Paul Le Tran
plt377
17/11/21
ECO388E

Problem Set 2

1a).

$\widehat{\theta_0} = -0.7080; SE\left(\widehat{\theta_0}\right) = 0.3787;$

$\widehat{\theta_1} = -0.0092; SE\left(\widehat{\theta_1}\right) = 0.0058;$

$\widehat{\theta_2} = -0.1040; SE\left(\widehat{\theta_2}\right) = 0.0212.$

1b).

We see that for a woman of mean age and education, the effect of an additional year of education on the probability of her working is roughly 4.1%.

1c).

$\widehat{\theta_0} = -1.1322; SE\left(\widehat{\theta_0}\right) = 0.6144;$

$\widehat{\theta_1} = -0.0152; SE\left(\widehat{\theta_1}\right) = 0.0094;$

$\widehat{\theta_2} = 0.1681; SE\left(\widehat{\theta_2}\right) = 0.0349.$

Observe that the marginal effects based on the average of the data for both models have the same analytical structure: The marginal probability effect of regressor $x_j = \phi(X'\theta) * \theta_j$. Recall that the tails of the logit distribution are "fatter" than those of the probit distribution. As a result, the marginal effect of regressor x_j under the logit distribution will be greater than if under the probit distribution, since the PDF under logit will be larger in magnitude. This ultimately results in the estimated coefficients under logit to be larger in magnitude than those estimated under probit.

However, the overall curvature of both distributions are similar. This is because the structure of the model (i.e., the relation between the variables and the error term) has not change. As a result, the ratio amongst coefficients under probit and logit don't change much at all.

2a).

Recall that x_{2i} is women's education level (in years). One economic reason how x_{2i} could be correlated with ϵ_i is that ϵ_i could be capturing the wage a woman wants from working. This is because women who want a higher wage may choose more education. Therefore, if women make this decision simultaneously regarding the two variables, you could expect the correlation between x_{2i} and ϵ_i to put upward bias on coefficient θ_2.

2b).

There is no question for this part.

2c).

From part 2b, we assume that a woman's educational level depends on x_{1i}, z_i, ϵ_i. From the system of two equations, we assume that unobservables ϵ_i, η_i are jointly normal and are independent of x_{1i}, z_i. Observe that

$$cov(\epsilon\_\{i\}, \eta\_\{i\}) = \rho = E[\eta\_\{i\} * \epsilon\_\{i\}] - E[\eta\_i] * E[\epsilon\_\{i\}] = E[\eta\_\{i\} * \epsilon\_\{i\}]$$

By the jointly normal assumption.

ρ relates to the possible endogeneity of x_{2i} through expectations. Specifically, with the given system of two equations, we know that $E[\mathrm{x}_{2i} * \epsilon_i] = E[\eta_i * \epsilon_i] = \rho$. Therefore, we can see that if x_{2i} is exogenous, we must have $\rho = 0$. But since x_{2i} is assumed to be endogenous, $\rho \neq 0$. Additionally, I expect $\rho > 0$. This is because the educational level of one's parents (z_i) should e positively correlated with x_{2i}. This is because a parent often serves as a role model to their children.

Finally, the assumption that z_i does not directly determine y_i (i.e., z_i affects y_i only through x_{2i}) is the exclusion principle. This is one of the three principles that make for a "good/robust" instrument in IV estimation. Without the exclusion principle, θ_2 would not be identified non-parametrically.

2d).

$\widehat{\theta_0} = -0.3915; SE\left(\widehat{\theta_0}\right) = 0.6587;$

$\widehat{\theta_1} = -0.0100; SE\left(\widehat{\theta_1}\right) = 0.0059;$

$\widehat{\theta_2} = 0.0808; SE\left(\widehat{\theta_2}\right) = 0.0448;$

$\widehat{\theta_3} = 9.1254; SE\left(\widehat{\theta_3}\right) = 0.5284;$

$\widehat{\theta_4} =$ -0.0031; $SE\left(\widehat{\theta_4}\right) = 0.0096;$

$\widehat{\theta_5} = 0.3649; SE\left(\widehat{\theta_5}\right) = 0.0229;$

$\sigma_\eta = 1.9799; SE\left(\sigma_\eta\right) = 0.0433;$

$\rho = -0.1173; SE(\rho) = 0.1935.$

2e).

