UW CSSS/POLS 512:
Time Series and Panel Data for the Social Sciences

Panel Data Models with Few Time Periods

Christopher Adolph

Department of Political Science
and
Center for Statistics and the Social Sciences
University of Washington, Seattle

Consider this panel model:

$$y_{it} = \phi y_{i,t-1} + \mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}$$

Remove the fixed effects α_i by differencing (the within estimator)

$$y_{it} - \bar{y}_i = \phi(y_{i,t-1} - \bar{y}_i) + (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

What happens if we estimate this with LS?

Notice estimates of $\bar{y}_i$ must be correlated with the error term $\varepsilon_{it} - \bar{\varepsilon}_i$

Consider this panel model:

$$y_{it} = \phi y_{i,t-1} + \mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}$$

Remove the fixed effects α_i by differencing (the within estimator)

$$y_{it} - \bar{y}_i = \phi(y_{i,t-1} - \bar{y}_i) + (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

What happens if we estimate this with LS?

Notice estimates of $\bar{y}_i$ must be correlated with the error term $\varepsilon_{it} - \bar{\varepsilon}_i$

Why? Consider the unit-average version of the model:

$$\begin{aligned}
\bar{y}_i &= \phi\bar{y}_i + \bar{\mathbf{x}}_i\boldsymbol{\beta} + \bar{\varepsilon}_i \\
\bar{y}_i - \phi\bar{y}_i &= \bar{\mathbf{x}}_i\boldsymbol{\beta} + \bar{\varepsilon}_i \\
\bar{y}_i &= \frac{\bar{\mathbf{x}}_i\boldsymbol{\beta} + \bar{\varepsilon}_i}{1-\phi}
\end{aligned}$$

Bias in AR(1) Panels with fixed effects & small T

$$y_{it} - \bar{y}_i = \phi(y_{i,t-1} - \bar{y}_i) + (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

When a regressor and the error are correlated, estimates are biased and inconsistent

In this case estimates are inconsistent in N: $\hat{\phi} - \phi \not\to 0$ as $N \to \infty$

Bias in AR(1) Panels with fixed effects & small T

$$y_{it} - \bar{y}_i = \phi(y_{i,t-1} - \bar{y}_i) + (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

When a regressor and the error are correlated, estimates are biased and inconsistent

In this case estimates are inconsistent in N: $\hat{\phi} - \phi \not\to 0$ as $N \to \infty$

However, estimates are still consistent in T: $\hat{\phi} - \phi \to 0$ as $T \to \infty$

Bias in AR(1) Panels with fixed effects & small T

$$y_{it} - \bar{y}_i = \phi(y_{i,t-1} - \bar{y}_i) + (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

When a regressor and the error are correlated, estimates are biased and inconsistent

In this case estimates are inconsistent in N: $\hat{\phi} - \phi \not\to 0$ as $N \to \infty$

However, estimates are still consistent in T: $\hat{\phi} - \phi \to 0$ as $T \to \infty$

The degree of bias is of order $1/T$, so for small T, this is a big problem

Just how much bias in $\hat{\phi}$? in $\hat{\beta}$? in long run effects?

Bias in AR(1) Panels with fixed effects & small T

Nickell (1981) characterizes the degree of bias in $\hat{\phi}$:

$$\hat{\phi} - \phi \rightarrow \frac{\frac{-(1+\phi)}{T-1} \times \left[1 - \frac{1}{T} \times \frac{1-\phi^T}{1-\phi}\right]}{1 - \frac{2\phi}{(1-\phi)(T-1)} \times \left[1 - \frac{1}{T} \times \frac{1-\phi^T}{1-\phi}\right]} \quad \text{as} \quad N \rightarrow \infty$$

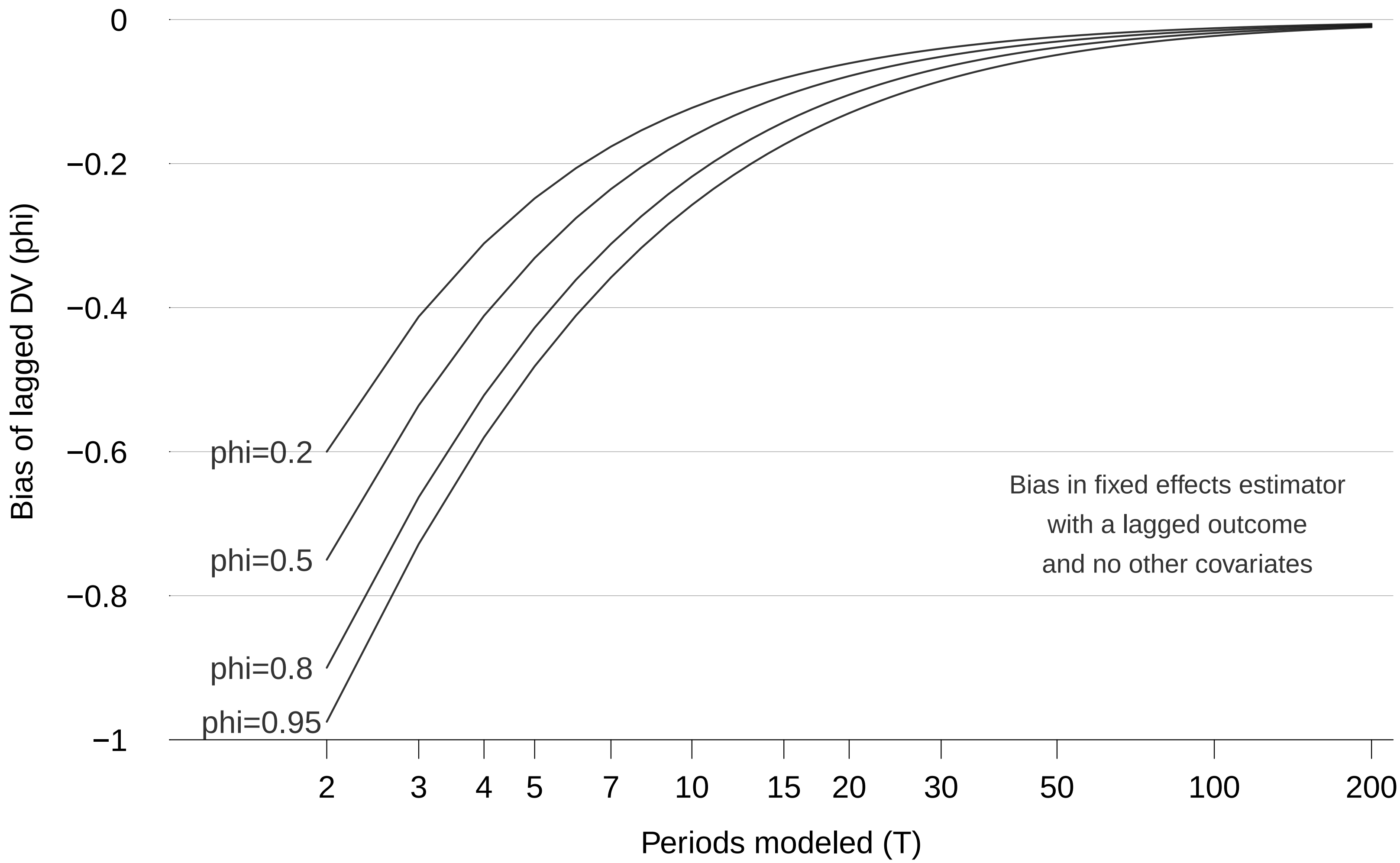
As noted, the bias in $\hat{\phi}$ is $O(1/T)$, and for "reasonably large T" is

$$\hat{\phi} - \phi \approx \frac{-(1+\phi)}{T-1} \quad \text{as} \quad N \rightarrow \infty$$

Upshot: $\hat{\phi}$ will always be underestimated in dynamic fixed effects models, but this bias vanishes as T increases

A picture of some example cases helps clarify where the dangers lie

But to understand the practical implications of Nickell bias we need Monte Carlo



Bias in $\hat{\phi}$ can be substantial, especially for small T or large ϕ

What about models with covariates? Bias in $\hat{\beta}$? Bias in long run effects, $\hat{\beta}/(1 - \hat{\phi})$?

These are the real quantities of interest, not ϕ!

Some Monte Carlo evidence

When does the FE estimator significantly bias the actual quantities of interest: the estimates of short- and long-run effects of covariates?

I conducted a Monte Carlo experiment with the following true model, burned in 100 periods:

$$y_{it} = \phi y_{i,t-1} + \beta x_{it} + \alpha_i + \varepsilon_{it} \qquad \varepsilon_{it} \sim \mathcal{N}(0, \sigma^2_\varepsilon)$$
$$x_{it} = \gamma x_{i,t-1} + \xi_{it} \qquad \xi_{it} \sim \mathcal{N}(0, \sigma^2_\xi)$$

where $\gamma = 0.5$, $\sigma^2_\varepsilon = 1$, and $\sigma^2_\xi = 2$, similar to Judson and Owen (1999)

I consider 100 to 1000 replications of every combination of the following:

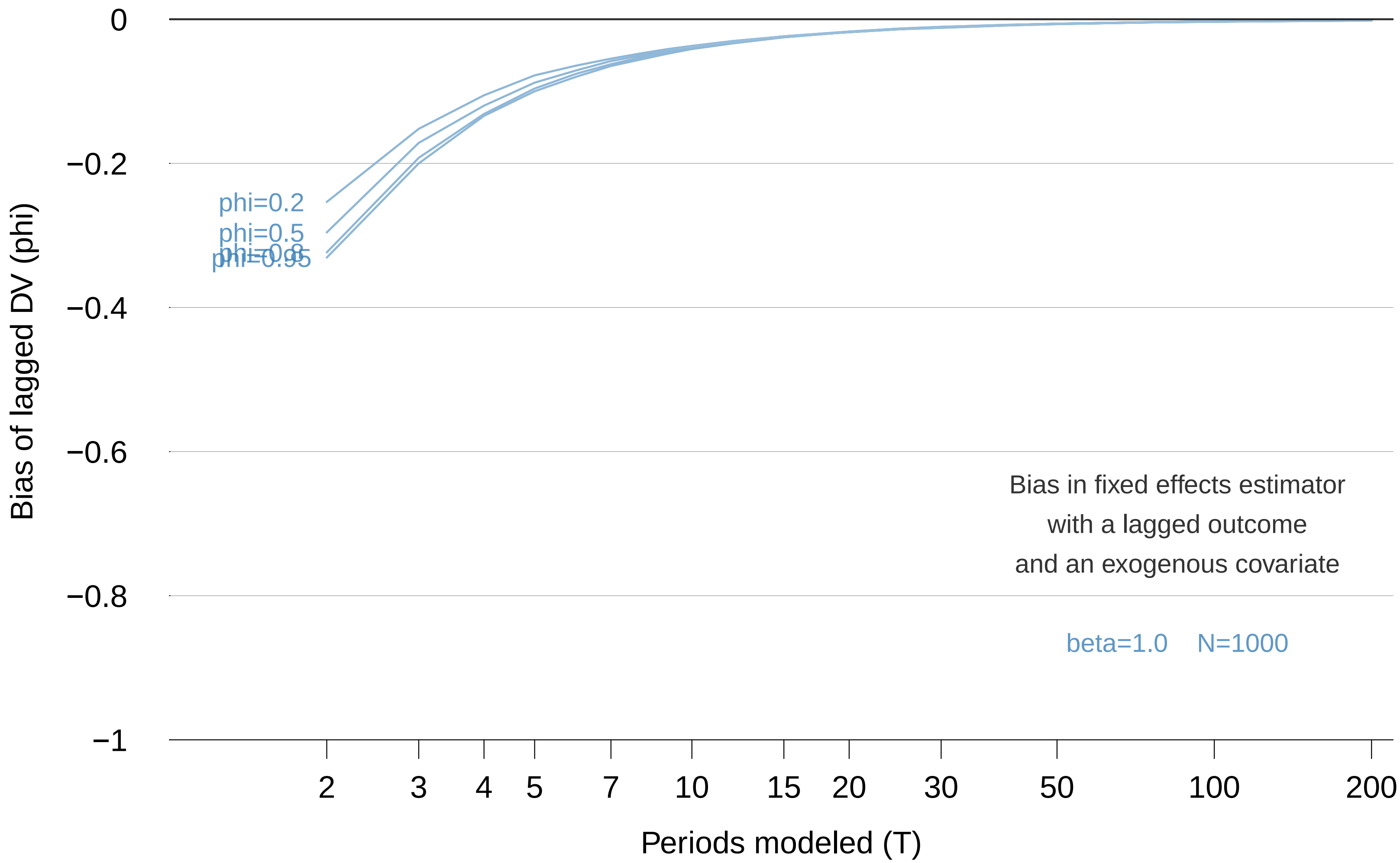
$\beta \in \{0.2, 1.0\}$ *weaker effects should increase bias*

$\phi \in \{0.2, 0.5, 0.8, 0.95\}$ *stronger correlation should increase bias*

$N \in \{30, 1000\}$ *makes no difference*

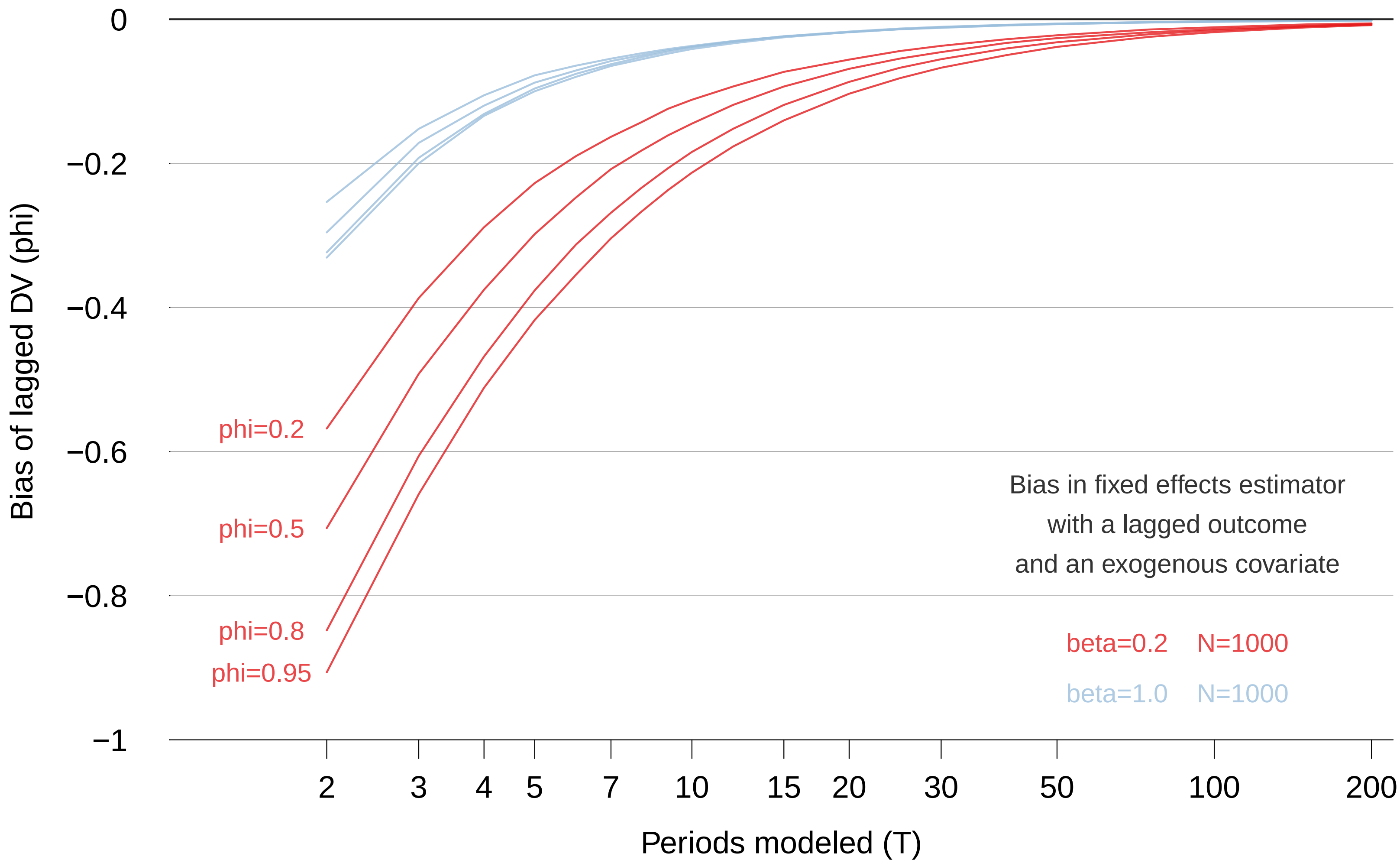
$T \in \{2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 25, 30, 40, 50, 75, 100, 150, 200\}$

fewer periods analyzed should increase bias



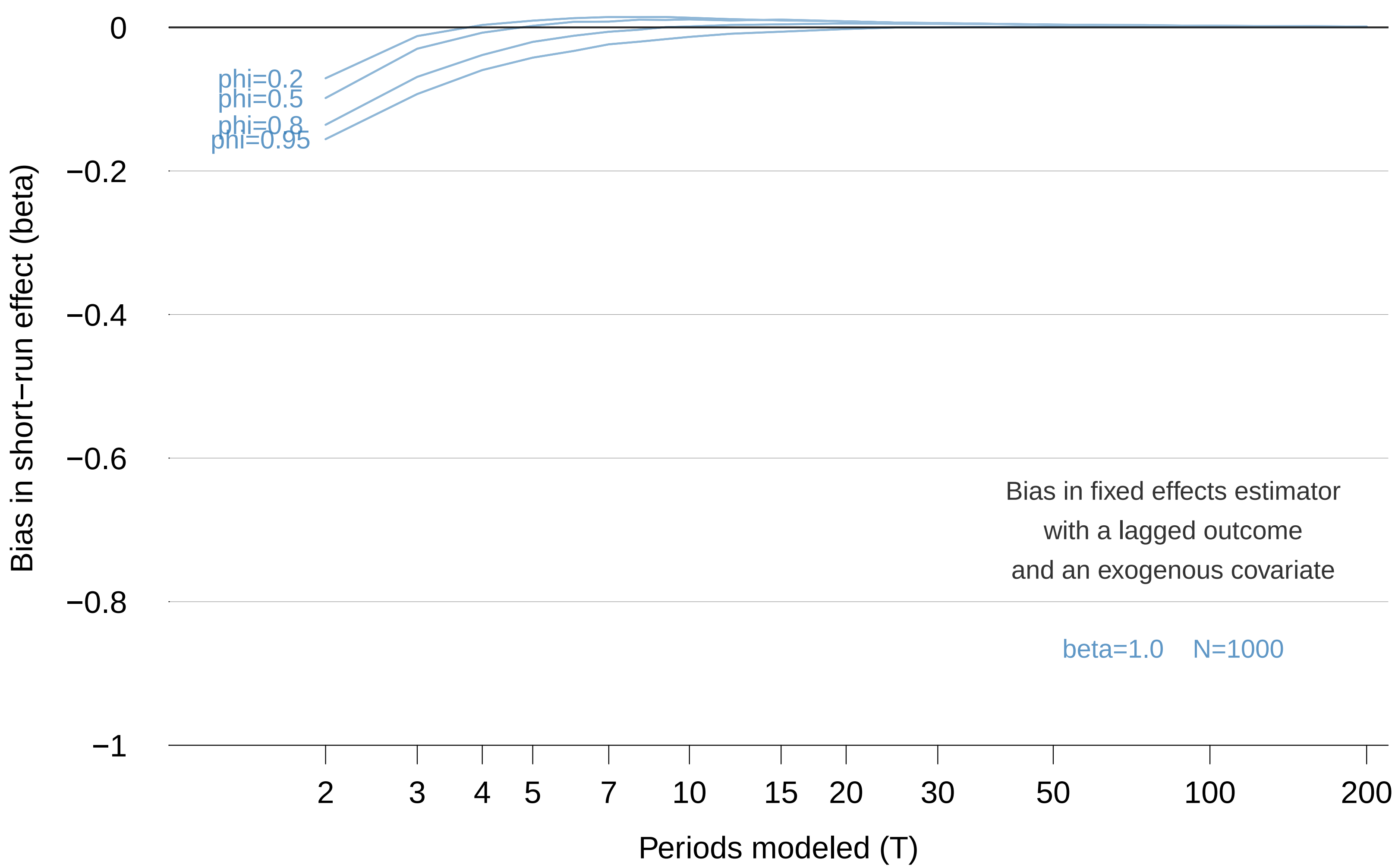
Bias in $\hat{\phi}$ for models with $N = 1000$, $\beta = 1$, and all values of ϕ and T

