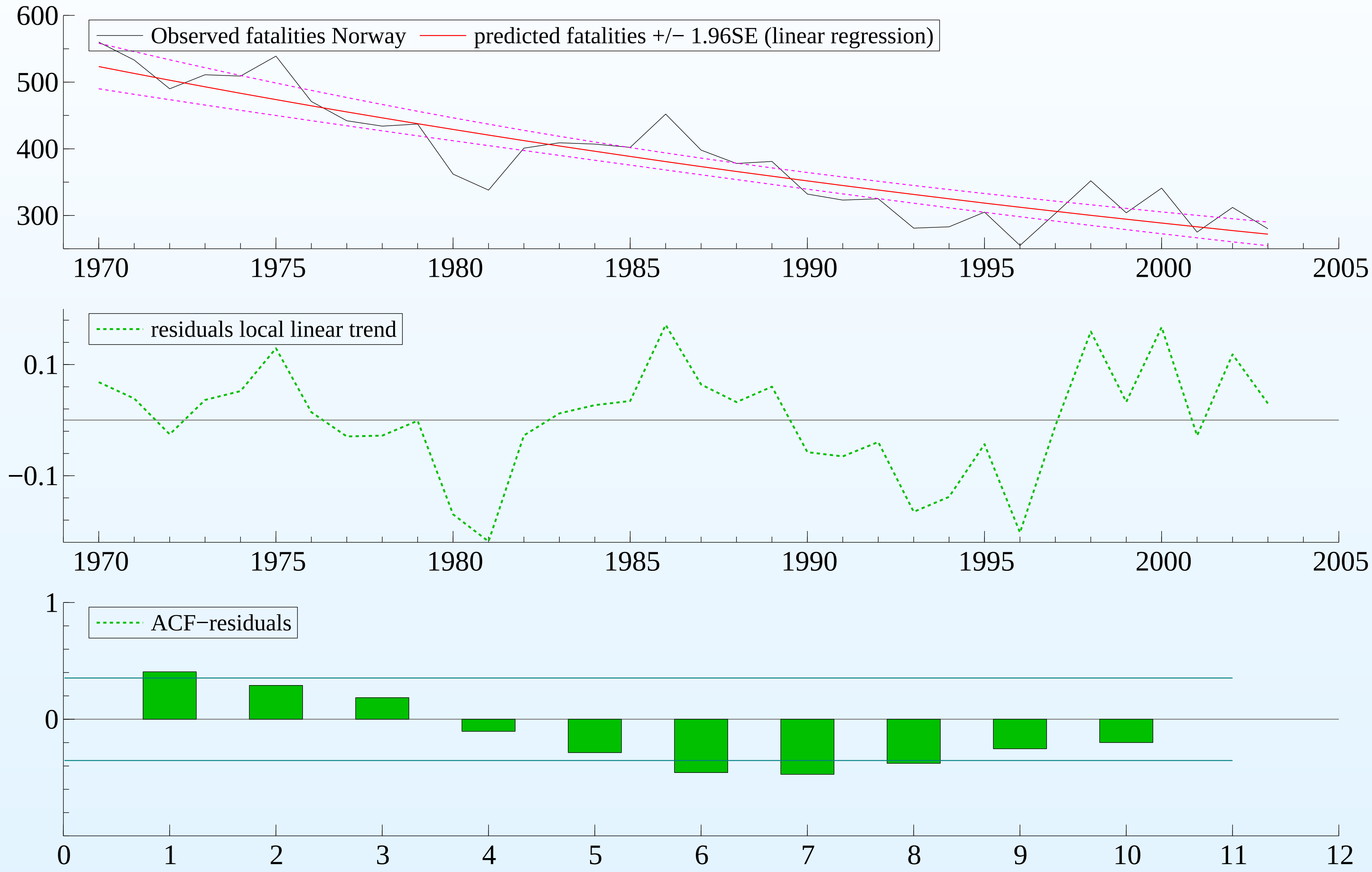
  where a is the intercept and b is a regression coefficient: the classical linear regression model

- Similarly, we could deal with the seasonal deviations γ_t from the trend by regressing the dependent variable y_t on $s - 1$ dummy variables, where s is the periodicity of the seasonal (in this case: 12)

Deterministic approach to time series

- With time series data this is generally not a good idea: the observations $y_1, y_2, \ldots, y_n$ of a time series are not independent, and the residuals of such a *deterministic* model will therefore not satisfy the assumption of independence either
- Dependent observations (residuals) are also said to be *serially correlated*
- Why is this so important?
- Serially correlated residuals result in standard errors that are either too small or too large
- Since standard errors are crucial for statistical tests of regression weights and the construction of confidence intervals, these tests and intervals will not reflect the true situation

Example of serial correlation



Stochastic approach to time series

- We adopt the structural time series framework of Harvey (1989) and Durbin and Koopman (2001) (see also Commandeur and Koopman (2007)) by allowing the intercept a and the regression weight b both to follow a *random walk*:

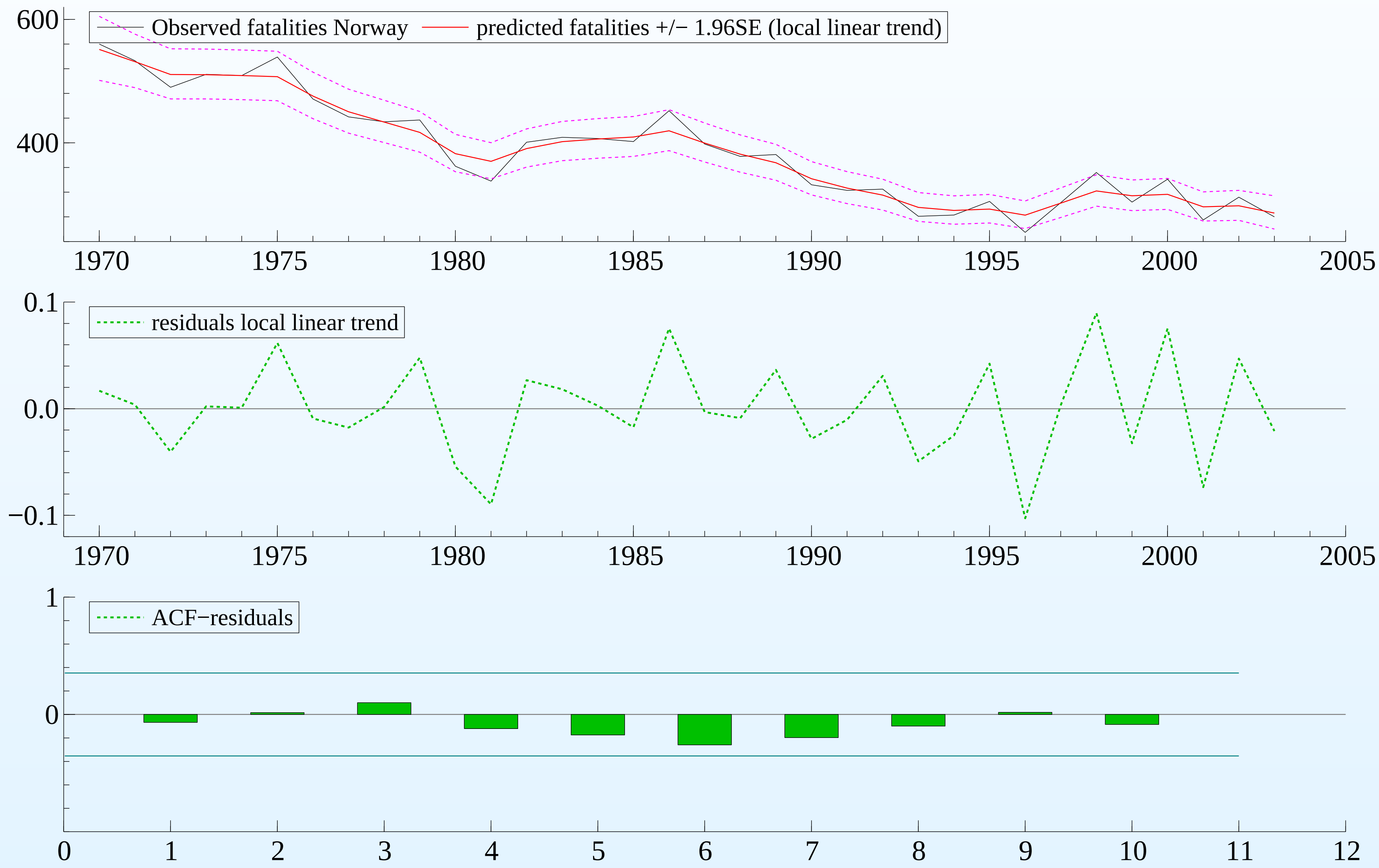
$$\begin{aligned} \mu_{t+1} &= \mu_t + b_t + \xi_t, & \xi_t &\sim \mathrm{NID}(0, \sigma^2_\xi) \\ b_{t+1} &= b_t + \zeta_t, & \zeta_t &\sim \mathrm{NID}(0, \sigma^2_\zeta) \end{aligned}$$

- The result is that the intercept μ_t and the regression coefficient b_t are allowed to vary over time, thus “absorbing” the serial correlation in the observed time series; it is then said that the intercept (level) and regression weight (slope) are treated *stochastically*

Stochastic approach

- The seasonal γ_t – if any is required – can also be subjected to a random walk
- By fixing σ^2_ξ and σ^2_ζ on zero we again obtain the classical (deterministic) linear regression model!