Given our t-statistic (with $df \ = \ 752$) of -16.6327 and associated p-value of essentially zero, we reject the null hypothesis at an α level of 0.05. As a result, we can say that it's possible x_{2i} is correlated with ϵ_i and thus, x_{2i} is endogenous.

2f).

Observe that τ_i is an unobservable that introduces heterogeneity in the response of education to changes in a woman's age. In other words, τ_i is essentially a shock that results in the impact of age on education to be different amongst women.

2g).

Note that we are choosing to integrate out τ_i. Similar to the model in part 2d, we can write the desired joint likelihood as:

$$\begin{aligned} p(y_i, x_{2i}|x_{1i}, z_i; \theta) &= p(x_{2i}|x_{1i}, z_i; \theta) * p(y_i|x_{1i}, z_i; \theta) \\ &= p(x_{2i}|x_{1i}, z_i; \theta) * p(y_i|x_{1i}, x_{2i}, z_i, \eta_i; \theta). \end{aligned}$$

The last equality holds because like before, the first stage equation is still invertible. In other words, this holds because $p(x_{2i}|x_{1i}, z_i; \theta) = p(x_{2i}|x_{1i}, z_i, \eta_i; \theta)$. This means even with the additional unobservable τ_i, we do not need to integrate out to find $p(x_{2i}|x_{1i}, z_i; \theta)$.

For $p(x_{2i}|x_{1i}, z_i; \theta), x_{2i}$ is still invertible in η_i. As a result, we can write the following expression:

$$p(x_{2i}|x_{1i}, z_i; \theta) = p(x_{2i}|x_{1i}, z_i, \eta_i; \theta).$$

However, we see that the RHS can only be found if we partially integrate out an unobservable. As mentioned in the beginning of this part's answer, we choose to integrate out τ_i:

$$p(x_{2i}|x_{1i}, z_i, \eta_i; \theta) = \int p(x_{2i}|x_{1i}, z_i, \eta_i, \tau_i; \theta) * p(\tau_i) d\tau_i.$$

We are able to use the unconditional probability $p(\tau_i)$ because of the assumption that τ_i is independent of $x_{1i}, x_{2i}, z_i, \eta_i, \epsilon_i$. Finally, because x_{2i} is invertible in η_i, we apple the ToRV:

$$\begin{aligned} p(x_{2i}|x_{1i}, z_i, \eta_i; \theta) &= \int p(\eta_i|x_{1i}, z_i, x_{2i}, \tau_i; \theta) * |det(d\eta_i/dy_i)| * p(\tau_i) d\tau_i \\ &= \frac{1}{S} \sum_{j}^{S} \phi(\eta_i/\sigma_\eta)/\, \sigma_\eta, \end{aligned}$$

where

$\eta_i = x_{2i} - \theta_3 - (\theta_4 + \sigma_\tau * \tau_i^j) x_{1i} - \theta_5 z_i;$

S represents the number of simulated draws we take of τ_i and $j = 1, \dots, S$ is the simulated draw itself.

We can finally write the probability as:

$$p(x_{2i}|x_{1i}, z_i; \theta) = \frac{1}{S}\sum_{j}^{S} \phi\big(x_{2i} - \theta_3 - \big(\theta_4 + \sigma_\tau * \tau_i^j\big)x_{1i} - \theta_5 * z_i\big)/\sigma_\eta.$$

To find an expression for $p(y_i|x_{1i}, x_{2i}, z_i, \eta_i; \theta)$, we again partially integrate out τ_i:

$$p(y_i|x_{1i}, x_{2i}, z_i, \eta_i; \theta) = \int p(y_i|x_{1i}, x_{2i}, z_i, \eta_i, \tau_i; \theta) * p(\tau_i)d\tau_i.$$

We are able to use the unconditional probability $p(\tau_i)$ because of the assumption that τ_i is independent of $x_{1i}, x_{2i}, z_i, \eta_i, \epsilon_i$. Simulating this integral gives us:

$$p(y_i|x_{1i}, x_{2i}, z_i, \eta_i; \theta) = \frac{1}{S} * \sum_{j}^{S} p\big(y_i|x_{1i}, x_{2i}, z_i, \tau_i^j, \eta_i; \theta\big).$$

