A Course in Applied Econometrics
Lecture 4: Linear Panel Data Models, II

Jeff Wooldridge
IRP Lectures, UW Madison, August 2008

5. Estimating Production Functions Using Proxy Variables

- Common approaches to production function estimation using firm-level panel data: fixed effects and first differencing. Typically, one assumes a Cobb-Douglas production function with additive firm heterogeneity.
- Problem: FE and FD estimators assume strict exogeneity of the inputs, conditional on firm heterogeneity; see, for example, Wooldridge (2002). Generally rules out the possibility that inputs are chosen in response to curren or past productivity shocks, a severe restriction on firm behavior.

- Instrumental variables methods can be used to relax the strict exogeneity assumption: lagged inputs as IVs after differencing or quasi-differencing. [Holtz-Eakin, Newey, and Rosen (1988), Arellano and Bover (1995), Blundell and Bond (2000).]
- Unfortunately, differencing removes much of the variation in the explanatory variables and can exacerbate measurement error in the inputs. Often, the instruments available after differencing often are only weakly correlated with the differenced explanatory variables. Some of the extra moment conditions discussed in Section 4 can help.

- Olley and Pakes (1996) (OP) suggest a different approach. Rather than allow for time-constant firm heterogeneity, OP show how investment can be used as a proxy variable for unobserved, time-varying productivity. Specifically, productivity can be expressed as an unknown function of capital and investment (when investment is strictly positive). OP present a two-step estimation method where, in the first stage, semiparametric methods are used to estimate the coefficients on the variable inputs. In a second step, the parameters on capital inputs can be identified under assumptions on the dynamics of the productivity process.

- Levinsohn and Petrin (2003) (LP) suggest using intermediate inputs to proxy for unobserved productivity. Two-step estimation.
- In implementing LP (or OP), convenient to assume that unknown functions are well approximated by low-order polynomials. Petrin, Poi, and Levinsohn (2004) (PPL) suggest third-degree polynomials. This leads to estimated parameters that are very similar to locally weighted estimation.
- A unified approach that can be applied to various situations, including Ackerberg, Caves, and Frazer (2006) (ACF): estimate two equations simultaneously. Simplifies inference, more efficient, provides insights into identification.

- Set up as a two-equation system for panel data with the same dependent variable, but where the set of instruments differs across equation, as in Wooldridge (1996).
- Write a production function for firm i in time period t as

$$y_{it} = \alpha + \mathbf{w}_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\gamma} + v_{it} + e_{it}, t = 1,\ldots,T, \tag{1}$$

where

y_{it} = natural logarithm of the firm's output
$\mathbf{w}_{it}$ = $1 \times J$ vector of variable inputs (labor)
$\mathbf{x}_{it}$ = $1 \times K$ vector of observed state variables (capital)

- The sequence $\{v_{it} : t = 1,\ldots,T\}$ is unobserved productivity, and $\{e_{it} : t = 1,2,\ldots,T\}$ is a sequence of shocks.
- Key implication of the theory underlying OP and LP: for some function $g(\cdot,\cdot)$,

$$v_{it} = g(\mathbf{x}_{it}, \mathbf{m}_{it}), t = 1,\ldots,T, \tag{2}$$

where $\mathbf{m}_{it}$ is a $1 \times M$ vector of proxy variables. In OP, $\mathbf{m}_{it}$ consists of investment (investment in OP, intermediate inputs in LP). In OP, representation (2) involves inverting a relationship relating investment and productivity and capital, but only for strictly positive investment; in LP, it is inverting a relationship between intermediate inputs and productivity and capital.

- For simplicity, assume $g(\cdot,\cdot)$ is time invariant. Under the assumption

$$E(e_{it}|\mathbf{w}_{it}, \mathbf{x}_{it}, \mathbf{m}_{it}) = 0, t = 1,2,\ldots,T, \tag{3}$$

we have the following regression function:

$$\begin{aligned} E(y_{it}|\mathbf{w}_{it}, \mathbf{x}_{it}, \mathbf{m}_{it}) &= \alpha + \mathbf{w}_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\gamma} + g(\mathbf{x}_{it}, \mathbf{m}_{it}) \\ &\equiv \mathbf{w}_{it}\boldsymbol{\beta} + \mathbf{h}(\mathbf{x}_{it}, \mathbf{m}_{it}), t = 1,\ldots,T, \end{aligned} \tag{4}$$

where $h(\mathbf{x}_{it}, \mathbf{m}_{it}) \equiv \alpha + \mathbf{x}_{it}\boldsymbol{\gamma} + g(\mathbf{x}_{it}, \mathbf{m}_{it})$. Since $g(\cdot,\cdot)$ is allowed to be a general function – in particular, linearity in $\mathbf{x}$ is a special case – $\boldsymbol{\gamma}$ (and the intercept, α) are clearly not identified from (4).