Example results of stochastic approach



Diagnostic tests

- The assumptions of independence, homoscedasticity, and normality of the model residuals can all be checked with formal diagnostic tests
- The formal test for independence (Box-Ljung) confirms that the residuals of the deterministic model are serially correlated, while those of the stochastic model are not

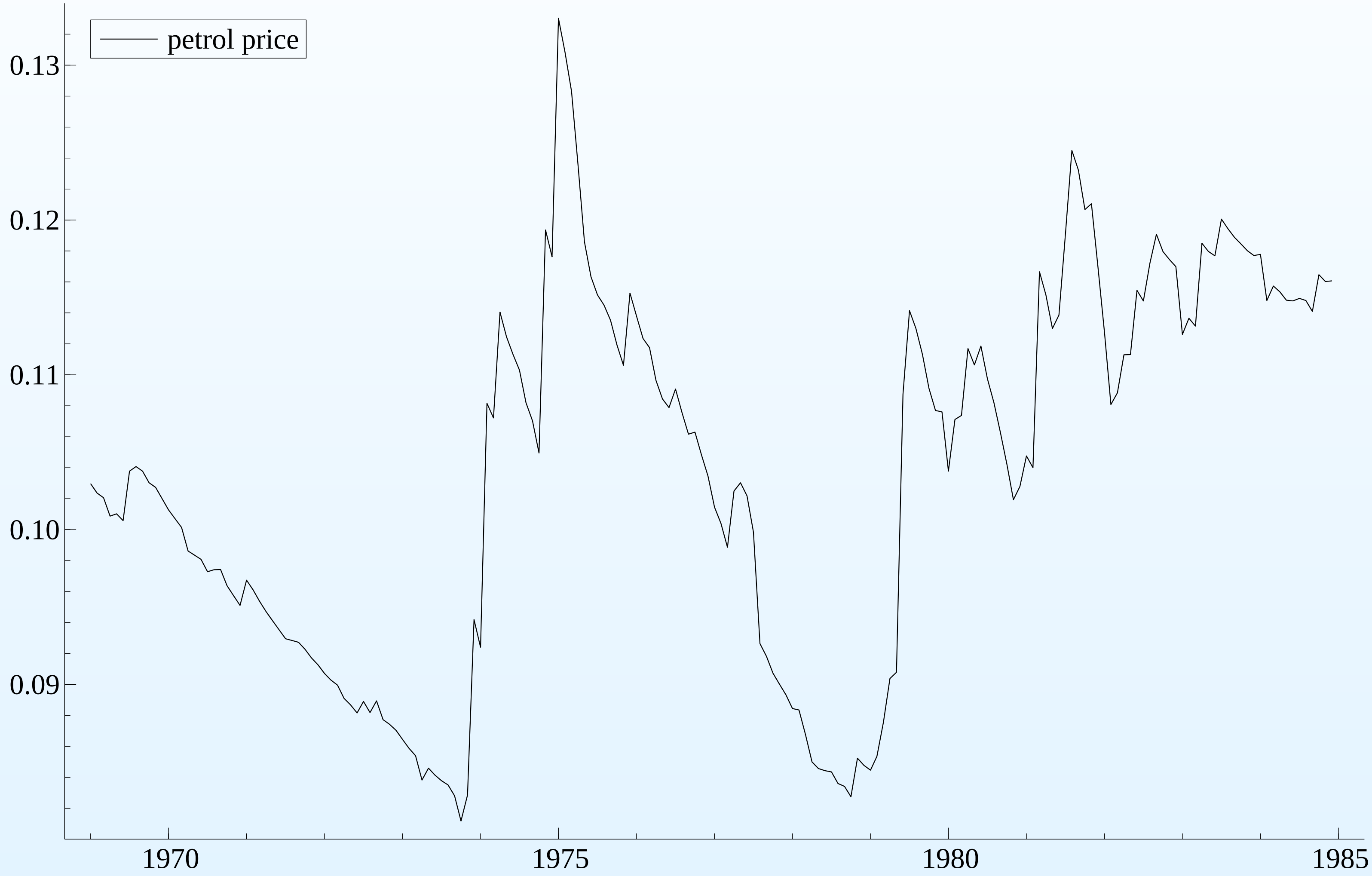
Observational studies based on time series

- The basic *explanatory* model is

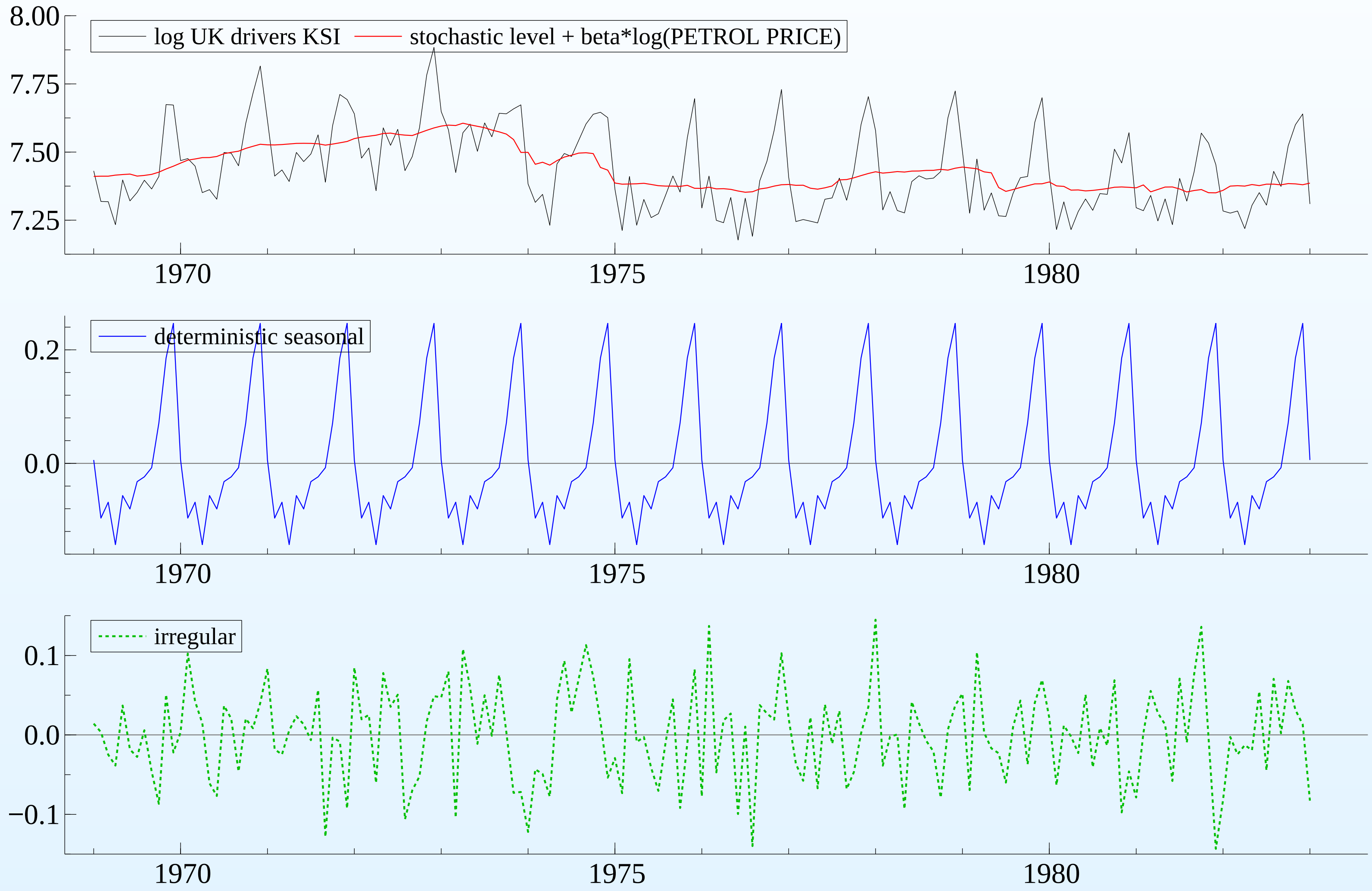
$$y_t = \mu_t + \gamma_t + \beta x_t + \varepsilon_t, \qquad \varepsilon_i \sim \mathrm{NID}(0, \sigma^2_\varepsilon)$$

  for time points $t = 1, \ldots, n$, where x_t is the logarithm of the monthly price of petrol, and β is an unknown regression coefficient
- We apply this model to the logarithm of the UK drivers KSI series, but only up to February 1983
- We then use the results of this model to *forecast* the series at and after February 1983; this provides an indication of what to expect for February 1983-December 1984 had the seat belt law *not* been introduced