We must now modify our κ_i equation that is in the second stage equation due to the structure of x_{2i} changing. Essentially, this just means that κ_i now includes τ_i:

$$\kappa_i = \theta_0 + \theta_2 * \theta_3 + \Big(\theta_1 + \theta_2 * \big(\theta_4 + \sigma_\tau * \tau_i^j\big)\Big) * x_{1i} + \theta_2 * \theta_5 * z_i + \theta_2 * \eta_i.$$

We finally obtain our simulated (log-)likelihood objective function:

$$L = \sum_{i}^{N} ln\big(p(x_{2i}|x_{1i}, z_i; \theta)\big) + ln\big(p(y_i|x_{1i}, x_{2i}, z_i, \eta_i; \theta)\big),$$

where

$$\begin{aligned} p(y_i|x_{1i}, x_{2i}, z_i, \eta_i; \theta) \\ &= \frac{1}{S}\sum_{j}^{S} y_i * \Phi\Big(\big(\kappa_i + \big(\rho/\sigma_\eta\big) * \eta_i\big)/(1-\rho^2)\Big) + (1 - y_i) \\ &* \Phi\Big(\big(\kappa_i + \big(\rho/\sigma_\eta\big) * \eta_i\big)/(1-\rho^2)\Big); \end{aligned}$$

$$\eta_i = x_{2i} - \theta_3 - \big(\theta_4 + \sigma_\tau * \tau_i^j\big)x_{1i} - \theta_5 * z_i;$$

$$\kappa_i = \theta_0 + \theta_2 * \theta_3 + \Big(\theta_1 + \theta_2 * \big(\theta_4 + \sigma_\tau * \tau_i^j\big)\Big) * x_{1i} + \theta_2 * \theta_5 * z_i + \theta_2 * \eta_i;$$

S represents the number of simulated draws we take of τ_i and $j = 1, \dots, S$ is the simulated draw itself.

2h).

Observe that we now have nine parameters in our model: $\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \rho, \sigma_\eta, \sigma_\tau$. We aim to choose at least nine moments for MSM estimation. Furthermore, observe that our model is basically an application of the general case of endogeneity in non-invertible models section of class. Because of this, it is simplest to form generic moments using the "reduced form" of our model:

Let $\widetilde{y_i} = (y_i, x_{2i}), \widetilde{\epsilon_i} = (\epsilon_i, \eta_i, \tau_i)$. Then we can write the following "generic" moments:

i. $\widetilde{y_i} - E[\widetilde{y_i}|x_{1i}, z_i; \theta]$;
ii. $(\widetilde{y_i} - E[\widetilde{y_i}|x_{1i}, z_i; \theta]) * x_{1i}$;
iii. $(\widetilde{y_i} - E[\widetilde{y_i}|x_{1i}, z_i; \theta]) * z_i$;
iv. $\widetilde{y_i^2} - E\left[\widetilde{y_i^2}\middle|x_{1i}, z_i; \theta\right]$.

Translating the "generic" moments of the "reduced form" to the original model yield the following eight moments:

1. $y_i - E[y_i|x_{1i}, z_i; \theta]$;
2. $(y_i - E[y_i|x_{1i}, z_i; \theta]) * x_{1i}$;
3. $(y_i - E[y_i|x_{1i}, z_i; \theta]) * z_i$;
4. $x_{2i} - E[x_{2i}|x_{1i}, z_i; \theta]$;
5. $(x_{2i} - E[x_{2i}|x_{1i}, z_i; \theta]) * x_{1i}$;
6. $(x_{2i} - E[x_{2i}|x_{1i}, z_i; \theta]) * z_i$;
7. $y_i^2 - E[y_i^2|x_{1i}, z_i; \theta]$;
8. $x_{2i}^2 - E[x_{2i}^2|x_{1i}, z_i; \theta]$.

Moments seven and eight (iv) will help in identifying parameters ρ, σ_τ. This is because the squared terms capture more of the endogeneity effect and variance. Recall our independence assumption about τ_i. From it, we can state that $E[\tau_i * x_{1i}] = E[\tau_i * z_i] = 0$. Therefore, using the idea that squared terms capture more information about variance-related parameters, the moments nine and 10 should help identify σ_τ:

9. $E[\tau_i * x_{1i}^2] = 0$;
10. $E[\tau_i * z_i^2] = 0$.

2i).

