Essex Summer School in Social Science Data Analysis
Panel Data Analysis for Comparative Research

IN-SAMPLE SIMULATION FOR PANEL DATA MODELS

Christopher Adolph

Department of Political Science *and*

Center for Statistics and the Social Sciences

University of Washington, Seattle

In-Sample Simulation for Panel Data Models

Dynamic simulation for panels is tricky, but just as valuable as for cross-sections

Dealing with panel simulation is why I started writing `simcf` functions;
I use these techniques extensively in my own work

In-sample simulation of panel models is especially useful / interesting / persuasive:

1. Make counterfactuals more factual and sample-specific
2. Avoids creating "average" cases that could never occur in reality
   - hypothetical people who are 52% female
   - hypothetical countries that score 0.2 on whether they are former colonies...

If you've ever asked:

"In simulations, why hold everything else constant at its mean?"

this lecture is for you

Plan of Lecture

Review of basic simulation concepts

Notation and concepts for in-sample simulation

Notation and concepts for dynamic panel simulation

Application to cigarette taxes panel GMM model

Review of basic simulation concepts

Suppose we have estimated a model of the form

$$y_{it} = \tau_t + y_{i,t-1}\phi + x_{it}\beta + \epsilon_{it}$$

Let $\theta = \text{vec}(\tau, \phi, \beta)$, the vector of parameters to estimate

Maximum Likelihood and Generalized Method of Moments estimates are asymptotically Normal, so for MLE and GMM θ converges in distribution:

$$f_\theta(\theta) \rightarrow \text{Multivariate Normal}\left(\hat{\theta}, \text{V}(\hat{\theta})\right) \quad \text{as} \quad n \rightarrow \infty$$

If we draw `sims=1000` or `sims=10000` vectors $\tilde{\theta}$ from this distribution,
the variance across those vectors captures the uncertainty in our parameters

Because we relied on asymptotic normality, there may be some bias in small samples

Review of basic simulation concepts

Suppose we have estimated a model of the form

$$y_{it} = \tau_t + y_{i,t-1}\phi + x_{it}\beta + \epsilon_{it}$$

Let $\theta = \text{vec}(\tau, \phi, \beta)$, the vector of parameters to estimate

For θ estimated by a Bayesian method
the model reports the posterior distribution of θ, $p(\theta)$ even in small samples

Bayesian models estimated by Markov chain Monte Carlo (MCMC),
report the parameters as a set of `sims` draws from their posterior distributions

The variance across these draws captures the uncertainty in our parameters

By computing the posterior directly, Bayes avoids bias, even in small samples

Review of basic simulation concepts

Suppose we have estimated a model of the form

$$y_{it} = \tau_t + y_{i,t-1}\phi + x_{it}\beta + \epsilon_{it}$$

Let $\theta = \text{vec}(\tau, \phi, \beta)$, the vector of parameters to estimate

Let $\tilde{\theta}$ be a random draw of the parameter vector
taken from the model's predictive or posterior distribution

We can simulate the expected value of y given a hypothetical x^{hyp} using $\tilde{y} = f(\tilde{\theta}, x^{\text{hyp}})$

Variation across the `sims` different version of $\tilde{y}|x^{\text{hyp}}$ captures the uncertainty in model parameters, as does any *quantity of interest* we construct from $\tilde{y}|x^{\text{hyp}}$

This allows us to construct model-based confidence intervals
around quantities of interest calculated from θ

Review of basic simulation concepts

The most common such quantities of interest are:

- the expected value of y_{it} given x_{it}^{hyp}
- the expected (or "first") difference between y_{it} given x_{it}^{hyp} and y_{it} given x_{it}^{base}

We can report the mean and 95% confidence interval of quantities of interest simply by computing averages and quantiles of the vector of simulates

Typically, we will hold fixed all but one covariate in the vector x_{it};
refer to the fixed covariates as having values h,
and the manipulated covariate as having counterfactual value c

Through careful selection of c and h,
we can use post-estimation simulation to answer substantive questions
while reporting the uncertainty of those answers

Notation & Concepts for In-Sample Simulation

We are interested in expected values and first differences conditioned on the model, observed data, and counterfactual scenarios (ignoring panel structure for now)

$$\hat{y}|\hat{\theta}, \boldsymbol{h}, \mathbf{c} \quad = \quad \mathbb{E}(y|\theta, \boldsymbol{h}, \mathbf{c})$$

indicates the expected value of y given estimated parameters $\hat{\theta}$, observed data h, and counterfactual assumptions c

The "first difference" version of this expected value is

$$\widehat{y_1 - y_0}|\hat{\theta}, \boldsymbol{h}, \mathbf{c}_1, \mathbf{c}_0 \quad = \quad \mathbb{E}(y_1 - y_0|\theta, \boldsymbol{h}, \mathbf{c}_1, \mathbf{c}_0)$$

Principle 1: Narrow counterfactuals.