• Equation (4) appears to identify $\boldsymbol{\beta}$. However, this need not be true, particularly when $\mathbf{m}_{it}$ contains intermediate inputs. As shown by Ackerberg, Caves, and Frazer (2006) (ACF), if labor inputs are chosen at the same time as intermediate inputs, there is a fundamental identification problem in (4): $\mathbf{w}_{it}$ is a deterministic function of $(\mathbf{x}_{it}, \mathbf{m}_{it})$, which means $\boldsymbol{\beta}$ is nonparametrically unidentified.

• To make matters worse, ACF show that $\mathbf{w}_{it}$ actually drops out of (4) when the production function is Cobb-Douglas.

• Better to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ together. Assume

$$E(e_{it}|\mathbf{w}_{it}, \mathbf{x}_{it}, \mathbf{m}_{it}, \mathbf{w}_{i,t-1}, \mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1}, \ldots, \mathbf{w}_{i1}, \mathbf{x}_{i1}, \mathbf{m}_{i1}) = 0, t = 1,2,\ldots,T. \quad (5)$$

This allows for serial dependence in the idiosyncratic shocks $\{e_{it} : t = 1,2,\ldots,T\}$ because neither past values of y_{it} nor e_{it} appear in the conditioning set.

• Also restrict the dynamics in the productivity process:

$$\begin{aligned} E(v_{it}|\mathbf{x}_{it}, \mathbf{w}_{i,t-1}\mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1}, \ldots) &= E(v_{it}|v_{i,t-1}) \\ &= f(v_{i,t-1}) \equiv f[g(\mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1})], \end{aligned} \quad (6)$$

where the latter equivalence holds for some $f(\cdot)$ because $v_{i,t-1} = g(\mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1})$.

• The variable inputs in $\mathbf{w}_{it}$ are allowed to be correlated with the innovations a_{it} in $v_{it} = f(v_{i,t-1}) + a_{it}$, but (6) means that $\mathbf{x}_{it}$, past $(\mathbf{w}_{it}, \mathbf{x}_{it}, \mathbf{m}_{it})$, and functions of these are uncorrelated with a_{it}.

• Plugging into (1) gives

$$y_{it} = \alpha + \mathbf{w}_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\gamma} + f[g(\mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1})] + a_{it} + e_{it}. \quad (7)$$

• Now, we can specify the two equations that identify $(\boldsymbol{\beta}, \boldsymbol{\gamma})$:

$$y_{it} = \alpha + \mathbf{w}_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\gamma} + g(\mathbf{x}_{it}, \mathbf{m}_{it}) + e_{it}, t = 1,\ldots,T \quad (8)$$

and

$$y_{it} = \alpha + \mathbf{w}_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\gamma} + f[g(\mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1})] + u_{it}, t = 2,\ldots,T, \quad (9)$$

where $u_{it} \equiv a_{it} + e_{it}$.

• Importantly, the available orthogonality conditions differ across these two equations. In (8), the orthogonality condition on the error is given by (5). The orthogonality conditions for (9) are

$$E(u_{it}|\mathbf{x}_{it}, \mathbf{w}_{i,t-1}\mathbf{x}_{i,t-1}, \mathbf{m}_{i,t-1}, \ldots, \mathbf{w}_{i1}, \mathbf{x}_{i1}, \mathbf{m}_{i1}) = 0, t = 2,\ldots,T. \quad (10)$$

In other words, in (8) and (9) we can use the contemporaneous state (capital) variables, $\mathbf{x}_{it}$, any lagged inputs, and functions of these, as instrumental variables. In (8) we can further add the elements of $\mathbf{m}_{it}$ (investment or intermediate inputs).