The SML estimator has several advantages over the MSM estimator. For one, the SML estimator is always efficient in that all the necessary moments are included in the estimation. In contrast, the MSM estimator is only efficient if we choose the optimal moments function. Otherwise, it will not be so. Additionally, the SML estimator in part 2g has the advantage in that we only need to simulate draws for τ_i.

The MSM estimator has the big advantage in that it will remain consistent holding the number of simulated draws we take of the error terms integrated out constant. In contrast, the SML estimator requires the set of simulated draws to approach infinity in order for the estimator to be consistent.

2j).

As seen in class, the simulated moments in part 2h are not smooth in θ. In order to obtain smoothness, we will simulate our MSM objective function using importance sampling. Recall that plugging in the first-stage equation into the second-stage equation yields the model single-equation form:

$y_i = I(\kappa_i + \epsilon_i > 0)$

$\Rightarrow E[y_i|x_{1i}, z_i; \theta] = \int \int I(\kappa_i + \epsilon_i > 0) * p(\tau_i) * p(\epsilon_i|\eta_i) * d\tau_i * d\epsilon_i,$

where

$\kappa_i = \theta_0 + \theta_2 * \theta_3 + \left(\theta_1 + \theta_2 * \left(\theta_4 + \sigma_\tau * \tau_i^j\right)\right) * x_{1i} + \theta_2 * \theta_5 * z_i + \theta_2 * \eta_i.$

Consider the change of variables $u_i = \kappa_i + \epsilon_i$. Our model is then:

$y_i = I(u_i > 0)$

$\Rightarrow E[y_i|x_{1i}, z_i; \theta] = \int I(u_i > 0) * \left(p\left(u_i|x_{1i}, z_i; \sigma_\tau, \sigma_\eta, \theta\right)/h(u_i)\right) * h(u_i) * du_i.$

Recall from class that we want $h(u_i)$ to be some density that doesn't depend on θ. If we only care about smoothing, let $h(u_i)$ be density $p(u_i| \dots)$ evaluated at initial guess on σ_τ, θ. We then get our simulation for $E[y_i|x_{1i}, z_i; \theta]$:

$$E[y_i|x_{1i}, z_i; \theta] = \frac{1}{S}\sum_{j}^{S} I\left(u_i^j > 0\right) * \left(p\left(u_i^S|x_{1i}, z_i; \sigma_\tau, \sigma_\eta, \theta\right)/h\left(u_i^S\right)\right),$$

where S represents the number of simulated draws we take of τ_i and $j = 1, \dots, S$ is the simulated draw itself.

2k).

With the possibility that our discrete choice model has endogeneity, we should think of bouth labour force participation and education level being jointly and simultaneously determined by our system of equations. This means that η_i is an unobservable that captures this joint decision. However, recall that τ_i introduces heterogeneity in the response of education to changes in a woman's age. This means that τ_i exerts influence on the solving/determination of a woman's education in our system of equations. As a result, it's possibly unreasonable to assume that τ_i is independent of η_i.

2l).

Stata's "ivprobit" command requires the following model assumptions to produce consistent estimates:

1. $(\eta_i, \epsilon_i) \sim$ Multivariate normal, for all i;

2. ϵ_i is homoscedastic;
3. The endogenous regressors (i.e., x_{2i}) are continuous.

ivprobit wouldn't be able to estimate our model in part 2f primarily because the assumption regarding homoscedasticity would be violated. This is because estimates for σ_τ are determined in part by x_{1i}. With our model being in single-equation form, we clearly see homoscedasticity is violated.

3a).

Observe that our model has both state dependence and unobserved heterogeneity. Because there is correlation in each individual's decisions over time, MLE needs to use the joint probability of consumer i's decisions in all time periods (conditioned on prices, initial decision, and parameters). This means the likelihood of the data for consumer i is:

$$L_i = p\left(y_{i,1}, \dots, y_{i,T} \middle| p_{i,1}, \dots, p_{i,T}, y_{i,0}; \theta\right).$$

Integrating out α_i, our likelihood becomes:

$$L_i = \int p\left(y_{i,1}, \dots, y_{i,T} \middle| p_{i,1}, \dots, p_{i,T}, y_{i,0}, \alpha_i; \theta\right) * p(\alpha_i).$$

Observe that $U_{i,t}$ only depends on $\epsilon_{i,t}$ if we know $p_{i,t}, y_{i,t-1}, \alpha_i$. Furthermore, focus on the probability

$$p\left(I\left(U_{i,t} > 0\right) \middle| p_{i,t}, y_{i,t-1}, \alpha_i; \theta\right).$$