Include as few elements as possible in c, moving them to *h* where possible

Notation & Concepts for In-Sample Simulation

Expected values: $\hat{y}|\hat{\theta}, \boldsymbol{h}, \mathbf{c}$ First differences: $\widehat{y_1 - y_0}|\hat{\theta}, \boldsymbol{h}, \mathbf{c}_1, \mathbf{c}_0$

Earlier in the course, we let $\boldsymbol{h} = \bar{\boldsymbol{h}}$, constructing scenarios in which we have the "average case" except for some counterfactual condition $\mathbf{c}_1$

Usually fairly close to the average EV or FD across the sample, but we can do better

Notation & Concepts for In-Sample Simulation

Expected values: $\hat{y}|\hat{\theta}, h, c$ First differences: $\widehat{y_1 - y_0}|\hat{\theta}, h, c_1, c_0$

Earlier in the course, we let $h = \bar{h}$, constructing scenarios in which we have the "average case" except for some counterfactual condition c_1

Usually fairly close to the average EV or FD across the sample, but we can do better

In-sample alternative: compute $\hat{y}_i$ or $\widehat{y_{i1} - y_{i0}}$ for each unit i, changing c_0 to c_1 while holding h_i constant at the factual levels observed in unit i

If the model is nonlinear or conditional in some way –
either y or c involves an nonlinear transformation or an interaction with h –
then the mean of the in-sample simulations will *not* equal the mean "case":

$$\text{mean}(\hat{y}_i|\hat{\theta}, h_i, c) \quad \neq \quad \hat{y}|\hat{\theta}, \bar{h}, c$$

Notation & Concepts for In-Sample Simulation

Expected values: $\hat{y}|\hat{\theta}, h, c$ First differences: $\widehat{y_1 - y_0}|\hat{\theta}, h, c_1, c_0$

Earlier in the course, we let $h = \bar{h}$, constructing scenarios in which we have the "average case" except for some counterfactual condition c_1

Usually fairly close to the average EV or FD across the sample, but we can do better

In-sample alternative: compute $\hat{y}_i$ or $\widehat{y_{i1} - y_{i0}}$ for each unit i, changing c_0 to c_1 while holding h_i constant at the factual levels observed in unit i

If the model is nonlinear or conditional in some way –
either y or c involves an nonlinear transformation or an interaction with h –
then the mean of the in-sample simulations will *not* equal the mean "case":

$$\begin{aligned} \text{mean}(\hat{y}_i|\hat{\theta}, h_i, c) &\neq \hat{y}|\hat{\theta}, \bar{h}, c \\ \text{mean}(\widehat{y_{i1} - y_{i0}}|\hat{\theta}, h_i, c_1, c_0) &\neq \widehat{y_1 - y_0}|\hat{\theta}, \bar{h}, c_1, c_0 \end{aligned}$$

Notation & Concepts for In-Sample Simulation

Upshot: In *many* models, computing $\text{mean}(\widehat{y_{i1} - y_{i0}})$ gives better estimates of the effect of switching c_0 to c_1 in your sample

The same point applies to simulated expected values,
simulated relative risks, and other QoIs

To the extent you care about the sample,
in-sample simulation is also more interesting than modeling the average case

Principle 2: In-sample counterfactuals.
For the "factual" values in your simulation, use h_i instead of $\bar{h}$,
then average or sum simulated quantities of interest over the sampled units i

Notation & Concepts for In-Sample Simulation

Principle 2: In-sample counterfactuals.
For the "factual" values in your simulation, use h_i instead of $\bar{h}$,
then average or sum simulated quantities of interest over the sampled units i

Issue 1. *We can't "sum" confidence intervals directly.*

You might suppose we could just simulate the point estimate and confidence interval for each unit, then "add them up"

But we can't combine confidence intervals that easily

Instead, we need to aggregate draws for each unit, $\tilde{y}_{i1} - \tilde{y}_{i0}$,
thus directly simulating the aggregate quantity of interest

Notation & Concepts for In-Sample Simulation

Principle 2: In-sample counterfactuals.
For the "factual" values in your simulation, use h_i instead of $\bar{h}$,
then average or sum simulated quantities of interest over the sampled units i

Issue 1. *We can't "sum" confidence intervals directly.*

If the quantity of interest is the **average** first difference (or EV, etc.) across units, simulate

$$\widetilde{\bar{y}_1 - \bar{y}_0} = \sum_{i=1}^{n} (\tilde{y}_{i1} - \tilde{y}_{i0})/n$$

Appropriate for quantities of interest that are rates or indexes:
- the unemployment rate
- packs smoked per day per capita
- the degree of democracy...

Principle 2: In-sample counterfactuals.
For the "factual" values in your simulation, use h_i instead of $\bar{h}$,
then average or sum simulated quantities of interest over the sampled units i

Issue 1. *We can't "sum" confidence intervals directly.*

If the quantity of interest is the **sum** of first differences (or EV, etc.) across units, simulate

$$\widetilde{\sum y_1 - \sum y_0} = \sum_{i=1}^{n} (\tilde{y}_{i1} - \tilde{y}_{i0})$$

Appropriate for quantities of interest that are counts:
- the total number of homicides
- the total amount of spending...

(But if your outcome is a count, may need a count model rather than linear regression)

Notation & Concepts for In-Sample Simulation

Principle 2: In-sample counterfactuals.
For the "factual" values in your simulation, use h_i instead of $\bar{h}$,
then average or sum simulated quantities of interest over the sampled units i

Issue 2. *We often want weighted averages across units.*

If we are averaging simulated regions to calculate country level QoIs,
we typically use regional population as weights w_i, so we simulate

$$\widetilde{\bar{y}_1 - \bar{y}_0} = \sum_{i=1}^{n} w_i(\tilde{y}_{it1} - \tilde{y}_{it0}) / \sum_{i=1}^{n} w_i$$

If the weights are known (the usual case), then the weighted average above retains the good statistical properties of the unweighted average

Notation & Concepts for In-Sample Simulation

We can summarize the set of `sims` values of $\widetilde{\bar{y}_1 - \bar{y}_0}$ as usual, via quantiles and means