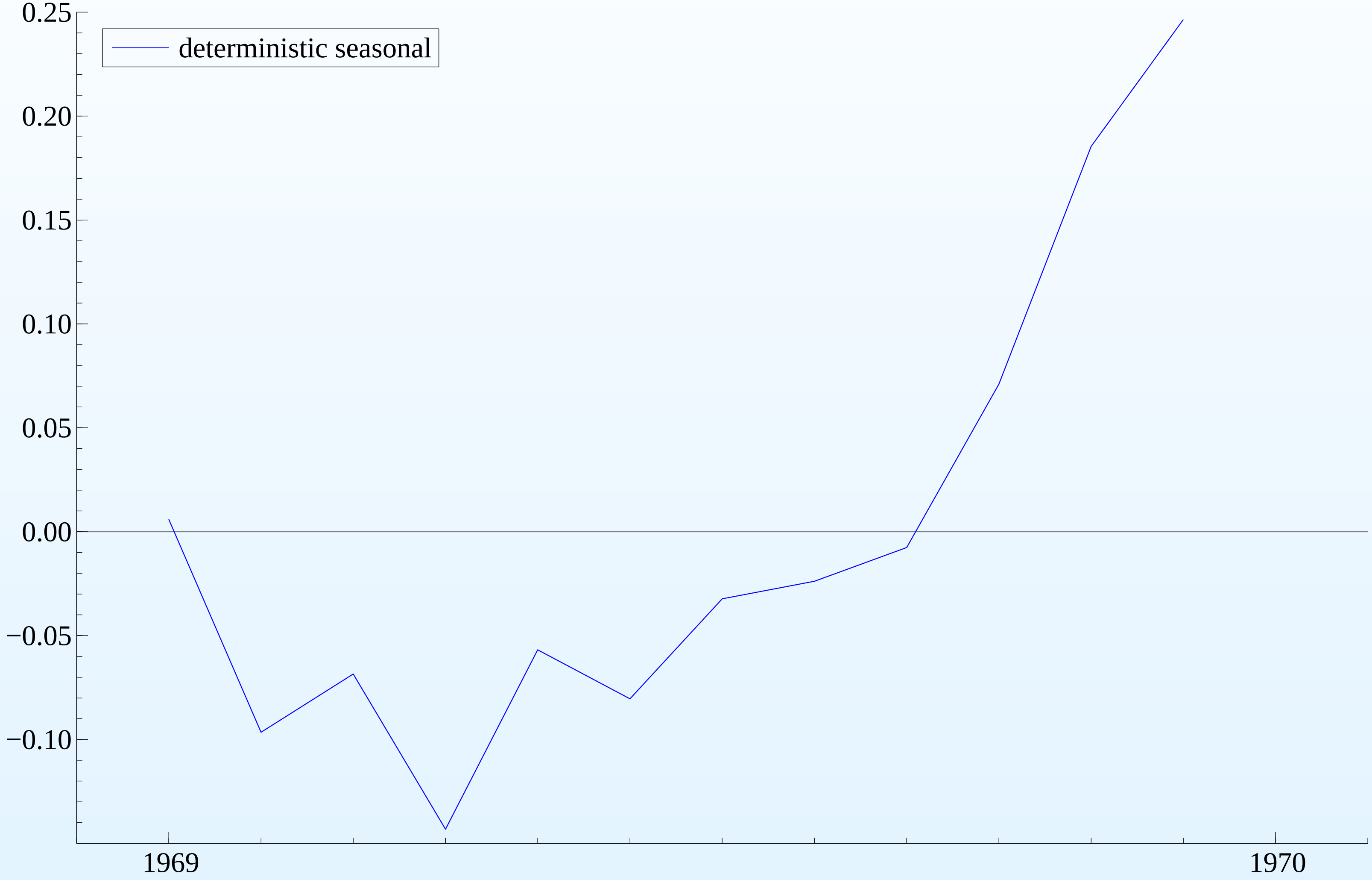
Explanatory variable petrol price



Up to and including January 1983: stochastic level



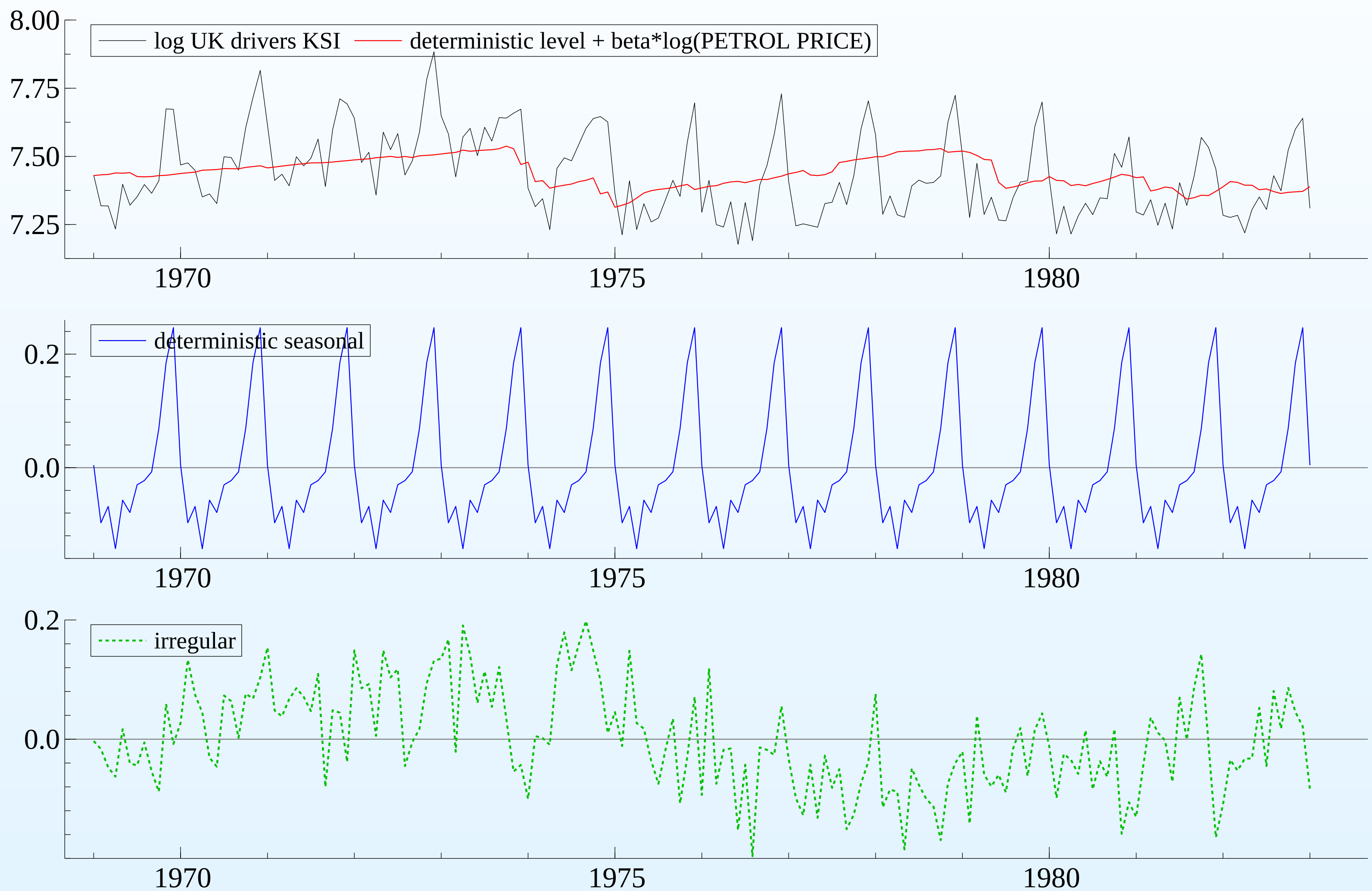
Zoom of the seasonal



Effect of log petrol price, stochastic model

- The regression weight for log petrol price is $\beta = -0.2921$
- Its standard error equals 0.0981 yielding a t-value of $-0.2921/0.0981 = -2.9780$, which is significant at the 5% level
- Since the time series for the UK drivers KSI and for petrol price are both in logs, a 1% increase in petrol price yields a 0.29% decrease in the number of drivers KSI

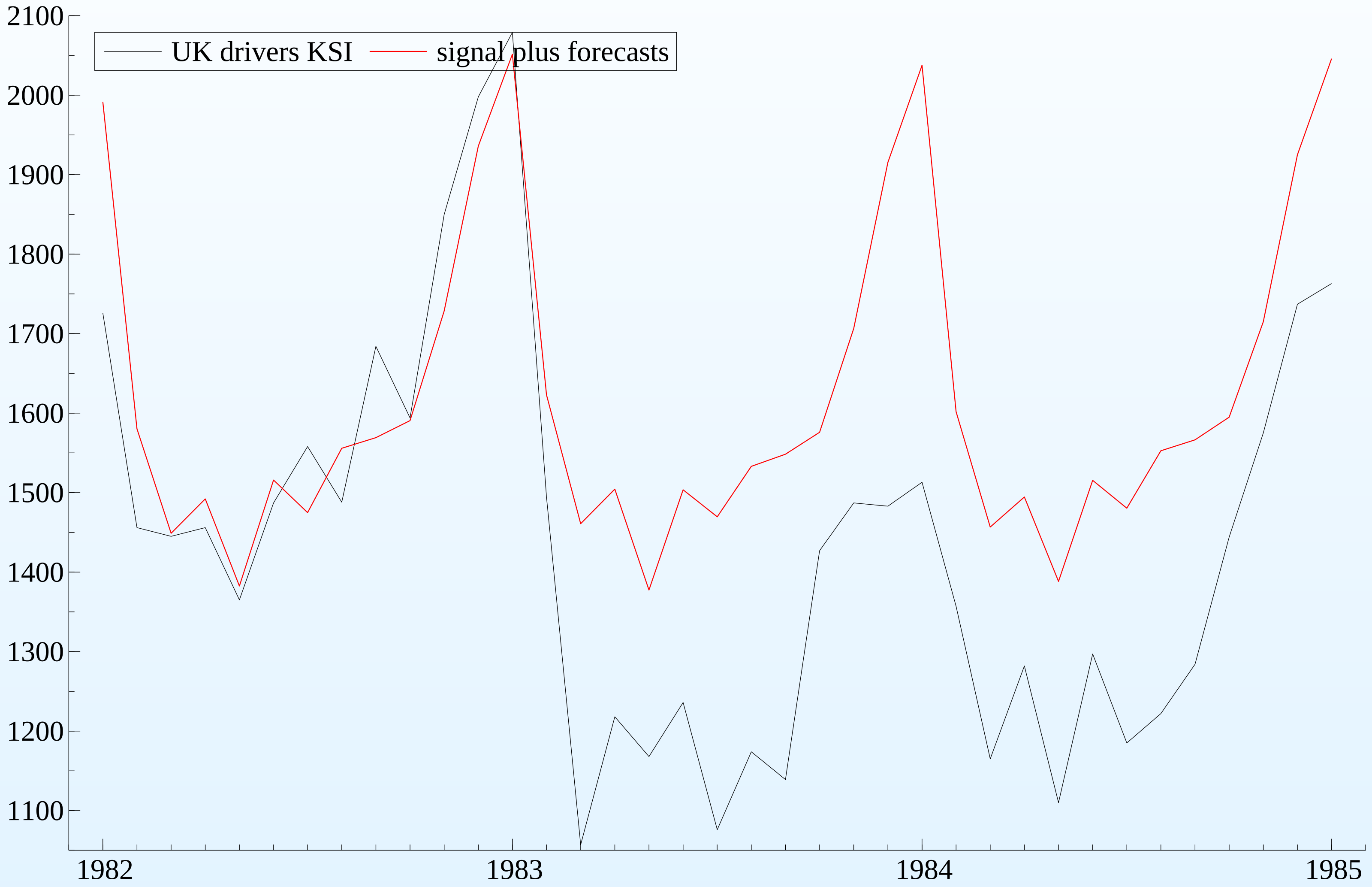
Up to and including January 1983: deterministic level



