

## Panel VARS and Local Projections<sup>1</sup>

Having studied both vector autoregressions and panel data models, we now consider a pair of extensions. Both are especially useful if we have a panel with a high enough  $T$  so that a VAR type approach is possibly preferable to panel data approaches that include either lags or time fixed effects. The first approach is the *panel VAR*. The second is the *local projection method*.

### *Panel VARs*

Much of this section follows this paper, which provides a general discussion of PVARs, Stata files to run them, and some good examples:

Abrigo, Michael R.M. and Inessa Love. February 2015. “Estimation of Panel Vector Autoregression in Stata: a Package of Programs.” *Working Papers*.

Suppose that we have a panel that, without loss of generality, is balanced. There are  $N$  cross sections, and  $T$  periods. For the same reasons that we wish to run a VAR with a single cross section (estimating lagged effects and dealing with endogeneity), we wish to run a panel VAR with these data. For each cross section  $y_{it}$  is a  $k \times 1$  vector with  $k$  endogenous variables. We may also include exogenous variables in the specification, but because they do not substantially affect the analysis that follows, I exclude them.

The most obvious approach may be to simply to estimate a single VAR that includes all variables for all cross sections:

$$y_{1t} = \alpha_1 + \beta_{1,1,1}y_{1t-1} + \beta_{1,2,1}y_{2t-1} + \dots + \beta_{1,N,1}y_{Nt-1} \dots + \beta_{1,1,P}y_{1t-p} \dots + \beta_{1,N,P}y_{Nt-p} + u_{1,t} \quad (1)$$

$$y_{2t} = \alpha_2 + \beta_{2,1,1}y_{1t-1} + \beta_{2,2,1}y_{2t-1} + \dots + \beta_{2,N,1}y_{Nt-1} \dots + \beta_{2,1,P}y_{1t-p} \dots + \beta_{2,N,P}y_{Nt-p} + u_{2,t} \quad (2)$$

$$\dots \quad (3)$$

$$y_{Nt} = \alpha_N + \beta_{N,1,1}y_{1t-1} + \beta_{N,2,1}y_{2t-1} + \dots + \beta_{N,N,1}y_{Nt-1} \dots + \beta_{N,1,P}y_{1t-p} \dots + \beta_{N,N,P}y_{Nt-p} + u_{N,t} \quad (4)$$

---

<sup>1</sup>These are undergraduate lecture notes. They do not represent academic work. Expect typos, sloppy formatting, and occasional (possibly stupefying) errors.

Recall that  $y_{1t}$  and  $\alpha_1$  include  $k$  elements. We thus have  $kN$  endogenous variables. Each  $\beta$  also includes  $k^2$  elements. If there are  $p$  lags in the VAR, then there will be  $Np$  such matrices. There are thus  $Npk^2 + k^2$  regression coefficients. So we run our specification and we get the following:



This is an example of the *curse of dimensionality*.  $Npk + k$  will almost certainly be far larger than  $NT$ , our sample size, rendering this approach infeasible.

The inclusion of variables from different cross sections in a PVAR (as shown above) is known as “dynamic interdependencies.” To be able to practically estimate a PVAR, we must be willing to make assumptions that eliminate most or all of these dynamic interdependencies. The process is not unlike when we imposed identifying restrictions to calculate our orthogonalized IRFs when doing ordinary VARs.

*Love and Zicchino (2006)*

We will focus on the approach of Love and Zicchino (2006) for two reasons. First, its pretty easy. Second, it is the only method that has Stata files publicly available. This approach imposes that the slope coefficients (the  $\beta$ s from earlier) are the same for all cross sections. It thus shuts down most of the possible dynamic interdependencies and instead is conceptually closer to taking an average (though not mathematically identical) of the slope coefficients:

$$y_{it} = A_1 y_{it-1} + A_2 y_{it-2} + \dots A_p y_{it-p} + u_{it} + e_{it} \quad (5)$$

where  $e_{it}$  is a vector of error terms and  $u_{it}$  are fixed effects that may sometimes be excluded depending on the application.

A big advantage of this approach is that it results in a number of parameters to estimate that may be feasible. Each  $A$  includes  $k^2$  elements. Each  $u_{it}$  includes  $k$  elements as well. There are thus  $k^2p + kN$  parameters to be estimated and it is far more likely that this approach will be feasible.