• When (8) does not identify $\boldsymbol{\beta}$, (9) would still generally identify $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ provided we have the orthogonality conditions in (10). Effectively, $\mathbf{x}_{it}$, $\mathbf{x}_{i,t-1}$, and $\mathbf{m}_{i,t-1}$ act as their own instruments and $\mathbf{w}_{i,t-1}$ acts as an instrument for $\mathbf{w}_{it}$. But better to use both equations.

• Equation (9) can be estimated by an instrumental variables version of Robinson's (1988) estimator to allow f and g to be completely unspecified.

• Simpler approach that allows (8) to provide identifying information about the parameters: approximate $g(\cdot,\cdot)$ and $f(\cdot)$ in (8) and (9) by low-order polynomials, say, up to order three. If x_{it} and m_{it} are both scalars, $g(x,m)$ is linear in terms of the form $x^p m^q$, where p and q are nonnegative integers with $p+q \leq 3$. More generally, $g(\mathbf{x},\mathbf{m})$ contains all polynomials of order three or less. In any case, assume that we can write

$$g(\mathbf{x}_{it},\mathbf{m}_{it}) = \lambda_0 + \mathbf{c}(\mathbf{x}_{it},\mathbf{m}_{it})\boldsymbol{\lambda} \tag{11}$$

for a $1 \times Q$ vector of functions $\mathbf{c}(\mathbf{x}_{it},\mathbf{m}_{it})$. The function $\mathbf{c}(\mathbf{x}_{it},\mathbf{m}_{it})$ contains at least $\mathbf{x}_{it}$ and $\mathbf{m}_{it}$ separately, since a linear version of $g(\mathbf{x}_{it},\mathbf{m}_{it})$ should always be an allowed special case.

• Assume that $f(\cdot)$ can be approximated by a polynomial in v:

$$f(v) = \rho_0 + \rho_1 v + \ldots + \rho_G v^G. \tag{12}$$

• Given the functions in (11) and (12), we now have

$$y_{it} = \alpha_0 + \mathbf{w}_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\gamma} + \mathbf{c}_{it}\boldsymbol{\lambda} + e_{it}, t = 1,\ldots,T \tag{13}$$

and

$$y_{it} = \eta_0 + \mathbf{w}_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\gamma} + \rho_1(\mathbf{c}_{i,t-1}\boldsymbol{\lambda}) + \ldots + \rho_G(\mathbf{c}_{i,t-1}\boldsymbol{\lambda})^G + u_{it}, t = 2,\ldots,T, \tag{14}$$

where α_0 and η_0 are new intercepts and $\mathbf{c}_{it} \equiv \mathbf{c}(\mathbf{x}_{it},\mathbf{m}_{it})$.

• Can specify instrumental variables (IVs) for each of these two equations. The most straightforward choice of IVs for (13) is simply

$$\mathbf{z}_{it1} \equiv (1,\mathbf{w}_{it},\mathbf{x}_{it},\mathbf{c}_{it}^o), \tag{15}$$

where $\mathbf{c}_{it}^o$ is $\mathbf{c}_{it}$ but without $\mathbf{x}_{it}$. The choice in (15) corresponds to the regression analysis in OP and LP for estimating $\boldsymbol{\beta}$ in a first stage.

• Under (5), any nonlinear function of $(\mathbf{w}_{it},\mathbf{x}_{it},\mathbf{c}_{it}^o)$ is also a valid IV, as are all lags and all functions of these lags. Adding a lag could be useful for generating overidentifying restrictions to test the model assumptions.

• Instruments for (14) would include $(\mathbf{x}_{it}, \mathbf{w}_{i,t-1}, \mathbf{c}_{i,t-1})$ and, especially if $G > 1$, nonlinear functions of $\mathbf{c}_{i,t-1}$ (probably low-order polynomials). Lags more than one period back are valid, too – say, one lag:

$$\mathbf{z}_{it2} = (1, \mathbf{x}_{it}, \mathbf{w}_{i,t-1}, \mathbf{c}_{i,t-1}, \mathbf{q}_{i,t-1}), \quad (16)$$

where $\mathbf{q}_{i,t-1}$ is a set of nonlinear functions of $\mathbf{c}_{i,t-1}$, probably consisting of low-order polynomials.