Effect of log petrol price, deterministic model

- The regression weight for log petrol price is now $\beta = -0.4533$
- Its standard error is 0.0578 yielding a t-value of $-0.4533/0.0578 = -7.8410$, which is even more significant than in the stochastic model
- Note the smaller standard error...

Forecasts for Febr.1983-Dec.1984



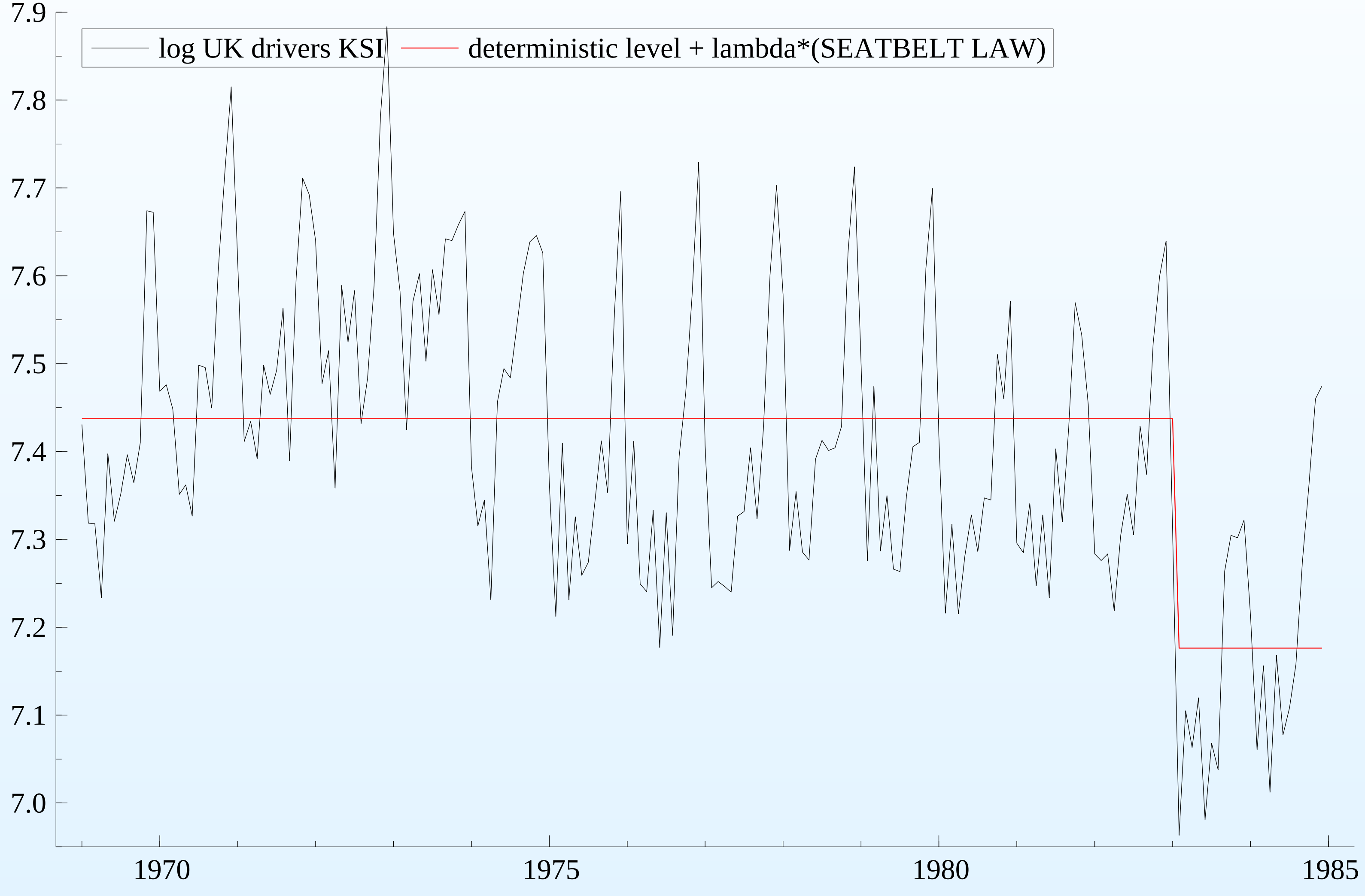
Intervention analysis

In order to explicitly evaluate the effect of the seat belt law, we extend the previous explanatory model as follows

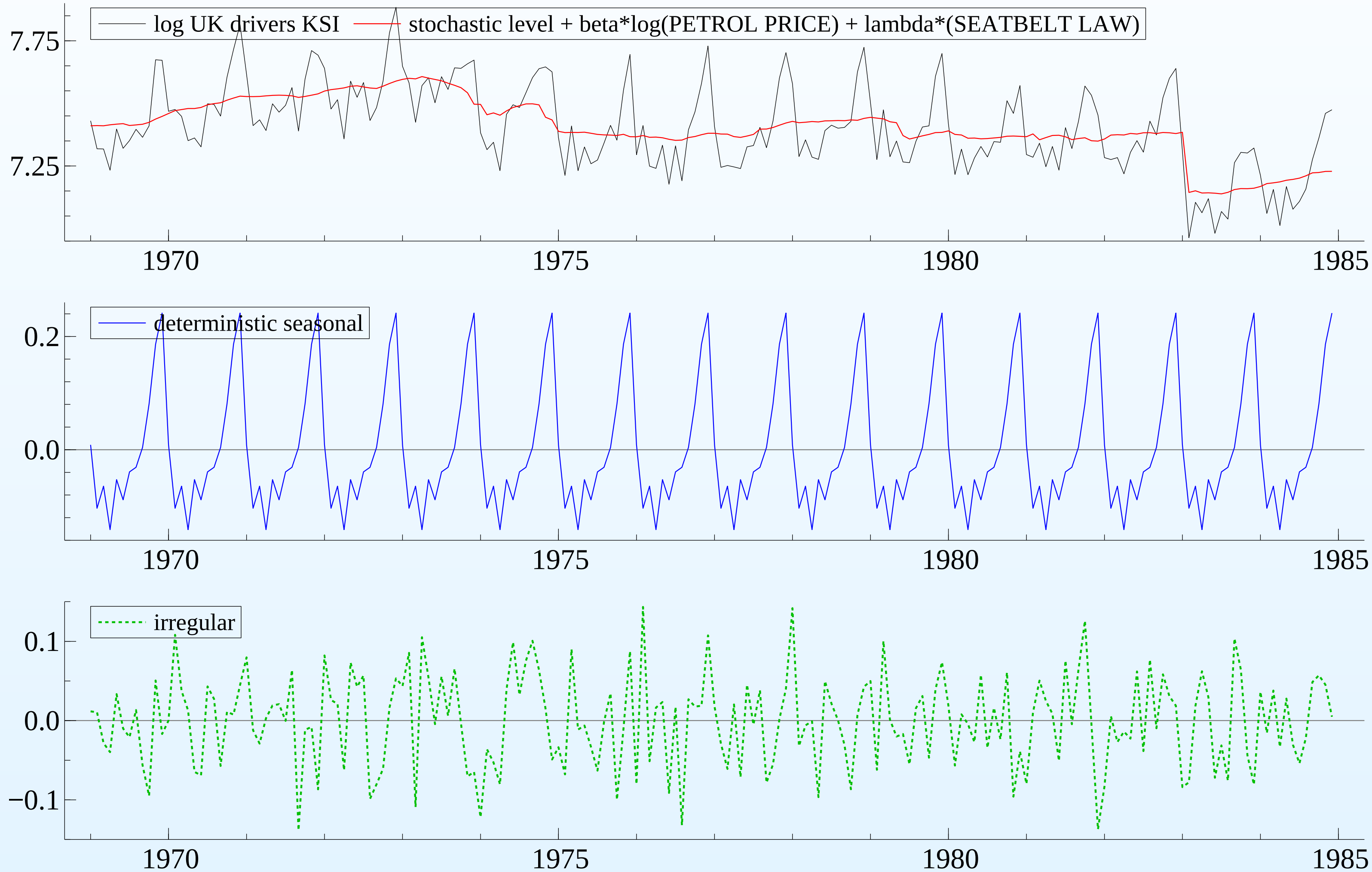
$$y_t = \mu_t + \gamma_t + \lambda w_t + \beta x_t + \varepsilon_t, \qquad \varepsilon_i \sim \mathrm{NID}(0, \sigma_\varepsilon^2)$$

where w_t is a dummy *level intervention* variable consisting of zeroes for the 169 time points up to the introduction of the seat belt law in February 1983, and of ones for the 23 time points at and after its introduction, and λ is an unknown regression coefficient.