Less bias than asymptotic results from Nickell (1981) with no covariate



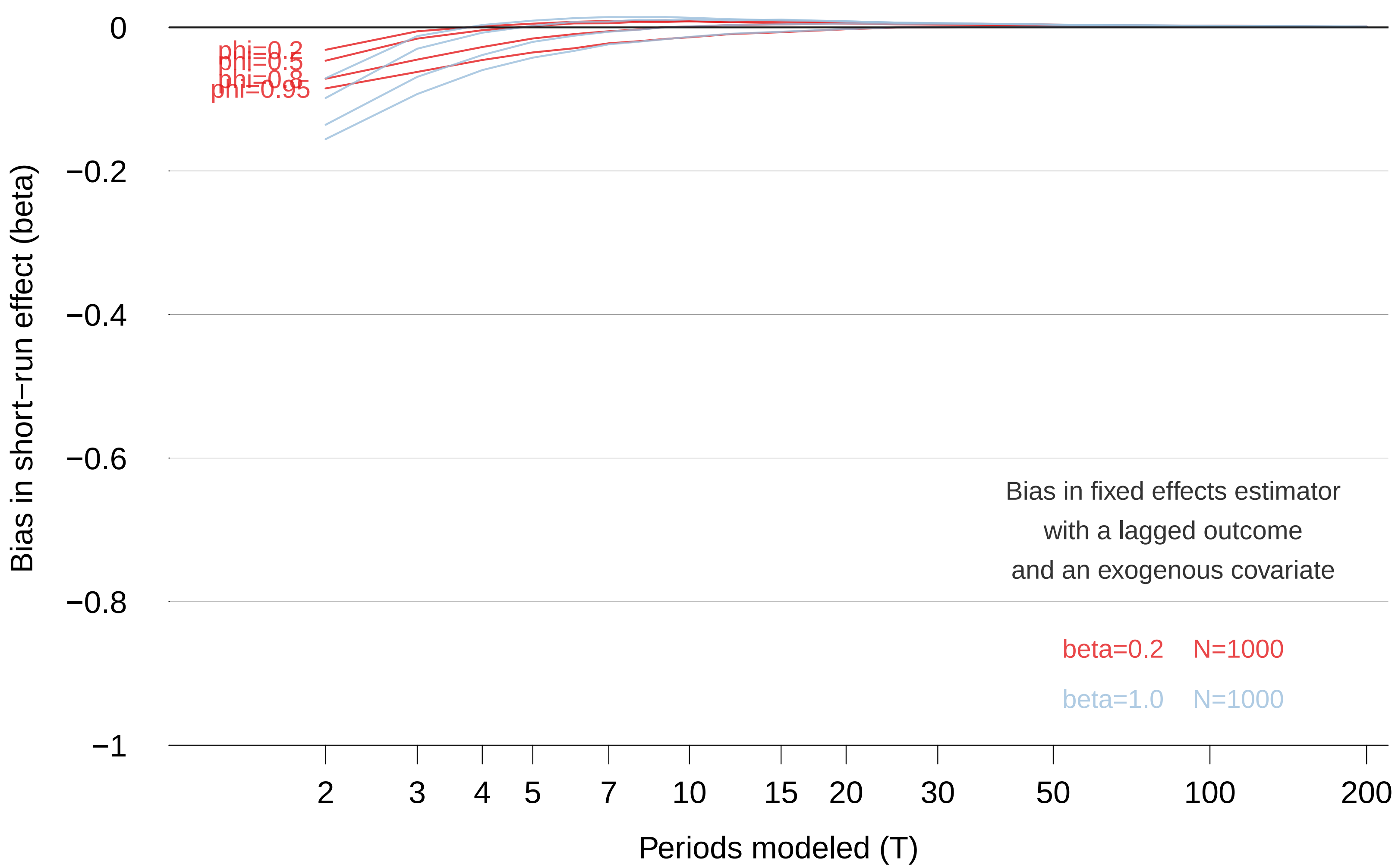
Weaker exogenous covariates *increase* bias in in $\hat{\phi}$

Intuition: x has weaker signal-noise ratio, leading to poorer estimates



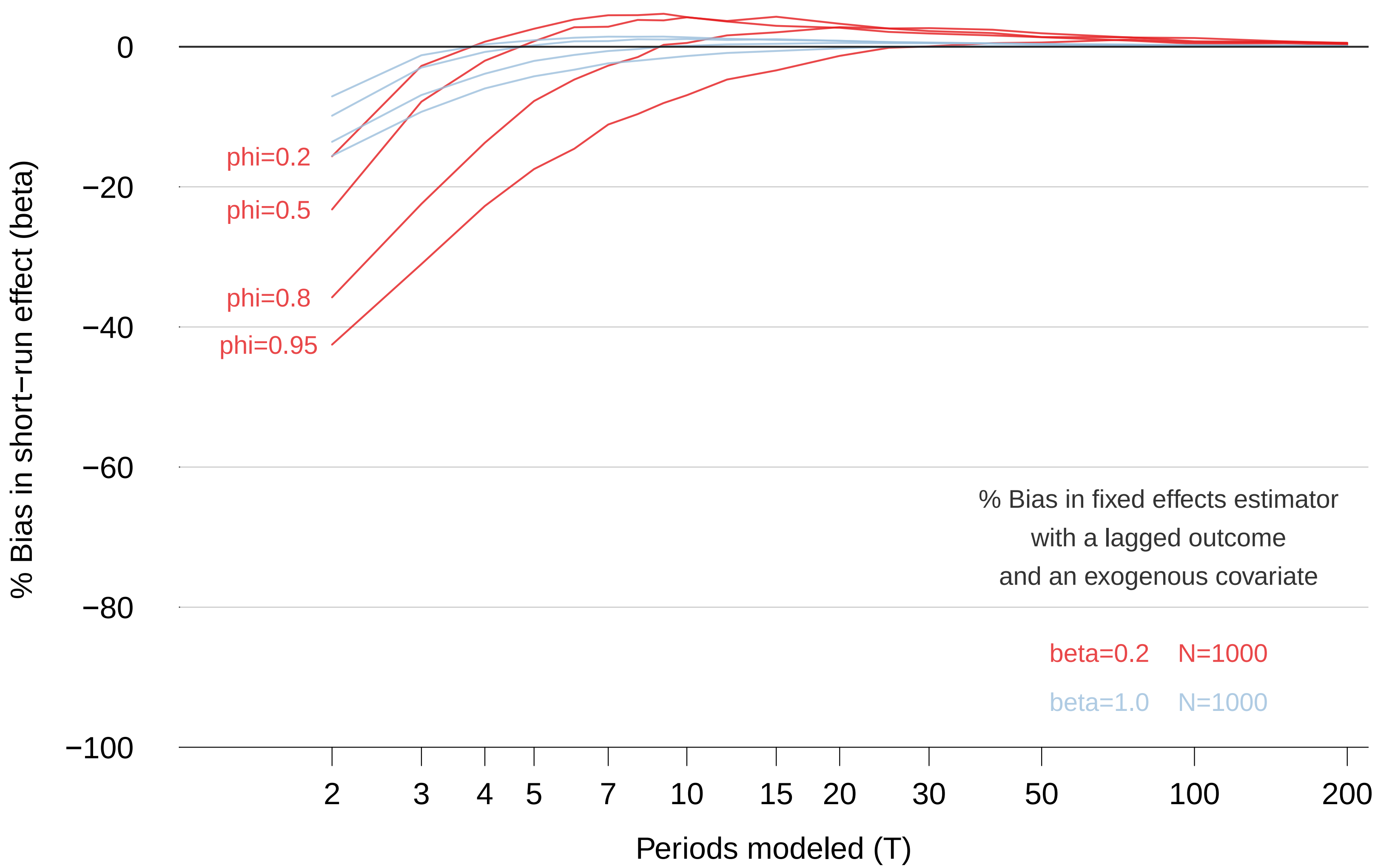
Nickell bias in estimates of β, the short run effect of x, is much smaller

Often seems negligible, unless T is very small *and* y is very serially correlated



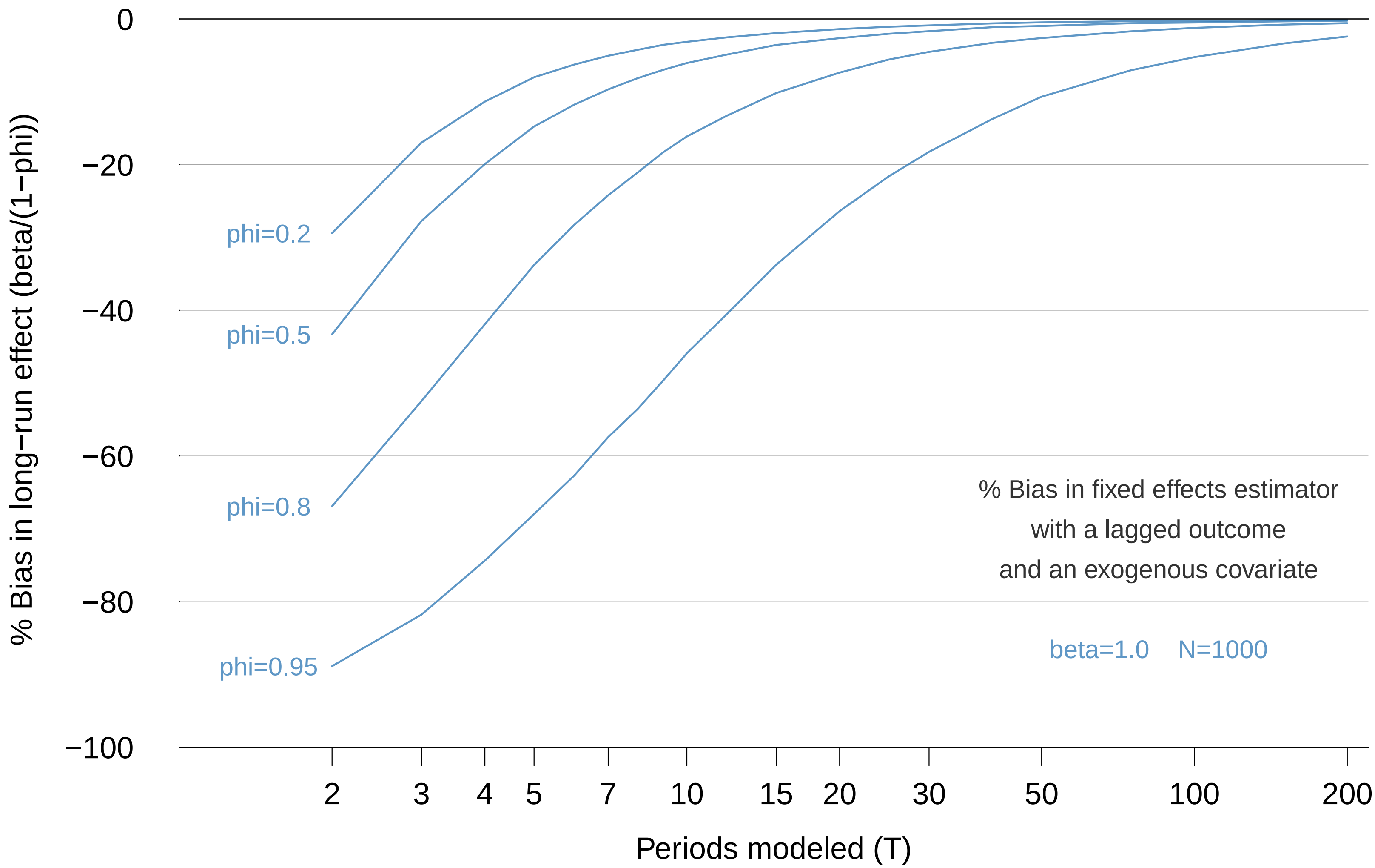
Now consider a weaker short-run relationship between x and y (true $\beta = 0.2$)

At first, you might think bias in $\hat{\beta}$ goes down when the true value of β is small. . .



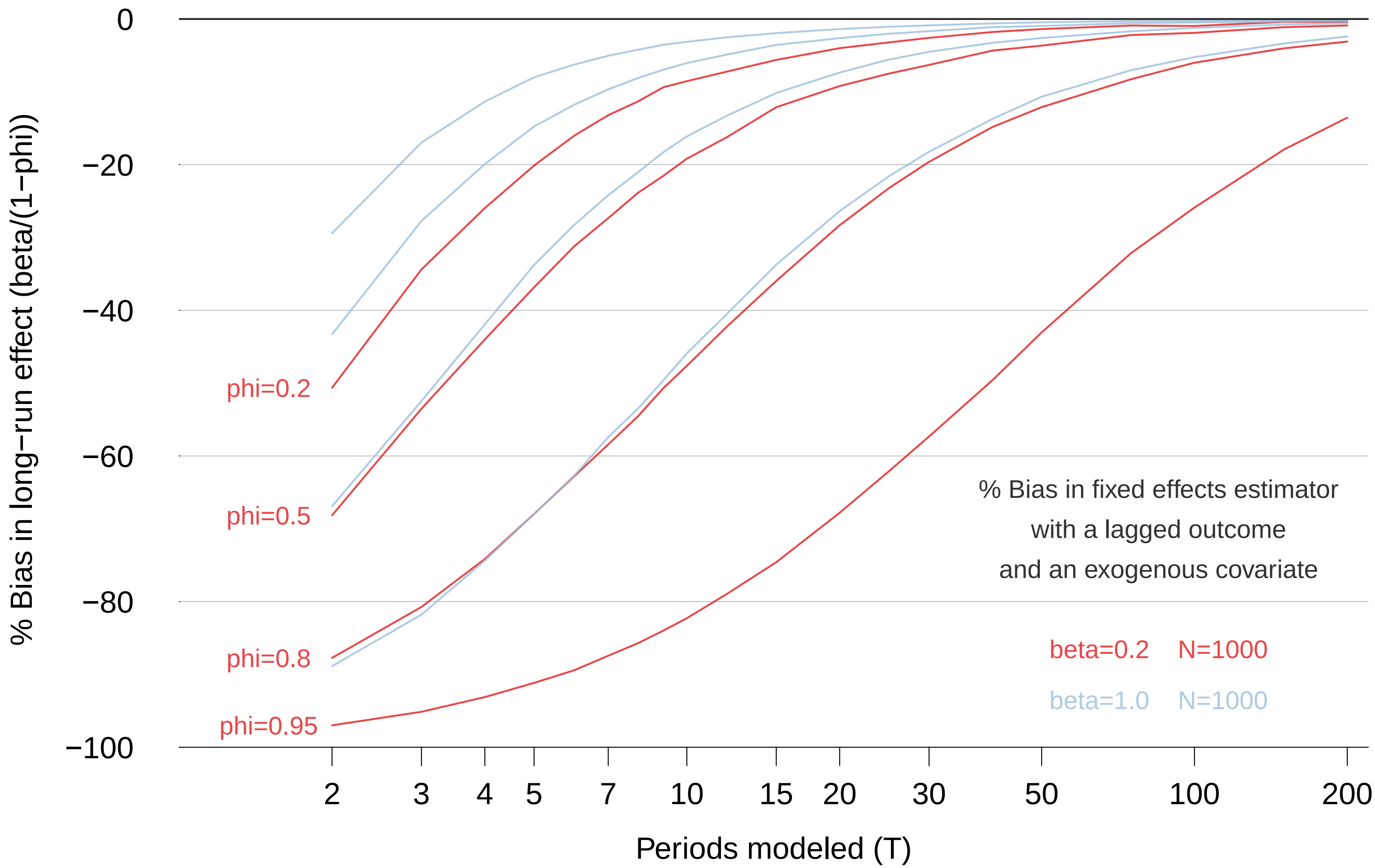
At first, you might think bias in $\hat{\beta}$ goes down when the true value of β is small. . .