• Total of $2 + J + K + Q + G$ parameters in (14). $(\mathbf{x}_{it}, \mathbf{w}_{i,t-1}, \mathbf{c}_{i,t-1})$ act as their own instruments, and then we would include enough nonlinear functions in $\mathbf{q}_{i,t-1}$ to identify $\rho_1, \ldots, \rho_G$.

• A sensible choice for the instrument matrix for the two equations: for each (i, t),

$$\mathbf{Z}_{it} \equiv \begin{pmatrix} (\mathbf{w}_{it}, \mathbf{c}_{it}, \mathbf{z}_{it2}) & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{it2} \end{pmatrix}, t = 2, \ldots, T. \quad (17)$$

This choice makes it clear that all instruments available for (15) are also valid for (16), and we have some additional moment restrictions in (15).

• GMM estimation of all parameters in (13) and (14) is straightforward. For each $t > 1$, define a 2×1 residual function as

$$\mathbf{r}_{it}(\boldsymbol{\theta}) = \begin{pmatrix} y_{it} - \alpha_0 - \mathbf{w}_{it}\boldsymbol{\beta} - \mathbf{x}_{it}\boldsymbol{\gamma} - \mathbf{c}_{it}\boldsymbol{\lambda} \\ y_{it} - \eta_0 - \mathbf{w}_{it}\boldsymbol{\beta} - \mathbf{x}_{it}\boldsymbol{\gamma} - \rho_1(\mathbf{c}_{i,t-1}\boldsymbol{\lambda}) - \ldots - \rho_G(\mathbf{c}_{i,t-1}\boldsymbol{\lambda})^G \end{pmatrix}, \quad (18)$$

so that

$$E[\mathbf{Z}_{it}'\mathbf{r}_{it}(\boldsymbol{\theta})] = \mathbf{0}, t = 2, \ldots, T. \quad (19)$$

• Wooldridge (2008, Economics Letters) contains more details.

• Interestingly, in one leading case – namely, that productivity follows a random walk with drift – the moment conditions are linear in the parameters. Using $G = 1$ and $\rho_1 = 1$, the residual functions become $r_{it1}(\boldsymbol{\theta}) = y_{it} - \alpha_0 - \mathbf{w}_{it}\boldsymbol{\beta} - \mathbf{x}_{it}\boldsymbol{\gamma} - \mathbf{c}_{it}\boldsymbol{\lambda}$ and $r_{it2}(\boldsymbol{\theta}) = y_{it} - \eta_0 - \mathbf{w}_{it}\boldsymbol{\beta} - \mathbf{x}_{it}\boldsymbol{\gamma} - \mathbf{c}_{i,t-1}\boldsymbol{\lambda}$. So write

$$\mathbf{y}_{it} = \mathbf{X}_{it}\boldsymbol{\theta} + \mathbf{r}_{it}$$

where $\mathbf{y}_{it}$ is the 2×1 vector with y_{it} in both elements,

$$\mathbf{X}_{it} = \begin{pmatrix} 1 & 0 & \mathbf{w}_{it} & \mathbf{x}_{it} & \mathbf{c}_{it} \\ 0 & 1 & \mathbf{w}_{it} & \mathbf{x}_{it} & \mathbf{c}_{i,t-1} \end{pmatrix}, \quad (20)$$

and $\boldsymbol{\theta} = (\alpha_0, \eta_0, \boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\lambda}')'$. $\mathbf{Z}_{it}$ as in (17).

6. Pseudo Panels from Pooled Cross Sections

• It is important to distinguish between the population model and the sampling scheme. We are interested in estimating the parameters of

$$y_t = \eta_t + \mathbf{x}_t\boldsymbol{\beta} + f + u_t,\ t = 1,\ldots,T, \tag{21}$$

which represents a population defined over T time periods.

• Normalize $E(f) = 0$. Assume all elements of $\mathbf{x}_t$ have some time variation. To interpret $\boldsymbol{\beta}$, contemporaneous exogeneity conditional on f:

$$E(u_t|\mathbf{x}_t, f) = 0, t = 1,\ldots,T. \tag{22}$$

But, the current literature does not even use this assumption. We will use an implication of (22):

$$E(u_t|f) = 0, t = 1,\ldots,T. \tag{23}$$

Because f aggregates all time-constant unobservables, we should think of (22) as implying that $E(u_t|g) = 0$ for any time-constant variable g, whether unobserved or observed.