Illustration of level intervention



Results of intervention analysis

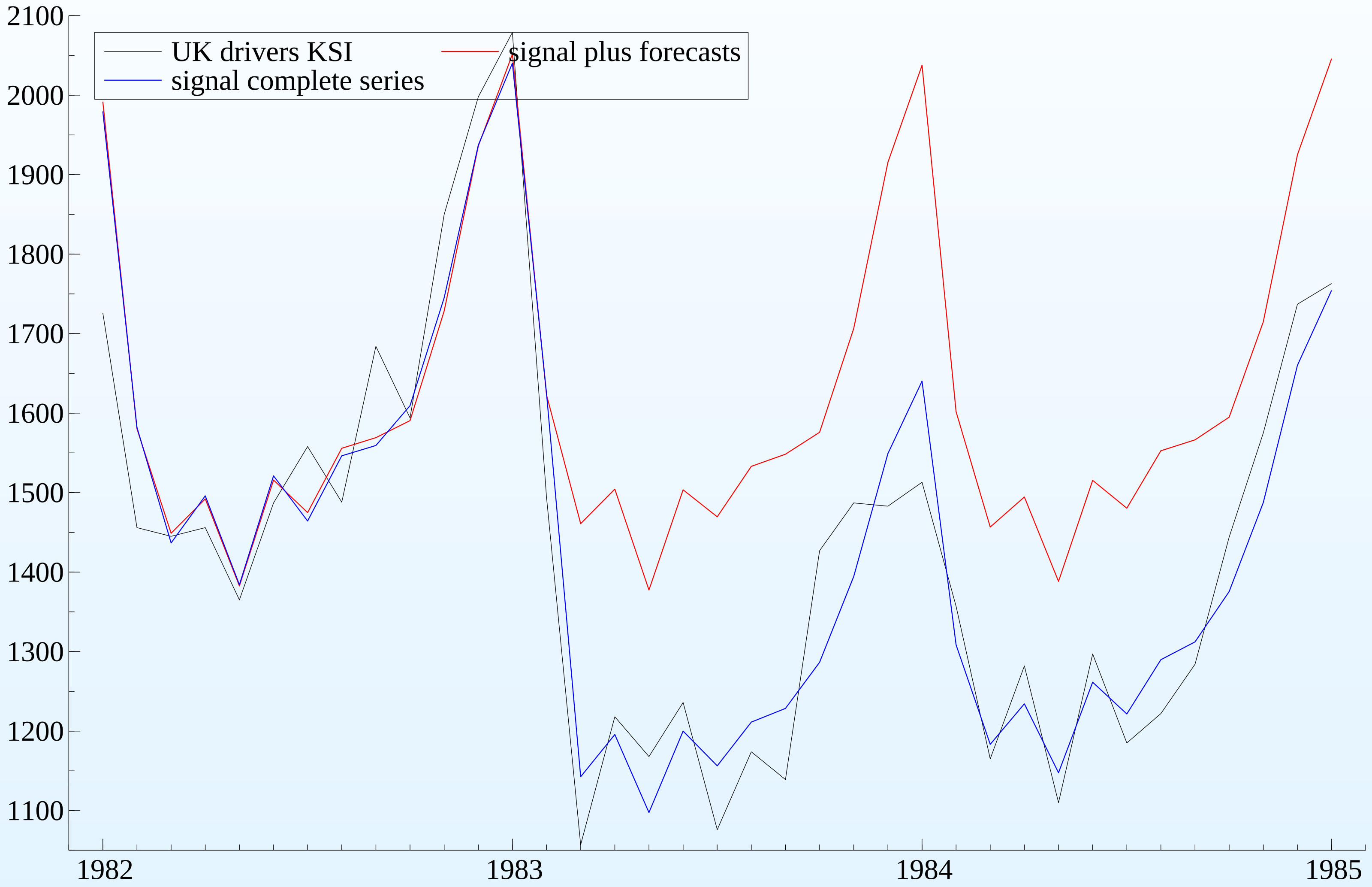


Effects of log petrol price and seat belt law

	coefficient	standard error	t-value	p-value
β	-0.2767	0.0984	-2.8122	0.0054
λ	-0.2376	0.0465	-5.1154	$7.7e-007$

- The regression weight for log petrol price is $\beta = -0.2767$: a 1% increase in petrol price yields a 0.28% decrease in the number of drivers KSI
- The regression weight for the seat belt law is $\lambda = -0.2376$: its introduction is associated with a $100(e^{-0.2376} - 1) = -21.1\%$ change in the number of drivers KSI

Summarizing so far



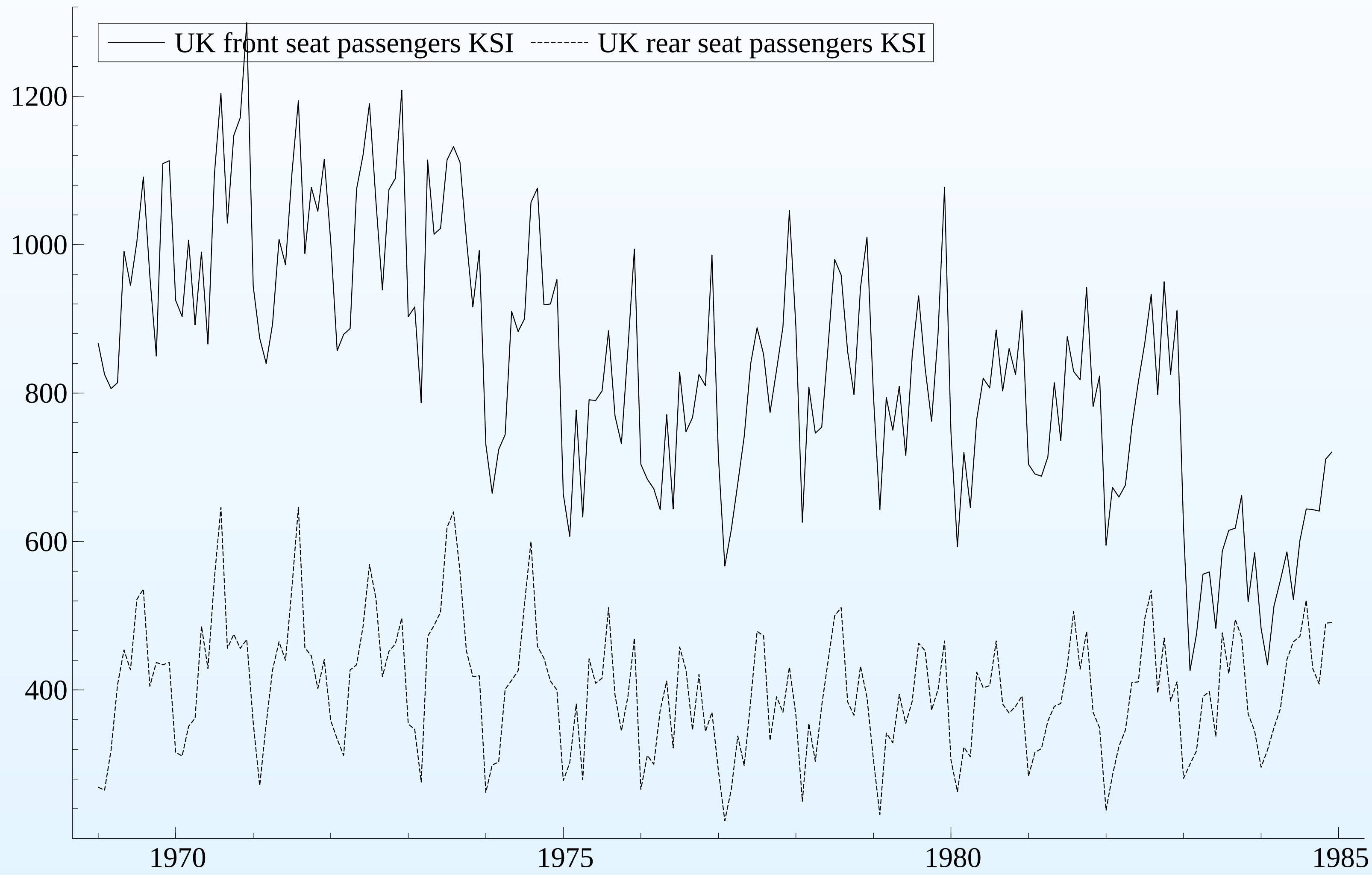
Conclusions

- Controlling for changes in petrol price (resulting in more careful driving?), the introduction of the seat belt in the UK is associated with a 21.1% decrease in the number of car drivers KSI
- After January 1983, there is a difference between what we would expect had the law not been introduced (based on the forecasts), and the observed counts
- Did we prove that the law *caused* this reduction? No!
- But at least we tried to establish what would have happened without the law being introduced, and controlled for seasonal fluctuations, for changes in petrol price, and for a general decreasing trend
- By using multiple observations in the pre-period we reduced the chances of a possible “regression to the mean” effect

Second approach: using a reference group

- To emulate a randomized controlled trial more closely, we can also use a *control or reference group*, if available (see also Hauer, 1997)
- This reference group should preferably be as similar as possible to the treatment group, *except that it is not subject to the measure to be evaluated*
- For the UK, monthly time series are available both for the numbers of *front seat* car passengers KSI, and for the numbers of *rear seat* car passengers KSI
- The seat belt law of February 1983 applied to drivers and front seat car passengers only, not to rear seat car passengers
- We therefore handle the front seat car passengers KSI as *treatment series*, and the rear seat car passengers KSI as *reference or control series*

UK front and rear seat passengers KSI



(A less naive approach)

	1982(2)-1983(1)	1983(2)-1984(1)	Total
front seat	9482	6568	16050
rear seat	4749	4618	9367
Total	14231	11186	25417

We obtain a χ^2 of 168.5 ($p < 0.01$). Without the intervention the change in treatment and reference group would be the same, and we would expect $9482 \times (4618/4749) = 9220$ front seat passengers KSI in the post-period. This implies a $100 \times [(6568 - 9220)/6568] = -40\%$ reduction as a result of the seat belt law.