Combining *narrow* **counterfactuals with** *in-sample* **context**
yields better estimates of the effect of the variable of interest

If we have confidence that our model estimates a causal effect,
then in effect, we are using an aggregate of local average treatment effects (LATEs)
to obtain a better estimate of the sample average treatment effect (SATE)

Even without a causal interpretation,
this is a useful way to summarize what the model is doing in terms of the sample

Finally, this is a bridge between pooled models and hierarchical models:
by simulating each group in the hierarchy separately, then combining results,
we have quantites of interest comparable to those from pooled estimation

Notation and concepts for dynamic panel simulation

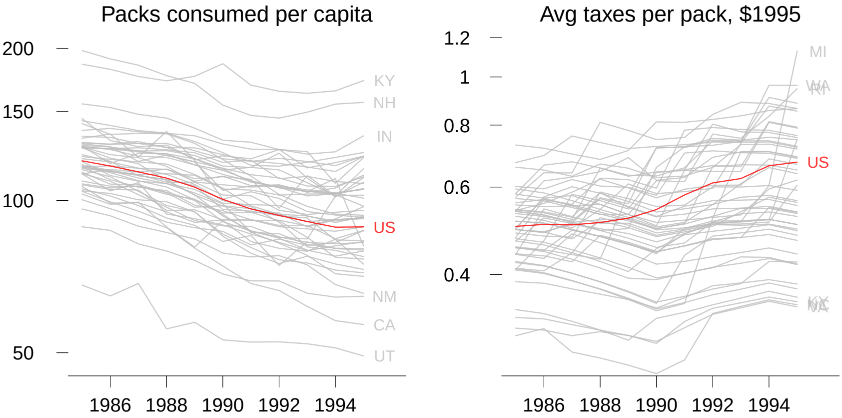
To apply the concept of in-sample simulation to dynamic models of panel data, it helps to have a running example

We will return to the cigarette tax example developed last week

Recall that we have data from 48 US states over the years 1985–1995 on:

- packs_{it}: packs of cigarettes consumed per capita
- price_{it}: average price of cigarettes, in cents
- tax_{it}: total taxes per pack, in cents
- income_{it}: average household income, in \$k

We have panel data, allow us to isolate the effect of taxes on consumption from all state- and year-invariant characteristics



Goal this time around:
develop simulations as fine-grained as the data itself
and present in a parallel graphical format

Notation and concepts for dynamic panel simulation

We treat the effect of taxes as analogous to a price increase;
because responses to price changes are elasticities, we need a log-log model

Because T is fairly small, we worry about Nickell bias,
so we will need to instrument using earlier lags of the outcome

We fit a log-log model of packs as a function of the effective price
using difference panel GMM with year effects (state effects are differenced out):

$$\Delta\text{log(packs)}_{it} = \Delta\tau_t + \Delta\text{log(packs)}_{i,t-1}\phi + \text{vec}(\Delta\text{price}_{it} + \Delta\text{tax}_{it}, \Delta\text{x}_{it})\beta + \Delta\epsilon_{it}$$

Let's predict packs consumed per capita in 1993, 1994, and 1995 iterated forward from 1992 after a hypothetical cigarette tax increase in every state starting in 1993

This requires us to condition on both data and parameters from the model above

Conditional forecasts iterated by period

$\widehat{\text{packs}}_{1993}|$

$\widehat{\text{packs}}_{1994}|$

$\widehat{\text{packs}}_{1995}|$

What must we condition on to construct $\widehat{\text{packs}}_{1993}$ from the model?

Conditional forecasts iterated by period

$$\widehat{\text{packs}}_{1993}|\hat{\beta}, \hat{\phi}, \hat{\tau}_{1993}, \hat{\tau}_{1992},$$

$$\widehat{\text{packs}}_{1994}|$$

$$\widehat{\text{packs}}_{1995}|$$

What must we condition on to construct $\widehat{\text{packs}}_{1993}$ from the model?

These terms are either model parameters or functions of them, and must be drawn from their predictive distributions

Conditional forecasts iterated by period

$$\widehat{\text{packs}}_{1993} | \hat{\beta}, \hat{\phi}, \hat{\tau}_{1993}, \hat{\tau}_{1992}, \overline{\text{packs}}_{1992}, \overline{\text{packs}}_{1991}, \bar{\text{x}}_{1993}, \text{tax}^{\text{hyp}}_{1993}$$

$$\widehat{\text{packs}}_{1994} |$$

$$\widehat{\text{packs}}_{1995} |$$

What must we condition on to construct $\widehat{\text{packs}}_{1993}$ from the model?

These terms are either model parameters or functions of them,
and must be drawn from their predictive distributions

These terms are data, and can be taken either from observation or as counterfactuals

Conditional forecasts iterated by period

$$\widehat{\text{packs}}_{1993} | \hat{\beta}, \hat{\phi}, \hat{\tau}_{1993}, \hat{\tau}_{1992}, \overline{\text{packs}}_{1992}, \overline{\text{packs}}_{1991}, \bar{\text{x}}_{1993}, \text{tax}^{\text{hyp}}_{1993}$$

$$\widehat{\text{packs}}_{1994} |$$

$$\widehat{\text{packs}}_{1995} |$$

Why do the forecast levels of packs depend on two lags of that level?