But as a percentage of the true short-run effect, bias increases substantially



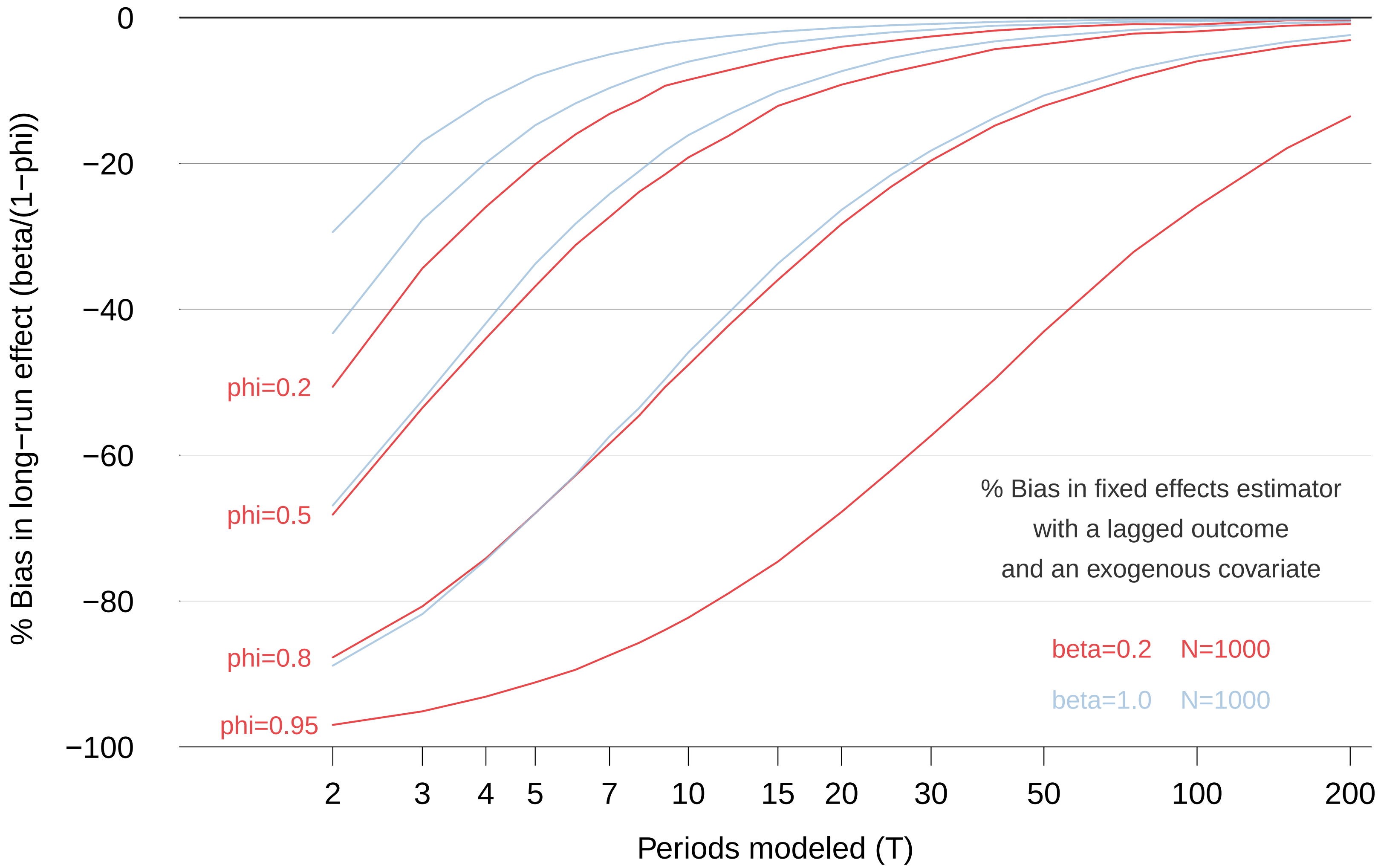
Recall that for an AR(1) process, the long-run effect of x on y is $\beta/(1-\phi)$

Because Nickell bias is more severe for $\hat{\phi}$ than $\hat{\beta}$,
it has a bigger effect on estimates of long-run relationships than short-run ones



This gets worse when the true relationship between x and y is weaker and/or serial correlation in y is strong

Harder in these cases to disentangle dynamic relationships, especially with small T



In panels with small T and/or high ϕ, we tend to distrust long-run estimates anyway

But because we need $\hat{\phi}$ to estimate relationships iterated over two or more periods, this is still a worry for careful applied work with small T and/or large ϕ

Bias in Panels with small T: Summarizing the evidence

Results for small $N = 30$ were identical – issue in Nickell bias is T, not N

In terms of bias, clear that small T dynamic panels are problematic for fixed effects models

Similar results for efficiency (not shown)

Monte Carlo results are sensitive to conditions; you could run your own

Code is on the course site and highly customizable

Not considered: multiple covariates, period effects, more complex patterns of correlation across time or units, heteroskedasticity, etc.

A huge literature in econometrics considers this topic

Bias in Panels with small T: Solutions?

Things that **don't** help:

- More units N
- Purging serial correlation in ε
- Getting the specification of x right

What would help:

One way to deal with correlated covariates & errors is with *instrumental* variables

If x_{it} and ε_{it} are correlated, find some z_{it} which is correlated with x_{it} but not ε_{it}

z_{it} is then an *instrument* for x_{it}

What is an instrument?

In words, an instrument z_{it}

- explains part of x_{it},
- but does not *otherwise* explain y_{it},
- so it does not belong in our model,
- but can be used to distinguish the part of x_{it} that influences y_{it} from the part that influences ε_{it}

Background: IV estimation

Consider a bivariate regression

$$y_i = \beta x_i + \varepsilon_i$$

We condition on x_i and take expectations,
assuming no correlation of the error with x_i,

$$\mathrm{E}(y_i|x_i) = \mathrm{E}(\beta x_i|x_i) + \mathrm{E}(\varepsilon_i|x_i)$$

Background: IV estimation

Consider a bivariate regression

$$y_i = \beta x_i + \varepsilon_i$$

We condition on x_i and take expectations,
assuming no correlation of the error with x_i,

$$\mathrm{E}(y_i|x_i) = \mathrm{E}(\beta x_i|x_i) + \mathrm{E}(\varepsilon_i|x_i)$$
$$\mathrm{E}(y_i|x_i) = \beta x_i + 0$$

Background: IV estimation

Consider a bivariate regression

$$y_i = \beta x_i + \varepsilon_i$$

We condition on x_i and take expectations,
assuming no correlation of the error with x_i,

$$\begin{aligned}
\mathrm{E}(y_i|x_i) &= \mathrm{E}(\beta x_i|x_i) + \mathrm{E}(\varepsilon_i|x_i) \\
\mathrm{E}(y_i|x_i) &= \beta x_i + 0 \\
\mathrm{E}(y_i|x_i) &= \beta x_i
\end{aligned}$$

Background: IV estimation

What if x_i and ε_i are correlated?

$$\hat{\beta}_{\mathrm{LS}} = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

Background: IV estimation

What if x_i and ε_i are correlated?

$$\begin{aligned}\hat{\beta}_{\mathrm{LS}} &= \frac{\sum_i x_i y_i}{\sum_i x_i^2} \\ &= \frac{\sum_i x_i(\beta x_i + \varepsilon_i)}{\sum_i x_i^2}\end{aligned}$$

Background: IV estimation

What if x_i and ε_i are correlated?

$$\begin{aligned}\hat{\beta}_{\mathrm{LS}} &= \frac{\sum_i x_i y_i}{\sum_i x_i^2} \\ &= \frac{\sum_i x_i(\beta x_i + \varepsilon_i)}{\sum_i x_i^2} \\ &= \frac{\sum_i x_i^2 \beta + \sum_i x_i \varepsilon_i}{\sum_i x_i^2}\end{aligned}$$

Background: IV estimation

What if x_i and ε_i are correlated?

$$\begin{aligned}
\hat{\beta}_{\mathrm{LS}} &= \frac{\sum_i x_i y_i}{\sum_i x_i^2} \\
&= \frac{\sum_i x_i(\beta x_i + \varepsilon_i)}{\sum_i x_i^2} \\
&= \frac{\sum_i x_i^2 \beta + \sum_i x_i \varepsilon_i}{\sum_i x_i^2} \\
&= \frac{\sum_i x_i^2 \beta}{\sum_i x_i^2} + \frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2}
\end{aligned}$$

Background: IV estimation

What if x_i and ε_i are correlated?

$$\begin{aligned}
\hat{\beta}_{\text{LS}} &= \frac{\sum_i x_i y_i}{\sum_i x_i^2} \\
&= \frac{\sum_i x_i(\beta x_i + \varepsilon_i)}{\sum_i x_i^2} \\
&= \frac{\sum_i x_i^2 \beta + \sum_i x_i \varepsilon_i}{\sum_i x_i^2} \\
&= \frac{\sum_i x_i^2 \beta}{\sum_i x_i^2} + \frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2} \\
&= \beta + \frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2}
\end{aligned}$$

LS is biased

Background: IV estimation

Now try an instrument z_i for x_i

Condition on z_i and take expectations:

$$\mathrm{E}(y_i|z_i) = \mathrm{E}(\beta x_i|z_i) + \mathrm{E}(\varepsilon_i|z_i)$$

Background: IV estimation

Now try an instrument z_i for x_i

Condition on z_i and take expectations:

$$\mathrm{E}(y_i|z_i) = \mathrm{E}(\beta x_i|z_i) + \mathrm{E}(\varepsilon_i|z_i)$$

We assumed that $\mathrm{E}(\varepsilon_i|z_i) = 0$, so we can write

$$\mathrm{E}(y_i|z_i) = \mathrm{E}(\beta x_i|z_i)$$

Background: IV estimation

Now try an instrument z_i for x_i

Condition on z_i and take expectations:

$$\mathrm{E}(y_i|z_i) = \mathrm{E}(\beta x_i|z_i) + \mathrm{E}(\varepsilon_i|z_i)$$

We assumed that $\mathrm{E}(\varepsilon_i|z_i) = 0$, so we can write

$$\mathrm{E}(y_i|z_i) = \mathrm{E}(\beta x_i|z_i)$$
$$\mathrm{E}(y_i|z_i) = \beta \mathrm{E}(x_i|z_i)$$

Background: IV estimation

Now try an instrument z_i for x_i

Condition on z_i and take expectations:

$$\mathrm{E}(y_i|z_i) = \mathrm{E}(\beta x_i|z_i) + \mathrm{E}(\varepsilon_i|z_i)$$

We assumed that $\mathrm{E}(\varepsilon_i|z_i) = 0$, so we can write

$$\begin{aligned}\mathrm{E}(y_i|z_i) &= \mathrm{E}(\beta x_i|z_i)\\ \mathrm{E}(y_i|z_i) &= \beta\mathrm{E}(x_i|z_i)\\ \beta &= \frac{\mathrm{E}(y_i|z_i)}{\mathrm{E}(x_i|z_i)}\end{aligned}$$

Solving for β will give us the IV estimator, $\hat{\beta}_{\mathrm{IV}}$

Background: IV estimation

This way of finding the IV estimator is known as the method of moments

So called because we are working just in the expectations of our variables, not with their complete probability distributions

An alternative to maximum likelihood estimation

Background: IV estimation

The IV estimator is consistent if z_i is an instrument:

$$\beta = \frac{\mathrm{E}(y_i|z_i)}{\mathrm{E}(x_i|z_i)}$$

Background: IV estimation

The IV estimator is consistent if z_i is an instrument:

$$\beta = \frac{\mathrm{E}(y_i|z_i)}{\mathrm{E}(x_i|z_i)}$$

$$\hat{\beta}_{\mathrm{IV}} = \frac{\sum_i z_i y_i}{\sum_i z_i x_i}$$

Background: IV estimation

The IV estimator is consistent if z_i is an instrument:

$$\beta = \frac{\mathrm{E}(y_i|z_i)}{\mathrm{E}(x_i|z_i)}$$

$$\hat{\beta}_{\mathrm{IV}} = \frac{\sum_i z_i y_i}{\sum_i z_i x_i}$$

$$= \frac{\sum_i z_i(\beta x_i + \varepsilon_i)}{\sum_i z_i x_i}$$

Background: IV estimation

The IV estimator is consistent if z_i is an instrument:

$$\beta = \frac{\mathrm{E}(y_i|z_i)}{\mathrm{E}(x_i|z_i)}$$

$$\hat{\beta}_{\mathrm{IV}} = \frac{\sum_i z_i y_i}{\sum_i z_i x_i}$$

$$= \frac{\sum_i z_i(\beta x_i + \varepsilon_i)}{\sum_i z_i x_i}$$

$$= \frac{\sum_i z_i x_i \beta + \sum_i z_i \varepsilon_i}{\sum_i z_i x_i}$$

Background: IV estimation

The IV estimator is consistent if z_i is an instrument:

$$\beta = \frac{\mathrm{E}(y_i|z_i)}{\mathrm{E}(x_i|z_i)}$$

$$\hat{\beta}_{\mathrm{IV}} = \frac{\sum_i z_i y_i}{\sum_i z_i x_i}$$

$$= \frac{\sum_i z_i(\beta x_i + \varepsilon_i)}{\sum_i z_i x_i}$$

$$= \frac{\sum_i z_i x_i \beta + \sum_i z_i \varepsilon_i}{\sum_i z_i x_i}$$

$$= \frac{\sum_i z_i x_i \beta}{\sum_i z_i x_i} + \frac{\sum_i z_i \varepsilon_i}{\sum_i z_i x_i}$$

Background: IV estimation

The IV estimator is consistent if z_i is an instrument:

$$\begin{aligned}
\beta &= \frac{\mathrm{E}(y_i|z_i)}{\mathrm{E}(x_i|z_i)} \\
\hat{\beta}_{\mathrm{IV}} &= \frac{\sum_i z_i y_i}{\sum_i z_i x_i} \\
&= \frac{\sum_i z_i(\beta x_i + \varepsilon_i)}{\sum_i z_i x_i} \\
&= \frac{\sum_i z_i x_i \beta + \sum_i z_i \varepsilon_i}{\sum_i z_i x_i} \\
&= \frac{\sum_i z_i x_i \beta}{\sum_i z_i x_i} + \frac{\sum_i z_i \varepsilon_i}{\sum_i z_i x_i} \\
&= \beta + \frac{\sum_i z_i \varepsilon_i}{\sum_i z_i x_i} \\
&= \beta
\end{aligned}$$

(Still biased in small samples if z is correlated with y conditional on x)

IV estimation for Panel

Return to our panel model,

$$y_{it} - \bar{y}_i = \phi(y_{i,t-1} - \bar{y}_i) + (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

Now take first differences of all variables to obtain

$$\Delta y_{it} = \phi \Delta y_{i,t-1} + \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \Delta \varepsilon_{it}$$

Note that $\Delta\varepsilon_{it}$ is now MA(1)

This is an alternative way to sweep out fixed effects

Less biased estimates of fixed effects than the within estimator in small T panels

But like other fixed effects estimators with lagged DV,
this is biased when T is small

IV estimation for Panel

$$\Delta y_{it} = \phi \Delta y_{i,t-1} + \Delta \mathbf{x}_{it} \boldsymbol{\beta} + \Delta \varepsilon_{it}$$

But like other fixed effects estimators with lagged DV,
this is biased when T is small

If we had instruments for $\Delta y_{i,t-1}$,
we could correct the bias in estimation of ϕ and thus potentially β

We do have some *weak* instruments for $\Delta y_{i,t-1}$:
older lagged levels (like $y_{i,t-2}$) and lagged differences (like $\Delta y_{i,t-2}$)

That is, $\Delta y_{i,t-2}$ and $y_{i,t-2}$ help predict $\Delta y_{i,t-1}$, but not $\Delta \varepsilon_{i,t-1}$

Only weak instruments, though

Using these lags in IV estimation: Anderson-Hsiao estimator

Why are lagged levels and changes in y instruments?

Fixed Effects				Differenced Fixed Effects		
Dependent Variable	Most recent error	Error Term	lag of Dependent Variable	Dep Var	Error Term	lag of Dep Var
$y_{i,t-3} - \bar{y}_i$	$\varepsilon_{i,t-3}$	$\varepsilon_{i,t-3} - \bar{\varepsilon}_i$				
$y_{i,t-2} - \bar{y}_i$	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-2} - \bar{\varepsilon}_i$	$y_{i,t-3} - \bar{y}_i$	$\Delta y_{i,t-2}$	$\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$	
$y_{i,t-1} - \bar{y}_i$	$\varepsilon_{i,t-1}$	$\varepsilon_{i,t-1} - \bar{\varepsilon}_i$	$y_{i,t-2} - \bar{y}_i$	$\Delta y_{i,t-1}$	$\varepsilon_{i,t-1} - \varepsilon_{i,t-2}$	$\Delta y_{i,t-2}$
$y_{it} - \bar{y}_i$	ε_{it}	$\varepsilon_{it} - \bar{\varepsilon}_i$	$y_{i,t-1} - \bar{y}_i$	Δy_{it}	$\varepsilon_{it} - \varepsilon_{i,t-1}$	$\Delta y_{i,t-1}$

To see which lags of the differences and levels of $y_{it} - \bar{y}_i$

are instruments for the first lag of the difference, $\Delta y_{i,t-1}$,

we need to see which lags are correlated with $\Delta y_{i,t-1}$ but not Δy_{it}

Why are lagged levels and changes in y instruments?

Fixed Effects				Differenced Fixed Effects		
Dependent Variable	Most recent error	Error Term	lag of Dependent Variable	Dep Var	Error Term	lag of Dep Var
$y_{i,t-3} - \bar{y}_i$	$\varepsilon_{i,t-3}$	$\varepsilon_{i,t-3} - \bar{\varepsilon}_i$				
$y_{i,t-2} - \bar{y}_i$	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-2} - \bar{\varepsilon}_i$	$y_{i,t-3} - \bar{y}_i$	$\Delta y_{i,t-2}$	$\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$	
$y_{i,t-1} - \bar{y}_i$	$\varepsilon_{i,t-1}$	$\varepsilon_{i,t-1} - \bar{\varepsilon}_i$	$y_{i,t-2} - \bar{y}_i$	$\Delta y_{i,t-1}$	$\varepsilon_{i,t-1} - \varepsilon_{i,t-2}$	$\Delta y_{i,t-2}$
$y_{it} - \bar{y}_i$	ε_{it}	$\varepsilon_{it} - \bar{\varepsilon}_i$	$y_{i,t-1} - \bar{y}_i$	Δy_{it}	$\varepsilon_{it} - \varepsilon_{i,t-1}$	$\Delta y_{i,t-1}$

Which error terms are incorporated in Δy_{it}?