The bivariate time series model

- We apply the following bivariate explanatory structural time series model to these two time series:

$$
\begin{aligned}
y_t^{(1)} &= \mu_t^{(1)} + \gamma_t^{(1)} + \beta_t^{(1)} x_t^{(1)} + \lambda_t^{(1)} w_t^{(1)} + \varepsilon_t^{(1)} \\
y_t^{(2)} &= \mu_t^{(2)} + \gamma_t^{(2)} + \beta_t^{(2)} x_t^{(2)} + \lambda_t^{(2)} w_t^{(2)} + \varepsilon_t^{(2)},
\end{aligned}
$$

  for time points $t = 1, \ldots, n$, where (1) and (2) denote front and rear seat passengers, respectively
- This bivariate model opens up the possibility of allowing the residuals $\varepsilon_t^{(1)}$ and $\varepsilon_t^{(2)}$ to be *correlated*:

$$
H = \begin{bmatrix} \sigma^2_{\varepsilon^{(1)}} & \text{cov}(\varepsilon_t^{(1)}, \varepsilon_t^{(2)}) \\ \text{cov}(\varepsilon_t^{(1)}, \varepsilon_t^{(2)}) & \sigma^2_{\varepsilon^{(2)}} \end{bmatrix}
$$

The bivariate time series model

The same applies to the disturbances of the two level components (the time-varying intercepts):

$$Q = \begin{bmatrix} \sigma^2_{\xi^{(1)}} & \text{cov}(\xi_t^{(1)}, \xi_t^{(2)}) \\ \text{cov}(\xi_t^{(1)}, \xi_t^{(2)}) & \sigma^2_{\xi^{(2)}} \end{bmatrix},$$

and to the disturbances of the two slope and seasonal disturbances, if required.

Results regression coefficients

	coefficient	SE	t-value	p-value
front seat				
log(petrol price)	-0.3071	0.1067	-2.8791	0.0045
seat belt law	-0.3372	0.0495	-6.8157	$1.2e-010$
rear seat				
log(petrol price)	-0.0836	0.1123	-0.7443	0.4576
seat belt law	0.0021	0.0518	0.0405	0.9677

These results indicate a very significant $100(e^{-0.3372} - 1) = -28.6\%$ reduction in the number of front seat passengers KSI associated with the introduction of the seat belt law, while its effect on the reference series of rear seat passengers KSI is completely insignificant.

Results disturbance matrices

The variance matrix of the residuals is found to be

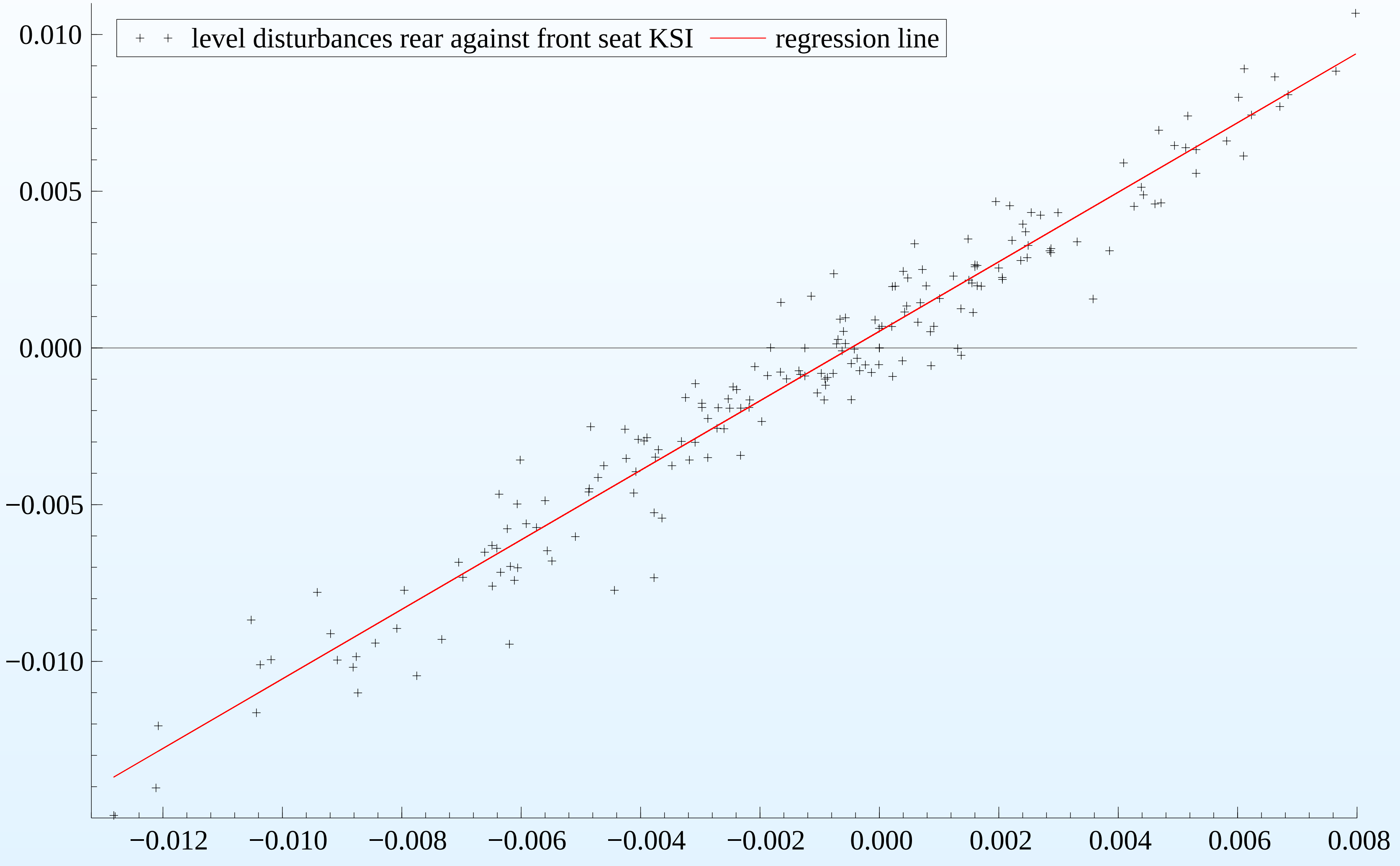
$$H = \begin{bmatrix} 0.0054 & 0.6439 \\ 0.0045 & 0.0085 \end{bmatrix},$$

while the estimates of the variance matrix of the level disturbances are

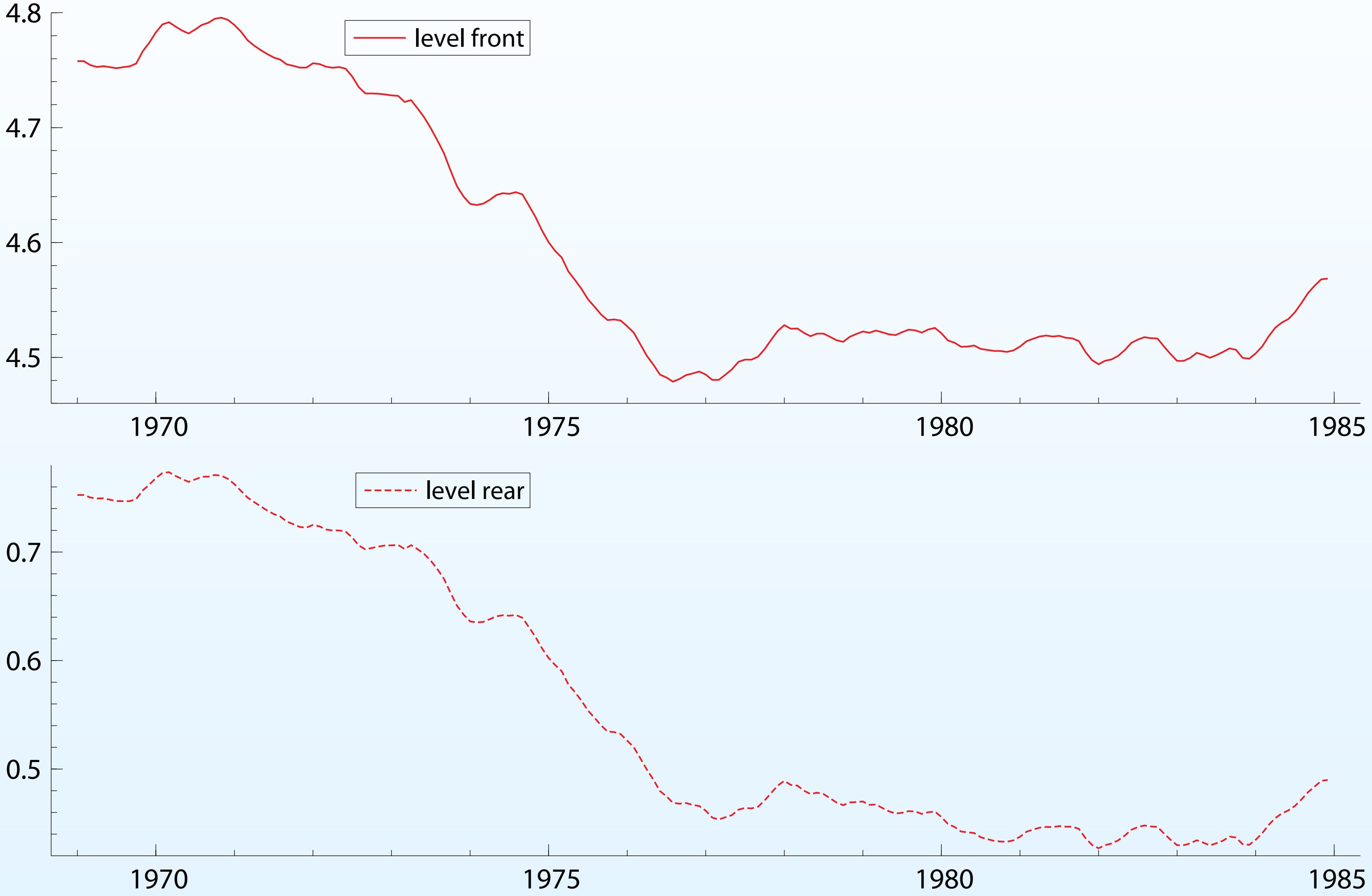
$$Q = \begin{bmatrix} 0.0003 & 0.9743 \\ 0.0002 & 0.0002 \end{bmatrix}.$$