The model includes a lagged difference, which is computed from these lags

In a moment, we will see how the forecast level of packs can be computed from the estimated model in differences

Conditional forecasts iterated by period

$$\widehat{\text{packs}}_{1993}|\hat{\beta}, \hat{\phi}, \hat{\tau}_{1993}, \hat{\tau}_{1992}, \overline{\text{packs}}_{1992}, \overline{\text{packs}}_{1991}, \bar{\text{x}}_{1993}, \text{tax}^{\text{hyp}}_{1993}$$

$$\widehat{\text{packs}}_{1994}|$$

$$\widehat{\text{packs}}_{1995}|$$

Why do the forecast levels of packs depend on two adjacent year effects?

A similar reason: all covariates in the model are differenced, including τ

So the net year effect for period t is $\tau_t - \tau_{t-1}$

Conditional forecasts iterated by period

$$\widehat{\text{packs}}_{1993}|\hat{\beta}, \hat{\phi}, \hat{\tau}_{1993}, \hat{\tau}_{1992}, \overline{\text{packs}}_{1992}, \overline{\text{packs}}_{1991}, \bar{x}_{1993}, \text{tax}^{\text{hyp}}_{1993}$$

$$\widehat{\text{packs}}_{1994}|\hat{\beta}, \hat{\phi}, \hat{\tau}_{1994}, \hat{\tau}_{1993}, \widehat{\text{packs}}_{1993}, \overline{\text{packs}}_{1992}, \bar{x}_{1994}, \text{tax}^{\text{hyp}}_{1994}$$

$$\widehat{\text{packs}}_{1995}|\hat{\beta}, \hat{\phi}, \hat{\tau}_{1995}, \hat{\tau}_{1994}, \widehat{\text{packs}}_{1994}, \widehat{\text{packs}}_{1993}, \bar{x}_{1995}, \text{tax}^{\text{hyp}}_{1995}$$

Why are the lags treated as "data" for the 1993 forecast,
and as "simulated parameters" for the 1995 forecast?

The key idea of dynamic forecasting is that predictions of period t depends on older periods like $t-1$, $t-2$, etc.

Conditional forecasts iterated by period

$$\widehat{\text{packs}}_{1993} | \hat{\beta}, \hat{\phi}, \hat{\tau}_{1993}, \hat{\tau}_{1992}, \overline{\text{packs}}_{1992}, \overline{\text{packs}}_{1991}, \bar{\mathbf{x}}_{1993}, \text{tax}^{\text{hyp}}_{1993}$$

$$\widehat{\text{packs}}_{1994} | \hat{\beta}, \hat{\phi}, \hat{\tau}_{1994}, \hat{\tau}_{1993}, \widehat{\text{packs}}_{1993}, \overline{\text{packs}}_{1992}, \bar{\mathbf{x}}_{1994}, \text{tax}^{\text{hyp}}_{1994}$$

$$\widehat{\text{packs}}_{1995} | \hat{\beta}, \hat{\phi}, \hat{\tau}_{1995}, \hat{\tau}_{1994}, \widehat{\text{packs}}_{1994}, \widehat{\text{packs}}_{1993}, \bar{\mathbf{x}}_{1995}, \text{tax}^{\text{hyp}}_{1995}$$

From the view of 1993, the levels of packs in 1991 and 1992 are data;
we take the average observed values in those years as lags

Predicting from the model forward from 1992 to 1995, the intervening values of packs in 1993 and 1994 are *forecasts* subject to model uncertainty

In `simcf`, `ldvsimev()`, `ldvsimfd()`, and `ldvsimrr()` will help keep track of this for us

Conditional forecasts iterated by period: unit-by-unit

$$\widehat{\text{packs}}_{i,1993} | \hat{\beta}, \hat{\phi}, \hat{\tau}_{1993}, \hat{\tau}_{1992}, \text{packs}_{i,1992}, \text{packs}_{i,1991}, \mathbf{x}_{i,1993}, \text{tax}^{\text{hyp}}_{i,1993}$$

$$\widehat{\text{packs}}_{i,1994} | \hat{\beta}, \hat{\phi}, \hat{\tau}_{1994}, \hat{\tau}_{1993}, \widehat{\text{packs}}_{i,1993}, \text{packs}_{i,1992}, \mathbf{x}_{i,1994}, \text{tax}^{\text{hyp}}_{i,1994}$$

$$\widehat{\text{packs}}_{i,1995} | \hat{\beta}, \hat{\phi}, \hat{\tau}_{1995}, \hat{\tau}_{1994}, \widehat{\text{packs}}_{i,1994}, \widehat{\text{packs}}_{i,1993}, \mathbf{x}_{i,1995}, \text{tax}^{\text{hyp}}_{i,1995}$$

If we compute forecasts for each unit,
we no longer need to pool either factual covariates or lags of the outcome

Produces a fine-grained set of simulations with only one counterfactual element – the change in taxes we are interested in

There are `units` $\times$ `periods forecast` of these expected values

Conditional forecasts iterated by period: unit-by-unit *simulates*

$$\widetilde{\text{packs}}_{i,1993} | \tilde{\beta}, \tilde{\phi}, \tilde{\tau}_{1993}, \tilde{\tau}_{1992}, \text{packs}_{i,1992}, \text{packs}_{i,1991}, \text{x}_{i,1993}, \text{tax}^{\text{hyp}}_{i,1993}$$

$$\widetilde{\text{packs}}_{i,1994} | \tilde{\beta}, \tilde{\phi}, \tilde{\tau}_{1994}, \tilde{\tau}_{1993}, \widetilde{\text{packs}}_{i,1993}, \text{packs}_{i,1992}, \text{x}_{i,1994}, \text{tax}^{\text{hyp}}_{i,1994}$$

$$\widetilde{\text{packs}}_{i,1995} | \tilde{\beta}, \tilde{\phi}, \tilde{\tau}_{1995}, \tilde{\tau}_{1994}, \widetilde{\text{packs}}_{i,1994}, \widetilde{\text{packs}}_{i,1993}, \text{x}_{i,1995}, \text{tax}^{\text{hyp}}_{i,1995}$$