• Deaton (1985) considered the case of independently sampled cross sections. Assume that the population for which (21) holds is divided into G groups (or cohorts). Common is birth year. For a random draw i at time t, let g_i be the group indicator, taking on a value in $\{1,2,\ldots,G\}$.

• By our earlier discussion,

$$E(u_{it}|g_i) = 0. \tag{24}$$

• Taking the expected value of (21) conditional on group membership and using only (24), we have

$$E(y_t|g) = \eta_t + E(\mathbf{x}_t|g)\boldsymbol{\beta} + E(f|g),\ t = 1,\ldots,T. \tag{25}$$

This is Deaton's starting point, and Moffitt (1993). If we start with (21) under (23), there is no "randomness" in (25). Later authors have left $u_{gt}^* = E(u_t|g)$ in the error term.

• Define the population means

$$\alpha_g = E(f|g),\ \mu_{gt}^y = E(y_t|g),\ \boldsymbol{\mu}_{gt}^{\mathbf{x}} = E(\mathbf{x}_t|g) \tag{26}$$

for $g = 1,\ldots,G$ and $t = 1,\ldots,T$. Then for $g = 1,\ldots,G$ and $t = 1,\ldots,T$, we have

$$\mu_{gt}^y = \eta_t + \boldsymbol{\mu}_{gt}^{\mathbf{x}}\boldsymbol{\beta} + \alpha_g. \tag{27}$$

• Equation (27) holds without any assumptions restricting the dependence between $\mathbf{x}_t$ and u_r across t and r. In fact, $\mathbf{x}_t$ can contain lagged dependent variables or contemporaneously endogenous variables. Should we be suspicious?

• Equation (27) looks like a linear regression model in the population means, μ_{gt}^{y} and $\boldsymbol{\mu}_{gt}^{\mathbf{x}}$. One can use a “fixed effects” regression to estimate η_t, α_g, and $\boldsymbol{\beta}$.

• With large cell sizes, N_{gt} (number of observations in each group/time period cell), better to treat as a minimum distance problem. One inefficient MD estimator is fixed effects applied to the sample means, based on the same relationship in the population:

$$\boldsymbol{\beta} = \left(\sum_{g=1}^{G}\sum_{t=1}^{T} \ddot{\boldsymbol{\mu}}_{gt}^{\mathbf{x}\prime}\ddot{\boldsymbol{\mu}}_{gt}^{\mathbf{x}}\right)^{-1}\left(\sum_{g=1}^{G}\sum_{t=1}^{T} \ddot{\boldsymbol{\mu}}_{gt}^{\mathbf{x}\prime}\mu_{gt}^{y}\right) \tag{28}$$

where $\ddot{\boldsymbol{\mu}}_{gt}^{\mathbf{x}}$ is the vector of residuals from the pooled regression

$$\boldsymbol{\mu}_{gt}^{\mathbf{x}} \text{ on } 1, d2, \ldots, dT, c2, \ldots, cG, \tag{29}$$

where dt denotes a dummy for period t and cg is a dummy variable for group g.

• From (28), clear that underlying population model cannot contain a full set of group/time interactions. We *could* allow this feature with individual-level data. Absense of full cohort/time effects in the population model is the key identifying restriction.

• $\boldsymbol{\beta}$ is not identified if we can write $\boldsymbol{\mu}_{gt}^{\mathbf{x}} = \boldsymbol{\lambda}_t + \boldsymbol{\omega}_g$ for vectors $\boldsymbol{\lambda}_t$ and $\boldsymbol{\omega}_g$, $t = 1,\ldots,T$, $g = 1,\ldots,G$. So, we must exclude a full set of group/time effects in the structural model but we need some interaction between them in the covariate means. Identification might still be weak if variation in $\{\ddot{\boldsymbol{\mu}}_{gt}^{\mathbf{x}} : t = 1,..,T, g = 1,\ldots,G\}$ is small: a small change in estimates of $\boldsymbol{\mu}_{gt}^{\mathbf{x}}$ can lead to large changes in $\hat{\boldsymbol{\beta}}$.