This approach makes PVARs more feasible than individual VARs for each cross section. Suppose, for example, that we have 10 countries and wish to run a VAR(2) with where  $T$  only equals 10. Even if we only include 2 endogenous variables, that leaves 6 parameters to be estimated for each country. And with only 20 observations, estimation is dubious at best.

Suppose, however, that we decide to run a PVAR instead. With fixed effects included, we still only have 24 parameters to be estimated, 4 if we are willing to dispense with them (whether this makes sense depends on the application). We now, however, have 200 observations making estimation far more promising.

This example also, however, illustrates the big limitation of this method. As always, our results are only as good as the assumptions that underpin them. If we believe that slope coefficients are highly heterogeneous or that dynamic interdependencies are critical, then our assumptions are poor and our results are invalid. We also lose the ability to compare an IRF for two different cross sectional entities that are included within a single PVAR.

There are two other complications. First, we have to deal with potential non-stationarity. We can do this by estimating the first difference of  $y_{it}$ . Often, however, it is better to instead use the *Helmert Transformation* instead of the first difference. The Helmert Transformation subtracts off the mean of all future observations.

At first glance, it seems as if we could estimate (5) with OLS, and then compute orthogonalized IRFs just as we did with regular VARs. The latter is true. For the former, however, we run into the problem that fixed effects is biased if we include lagged dependent variables as regressors. This bias disappears as  $T \rightarrow \infty$ , but remains large for fairly large finite samples. The solution is to run a GMM estimator that uses instruments to correct for the bias that comes from including additional (beyond the lags already included as part of the PVAR) lagged dependent variables. We follow Arellano and Bover (1995) who propose using lagged instruments and levels as instruments.<sup>2</sup> Because we use additional lags, we eliminate additional observations which may be problematic if  $T$  is low.

---

<sup>2</sup>Arellano, M. and O. Bover. 1995. "Another look at the instrumental variable estimation of error-components model." *Journal of Econometrics*."

### *Love's Stata files*

PVAR estimation is not feasible with Stata as you find it. Instead you must install Inessa Love's ado files. This is easy:

1. Go to her homepage and download the files.
2. Wherever Stata is installed, go to *ado/base/p* and simply paste them in.
3. If you use these files in a paper, you should acknowledge their use and thank her for making them publicly available on the title page.

The newest version of these ado files work just like the files for VARS except that you add a “p” to the front of the command. “pvar” thus does the estimation, “pvarsoc” allows you to choose the appropriate lag length using various information criteria, and “pirf” compute the impulse response functions.

### *A Quick Example*

In Section 4.1, the authors consider an example of women aged 14 to 26 in the National Longitudinal Survey data between 1968 and 1978:

```
pvarsoc wage hours, maxlag(3) pvaropts(instl(1/4))
```

Like “varsoc,” this command allows us to choose the lag length. The option, “pvaropts(instl(1/4))” tells Stata to use 4 lags as instruments in GMM. The results (page 19) suggest just one lag. The next lag conducts the estimation:

```
pvar wage hours, instl(1/4)
```

Because one lag is the default, the lags instruction is omitted. Stata omits any observation with missing data, including with the instruments. We can instead treat missing instruments as zeros using the “gmmstyle” option.

The authors use the “pvarstable, graph” command. This calculates the modulus of the PVARs eigenvalues and checks if they are inside the unit circle. The resulting IRFs will be stationary if and only if all of them are inside the unit circle. the “varstable” command does the same for VARs.

The authors also report a variance decomposition. A variance decomposition is the second most commonly reported result from a VAR or PVAR. This results states that after  $X$  periods,  $Y$  fraction of one variable’s variance is explained by shocks to another variable. In this example, of the variance in the wage that results from the shocks 10 periods ago, 94% comes from shocks to the wage and 6% comes from shocks to hours.

```
pvarfevd, mc(200) save("fevd.dta")
```

Finally, IRFs are computed in much the same way as for VARs:

```
pvarirf, mc(200) oirf byopt(yrescale)
```

The “byopt(yrescale)” option fixes that annoying scaling problem when we graph IRFs.