For these in-sample simulations, it is useful to retain the underlying simulates

These are what we actually iterate over to construct each forecast

At the unit level, there are `sims` $\times$ `units` $\times$ `periods forecast` of these simulates

Notation and concepts for dynamic panel simulation

Let's see how we compute a single set of simulated forecasts for unit *i* in our model of packs, which is both logged and differenced:

$$\Delta\widetilde{\text{log(packs)}}_{it} = \Delta\tilde{\tau}_t + \Delta\widetilde{\text{log(packs)}}_{i,t-1}\tilde{\phi} + \text{vec}(\Delta\text{price}_{it} + \Delta\text{tax}_{it}^{\text{hyp}}, \Delta x_{it})\tilde{\beta}$$

Notation and concepts for dynamic panel simulation

Let's see how we compute a single set of simulated forecasts for unit *i* in our model of packs, which is both logged and differenced:

$$\Delta\widetilde{\text{log(packs)}}_{it} = \Delta\tilde{\tau}_t + \Delta\widetilde{\text{log(packs)}}_{i,t-1}\tilde{\phi} + \text{vec}(\Delta\text{price}_{it} + \Delta\text{tax}_{it}^{\text{hyp}}, \Delta x_{it})\tilde{\beta}$$

$$\widetilde{\text{log(packs)}}_{it} - \widetilde{\text{log(packs)}}_{i,t-1} = \Delta\tilde{\tau}_t + \Delta\widetilde{\text{log(packs)}}_{i,t-1}\tilde{\phi} + \text{vec}(\Delta\text{price}_{it} + \Delta\text{tax}_{it}^{\text{hyp}}, \Delta x_{it})\tilde{\beta}$$

Notation and concepts for dynamic panel simulation

Let's see how we compute a single set of simulated forecasts for unit i in our model of packs, which is both logged and differenced:

$$\Delta\widetilde{\text{log(packs)}}_{it} = \Delta\tilde{\tau}_t + \Delta\widetilde{\text{log(packs)}}_{i,t-1}\tilde{\phi} + \text{vec}(\Delta\text{price}_{it} + \Delta\text{tax}_{it}^{\text{hyp}}, \Delta x_{it})\tilde{\beta}$$

$$\widetilde{\text{log(packs)}}_{it} - \widetilde{\text{log(packs)}}_{i,t-1} = \Delta\tilde{\tau}_t + \Delta\widetilde{\text{log(packs)}}_{i,t-1}\tilde{\phi} + \text{vec}(\Delta\text{price}_{it} + \Delta\text{tax}_{it}^{\text{hyp}}, \Delta x_{it})\tilde{\beta}$$

$$\widetilde{\text{log(packs)}}_{it} = \Delta\tilde{\tau}_t + \Delta\widetilde{\text{log(packs)}}_{i,t-1}\tilde{\phi} + \widetilde{\text{log(packs)}}_{i,t-1} + \text{vec}(\Delta\text{price}_{it} + \Delta\text{tax}_{it}^{\text{hyp}}, \Delta x_{it})\tilde{\beta}$$

Notation and concepts for dynamic panel simulation

Let's see how we compute a single set of simulated forecasts for unit i in our model of packs, which is both logged and differenced:

$$
\begin{aligned}
\Delta\widetilde{\text{log(packs)}}_{it} &= \Delta\tilde{\tau}_t + \Delta\widetilde{\text{log(packs)}}_{i,t-1}\tilde{\phi} + \text{vec}(\Delta\text{price}_{it} + \Delta\text{tax}_{it}^{\text{hyp}}, \Delta x_{it})\tilde{\beta} \\
\widetilde{\text{log(packs)}}_{it} - \widetilde{\text{log(packs)}}_{i,t-1} &= \Delta\tilde{\tau}_t + \Delta\widetilde{\text{log(packs)}}_{i,t-1}\tilde{\phi} + \text{vec}(\Delta\text{price}_{it} + \Delta\text{tax}_{it}^{\text{hyp}}, \Delta x_{it})\tilde{\beta} \\
\widetilde{\text{log(packs)}}_{it} &= \Delta\tilde{\tau}_t + \Delta\widetilde{\text{log(packs)}}_{i,t-1}\tilde{\phi} + \widetilde{\text{log(packs)}}_{i,t-1} \\
&\quad + \ \text{vec}(\Delta\text{price}_{it} + \Delta\text{tax}_{it}^{\text{hyp}}, \Delta x_{it})\tilde{\beta} \\
\widetilde{\text{packs}}_{it}|\text{tax}_{it}^{\text{hyp}} &= \exp\Big(\Delta\tilde{\tau}_t + \Delta\widetilde{\text{log(packs)}}_{i,t-1}\tilde{\phi} + \widetilde{\text{log(packs)}}_{i,t-1}|\text{tax}_{i,t-1}^{\text{hyp}} \\
&\quad + \text{vec}(\Delta\text{price}_{it} + \Delta\text{tax}_{it}^{\text{hyp}}, \Delta x_{it})\tilde{\beta}\Big)
\end{aligned}
$$