Because $\epsilon_{i,t}$ is iid, this means they are independent over time. Observe that this holds because the individual unobserved heterogeneity is integrated out and accounted for. As a result, we know the following holds:

$$p\left(I\left(U_{i,t} > 0\right) \middle| p_{i,t}, y_{i,t-1}, \alpha_i; \theta\right) \perp p\left(I\left(U_{i,t+1} > 0\right) \middle| p_{i,t+1}, y_{i,t}, \alpha_i; \theta\right).$$

Consider we also know that $y_{i,t} = I\left(U_{i,t} > 0\right)$, our relation simplifies as follows:

$$p\left(y_{i,t} \middle| p_{i,t}, y_{i,t-1}, \alpha_i; \theta\right) \perp p\left(y_{i,t+1} \middle| p_{i,t+1}, y_{i,t}, \alpha_i; \theta\right).$$

In other words, because $\epsilon_{i,t}$ are iid, when combined with conditioning on $p_{i,t}, y_{i,t-1}, \alpha_i$, and parameters, factoring the inner joint probability of our likelihood function for consumer i becomes

$$L_i = \int \left[\prod_{t=1}^{T} p\left(y_{i,t} \middle| p_{i,t}, y_{i,t-1}, \alpha_i; \theta\right)\right] * p(\alpha_i).$$

3b).

Consider the following likelihood function:

$$L_i = p\left(y_{i,1}, \dots, y_{i,T} \middle| p_{i,1}, \dots, p_{i,T}, y_{i,0}; \theta\right) = \prod_{t=1}^{T} p\left(y_{i,t} \middle| p_{i,t}, y_{i,t-1}; \theta\right).$$

This equality would not hold, meaning it would not be obtainable if we choose not to integrate out α_i. The reason for this is because if we choose not to integrate out the unobserved heterogeneity, the $\epsilon_{i,t}$ over time would be correlated with each other. This would result in the probabilities inside of the product to no longer be independent of each other.

3c).

Recall that $y_i = I(U_{i,t} > 0)$. That means the inner probability terms of the product are

$$p\left(y_{i,t} \middle| p_{i,t}, y_{i,t-1}, \alpha_i; \theta\right) = p\left(I(U_{i,t} > 0) \middle| p_{i,t}, y_{i,t-1}, \alpha_i; \theta\right).$$

When we observe $y_{i,t} = 1$, we know $U_{i,t} > 0$. This means we can write the RHS probability above as:

$$p\left(I(U_{i,t} > 0) \middle| p_{i,t}, y_{i,t-1}, \alpha_i; \theta\right) = \frac{exp\left(\theta_0 + \theta_1 * p_{i,t} + \theta_2 * y_{i,t-1} + \sigma_\alpha * \alpha_i\right)}{1 + exp\left(\theta_0 + \theta_1 * p_{i,t} + \theta_2 * y_{i,t-1} + s\backslash igma_\alpha * \alpha_i\right)}.$$

Similarly, when we observe $y_{i,t} = 0$, we know $U_{i,t} < 0$. Our RHS probability then becomes:

$$p\left(I(U_{i,t} < 0) \middle| p_{i,t}, y_{i,t-1}, \alpha_i; \theta\right) = \frac{1}{1 + exp\left(\theta_0 + \theta_1 * p_{i,t} + \theta_2 * y_{i,t-1} + \sigma_\alpha * \alpha_i\right)}.$$

Combining the two, the inner probability of the product in our likelihood function can therefore be expressed as:

$$\begin{aligned} p\left(y_{i,t} \middle| p_{i,t}, y_{i,t-1}, \alpha_i; \theta\right) \\ &= y_{i,t} * \frac{exp\left(\theta_0 + \theta_1 * p_{i,t} + \theta_2 * y_{i,t-1} + \sigma_\alpha * \alpha_i\right)}{1 + exp\left(\theta_0 + \theta_1 * p_{i,t} + \theta_2 * y_{i,t-1} + \sigma_\alpha * \alpha_i\right)} + \left(1 - y_{i,t}\right) \\ &* \frac{1}{1 + exp\left(\theta_0 + \theta_1 * p_{i,t} + \theta_2 * y_{i,t-1} + \sigma_\alpha * \alpha_i\right)}. \end{aligned}$$

Substituting this into our inner-probability-product will give us expression (3) of the likelihood function displayed in the problem set.