Why are lagged levels and changes in y instruments?

Fixed Effects				Differenced Fixed Effects		
Dependent Variable	Most recent error	Error Term	lag of Dependent Variable	Dep Var	Error Term	lag of Dep Var
$y_{i,t-3} - \bar{y}_i$	$\varepsilon_{i,t-3}$	$\varepsilon_{i,t-3} - \bar{\varepsilon}_i$				
$y_{i,t-2} - \bar{y}_i$	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-2} - \bar{\varepsilon}_i$	$y_{i,t-3} - \bar{y}_i$	$\Delta y_{i,t-2}$	$\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$	
$y_{i,t-1} - \bar{y}_i$	$\varepsilon_{i,t-1}$	$\varepsilon_{i,t-1} - \bar{\varepsilon}_i$	$y_{i,t-2} - \bar{y}_i$	$\Delta y_{i,t-1}$	$\varepsilon_{i,t-1} - \varepsilon_{i,t-2}$	$\Delta y_{i,t-2}$
$y_{it} - \bar{y}_i$	ε_{it}	$\varepsilon_{it} - \bar{\varepsilon}_i$	$y_{i,t-1} - \bar{y}_i$	Δy_{it}	$\varepsilon_{it} - \varepsilon_{i,t-1}$	$\Delta y_{i,t-1}$

Which error terms are incorporated in Δy_{it}?

The terms in yellow are differenced to make Δy_{it},

Why are lagged levels and changes in y instruments?

Fixed Effects				Differenced Fixed Effects		
Dependent Variable	Most recent error	Error Term	lag of Dependent Variable	Dep Var	Error Term	lag of Dep Var
$y_{i,t-3} - \bar{y}_i$	$\varepsilon_{i,t-3}$	$\varepsilon_{i,t-3} - \bar{\varepsilon}_i$				
$y_{i,t-2} - \bar{y}_i$	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-2} - \bar{\varepsilon}_i$	$y_{i,t-3} - \bar{y}_i$	$\Delta y_{i,t-2}$	$\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$	
$y_{i,t-1} - \bar{y}_i$	$\varepsilon_{i,t-1}$	$\varepsilon_{i,t-1} - \bar{\varepsilon}_i$	$y_{i,t-2} - \bar{y}_i$	$\Delta y_{i,t-1}$	$\varepsilon_{i,t-1} - \varepsilon_{i,t-2}$	$\Delta y_{i,t-2}$
$y_{it} - \bar{y}_i$	ε_{it}	$\varepsilon_{it} - \bar{\varepsilon}_i$	$y_{i,t-1} - \bar{y}_i$	Δy_{it}	$\varepsilon_{it} - \varepsilon_{i,t-1}$	$\Delta y_{i,t-1}$

Which error terms are incorporated in Δy_{it}?

The terms in yellow are differenced to make Δy_{it},

and so Δy_{it} includes *only* the difference of errors also in yellow

Why are lagged levels and changes in y instruments?

Fixed Effects				Differenced Fixed Effects		
Dependent Variable	Most recent error	Error Term	lag of Dependent Variable	Dep Var	Error Term	lag of Dep Var
$y_{i,t-3} - \bar{y}_i$	$\varepsilon_{i,t-3}$	$\varepsilon_{i,t-3} - \bar{\varepsilon}_i$				
$y_{i,t-2} - \bar{y}_i$	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-2} - \bar{\varepsilon}_i$	$y_{i,t-3} - \bar{y}_i$	$\Delta y_{i,t-2}$	$\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$	
$y_{i,t-1} - \bar{y}_i$	$\varepsilon_{i,t-1}$	$\varepsilon_{i,t-1} - \bar{\varepsilon}_i$	$y_{i,t-2} - \bar{y}_i$	$\Delta y_{i,t-1}$	$\varepsilon_{i,t-1} - \varepsilon_{i,t-2}$	$\Delta y_{i,t-2}$
$y_{it} - \bar{y}_i$	ε_{it}	$\varepsilon_{it} - \bar{\varepsilon}_i$	$y_{i,t-1} - \bar{y}_i$	Δy_{it}	$\varepsilon_{it} - \varepsilon_{i,t-1}$	$\Delta y_{i,t-1}$

Is the term in red, $\Delta y_{i,t-2}$, an instrument?

Why are lagged levels and changes in y instruments?

Fixed Effects				Differenced Fixed Effects		
Dependent Variable	Most recent error	Error Term	lag of Dependent Variable	Dep Var	Error Term	lag of Dep Var
$y_{i,t-3} - \bar{y}_i$	$\varepsilon_{i,t-3}$	$\varepsilon_{i,t-3} - \bar{\varepsilon}_i$				
$y_{i,t-2} - \bar{y}_i$	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-2} - \bar{\varepsilon}_i$	$y_{i,t-3} - \bar{y}_i$	$\Delta y_{i,t-2}$	$\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$	
$y_{i,t-1} - \bar{y}_i$	$\varepsilon_{i,t-1}$	$\varepsilon_{i,t-1} - \bar{\varepsilon}_i$	$y_{i,t-2} - \bar{y}_i$	$\Delta y_{i,t-1}$	$\varepsilon_{i,t-1} - \varepsilon_{i,t-2}$	$\Delta y_{i,t-2}$
$y_{it} - \bar{y}_i$	ε_{it}	$\varepsilon_{it} - \bar{\varepsilon}_i$	$y_{i,t-1} - \bar{y}_i$	Δy_{it}	$\varepsilon_{it} - \varepsilon_{i,t-1}$	$\Delta y_{i,t-1}$

Is the term in red, $\Delta y_{i,t-2}$, an instrument?

The terms in red are differenced to make $\Delta y_{i,t-2}$,

Why are lagged levels and changes in y instruments?

Fixed Effects				Differenced Fixed Effects		
Dependent Variable	Most recent error	Error Term	lag of Dependent Variable	Dep Var	Error Term	lag of Dep Var
$y_{i,t-3} - \bar{y}_i$	$\varepsilon_{i,t-3}$	$\varepsilon_{i,t-3} - \bar{\varepsilon}_i$				
$y_{i,t-2} - \bar{y}_i$	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-2} - \bar{\varepsilon}_i$	$y_{i,t-3} - \bar{y}_i$	$\Delta y_{i,t-2}$	$\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$	
$y_{i,t-1} - \bar{y}_i$	$\varepsilon_{i,t-1}$	$\varepsilon_{i,t-1} - \bar{\varepsilon}_i$	$y_{i,t-2} - \bar{y}_i$	$\Delta y_{i,t-1}$	$\varepsilon_{i,t-1} - \varepsilon_{i,t-2}$	$\Delta y_{i,t-2}$
$y_{it} - \bar{y}_i$	ε_{it}	$\varepsilon_{it} - \bar{\varepsilon}_i$	$y_{i,t-1} - \bar{y}_i$	Δy_{it}	$\varepsilon_{it} - \varepsilon_{i,t-1}$	$\Delta y_{i,t-1}$

Is the term in red, $\Delta y_{i,t-2}$, an instrument?

The terms in red are differenced to make $\Delta y_{i,t-2}$,

and so $\Delta y_{i,t-2}$ includes *only* the difference of errors in red

And there is no correlation between $\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$ and $\varepsilon_{it} - \varepsilon_{i,t-1}$

So $\Delta y_{i,t-2}$ is an instrument for $\Delta y_{i,t-1}$ in a model of Δy_{it}

Why are lagged levels and changes in y instruments?

Fixed Effects				Differenced Fixed Effects		
Dependent Variable	Most recent error	Error Term	lag of Dependent Variable	Dep Var	Error Term	lag of Dep Var
$y_{i,t-3} - \bar{y}_i$	$\varepsilon_{i,t-3}$	$\varepsilon_{i,t-3} - \bar{\varepsilon}_i$				
$y_{i,t-2} - \bar{y}_i$	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-2} - \bar{\varepsilon}_i$	$y_{i,t-3} - \bar{y}_i$	$\Delta y_{i,t-2}$	$\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$	
$y_{i,t-1} - \bar{y}_i$	$\varepsilon_{i,t-1}$	$\varepsilon_{i,t-1} - \bar{\varepsilon}_i$	$y_{i,t-2} - \bar{y}_i$	$\Delta y_{i,t-1}$	$\varepsilon_{i,t-1} - \varepsilon_{i,t-2}$	$\Delta y_{i,t-2}$
$y_{it} - \bar{y}_i$	ε_{it}	$\varepsilon_{it} - \bar{\varepsilon}_i$	$y_{i,t-1} - \bar{y}_i$	Δy_{it}	$\varepsilon_{it} - \varepsilon_{i,t-1}$	$\Delta y_{i,t-1}$

By the same logic, $y_{i,t-2} - \bar{y}_i$ and $y_{i,t-3} - \bar{y}_i$ are also instruments for $\Delta y_{i,t-1}$

If we only can use one set of instruments,
Arellano (1989) suggests these lagged levels are a better choice

IV estimation for Panel

Arellano & Bond noticed that within a dataset with N units and T periods, later observations have more lags available as instruments

In terms of differences *as instruments:*

For $T = 3$ differenced periods, there are no instruments

For $T = 4$ differenced periods, can create: $\Delta y_{i,t-2}$

For $T = 5$ differenced periods, can create: $\Delta y_{i,t-2}$, $\Delta y_{i,t-3}$

For $T = 6$ differenced periods, can create: $\Delta y_{i,t-2}$, $\Delta y_{i,t-3}$, $\Delta y_{i,t-4}$

So if we were to use lagged differences as instruments, we would be limited to panels of $T = 4$ or longer

But levels work better, so. . .

IV estimation for Panel

In terms of levels *as instruments:*

For $T = 2$ differenced periods, there are no instruments

For $T = 3$ differenced periods, can create: $y_{i,t-2}$

For $T = 4$ differenced periods, can create: $y_{i,t-2}$, $y_{i,t-3}$

For $T = 5$ differenced periods, can create: $y_{i,t-2}$, $y_{i,t-3}$, $y_{i,t-4}$

So we need to have at least 3 periods to use Arellano & Bond's estimator

Estimation is by Generalized Method of Moments (GMM),
because the usual IV regression can't handle the varying number of instruments

Basic idea of GMM for small T panel data: find, by iterative search,
the β's and ϕ's that minimize the distance between $\Delta^d y_{it}$ and $\mathrm{E}(\Delta^d y_{it})$,
instrumenting for $\Delta y_{i,t-1}$

GMM models allow a separate equation for each period,
so each could potentially have different numbers of instruments

IV estimation for Panel

But we have some left over instruments (the lagged differences). . .

For $T = 4$ differenced periods, can create: $\Delta y_{i,t-2}$

For $T = 5$ differenced periods, can create: $\Delta y_{i,t-2}$, $\Delta y_{i,t-3}$

For $T = 6$ differenced periods, can create: $\Delta y_{i,t-2}$, $\Delta y_{i,t-3}$, $\Delta y_{i,t-4}$

Can we use these to make the estimator better, assuming we have at least $T = 4$?

IV estimation for Panel

Two main options for panel GMM:

Arellano-Bond's **difference GMM** estimator uses Δy_{it} as the outcome and all available lagged levels as instruments in each period

Requires at least $T = 3$ or more, better as T grows
(but panel GMM isn't necessary at all for large T)

IV estimation for Panel

Two main options for panel GMM:

Arellano-Bond's **difference GMM** estimator uses Δy_{it} as the outcome and all available lagged levels as instruments in each period

Requires at least $T = 3$ or more, better as T grows
(but panel GMM isn't necessary at all for large T)

Arellano-Blover/Blundell-Bond's **system GMM** estimator adds in available the lagged difference $\Delta y_{i,t-2}$, etc., as additional instruments in a dual set of observations using the outcome y_{it}

As a result, system GMM models the levels (and differences) of y;
fixed effects are swept out of the *instruments*

(Assumption to do this: changes in instruments are uncorrelated with fixed effects)

System GMM is better for small T but requires at least $T = 4$

Both methods are sensitive to small changes in assumptions,
including the inclusion of too many weak instruments

IV estimation for Panel

Note that for either system or difference GMM,
we need to be careful with any AR(p) processes left *after* differencing

That is, if we have an ARIMA(p,1,0), $p > 0$,
very recent lags of y or Δy will not be instruments:

If $\Delta y_{i,t}$ is AR(1), will need to start one period earlier on instruments

I.e., if $\Delta y_{i,t}$ is AR(1), then first available instrument is $y_{i,t-3}$

if $\Delta y_{i,t}$ is AR(2), then first available instrument is $y_{i,t-4}$

etc.

In `R`, these estimators are available using `pgmm()` in the `plm` library

Note that `pgmm()` requires a special kind of data frame, created with `pdata.frame()`, so that it knows what the units and periods are

Dynamic Panel Example: Cigarette Taxes

Common application for DPD models *a la* Arellano-Bond: policy "experiments"

Idea: major public policies are usually not randomly assigned

But maybe most confounders are either unit-characteristics or period shocks

Cigarette taxes vary by state in the US:
Taxes discourage unhealthy behavior/raise revenue

But states vary in public attitudes towards state intervention and taxes

And over time, social acceptance of smoking is waning as (public) health consequences become better known

Dynamic Panel Example: Cigarette Taxes

Conditioning on state and year effects, and perhaps a few time-varying controls, maybe tax rates are exogenous to smoking behavior

(A big assumption, but hard to do better with observational data)

So a panel data model with state and year effects might isolate better estimates of effect of tax variation on smoking behavior

What kinds of omitted variables could still confound this model?

What kinds of variation are we exploiting to estimate the causal effect of taxes on smoking?

Dynamic Panel Example: Cigarette Taxes

What kinds of omitted variables could still confound this model?

Difference or system GMM could be confounded by:

Time and state-varying covariates correlated with both taxes and consumption

Dynamic Panel Example: Cigarette Taxes

What kinds of omitted variables could still confound this model?

Difference or system GMM could be confounded by:

Time and state-varying covariates correlated with both taxes and consumption

One candidate for a confounder:
Over time, social stigma/pressure not to smoke has risen in most states

This social pressure likely directly reduces smoking
and lowers public opposition to cigarette taxes and other anti-smoking policies

If people in different states come to socially punish smoking at different rates over time, tax effects are confounded with social forces/other state & local policies

Dynamic Panel Example: Cigarette Taxes

What kinds of omitted variables could still confound this model?

System GMM could also be confounded by:

Correlations between individual time-invariant characteristics
and changes over time in the rate of change in y (and hence the instruments),
after conditioning on observed covariates

Example: Over the study, most states reduce their tobacco consumption

After controlling for measured covariates x_{it},
suppose some states are early *leaders* in reduction and other states early *laggards*

System GMM will be biased if those early laggards later tend to reduce smoking faster (on the scale of y_{it}) than early leaders as a form of "catch-up"

Note we could look for evidence of this in the data,
and either controls or appropriate transformation of y_{it} might help

Dynamic Panel Example: Cigarette Taxes

What kinds of variation are we exploiting to estimate the causal effect of taxes on smoking?

Variation in legislative cycle?

Variation in budget crises?

Variation in parties in office?

Do these suggest other avenues of confounding?

Dynamic panel models narrow the scope of confounding, but do not eliminate it

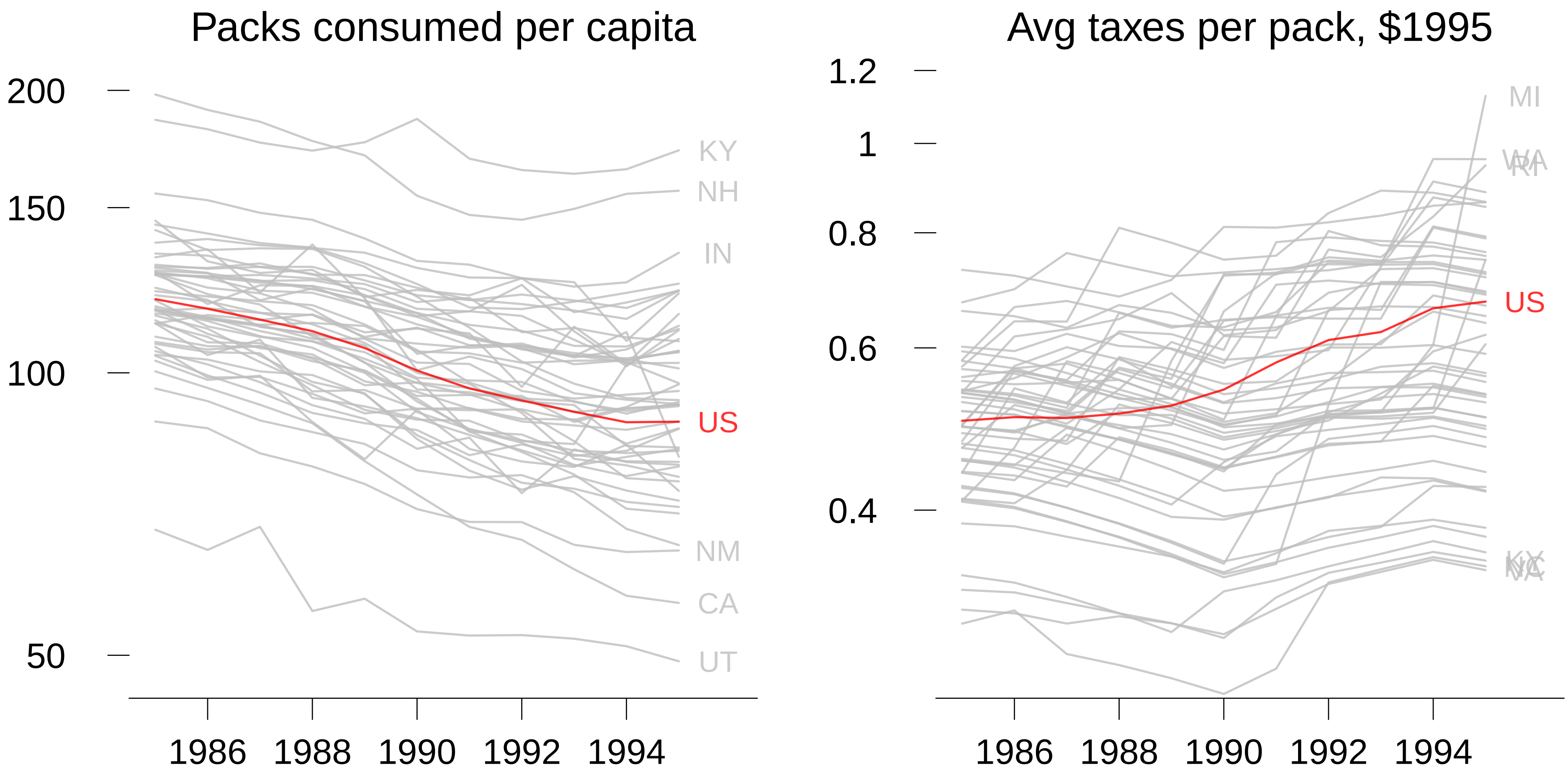
Dynamic Panel Example: Cigarette Taxes

A panel data model with state and year effects might isolate better estimates of effect of tax variation on smoking behavior

From Jonathan Gruber (MIT), we have state-year data (1985–1995) on:

- packs smoked per capita annually
- average price per pack
- average taxes per pack
- household income per capita (1995 dollars)

We will combine prices and taxes,
assuming tax increases operate like other price changes



Code for this display is in the sample script – complex but powerful

Consider displaying all of the data, even for panel

We see strong trends in smoking and in taxes in opposite directions
Is the connection causal or spurious?