This implies that the correlation between the level components themselves must also be high; this correlation is, in fact, 0.9827.

Plot of level disturbances: front versus rear



Plots of level components: front and rear



Second bivariate analysis

- Based on these results, we re-analyse the two series forcing the variance matrix of the level disturbances to have *rank one*, i.e., we impose the restriction that the level disturbances for the two series must be perfectly correlated
- We also remove the level intervention variable for the introduction of the seat belt law from the rear seat series

Second analysis: disturbance variance matrices

The variance matrix of the residuals is found to be

$$H = \begin{bmatrix} 0.0055 & 0.0044 \\ 0.0044 & 0.0088 \end{bmatrix},$$

while the estimates of the variance matrix of the level disturbances are

$$Q = \begin{bmatrix} 0.0002 & 1.0000 \\ 0.0002 & 0.0002 \end{bmatrix}.$$

Second analysis: disturbance variance matrices

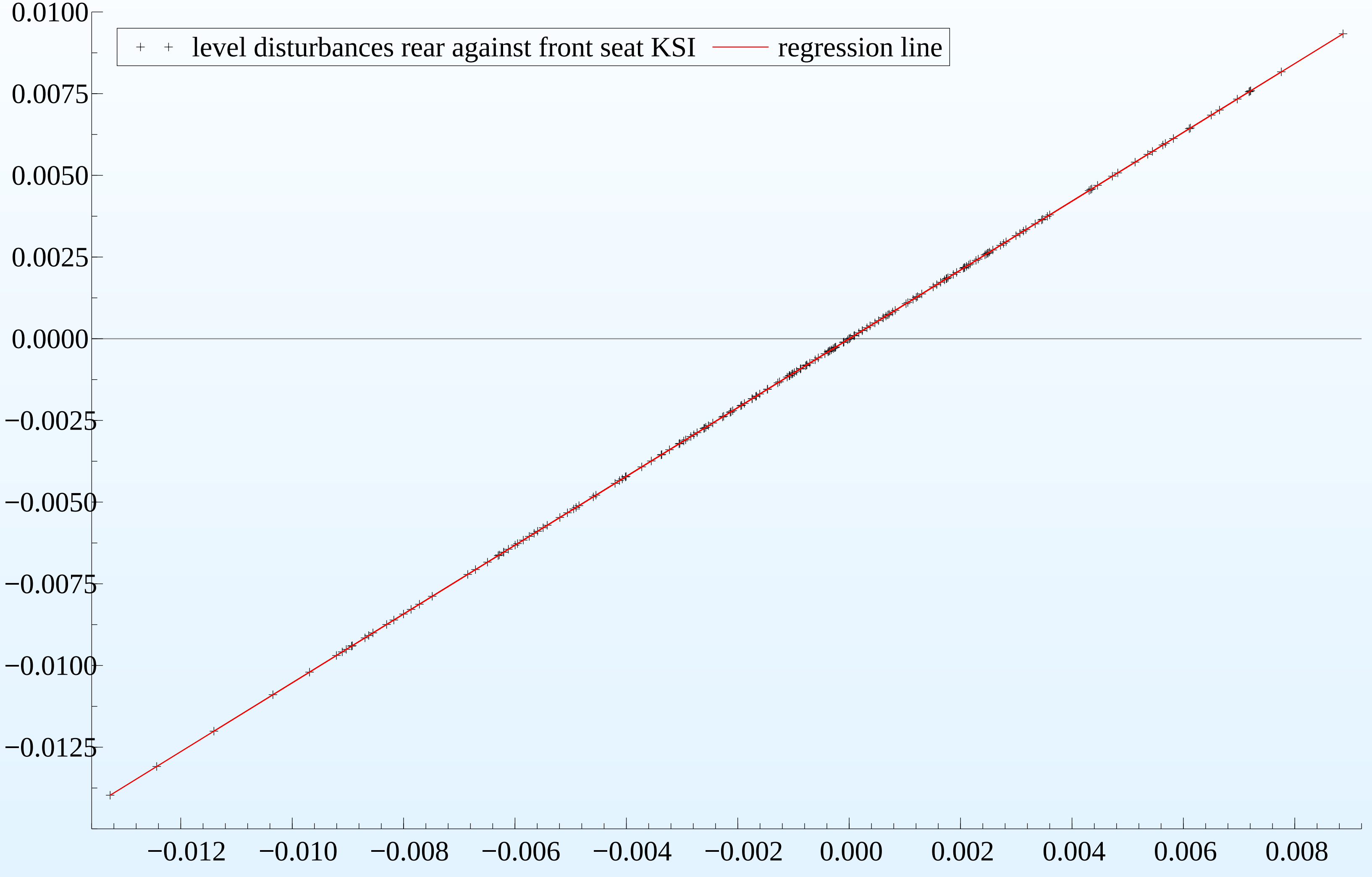
The level disturbances are now perfectly related through the following linear relationship:

$$\xi_t^{(1)} = 1.0529\xi_t^{(2)},$$

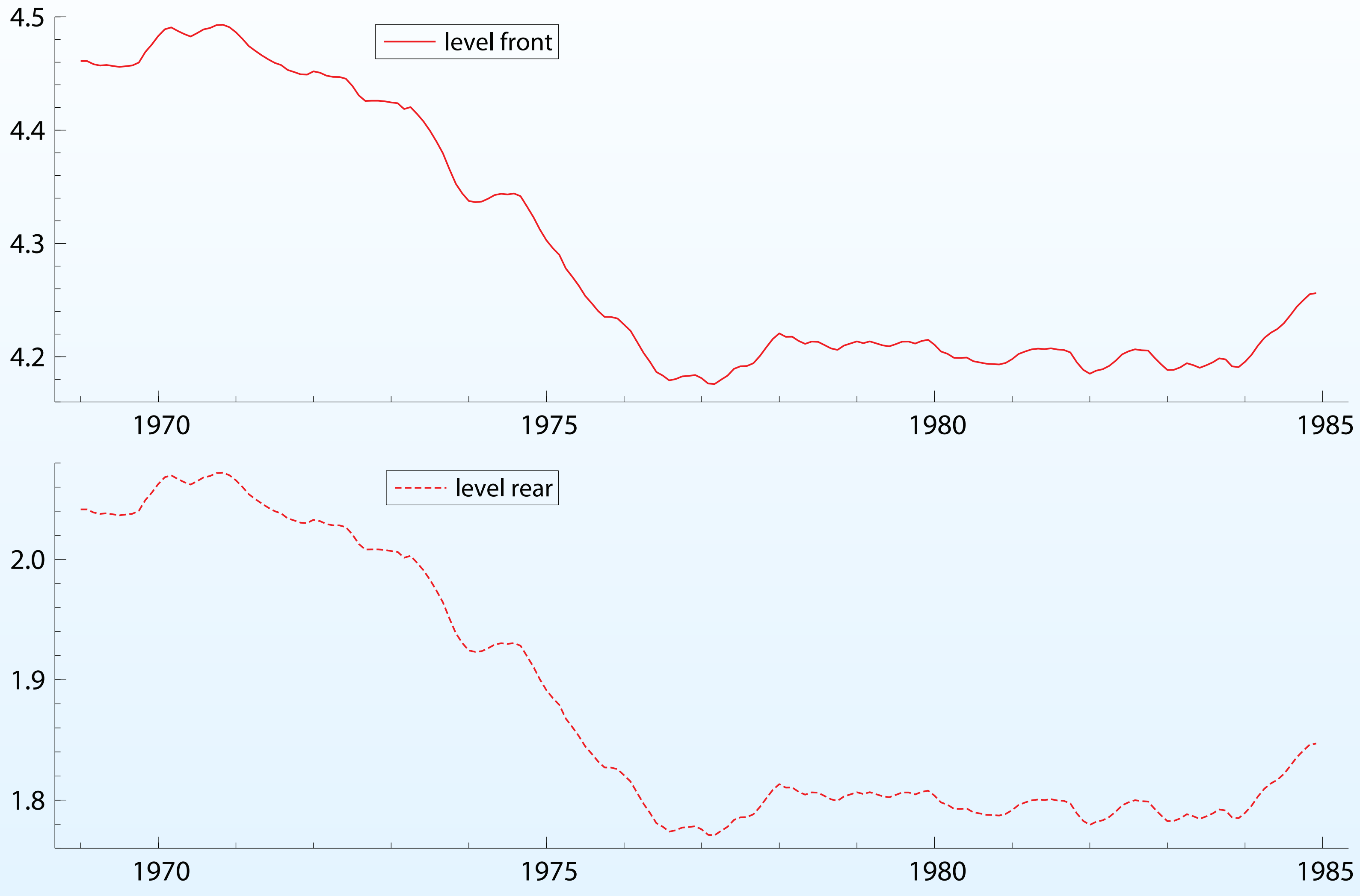
while the level components themselves are perfectly related by