3d).

Note that we are using $S = 100$ simulated draws. This is because after testing with 250, 500, and 1000 simulated draws, both the parameter estimates and standard errors are extremely similar. Furthermore, we are assuming that $\frac{\sqrt{N}}{S} \rightarrow 0$ as $N \rightarrow \infty$ in order to ignore simulation error when we compute standard errors.

$\widehat{\theta_0} = -0.7953; SE(\widehat{\theta_0}) = 0.1936;$

$\widehat{\theta_1} = 0.9169; SE(\widehat{\theta_1}) = 0.1558;$

$\widehat{\theta_2} = 0.5097; SE(\widehat{\theta_2}) = 0.1012;$

$\widehat{\sigma_\alpha} = 0.8919; SE(\widehat{\sigma_\alpha}) = 0.1172.$

3e).

With $df = 1998$, our t-stats for parameters $(\widehat{\theta_0}, \widehat{\theta_1}, \widehat{\theta_2}, \widehat{\sigma_\alpha})$ are $(-4.1078, 5.8837, 5.0390, 7.6107)$. The respective p-values (at $\alpha = 0.05$) are $(4.1564e-05, 4.6914e-09, 5.1009e-07, 4.1744e-14)$.

Recall that θ_2 captures the effect of state dependence and σ_α captures the effect of unobserved, individual heterogeneity. From our calculates above, we see that both parameters are statistically significant. However, σ_α is larger than θ_2. Therefore, we can say that unobserved, individual heterogeneity is possibly more important in explaining the correlation of $y_{i,t}$ over time than state dependence.

3f).

$\widehat{\theta_0} = -0.9485; SE(\widehat{\theta_0}) = 0.1520;$

$\widehat{\theta_1} = 0.7297; SE(\widehat{\theta_1}) = 0.1449;$

$\widehat{\theta_2} = 1.0118; SE(\widehat{\theta_2}) = 0.0765.$

With $df = 1998$, our t-stats for parameters $(\widehat{\theta_0}, \widehat{\theta_1}, \widehat{\theta_2})$ are $(-6.2380, 5.0371, 13.2285)$. The respective p-values (at $\alpha = 0.05$) are $(5.3935e-10, 5.1507e-07, 0)$.

When re-estimating the model without unobserved, individual heterogeneity, we see that the estimate of $\theta_{[2\}}$ increased by almost a factor of double (in terms of absolute magnitude). The reason for this is because α_i is now an omitted variable which is correlated with the previous state, $y_{i,t-1}$. As a result, the omitted variable bias puts upward pressure on the parameter estimates. This is especially true for θ_2, the parameter associated with state dependence.

3g).

If we use only linear probability models, one method to "crudely" test the null hypothesis that there is no state dependence would be to estimate the following 2SLS regression:

$$p(y_{i,t} = 1 | p_{i,t}, y_{i,t-1}) = \theta_0 + \theta_1 * p_{i,t} + \theta_2 * y_{i,t-1} + \epsilon_{i,t};$$

$$y_{i,t-1} = \theta_3 + \theta_4 * p_{i,t} + \theta_5 * p_{i,t-1} + \eta_{i,t}.$$

Our relevant exclusion principles needed for prices to be a valid instrument are that:

1. We need price in general to be exogenous from $y_{i,t}$.
2. We need $p_{i,t-1}$ to only affect $y_{i,t}$ through previous state $y_{i,t-1}$.

It should be mentioned that instead of doing 2SLS, Chamberlain mentions an even simpler way to test for the relevancy of state dependence. Specifically, one only needs to run a simple linear regression of $y_{i,t}$ on some general function of $p_{i,t}$ and $p_{i,t-1}$ (in some ways, this is the reduced form of our 2SlS regression above). If the parameter associated with $p_{i,t-1}$ is zero, then we know state dependence is not relevant. The simplest example of this is:

$$y_{i,t} = \theta_1 * p_{i,t} + \theta_2 * p_{i,t-1} + \epsilon_{i,t}.$$

The same exclusion principles apply:

1. We need price in general to be exogenous from $y_{i,t}$.
2. We need $p_{i,t-1}$ to only affect $y_{i,t}$ through previous state $y_{i,t-1}$.