Dynamic Panel Example: Cigarette Taxes

Specification 1 (Linear):

$$\text{packpc}_{it} = \alpha_i + \tau_t + \beta_1 \text{price}_{it} + \beta_2 \text{income}_{it} + \phi_1 \text{packpc}_{i,t-1} + \varepsilon_{it}$$

A linear specification is simple, but probably inappropriate

Dynamic Panel Example: Cigarette Taxes

Specification 1 (Linear):

$$\mathrm{packpc}_{it} = \alpha_i + \tau_t + \beta_1 \mathrm{price}_{it} + \beta_2 \mathrm{income}_{it} + \phi_1 \mathrm{packpc}_{i,t-1} + \varepsilon_{it}$$

A linear specification is simple, but probably inappropriate

Taxes affect elasticities:
a percentage increase in price should yield a constant percent change in consumption

Specification 2 (Log-Log):

$$\log \mathrm{packpc}_{it} = \alpha_i^* + \tau_t^* + \beta_1^* \log \mathrm{price}_{it} + \beta_2^* \log \mathrm{income}_{it} + \phi_1^* \log \mathrm{packpc}_{i,t-1} + \varepsilon_{it}^*$$

While β_1 is a slope, in a log-log model, β_1^* is an *elasticity*,
the percent change in packspc for a percent change in price

Dynamic Panel Example: Cigarette Taxes

Specification 1 (Linear), *After differencing out fixed effects*:

$$\Delta \text{packpc}_{it} = \Delta \tau_t + \beta_1 \Delta \text{price}_{it} + \beta_2 \Delta \text{income}_{it} + \phi_1 \Delta \text{packpc}_{i,t-1} + \Delta \varepsilon_{it}$$

A linear specification is simple, but probably inappropriate

Dynamic Panel Example: Cigarette Taxes

Specification 1 (Linear), *After differencing out fixed effects*:

$$\Delta \text{packpc}_{it} = \Delta\tau_t + \beta_1 \Delta \text{price}_{it} + \beta_2 \Delta \text{income}_{it} + \phi_1 \Delta \text{packpc}_{i,t-1} + \Delta\varepsilon_{it}$$

A linear specification is simple, but probably inappropriate

Taxes affect elasticities:
a percentage increase in price should yield a constant percent change in consumption

Specification 2 (Log-Log) , *After differencing out fixed effects*:

$$\Delta \log \text{packpc}_{it} = \Delta\tau_t^* + \beta_1^* \Delta \log \text{price}_{it} + \beta_2^* \Delta \log \text{income}_{it} + \phi_1^* \Delta \log \text{packpc}_{i,t-1} + \Delta\varepsilon_{it}^*$$

While β_1 is a slope, in a log-log model, β_1^* is an *elasticity*,
the percent change in packspc for a percent change in price

Dynamic Panel Example: Cigarette Taxes

Both specifications control for fixed intercepts *and* lagged outcomes

We have only 11 periods and need to sweep out fixed effects,
so Nickell bias is a concern

Let's use dynamic panel models (difference GMM and system GMM)

Dynamic Panel Example: Cigarette Taxes

Both specifications control for fixed intercepts *and* lagged outcomes

We have only 11 periods and need to sweep out fixed effects,
so Nickell bias is a concern

Let's use dynamic panel models (difference GMM and system GMM)

Model issues/assumptions to investigate:

1. Number/Weakness of instruments
2. Serial correlation of error terms – should be AR(1) only
3. Difference or system GMM (whether to use lagged differences as IV)
4. Use of year effects
5. Use of linear or log-log specification

Panel GMM is *very* sensitive to assumptions –
need to check GOF and substantive implications of each of the above

Dynamic Panel Models of Cigarette Consumption				
	1a	1b	1c	1d
Price	-0.18	-0.18	-0.18	-0.16
	0.03	0.03	0.03	0.03
Income	-0.48	-0.33	-0.26	0.11
	0.49	0.50	0.46	0.45
ϕ_1	0.64	0.65	0.66	0.70
	0.06	0.06	0.05	0.05
GMM	diff	diff	diff	diff
n of IVs	all	4	2	1
Year Effects				
State Effects	x	x	x	x
Sargan p	0.31	0.05	0.00	0.00
AR(2) errors p	0.57	0.59	0.62	0.72
N	48	48	48	48
T	11	11	11	11
NT used	432	432	432	432

First question: If we use a linear specification, difference GMM & no year effects, how many instruments do we need?

Cheat sheet

Sargan-Hansen test of overidentifying restrictions:

Exploits abundance of instruments (overidentification) to check whether model fits moments of data well; null is good fit, so low p indicates a poor fitting model

AR(2) serial correlation:

Panel GMM errors are AR(1) by construction;
anything higher violates model assumptions (and would require more lags)
high p indicates no evidence of AR(2) serial correlation

Both tests reported by `summary.pgmm()`

Dynamic Panel Models of Cigarette Consumption				
	1a	1b	1c	1d
Price	-0.18	-0.18	-0.18	-0.16
	0.03	0.03	0.03	0.03
Income	-0.48	-0.33	-0.26	0.11
	0.49	0.50	0.46	0.45
ϕ_1	0.64	0.65	0.66	0.70
	0.06	0.06	0.05	0.05
GMM	diff	diff	diff	diff
n of IVs	all	4	2	1
Year Effects				
State Effects	x	x	x	x
Sargan p	0.31	0.05	0.00	0.00
AR(2) errors p	0.57	0.59	0.62	0.72
N	48	48	48	48
T	11	11	11	11
NT used	432	432	432	432

Sargan tests show concern if we limit instruments to recent lags of $y_{i,t}$
Another worry: Sargan p biased $\rightarrow$ 1 as number of IVs increases. . .

Dynamic Panel Models of Cigarette Consumption				
	1a	1b	1c	1d
Price	-0.18	-0.18	-0.18	-0.16
	0.03	0.03	0.03	0.03
Income	-0.48	-0.33	-0.26	0.11
	0.49	0.50	0.46	0.45
ϕ_1	0.64	0.65	0.66	0.70
	0.06	0.06	0.05	0.05
GMM	diff	diff	diff	diff
n of IVs	all	4	2	1
Year Effects				
State Effects	x	x	x	x
Sargan p	0.31	0.05	0.00	0.00
AR(2) errors p	0.57	0.59	0.62	0.72
N	48	48	48	48
T	11	11	11	11
NT used	432	432	432	432

Still, results are little changed substantively

Dynamic Panel Models of Cigarette Consumption				
	1a	1b	1c	1d
Price	-0.18	-0.18	-0.18	-0.16
	0.03	0.03	0.03	0.03
Income	-0.48	-0.33	-0.26	0.11
	0.49	0.50	0.46	0.45
ϕ_1	0.64	0.65	0.66	0.70
	0.06	0.06	0.05	0.05
GMM	diff	diff	diff	diff
n of IVs	all	4	2	1
Year Effects				
State Effects	x	x	x	x
Sargan p	0.31	0.05	0.00	0.00
AR(2) errors p	0.57	0.59	0.62	0.72
N	48	48	48	48
T	11	11	11	11
NT used	432	432	432	432

Are we worried about unmodeled serial correlation?

Substantive interpretation

So what does this mean for our substantive question?

How much will smoking decline if we raise cigarette taxes?

Suppose we take the average state (which has about \$0.60/pack taxes in 1995), and double those taxes to \$1.20/pack

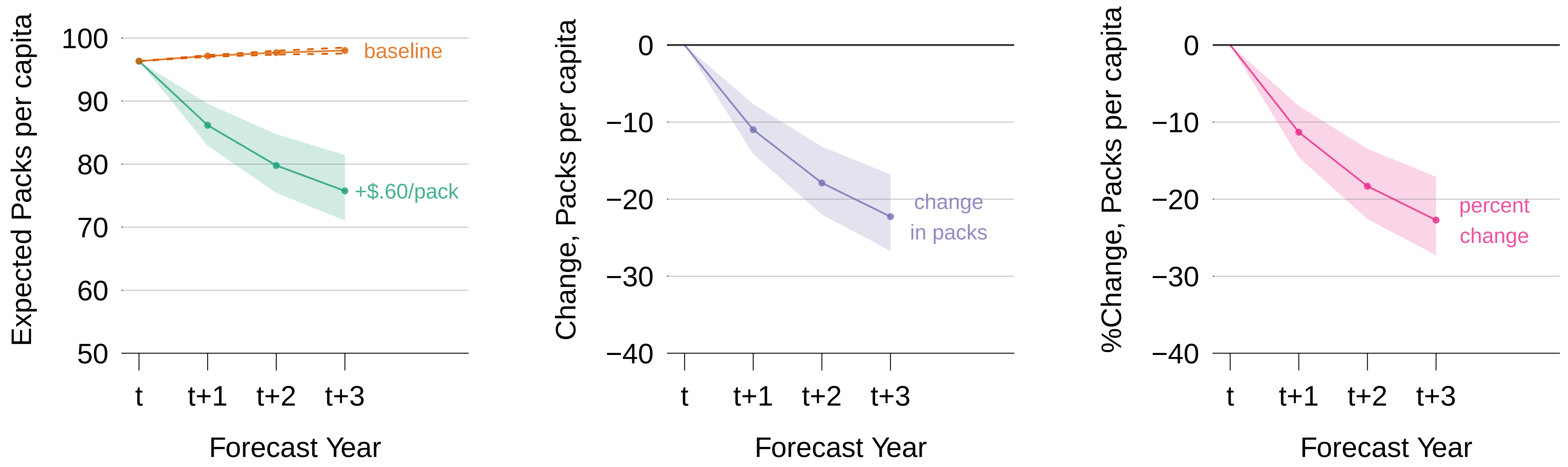
How much will smoking decline the next year?

Over the next three years?

Does this depend on our model?

What's the 95% confidence interval?

Cigarette Taxes & Consumption: 1a. Linear Difference GMM, Individual Effects



Once again, we turn to simulation,
applying our old friends `ldvsimev()`, `ldvsimfd()`, and `ldvsimrr()`

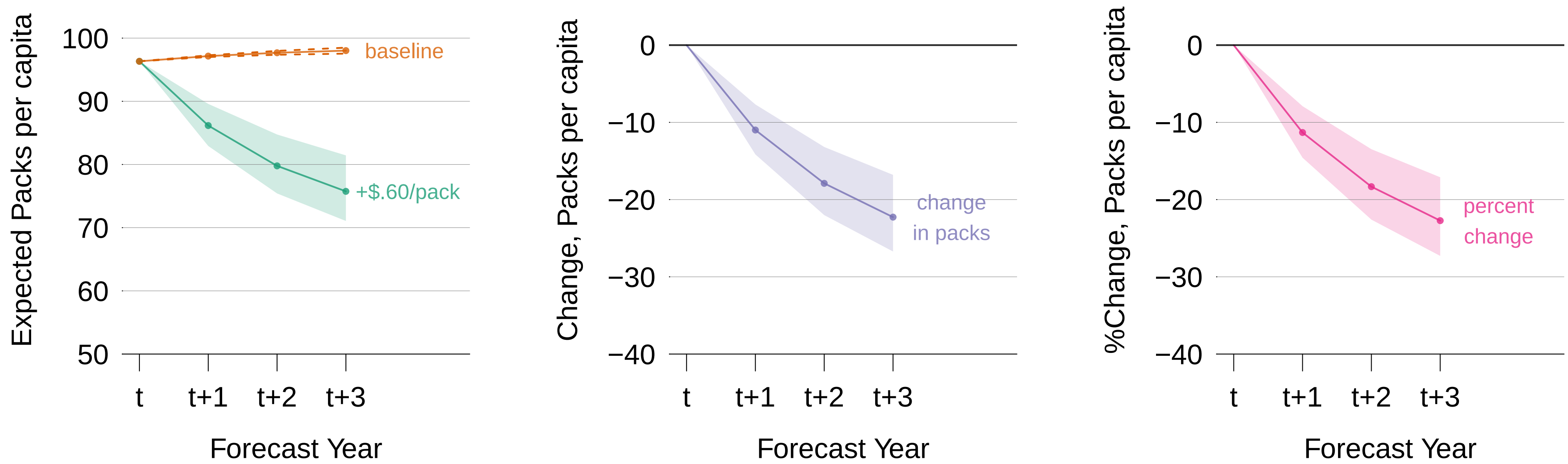
The effect of a doubled tax in the average state under model 1a is shown

Left plot: expected levels of smoking under the baseline and counterfactual

Middle plot: the *difference-in-difference* over time in packs consumed given a 60 cent increase in cigarette taxes

Right plot: percent change in packs consumed given the tax increase

Cigarette Taxes & Consumption: 1a. Linear Difference GMM, Individual Effects



Note: As you will see in section, coding these simulations is *very* painstaking

Panel GMM estimators are complex:
lots of choices of assumptions, and lots of different ways to deal with the data

Complex models: Sensitive to assumptions *and* coding errors!

Model checkpoint 1

1. Number/Weakness of instruments

not much difference here

2. Serial correlation of error terms – should be AR(1) only

model passes tests

Next to test. . .

3. Difference or system GMM (whether to use lagged differences as IV)
4. Use of year effects
5. Use of linear or log-log specification

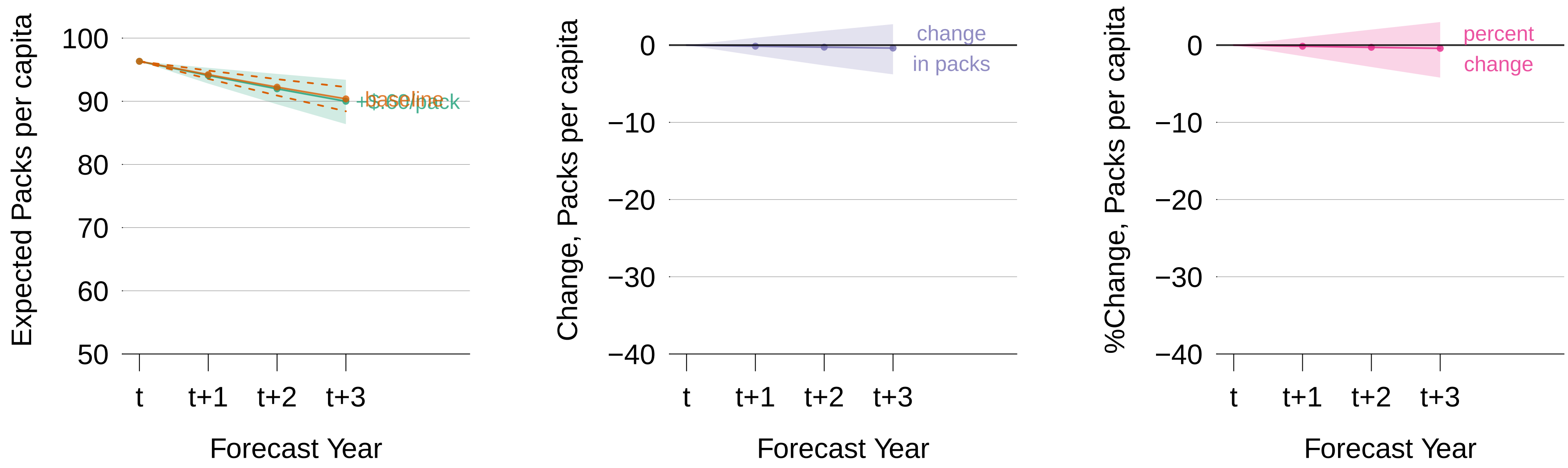
Dynamic Panel Models of Cigarette Consumption				
	1a	1e	1g	1h
Price	−0.18	0.00	−0.29	−0.10
	0.03	0.01	0.06	0.03
Income	−0.48	0.20	1.06	−0.02
	0.49	0.12	0.67	0.11
ϕ_1	0.64	0.94	0.25	0.91
	0.06	0.01	0.12	0.04
Intercept				23.28
				7.56
GMM	diff	sys	diff	sys
n of IVs	all	all	all	all
Year Effects			x	x
State Effects	x	x	x	x
Sargan p	0.31	0.74	0.39	0.81
AR(2) errors p	0.57	0.55	0.98	0.46
N	48	48	48	48
T	11	11	11	11
NT used	432	912	432	912

How do models 1a, 1e, 1g, and 1h differ in assumptions?

In goodness of fit?

In substantive implications?