We have produced one simulated value of packs per capita for a single state and period under the counterfactual scenario that the change in taxes are set to $\Delta\text{tax}_{it}^{\text{hyp}}$

Notation and concepts for dynamic panel simulation

$$\widetilde{\text{packs}}_{it}|\text{tax}_{it}^{\text{hyp}} = \exp\Big(\Delta\tilde{\tau}_t + \Delta\widetilde{\log(\text{packs})}_{i,t-1}\tilde{\phi} + \widetilde{\log(\text{packs})}_{i,t-1}|\text{tax}_{i,t-1}^{\text{hyp}}$$

$$+ \text{vec}(\Delta\text{price}_{it} + \Delta\text{tax}_{it}^{\text{hyp}}, \Delta x_{it})\tilde{\beta}\Big)$$

We can forecast forward to period $t+1$, $t+2$, etc., noting that we need to use the simulated consumption in each state as lags in successive forecasts

For each period t in state i, we obtain `sims=1000` simulated values of $\widetilde{\text{packs}}_{it}|\text{tax}_{it}^{\text{hyp}}$

The mean and central 95% interval of these simulates provide state-by-state estimates of smoking given taxes in period t:

$$\left\{\text{packs}_{it}^{\text{lower}}|\text{tax}_{it}^{\text{hyp}}, \quad \widehat{\text{packs}}_{it}|\text{tax}_{it}^{\text{hyp}}, \quad \text{packs}_{it}^{\text{upper}}|\text{tax}_{it}^{\text{hyp}}\right\}$$

Notation and concepts for dynamic panel simulation

To construct a first difference,
we need to separately simulate a baseline scenario;
e.g., suppose the tax base followed its historically observed levels over i and t, $\text{tax}_{it}^{\text{base}}$

Then we construct `sims` simulated differences for each unit and period:

$$\widetilde{\text{packs}}_{it}|\text{tax}_{it}^{\text{hyp}} - \widetilde{\text{packs}}_{it}|\text{tax}_{it}^{\text{base}}$$

Repeating this simulation `sims=1000` times yields the estimated first difference and its quantiles for each unit

If we believe that our model correctly identifies the causal effect of taxes,
then these `units` different quantities are Local Average Treatment Effects (LATEs)

Note: if we want to do anything interesting with these simulates,
its critical to use the same simulated parameters for each state and year

Notation and concepts for dynamic panel simulation

Now we have simulates of the first difference of packs in each state and year

We can aggregate these *simulates* across each state-year to compute in-sample estimates of the net nationwide effect of a tax increase passed in every state

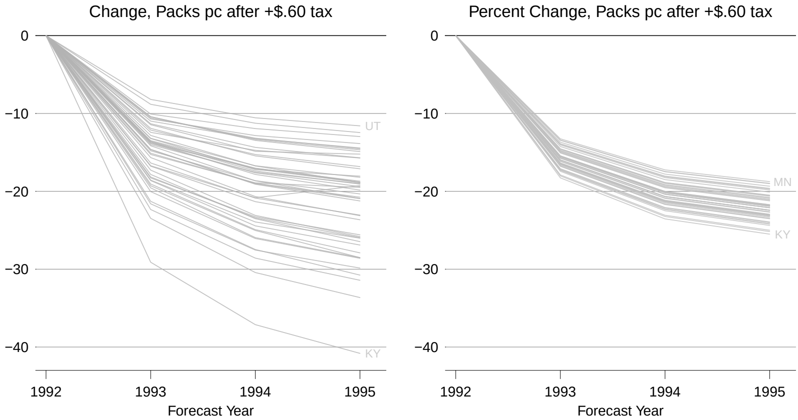
Obvious aggregation function in this example is the population-weighted average (this may vary based on substantive concerns):

$$\widetilde{\text{packs}_t}|\text{tax}_t^{\text{hyp}} - \widetilde{\text{packs}_t}|\text{tax}_t^{\text{base}} = \sum_{i=1}^{n} w_{it} \left(\widetilde{\text{packs}_{it}}|\text{tax}_{it}^{\text{hyp}} - \widetilde{\text{packs}_{it}}|\text{tax}_{it}^{\text{base}} \right) / \sum_{i=1}^{n} w_{it}$$

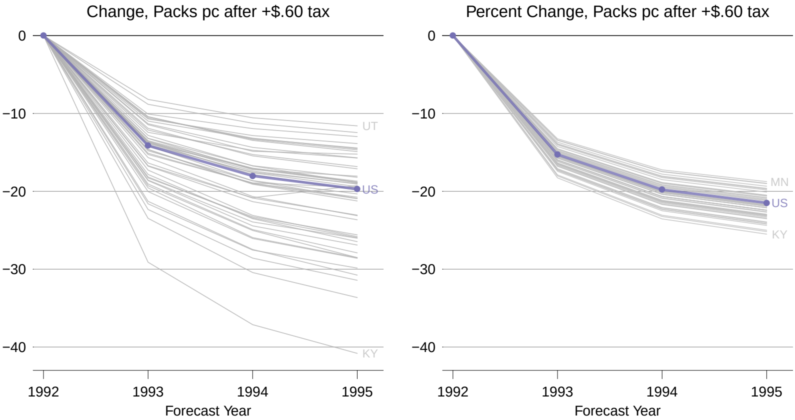
If we believe our model identifies the causal effect of taxes, this is the Sample Average Treatment Effect (SATE)

Summarize our `sims=1000` simulations with a mean & 95% CI, computed after taking the weighted average across states