## Local Projections

Jorda (2005) argues against the widespread use of VARS (including panel VARS).<sup>3</sup> Instead, he proposes that researchers use a *local projection* to obtain IRFs instead. Local projections can be very simple and are typically estimated using just OLS with robust standard errors. To illustrate how they work, we start with a simple univariate example. Suppose that we wish to generate an IRF for a shock to  $x_t$  on itself. The first step is simply to choose the size of the shock,  $\bar{u}$ . At the time of the shock,  $E[x_t]$ , the point estimate for the shock at the time of impact (which has been set as period  $t$ ) is then simply  $\bar{x} + \bar{u}$ , where  $\bar{x}$  is typically the mean.

The next step is to choose, using an information criteria, how many lags to include in the estimated autoregressive process. Suppose that we choose two lags. The next period of the IRF function is then obtained by regressing  $x_t$  on two lags of itself:

$$x_t = \alpha + \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_t \tag{6}$$

$E[x_{t+1}]$ , the IRF estimate for the period after the shock is then  $\alpha + \beta_1(\bar{x} + \bar{u}) + \beta_2\bar{x}$ , and the confidence interval may be computed using the standard errors of the regression coefficients. So far, this is no different than an ordinary autoregression. Were we doing a simple autoregression we would simply iterate forward to obtain  $E[x_{t+2}]$

---

<sup>3</sup>Jorda. O. 2005. “Estimation and Inference of Impulse Responses by Local Projections.” *American Economic Review*, 95(1): 161-182.

$$E[x_{t+2}] = \alpha + \beta_1 E[x_{t+1}] + \beta_2(\bar{x} + \bar{u}) \quad (7)$$

In words, we use the one period ahead expectation to form the two-period ahead expectation. The local projection method now departs from this approach. It instead forms  $E[x_{t+2}]$ , and each subsequent period, using a separate OLS regression. To obtain the IRF for  $t + 2$ , two periods after the shock, we now regress  $x_t$  on  $x_{t-2}$  and  $x_{t-3}$ .

$$x_t = \alpha + \beta_1 x_{t-2} + \beta_2 x_{t-3} + u_t \quad (8)$$

Again,  $E[x_{t+2}] = \alpha + \beta_1(\bar{x} + \bar{u}) + \beta_2 \bar{x}$  and it is again straightforward to compute the confidence interval. But now, the regression coefficients take different values. Each period gets its own regression. Continuing:

$$x_t = \alpha + \beta_1 x_{t-3} + \beta_2 x_{t-4} + u_t \quad (9)$$

and  $E[x_{t+3}] = \alpha + \beta_1(\bar{x} + \bar{u}) + \beta_2 \bar{x}$ .

We now consider the possible advantages of this approach:

1. It is very simple, only relying on simple OLS with robust standard errors. This is not a big deal in this univariate example, but local projections will be very useful to avoid dealing with some of the nastier issues that pop up in VARs.
2. We rarely know the true data generating process. In reality, the DGP is unlikely to be perfectly represented by a VAR. Suppose, for example, that (6) is misspecified.  $E[x_{t+1}]$  thus contains an additional source of error. Note that in (7),  $E[x_{t+1}]$  is used to compute  $E[x_{t+2}]$ . Thus any misspecification in  $E[x_{t+1}]$  is compounded forward. This is often cited as the main benefit. Local projections may be less vulnerable to misspecification because each period in the IRF gets its own regression instead of relying on the iteration of older, possibly misspecified, expectations.
3. Suppose that the model's dynamics are exotic. An IRF may, for example, display a hump shape pattern for nine periods followed by an abrupt spike in period 10. A conventional VAR can only capture this type of non-linearity if it includes many lags. Even a low order LP may pick this up, however. LPs may better capture non-linearities.

There are, however, disadvantages.

1. Suppose that a VAR is the correct specification. Then (7) is the correct way to form the two period ahead expectation. The local projection approach will be consistent but inefficient. With enough data, it will be able to replicate the IRF from the VAR method. But for small samples, there will be an efficiency loss. Jorda (2005) argues that in the examples he considers, this efficiency loss is small.
2. To get the  $n^{th}$  period ahead in an IRF using the LP method, we need at least  $n$  lags. If  $T$  is small, then this may cause us to run short on observations and the confidence intervals for distant periods in the IRFs may become large. This is not an issue using the VAR approach where even a VAR(1) can be used to compute an IRF far into the future.