• Estimation by nonseparable MD because $\mathbf{h}(\boldsymbol{\pi},\boldsymbol{\theta}) = \mathbf{0}$ are the restrictions on the structural parameters $\boldsymbol{\theta}$ given cell means $\boldsymbol{\pi}$ (Chamberlain, lecture notes). But given $\boldsymbol{\pi}$, conditions are linear in $\boldsymbol{\theta}$. After working it through, the optimal estimator is intuitive and easy to obtain. After “FE” estimation, obtain the residual variances within each cell, $\hat{\tau}_{gt}^{2}$, based on $y_{itg} - \mathbf{x}_{it}\check{\boldsymbol{\beta}} - \hat{\alpha}_g - \check{\eta}_t$, where $\check{\boldsymbol{\beta}}$ is the “FE” estimate, and so on.

• Define “regressors” $\hat{\boldsymbol{\omega}}_{gt} = (\hat{\boldsymbol{\mu}}_{gt}^{\mathbf{x}\prime}, \mathbf{d}_t, \mathbf{c}_g)$, and let $\hat{\mathbf{W}}$ be the $GT \times (K + T + G - 1)$ stacked matrix (where we drop, say, the time dummy for the first period.). Let $\hat{\mathbf{C}}$ be the $GT \times GT$ diagonal matrix with $\hat{\tau}_{gt}^{2}/(N_{gt}/N)$ down the diagonal.

• The optimal MD estimator, which is $\sqrt{N}$-asymptotically normal, is

$$\hat{\boldsymbol{\theta}} = (\hat{\mathbf{W}}'\hat{\mathbf{C}}^{-1}\hat{\mathbf{W}})^{-1}\hat{\mathbf{W}}'\hat{\mathbf{C}}^{-1}\hat{\boldsymbol{\mu}}^{y}. \tag{30}$$

As in separable cases, the efficient MD estimator looks like a "weighted least squares" estimator and its asymptotic variance is estimated as $(\hat{\mathbf{W}}'\hat{\mathbf{C}}^{-1}\hat{\mathbf{W}})^{-1}/N$.

• Bootstrapping to account for "weak" identification?

• Inoue (2008) obtains a different limiting distribution, which is stochastic, because he treats estimation of $\boldsymbol{\mu}_{gt}^{\mathbf{x}}$ and μ_{gt}^{y} asymmetrically.

• Deaton (1985), VN (1993), and Collado (1998), use a different asymptotic analysis: $GT \to \infty$ (Deaton) or $G \to \infty$, with fixed cell sizes.

• Allows for models with lagged dependent variables, but now the vectors of means contain redundancies. If

$$y_t = \eta_t + \rho y_{t-1} + \mathbf{z}_t\boldsymbol{\gamma} + f + u_t,\ E(u_t|g) = 0, \tag{31}$$

then the same moments are valid. But, now we would define the vector of means as $(\mu_{gt}^{y}, \boldsymbol{\mu}_{gt}^{\mathbf{z}})$, and appropriately pick off μ_{gt}^{y} in defining the moment conditions. We now have fewer moment conditions to estimate the parameters.

• The MD approach applies to extensions of the basic model. Random trend model (Heckman and Hotz, 1989):

$$y_t = \eta_t + \mathbf{x}_t\boldsymbol{\beta} + f_1 + f_2 t + u_t. \tag{32}$$

$$\mu_{gt}^{y} = \eta_t + \boldsymbol{\mu}_{gt}^{\mathbf{x}}\boldsymbol{\beta} + \alpha_g + \varphi_g t, \tag{33}$$

• We can even estimate models with time-varying factor loads on the heterogeneity:

$$y_t = \eta_t + \mathbf{x}_t\boldsymbol{\beta} + \lambda_t f + u_t, \tag{34}$$

$$\mu_{gt}^{y} = \eta_t + \boldsymbol{\mu}_{gt}^{\mathbf{x}}\boldsymbol{\beta} + \lambda_t\alpha_g. \tag{35}$$

• How can we use a stronger assumption, such as $E(u_t|\mathbf{z}_t, f) = \mathbf{0}$, $t = 1, \ldots, T$, for instruments $\mathbf{z}_t$, to more precisely estimate $\boldsymbol{\beta}$? Gives lots of potentially useful moment conditions:

$$E(\mathbf{z}_t' y_t|g) = \eta_t E(\mathbf{z}_t'|g) + E(\mathbf{z}_t'\mathbf{x}_t|g)\boldsymbol{\beta} + E(\mathbf{z}_t' f|g), \tag{36}$$

using $E(\mathbf{z}_t' u_t|g) = \mathbf{0}$.