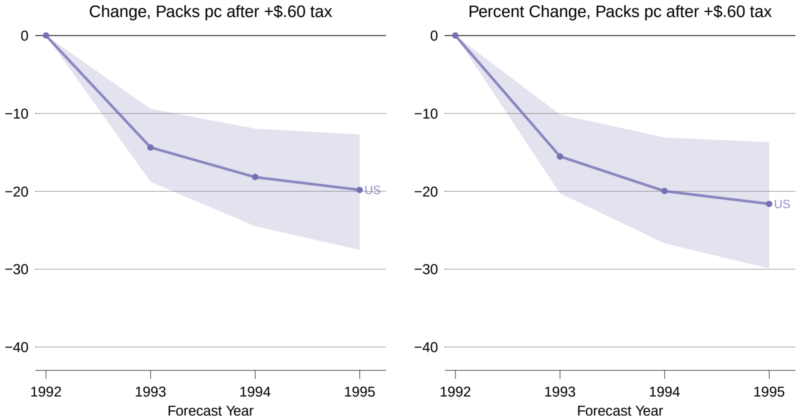
Let's apply these in-sample simulation techniques to the panel GMM estimates for the effect of cigarette taxes on smoking



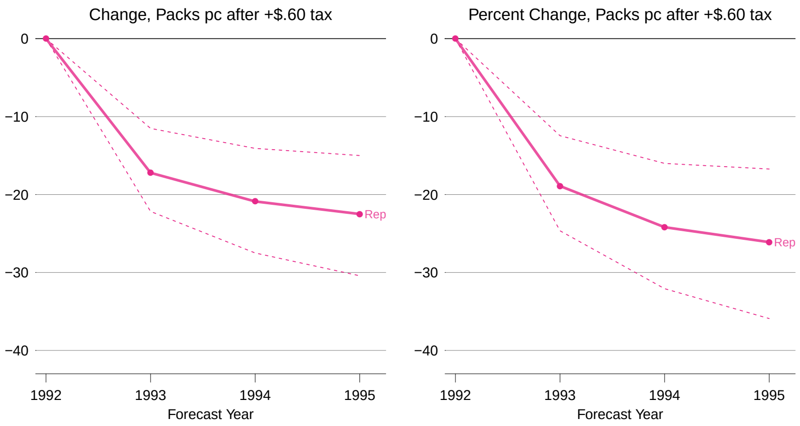
We start by simulating the in-sample behavior of smoking rates in each state, given that state's factual covariates
and a common $0.60 increase in taxes in 1993, maintained over time



Our estimates for the national effect of the tax increases
is the population-weighted average of state effects

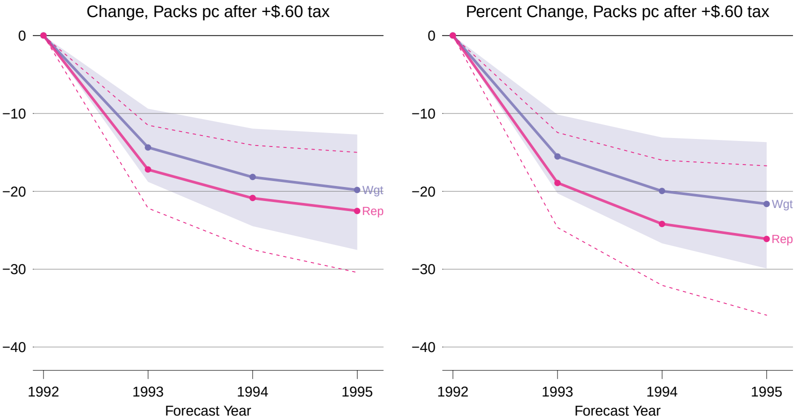


Naturally, all our results have confidence intervals (here, 95%);
we suppressed them to make the graphic easier to read when plotting all states

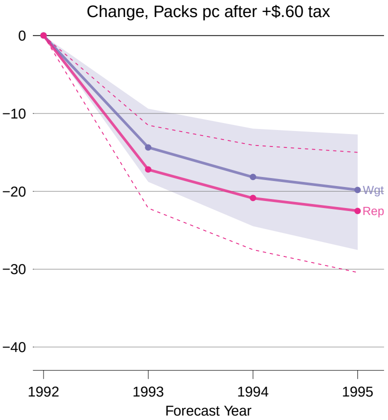


Earlier, we constructed our simulations for a "representative" or "average" state

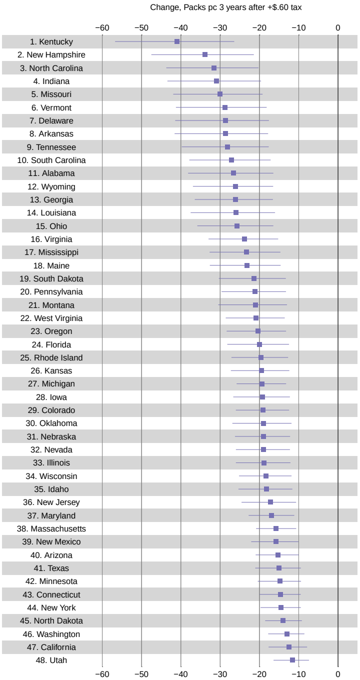
The "representative" estimate over the years 1993–1995 allows the tax rate and year effects to change, but keeps all covariates fixed at their annual means across states