We can now extend this example, with just a little added complexity, to the case of a VAR where  $x_t$  is now a vector of endogenous variables. This reveals another issue (which may be seen as a weakness or strength) with the LP method. There is less consensus with how to deal with the issue of contemporaneous causation. One approach is to risk endogeneity in the impact IRFs by including contemporaneous variables in the specification. Another is to exclude them which will rule out contemporaneous effects on shocks to other variables. This demonstrates another strength of the LP method, however. Not only does each period get its own regression, so does each variable in the system.

Suppose that we have a two variable system,  $A$  and  $B$ . Using the LP method, we can separately choose different lag lengths for  $A$  and  $B$  in each regression. There may be four distinct values. If we are confident that  $A$  can affect  $B$  contemporaneously, but not vice-versa, then we can include  $B$ 's current value in the regression for  $A$ .

### *An Example*

We now briefly discuss the following paper: Rana, G.A., and P. Shea. 2015. "Estimating the causal relationship between foreclosures and unemployment during the great recession." *Economics Letters*, Vol. 134. p. 90-93.

It seems clear that problems in the housing market were the proximate cause of the Great Recession. But it is less clear whether lower housing prices or higher foreclosures were the more important channel. There are mechanisms whereby both could increase unemployment, but macroeconomic theory mostly focuses on the former. We try to answer how housing caused the recession. Because this is a macroeconomic system, VAR type analysis is desirable with IRFs being the main type of result. We choose to use the LP approach.

The data are a panel of 11 U.S. states. The four variables are housing prices, effective mortgage rates, the U-3 unemployment rate, and foreclosures. The data are quarterly and run from 2003-2013. We determine that housing prices are I(2) while the other variables are all I(1).

Because we use the LP method, each period for each variable gets its own regression. We include the contemporaneous values of all variables in our specification (risking endogeneity in the impact period), and choose different lag structures for all four variables using adjusted  $R^2$ .

The main result is that higher foreclosures seem to be more important than lower housing prices. We write about the former:

The second column shows that higher foreclosures have a large, positive, and statistically significant effect on the U-3 unemployment rate. Integrating, the effect on the level of unemployment is a 6.4% increase after 3 years. This result supports the result of Marshall and Shea (2015) where mortgage default amplifies the business cycle. The third column shows that higher foreclosures also cause a large decline in housing prices.

About the latter, we add:

The second column shows that higher housing prices have only small effects on unemployment. The effect is only statistically significant in 3 of 12 periods, and integrating, we find only a cumulative decline in the unemployment rate of 0.36% after 3 years. This result shows that some of the rise in unemployment from a shock to foreclosures (from Fig. 2) occurs because more foreclosures cause lower housing prices which then causes more unemployment.

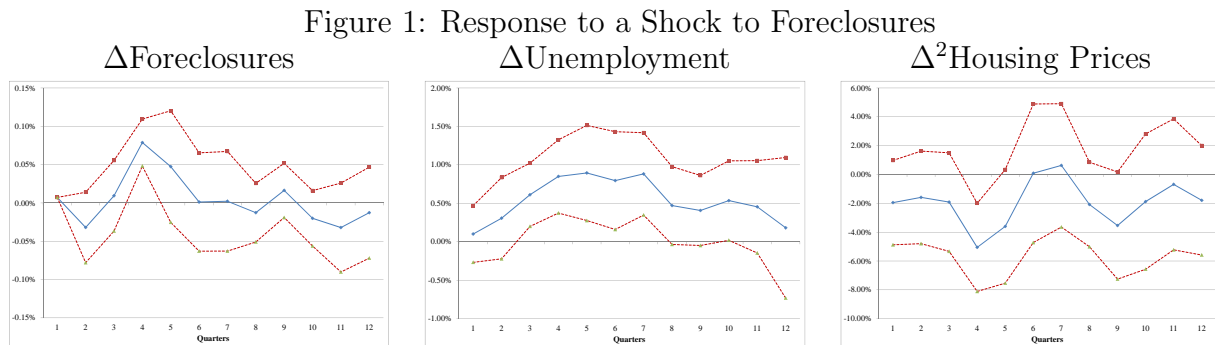




Figure 2: Response to a Shock to Housing Prices