$$\mu_t^{(1)} = 2.3112 + 1.0529\mu_t^{(2)}.$$

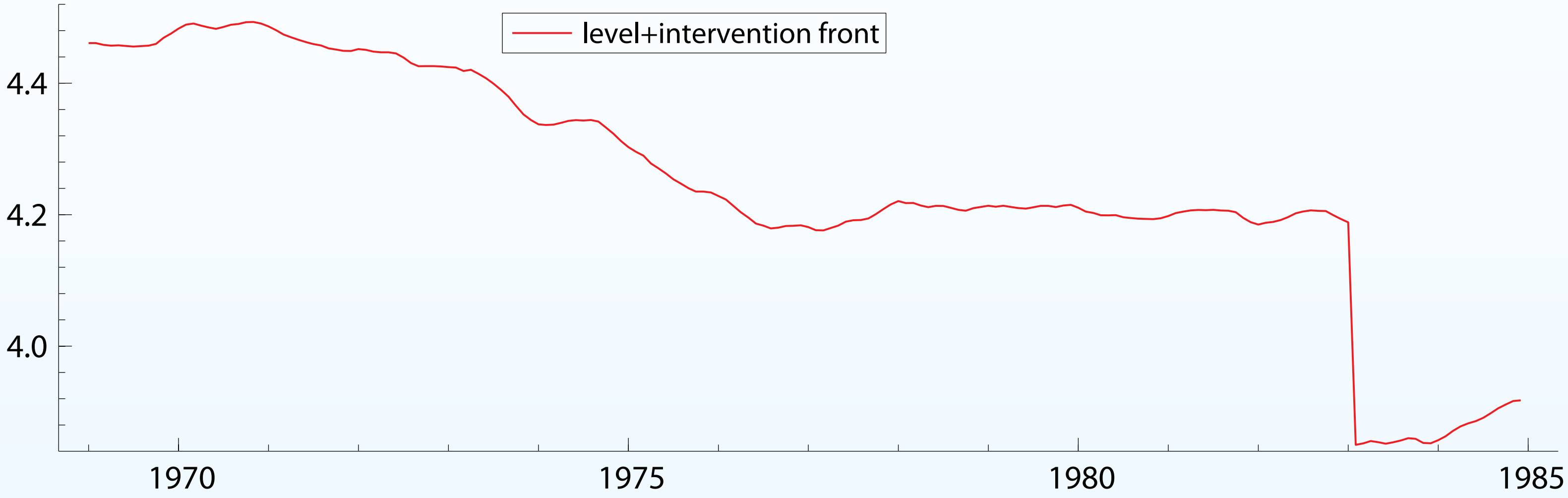
Plot of level disturbances: front versus rear

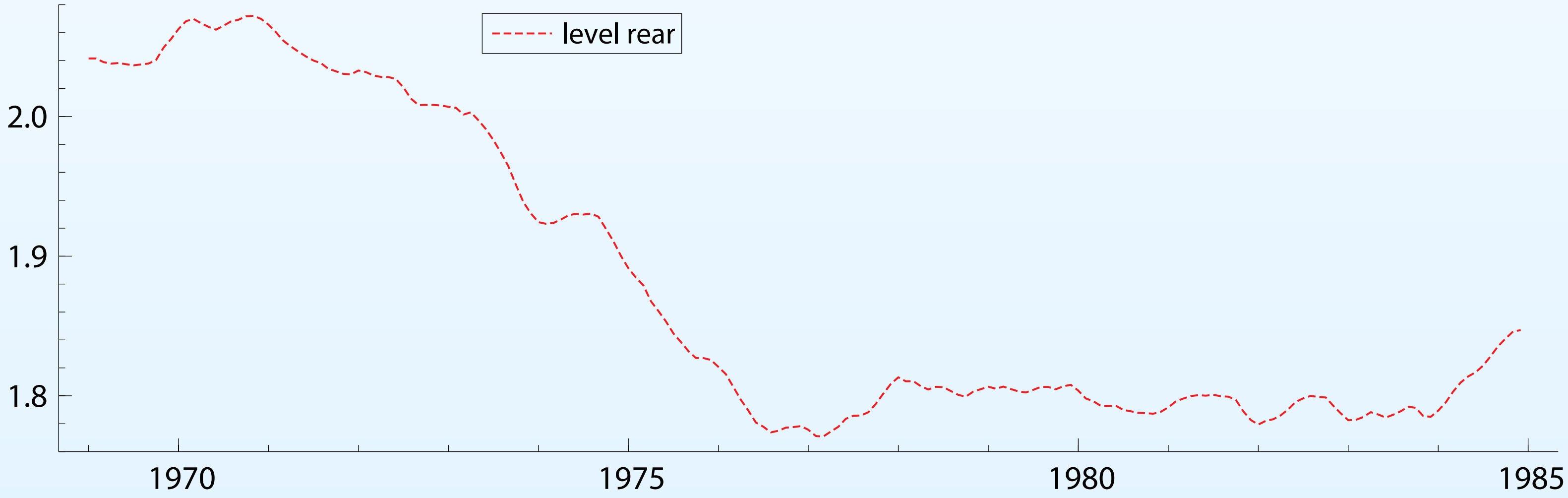


Plots of level components: front and rear



Level components: front plus intervention, and rear





Results of second analysis: regression coefficients

	coefficient	SE	t-value	p-value
front seat				
log(petrol price)	-0.3153	0.1037	-3.0410	0.0027
seat belt law	-0.3387	0.0196	-17.2850	$1.1e-040$
rear seat				
petrol price	-0.0819	0.1070	-0.7649	0.4453

In this analysis, the introduction of the seat belt law is again found to associated with a very significant change of $100(e^{-0.3387} - 1) = -28.7\%$ in the number of front seat passengers KSI.

Confidence limits

- The value of -17.2850 of the t-test for the effect of the seat belt law on the treatment series is remarkable: it is more than two times larger than the value of -6.8157 found in the previous bivariate time series analysis
- This example shows that the use of *common components* (in this case: levels) may provide more efficient inferences: smaller standard errors, and therefore a more precise measure of the intervention effect
- Specifically, the 95% confidence interval for $\lambda = -0.3372$ in the previous analysis is (-0.4342, -0.2402), while that for $\lambda = -0.3387$ in the current analysis is (-0.3771, -0.3003)

Further considerations

Objections to the use of rear seat passengers as a reference group:

- if, as the risk compensation theory implies, drivers wearing belts drive more recklessly, there will be more rear seat passengers injured
- rear seat passengers may sustain more serious injuries because those in the front are wearing seat belts and thus remain in their seats on impact
- passengers might have transferred, or been transferred, to the rear so that a belt need not be worn
- more rear seat passengers may have followed those in the front in wearing belts

However, in this study no statistical evidence is found to suggest that the law had any effect on rear seat passengers KSI.

Further considerations

Objections to the use of rear seat passengers as a reference group:

- there are dependencies between the treatment and reference group, since they travel in the same cars (Hauer, 1997)
- other reference groups could be considered, e.g.: counts of front seat passengers KSI on even days (treatment group) versus counts of rear seat passengers KSI on odd days (reference group); counts of front or rear passengers KSI observed in some other region than where the law was introduced (as reference group)

Conclusions

- The perfect observational study does not exist, especially when it comes to conclusions in terms of cause and effect
- However, this certainly does not imply that it should be avoided
- It is a good idea to perform different types of analysis to cross-validate effect estimates
- Should the results of several observational studies be available, a meta-analysis can be applied to obtain further proof/disproof of an effect (see Elvik and Vaa, 2004)

References

- Commandeur, J.J.F. & Koopman, S.J. (2007). An introduction to State Space Time Series Analysis. Oxford practical econometrics series, nr. 1. Oxford University Press, Oxford.
- Durbin, J. & Koopman, S.J. (2001). Time Series Analysis by State Space Methods. Oxford statistical science series, nr. 24. Oxford University Press, Oxford.
- Elvik, R. & Vaa, T. (2004). The handbook of road safety measures. Emerald Group Publishing Limited, Bingley.
- Harvey, A.C. (1989). Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge.

References

- Harvey, A.C. & Durbin, J. (1986). The effects of seat belt legislation on British road casualties: A case study in structural time series modelling. Journal of the Royal Statistical Society A, vol. 149, nr. 3, p. 187-27.
- Hauer, E. (1997). Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety. Pergamon, New York.