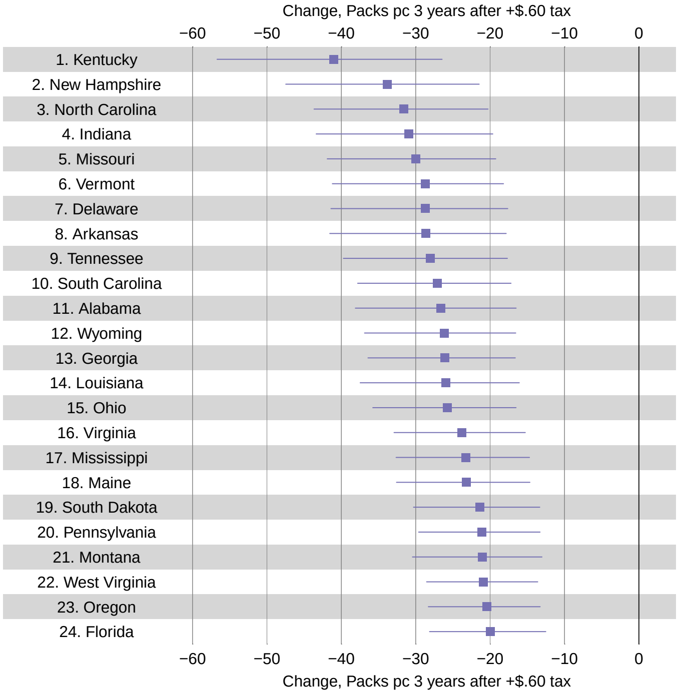


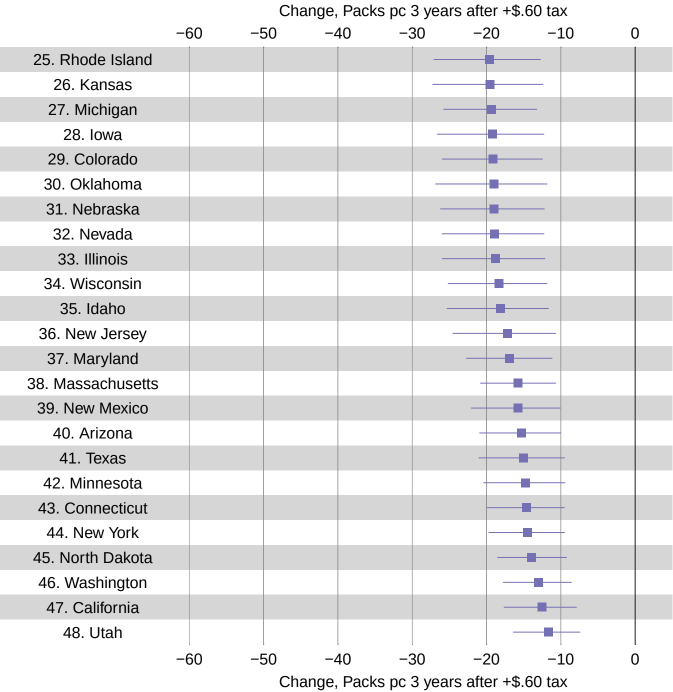
Relying on the representative case here
moderately over-estimate the national effect of taxes



Sign reversals or massive differences here are unlikely,
but the substative bias could still be important


Change, Packs pc 3 years after +$.60 tax
−60
−50
−40
−30
−20
−10
0
1. Kentucky
2. New Hampshire
3. North Carolina
4. Indiana
5. Missouri
6. Vermont
7. Delaware
8. Arkansas
9. Tennessee
10. South Carolina
11. Alabama
12. Wyoming
13. Georgia
14. Louisiana
15. Ohio
16. Virginia
17. Mississippi
18. Maine
19. South Dakota
20. Pennsylvania
21. Montana
22. West Virginia
23. Oregon
24. Florida
25. Rhode Island
26. Kansas
27. Michigan
28. Iowa
29. Colorado
30. Oklahoma
31. Nebraska
32. Nevada
33. Illinois
34. Wisconsin
35. Idaho
36. New Jersey
37. Maryland
38. Massachusetts
39. New Mexico
40. Arizona
41. Texas
42. Minnesota
43. Connecticut
44. New York
45. North Dakota
46. Washington
47. California
48. Utah


Change, Packs pc 3 years after +$.60 tax
−60
−50
−40
−30
−20
−10
0
1. Kentucky
2. New Hampshire
3. North Carolina
4. Indiana
5. Missouri
6. Vermont
7. Delaware
8. Arkansas
9. Tennessee
10. South Carolina
11. Alabama
12. Wyoming
13. Georgia
14. Louisiana
15. Ohio
16. Virginia
17. Mississippi
18. Maine
19. South Dakota
20. Pennsylvania
21. Montana
22. West Virginia
23. Oregon
24. Florida
−60
−50
−40
−30
−20
−10
0
Change, Packs pc 3 years after +$.60 tax



Change, Packs pc 3 years after +$.60 tax
−60
−50
−40
−30
−20
−10
0
25. Rhode Island
26. Kansas
27. Michigan
28. Iowa
29. Colorado
30. Oklahoma
31. Nebraska
32. Nevada
33. Illinois
34. Wisconsin
35. Idaho
36. New Jersey
37. Maryland
38. Massachusetts
39. New Mexico
40. Arizona
41. Texas
42. Minnesota
43. Connecticut
44. New York
45. North Dakota
46. Washington
47. California
48. Utah
−60
−50
−40
−30
−20
−10
0
Change, Packs pc 3 years after +$.60 tax