Cigarette Taxes & Consumption: 1e. Linear System GMM, Individual Effects



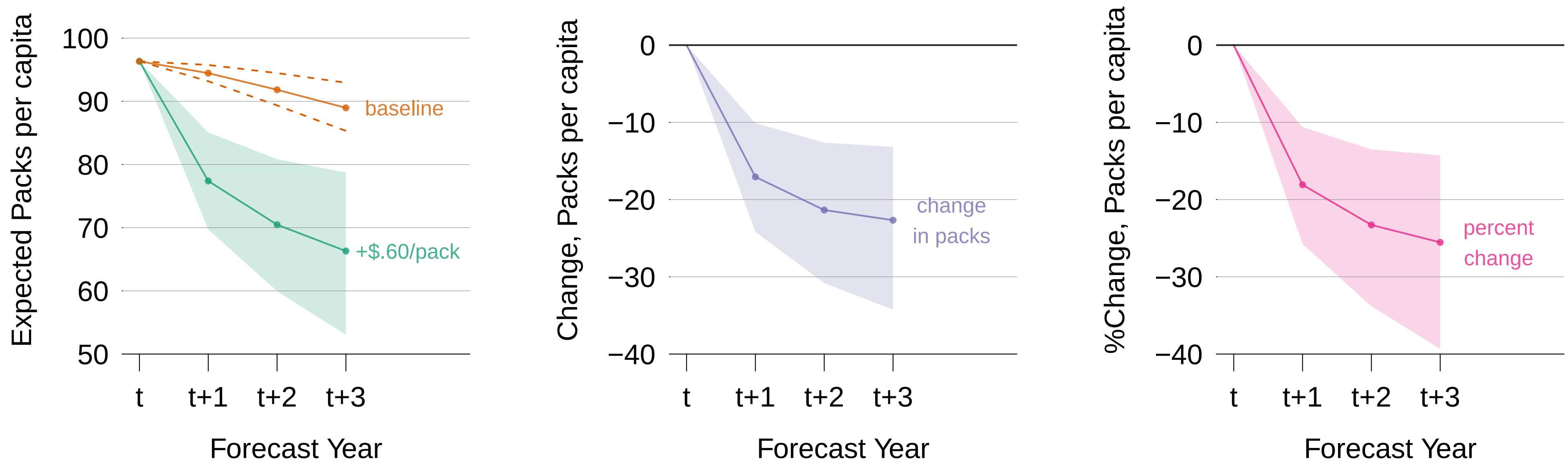
Model 1e (system GMM) exploits an additional kind of instrument:
lagged differences of y_{it} in addition to lagged levels

Goal is still coping with Nickell bias

Results are very sensitive:
In the linear, one-way model,
the effect of prices totally disappears in the system GMM estimates!

Maybe less important if we improve the rest of the model?

Cigarette Taxes & Consumption: 1g. Linear Difference GMM, Two−Way Effects



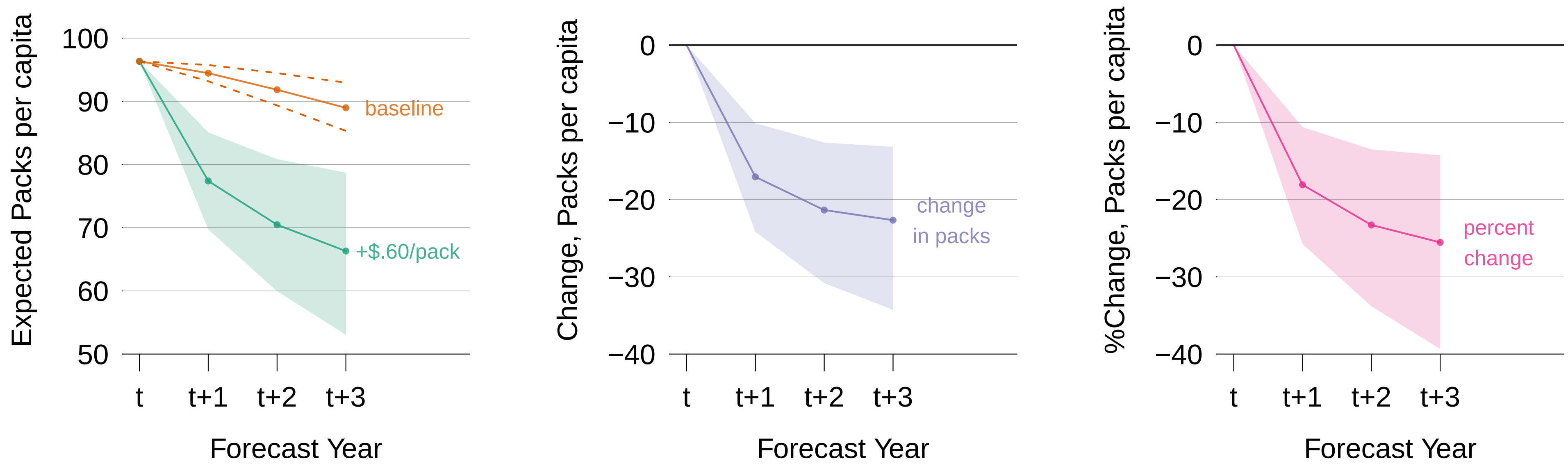
Model 1g adds year fixed effects to the difference estimator

Year fixed effects flexibly estimate the trend in smoking consumption

Strong reason to believe in an exogenous trend downwards due to health concerns and social pressure

Baseline smoking now shows a sharp decline, matching the behavior in the data

Cigarette Taxes & Consumption: 1g. Linear Difference GMM, Two−Way Effects

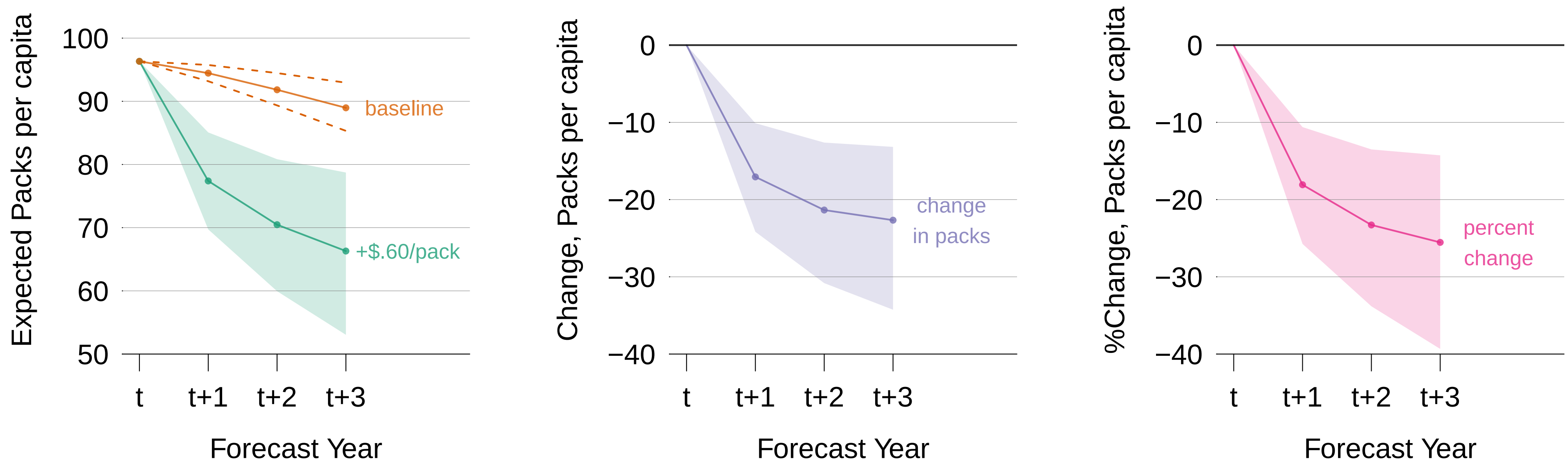


Two questions:

1. What trend do the year effects imply? How do we capture this in simulations? Could we substitute a linear trend?

2. These are big effects from a 60 cent tax. What would the model predict if taxes went up $2 dollars? Do you believe it? Is linearity reasonable here?

Cigarette Taxes & Consumption: 1g. Linear Difference GMM, Two−Way Effects

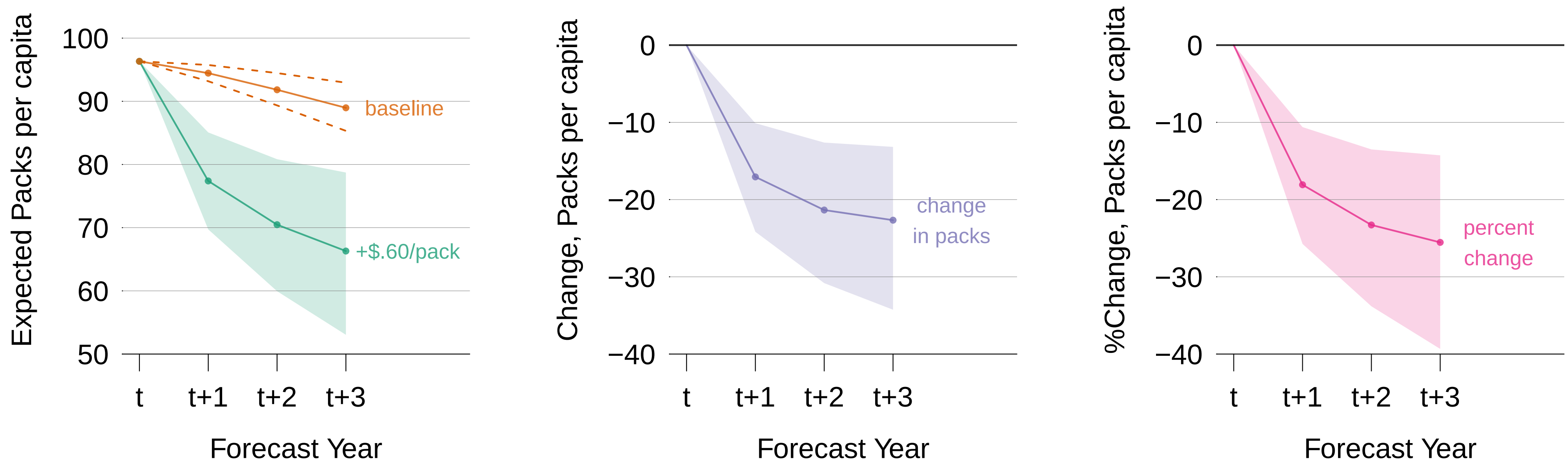


1. What trend do the year effects imply? How do we capture this in simulations? Could we substitute a linear trend?

We will return to this question in detail after settling on the rest of the specification (logs vs linear)

In these simulations, I assume that the near future follows a similar trend to that captured in recent year effects

Cigarette Taxes & Consumption: 1g. Linear Difference GMM, Two−Way Effects

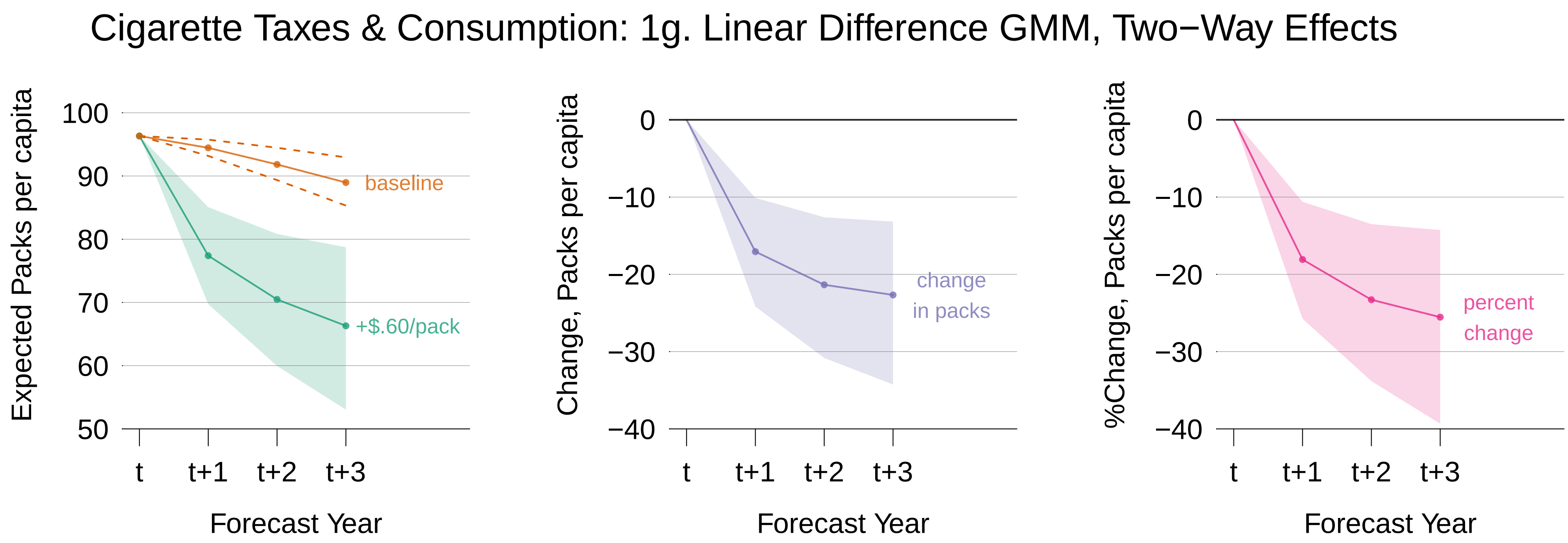


1. What trend do the year effects imply? How do we capture this in simulations? Could we substitute a linear trend?

A serious forecasting exercise would need to predict the *future* trend as well as model the effects of taxes

In our simulations, we just assume "recent" trends persist

This is already a big improvement over the models with no year effects, which tend to predict increased smoking over time from growing incomes!

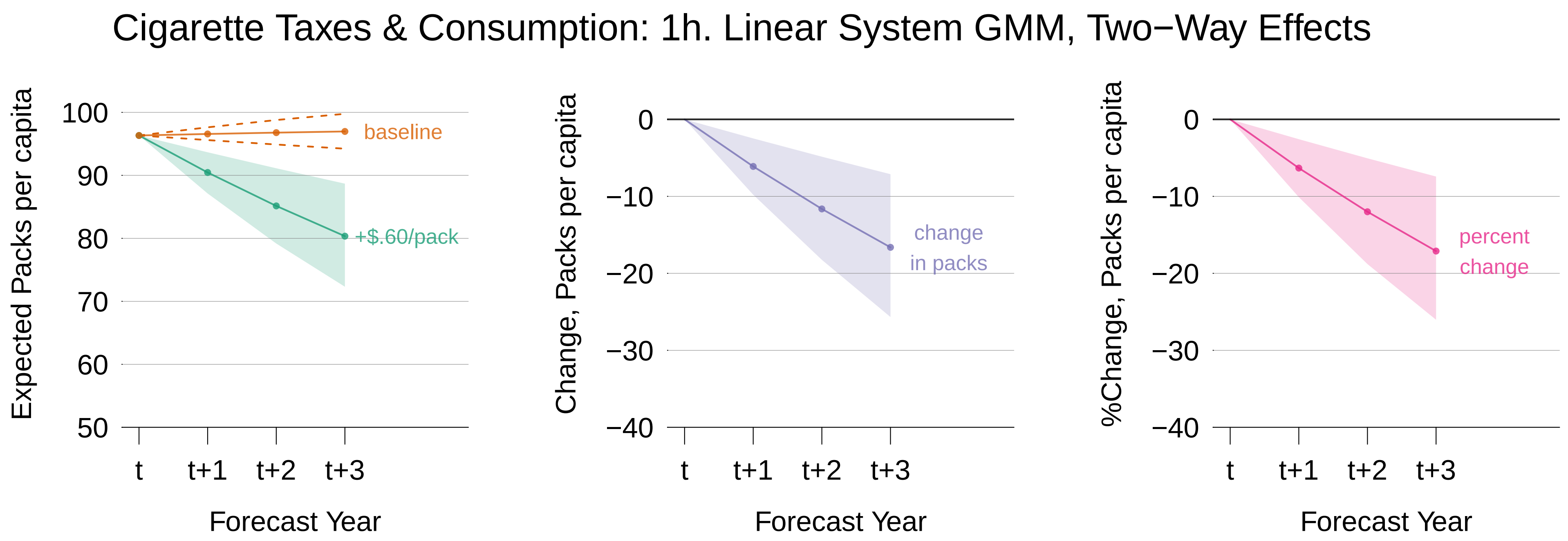


2. These are big effects from a 60 cent tax. What would the model predict if taxes went up $2 dollars? Do you believe it? Is linearity reasonable here?

A $2 dollar increase might reduce expected smoking to zero

That seems unreasonable –
some smokers will likely hold on even at very high taxes of $10 or more

Lower zero bound on smoking rate also reinforces the need for a log-log specification



Now we consider model 1h, linear system GMM with year effects

When we combine the year effects and the system GMM instruments, we find more agreement between difference and system GMM

But the magnitude of our effects is still sensitive to instrument choice; maybe if we log everything this will get better?

Model checkpoint 2

1. Number/Weakness of instruments

not much difference here

2. Serial correlation of error terms – should be AR(1) only

model passes tests

3. Difference or system GMM (whether to use lagged differences as IV)

substantive sensitivity, mitigated by better models

4. Use of year effects

strongly justified

Next to test. . .

5. Use of linear or log-log specification

Dynamic Panel Models of Cigarette Consumption				
	3a	3e	3g	3h
log Price	−0.31	0.01	−0.62	−0.17
	0.04	0.01	0.11	0.05
log Income	−0.07	−0.01	0.18	−0.01
	0.11	0.02	0.18	0.02
ϕ_1	0.67	0.99	0.33	0.95
	0.06	0.01	0.15	0.03
Intercept				1.09
				0.34
GMM	diff	sys	diff	sys
n of IVs	all	all	all	all
Year Effects			x	x
State Effects	x	x	x	x
Sargan p	0.39	0.75	0.50	0.85
AR(2) errors p	0.95	0.33	0.60	0.27
N	48	48	48	48
T	11	11	11	11
NT used	432	912	432	912

Interpreting elasticities

How do we interpret the coefficients in this table?

Because the model is logged in the outcome and covariates,
the coefficients are elasticities

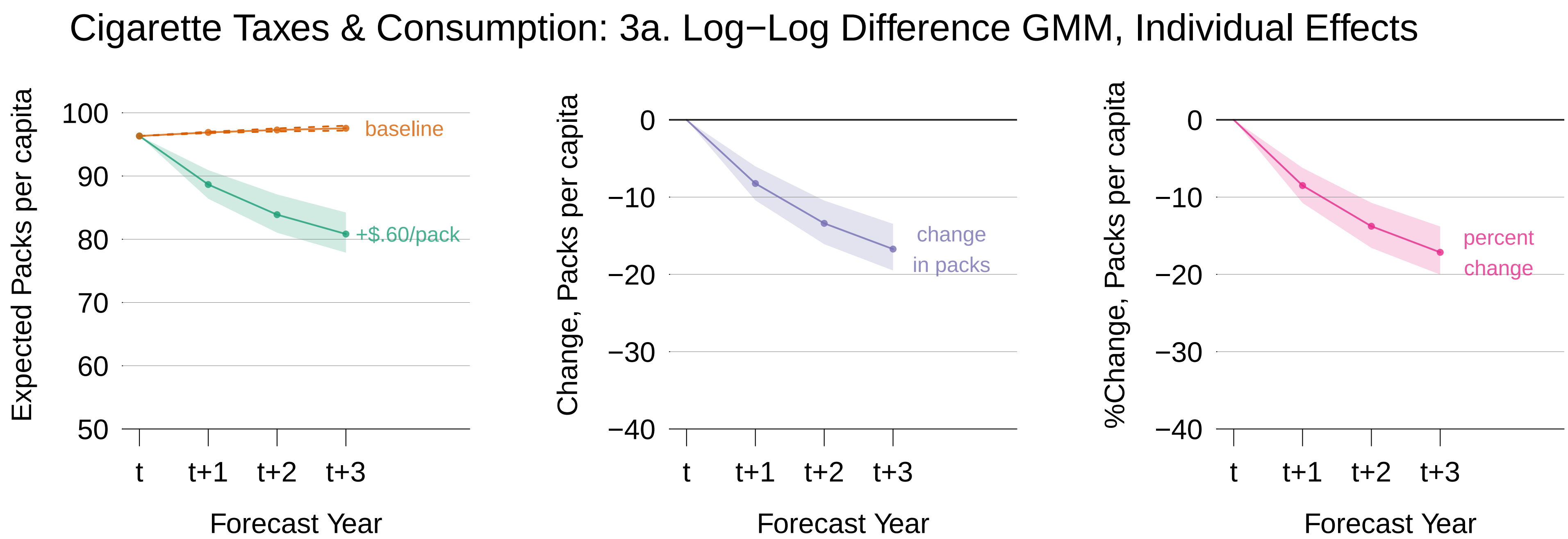
E.g., in model 3a, a 1% increase in taxes reduces consumption 0.31%

From the coefficients, it's clear results are sensitive to model assumptions

But note that the estimates of β_1 and ϕ_1 are negatively correlated

This suggests that sometimes there is a big immediate effect that doesn't grow,
while other times there is a small immediate effect that gets bigger over time

How different are these results, really?



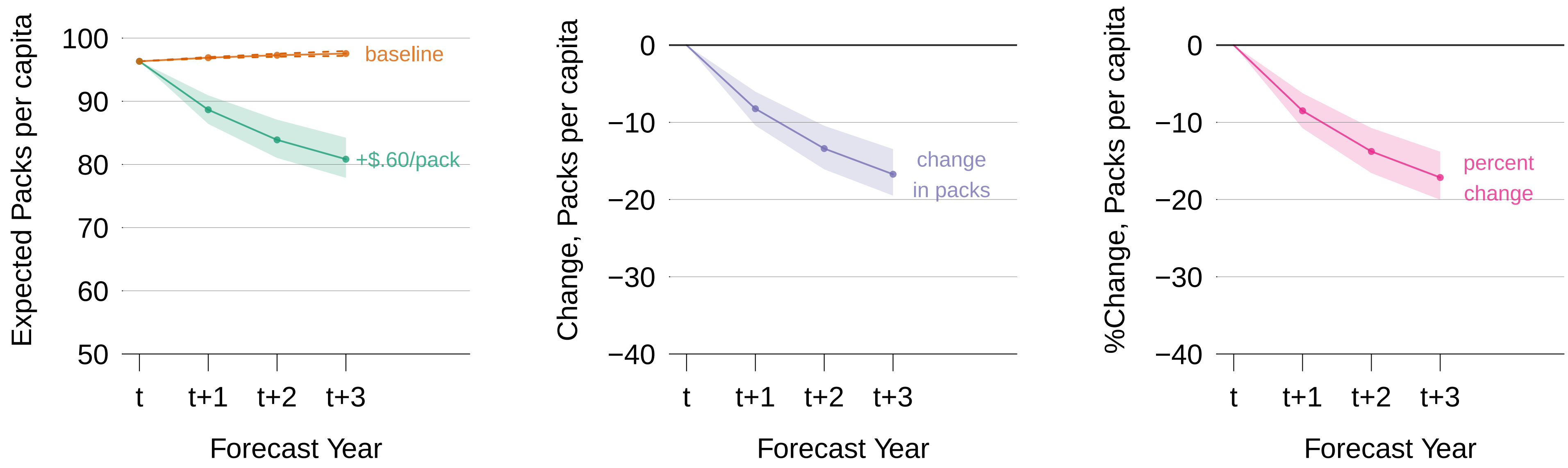
Now the effects of taxes are *nonlinear*

Reasonable: Otherwise a $2 tax increase would completely wipe out smoking; in reality, some would be willing to pay

(Intuition check: How should log-log estimates compare to linear ones for a $0.1 increase? A $10 increase?)

This also implies the same tax increase in different states would have different effects, depending on other covariates and the initial tax rate

Cigarette Taxes & Consumption: 3a. Log−Log Difference GMM, Individual Effects

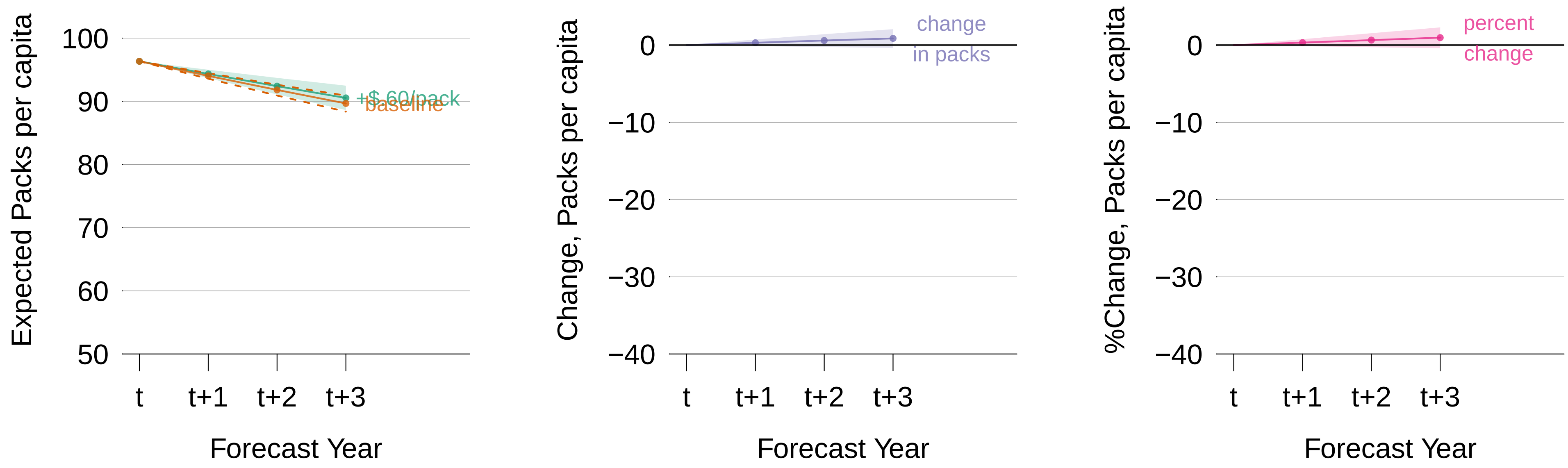


Constructing these simulations –
of a differenced, logged outcome from a complex model –
is *very* tricky

A single coding error will drastically change results

Not a case for simple copy and paste of `simcf` code:
think deeply about what each model is doing

Cigarette Taxes & Consumption: 3e. Log−Log System GMM, Individual Effects



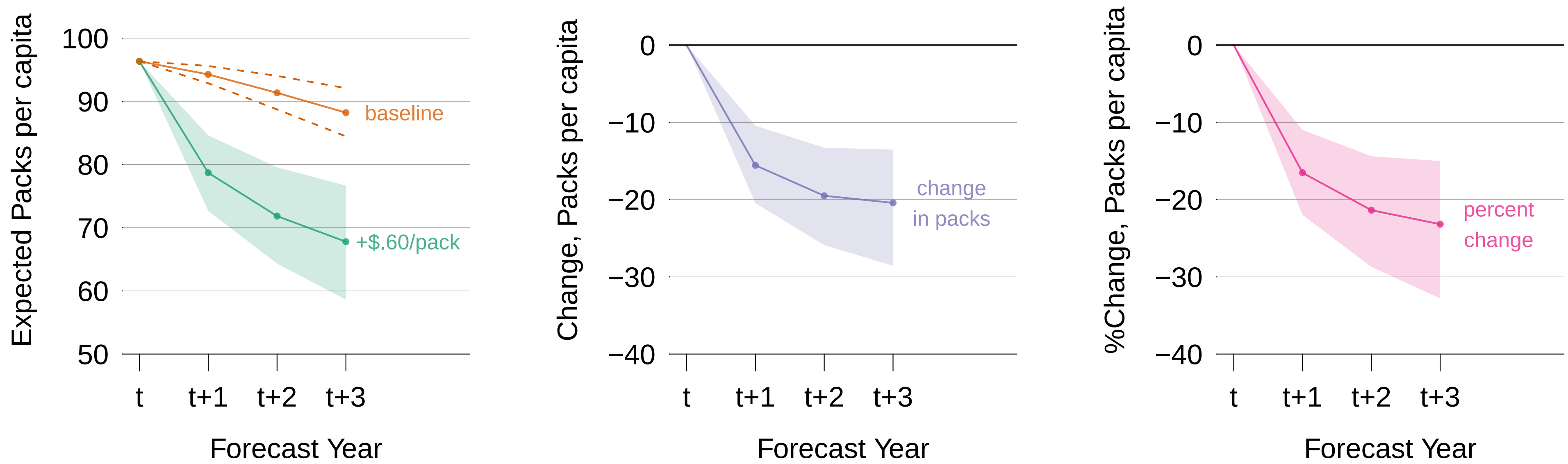
In a system GMM with no year effects, we find no effect of taxes

But we have strong reason to think there is a yearly trend,
so let's restore the year effects

System GMM is biased when change over time in y correlates with unit effects

Purging potential over-time confounding is thus even more important in this model

Cigarette Taxes & Consumption: 3g. Log−Log Difference GMM, Two−Way Effects



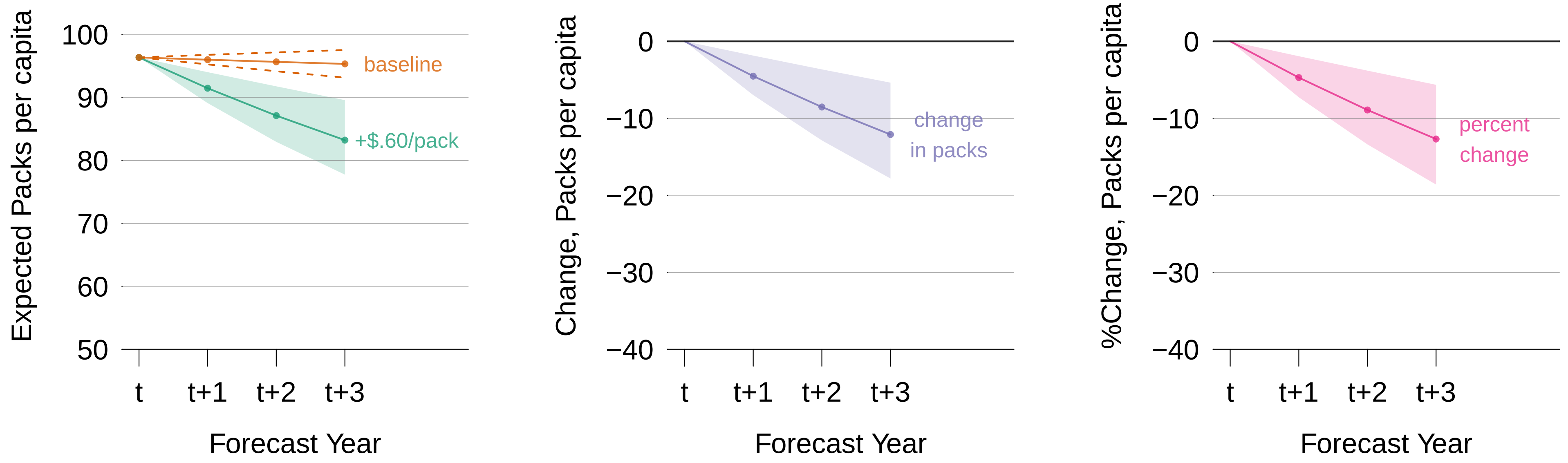
Adding year effects,
The *difference* GMM with year effects finds a big effect
that diminishes over time

More than 20% decline in total consumption after three years

Model assumptions and result are most plausible yet

This model is probably our best bet,
but let's see what happens when we add system GMM instruments

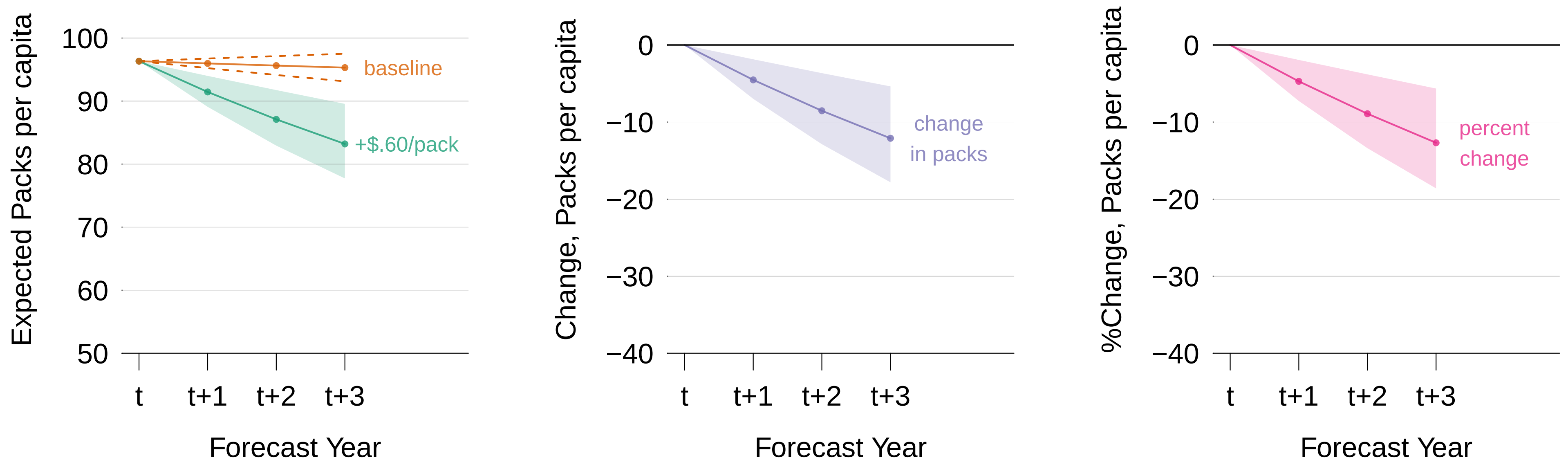
Cigarette Taxes & Consumption: 3h. Log−Log System GMM, Two−Way Effects



But the system GMM with year effects also find a substantial impact

While the immediate impact is different in these models,
over time they converge

Cigarette Taxes & Consumption: 3h. Log−Log System GMM, Two−Way Effects

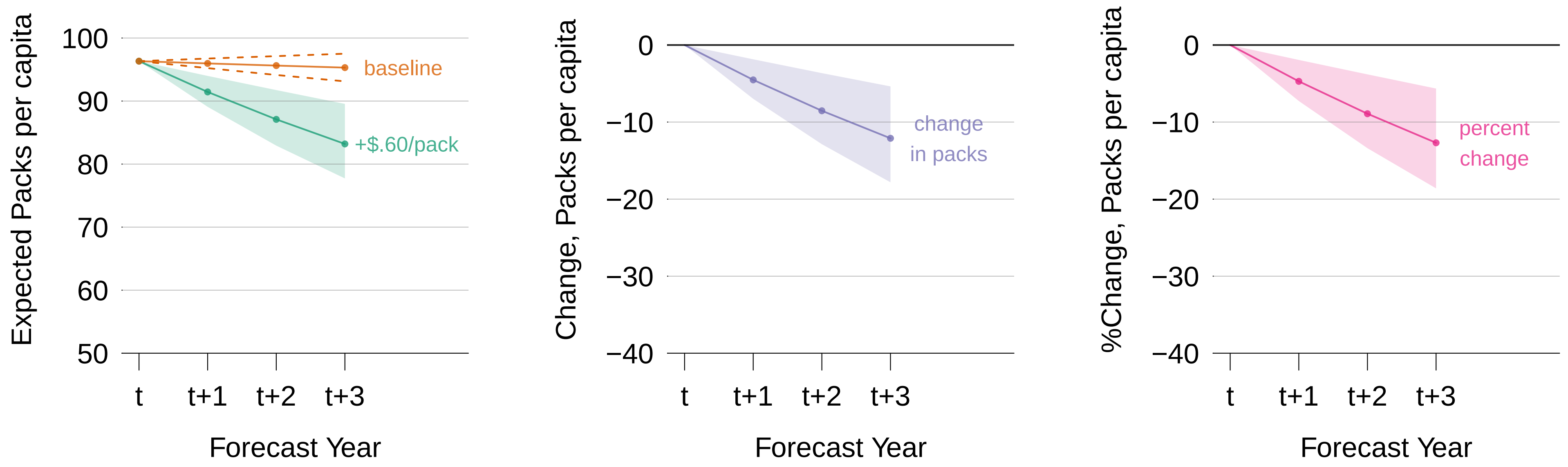


Controlling for years and logging the outcome helped remedy sources of bias in system GMM

E.g., on a linear scale, if all states are lowering smoking, but at different rates, then laggards must eventually catch up because of the zero lower bound

On a log-log scale, this is a smaller problem – leaders could maintain high percentage reductions even as levels get low, at least for a while

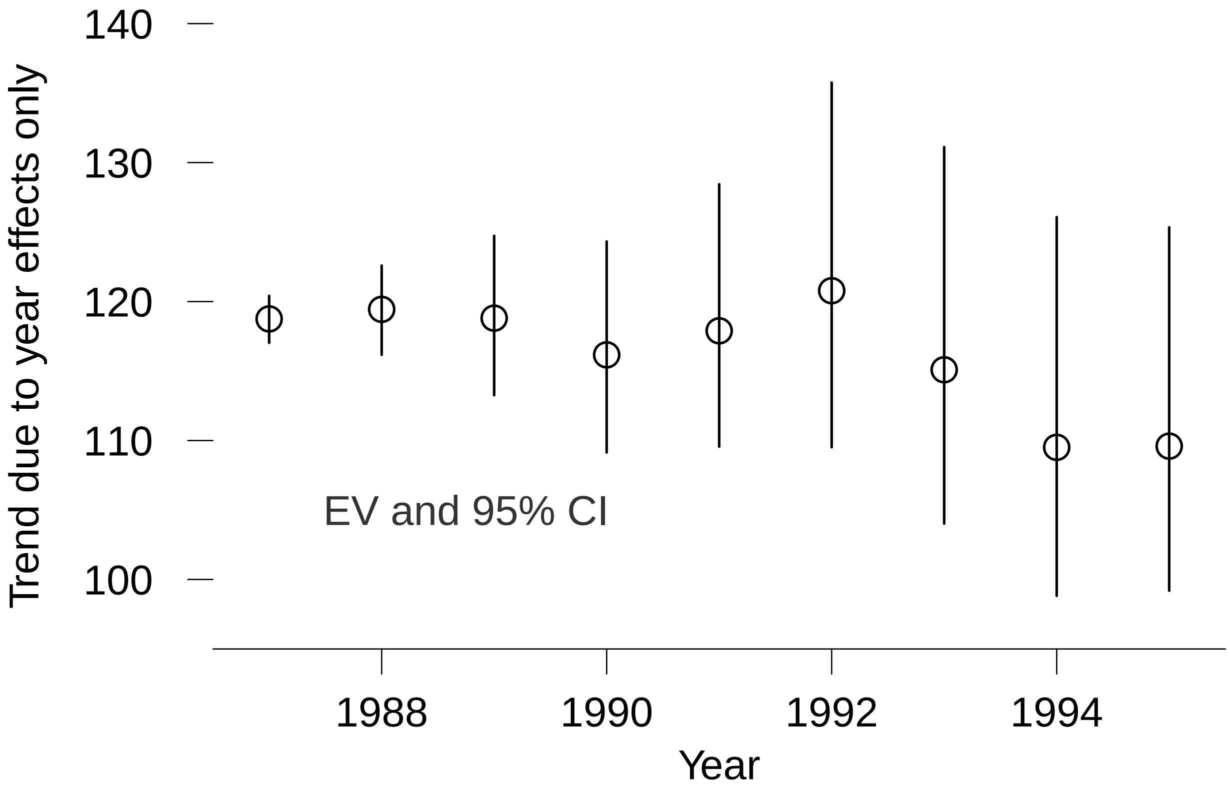
Cigarette Taxes & Consumption: 3h. Log−Log System GMM, Two−Way Effects



Panel GMM is very sensitive to assumptions –
lots to investigate even in the simplest model

(we should probably check in these final models – 1g and 1h – on the number of instruments again)

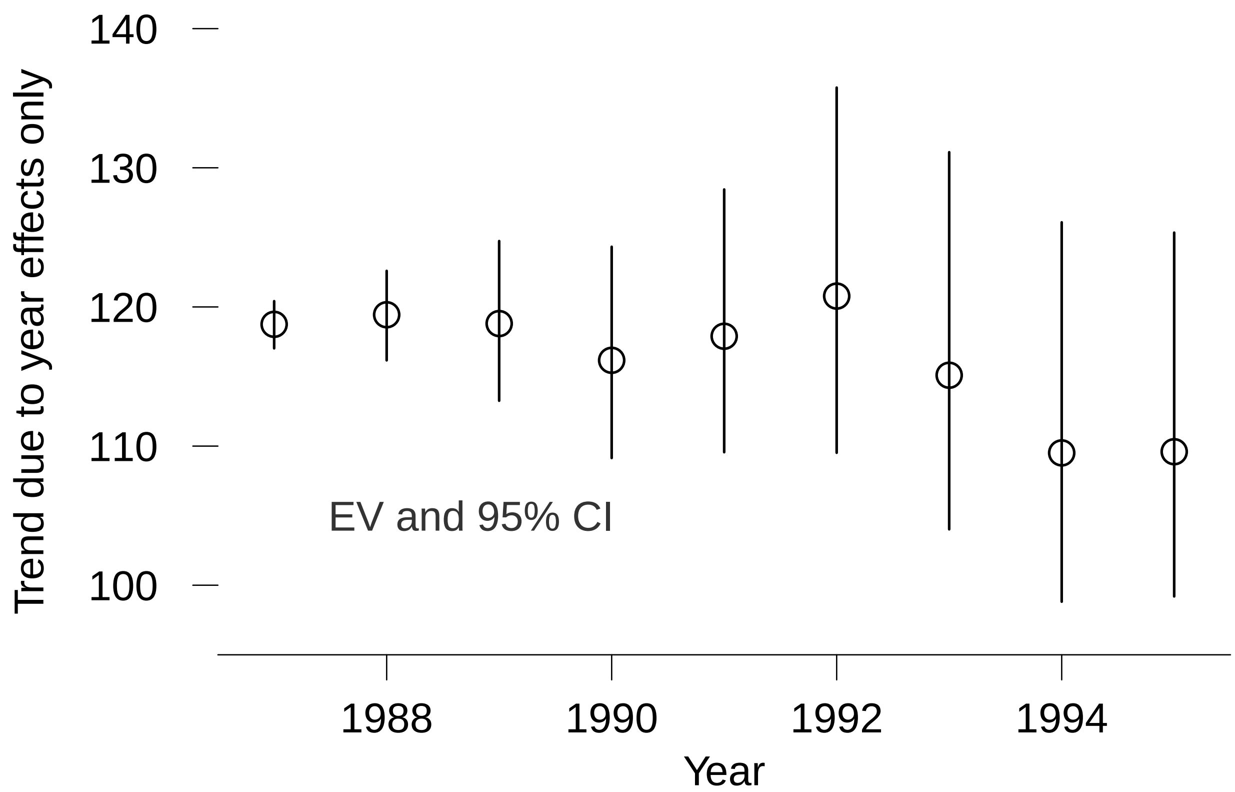
Very hard to tell whether differences are substantive without thinking hard about what the model means/doing some simulations



Aside: What trend do the year effects imply? How do we capture this in simulations? Could we substitute a linear trend?

Because the model is differenced, the year effects are differences of τ's;
for example, the 1993 effect is $\tau_{1993} - \tau_{1992}$

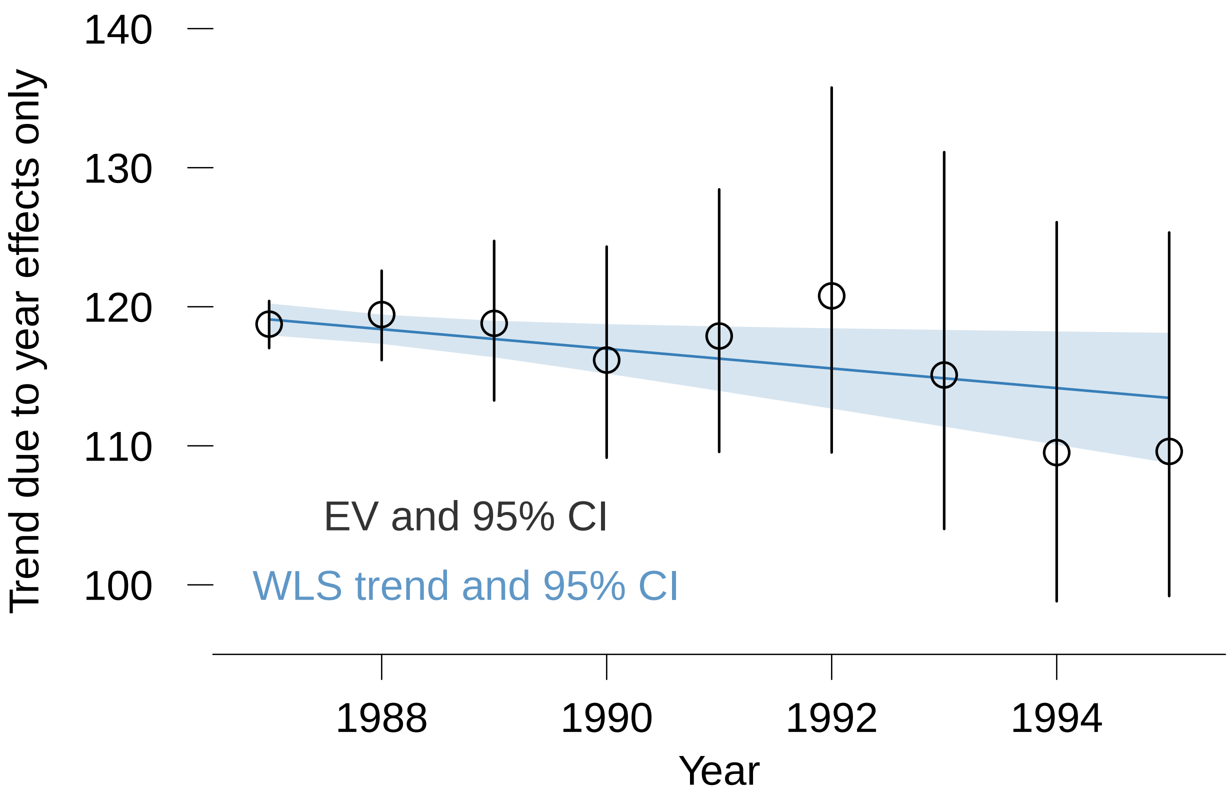
Because the model is also in logs,
it's easiest to see the year effects through simulation



Aside: What trend do the year effects imply? How do we capture this in simulations? Could we substitute a linear trend?

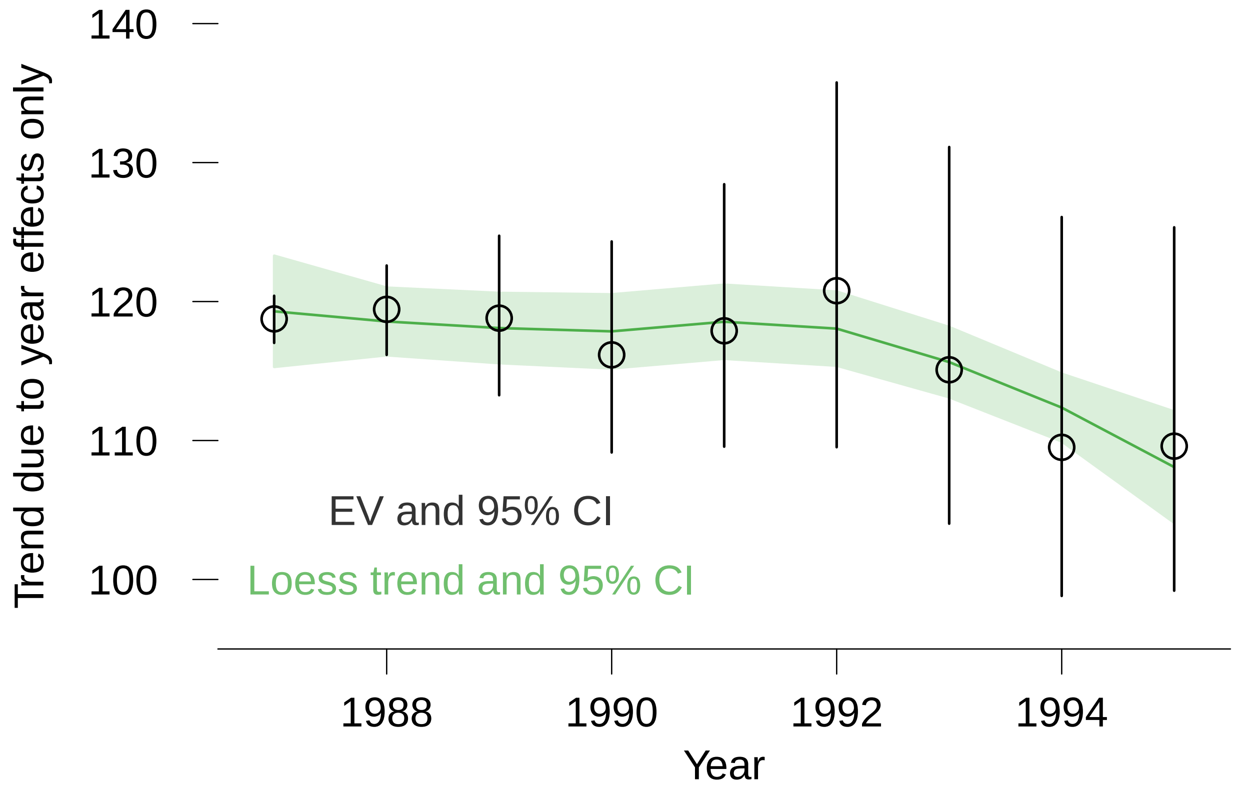
Here, I assume all covariates remained fixed at their 1992 levels,
then simulated the evolution of the level of smoking given the changing τ's over time

Hint of a downward trend?



2. What trend do the year effects imply? How do we capture this in simulations? Could we substitute a linear trend?

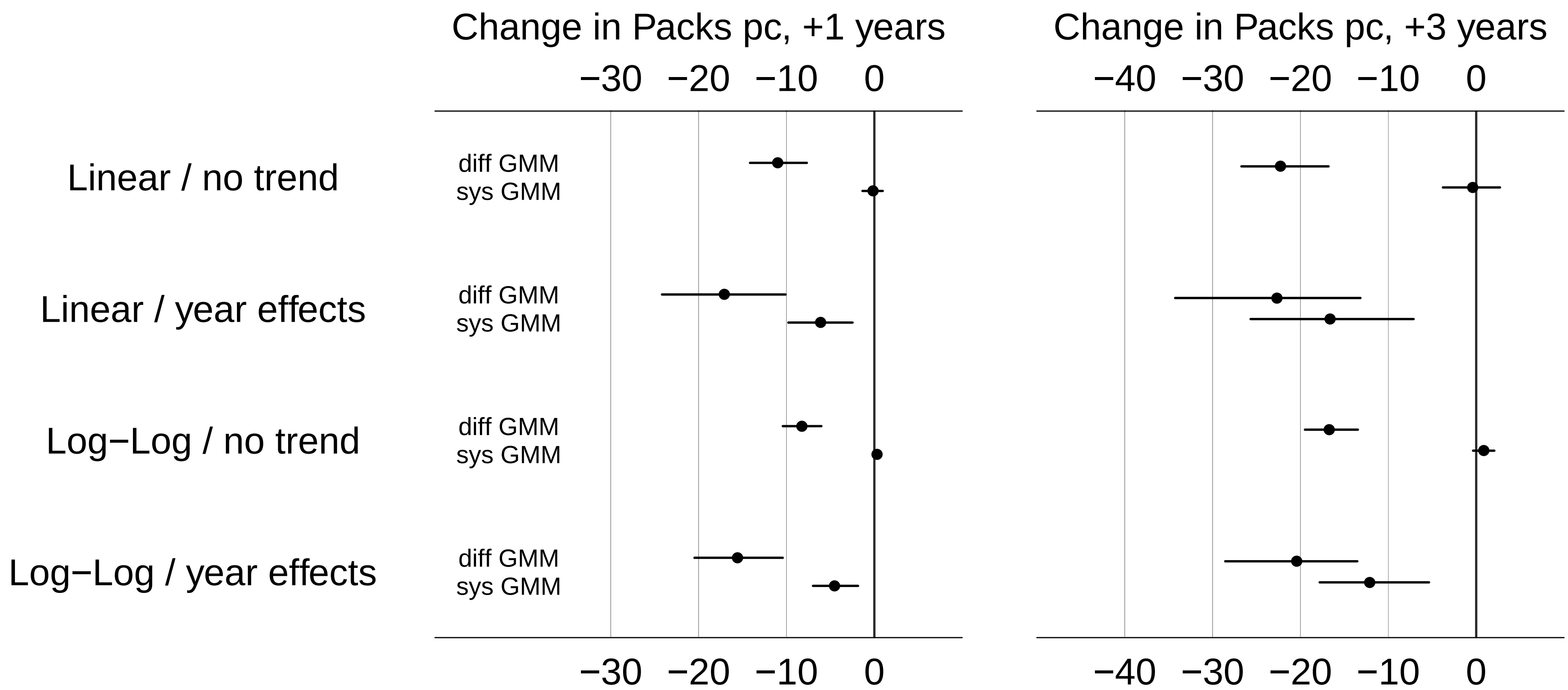
A weighted least squares fit (with standard errors of the estimates as weights) suggests there is a bit of a downward trend, but is it really linear?



2. What trend do the year effects imply? How do we capture this in simulations? Could we substitute a linear trend?

A loess fit suggests the trend was flat in the first part of the data,
and sharply downward over the last three years

In the simulations shown above,
we assume the average trend of the last three years persists in the near future;
e.g., we let the year effect be $1/3 \times (\tau_{1995} - \tau_{1992})$



Collecting results across models helps resolve these sensitivities

1. sys GMM is particularly sensitive: finds no effect under misspecification; agrees more with diff GMM year with years and logs included

2. Even under the preferred specification, effect size is sensitive to choice of instruments

Dynamic Panel Data Models: Final thoughts

More sensitive to assumptions than almost any other model

Poor behavior if outcome is nonstationary (or nearly so),
as instruments will be weaker (hard to predict a random walk)

Relying on a large pool of weak instruments introduces special challenges – important to check whether this is changing results

Lots of instruments $\rightarrow$ singular var-cov matrices
important to use generalized VC methods to compute standard errors

As we shall see, it's also important to use a heteroskedasticity and autocorrelation consistent VC matrix