Three Papers on Poisson Regression

Chencheng Fang, BGSE

July 8, 2023

1 Intro

This handout discusses three important papers on Poisson regression. KING, 1989 focuses on a Seemingly Unrelated Poisson Regression Model, considering the covariance between parameter estimates of two Poisson regressions. Hausman et al., 1984 talks about panel data Poisson models with random or fixed effects. Wooldridge, 1999 discusses a distribution-free estimation of some nonlinear panel data models including fixed effects Poisson model. Moreover, I explain in section 4 why there is no such a so-call Seemingly Unrelated Fixed Effects Poisson Model, which attempts to combine SUR Poisson in KING, 1989 and FE Poisson in Hausman et al., 1984.

2 KING, 1989

2.1 Setup

It is assumed y_{1i}^*, y_{2i}^* and U are all generated from Poisson distributions and independent at observation i. For observations i and j, all three random variables are also uncorrelated among themselves and each other.

$$
\begin{aligned}
y_{1i}^* &\sim \text{Poisson}(\lambda_{1i}) \\
y_{2i}^* &\sim \text{Poisson}(\lambda_{2i}) \\
U &\sim \text{Poisson}(\xi)
\end{aligned}
$$

The following two observed dependent variables y_{1i} and y_{2i} are defined.

$$
\begin{aligned}
y_{1i} &= y_{1i}^* + U \\
y_{2i} &= y_{2i}^* + U
\end{aligned}
$$

By the property that the sum of variables with Poisson distribution is still Poisson distribution (See Appendix 6.1), we have:

$$
\begin{aligned}
y_{1i} &= y_{1i}^* + U \sim \text{Poisson}(\lambda_{1i} + \xi) = \text{Poisson}(\theta_{1i}) \\
y_{2i} &= y_{2i}^* + U \sim \text{Poisson}(\lambda_{2i} + \xi) = \text{Poisson}(\theta_{2i})
\end{aligned}
$$

The covariance between y_{1i} and y_{2i} is $\text{cov}(y_{1i}^* + U, y_{2i}^* + U) = \text{var}(U) = \xi$

2.2 MLE

The conditional joint pdf for observation i:

$$
\begin{aligned}
P(y_{1i}, y_{2i} | \lambda_{1i}, \lambda_{2i}, \xi) &= P(y_{1i}^* + U, y_{2i}^* + U | \lambda_{1i}, \lambda_{2i}, \xi) \\
&= \sum_{j=0}^{y_{1i} \wedge y_{2i}} P(U = j, y_{1i}^* = y_{1i} - j, y_{2i}^* = y_{2i} - j) \\
&= \sum_{j=0}^{y_{1i} \wedge y_{2i}} P(U = j) P(y_{1i}^* = y_{1i} - j) P(y_{2i}^* = y_{2i} - j) \\
&= \exp(-\xi - \lambda_{1i} - \lambda_{2i}) \sum_{j=0}^{y_{1i} \wedge y_{2i}} \frac{\xi}{j!} \frac{\lambda_{1i}^{(y_{1i} - j)}}{(y_{1i} - j)!} \frac{\lambda_{2i}^{(y_{2i} - j)}}{(y_{2i} - j)!}
\end{aligned}
$$

Equivalently,

$$P(y_{1i}, y_{2i}|\theta_{1i}, \theta_{2i}, \xi) = \exp(\xi - \theta_{1i} - \theta_{2i}) \sum_{j=0}^{y_{1i} \wedge y_{2i}} \frac{\xi}{j!} \frac{(\theta_{1i} - \xi)^{(y_{1i}-j)}}{(y_{1i} - j)!} \frac{(\theta_{2i} - \xi)^{(y_{2i}-j)}}{(y_{2i} - j)!}$$

Now, suppose $\mathbb{E}(y_{1i}|X_{1i}) = \exp(x'_{1i}\beta_1)$ and $\mathbb{E}(y_{2i}|X_{2i}) = \exp(x'_{2i}\beta_2)$, by the property that $\mathbb{E}(y_{1i}|X_{1i}) = \theta_{1i}$ and $\mathbb{E}(y_{2i}|X_{2i}) = \theta_{2i}$, we have likelihood function over all observations:

$$\begin{aligned}
L &= \prod_{i=1}^{n} P(y_{1i}, y_{2i}|\beta_{1i}, \beta_{2i}, \xi) \\
&= \prod_{i=1}^{n} \exp(\xi - e^{x'_{1i}\beta_1} - e^{x'_{2i}\beta_2}) \sum_{j=0}^{y_{1i} \wedge y_{2i}} \frac{\xi}{j!} \frac{(x'_{1i}\beta_1 - \xi)^{(y_{1i}-j)}}{(y_{1i} - j)!} \frac{(x'_{2i}\beta_2 - \xi)^{(y_{2i}-j)}}{(y_{2i} - j)!}
\end{aligned}$$

Hence, log likelihood function is also attained. FOC and SOC of log likelihood can thus be derived. FOC needs to be equal to 0 and SOC needs to be negative, so we can have a unique maximizer.

2.3 Efficiency Gain

According to information equality, the asymptotic variance of our MLE estimator is the inverse of Fisher information. KING, 1989 proves that the difference between this asymptotic variance and the variance in separate exponential Poisson model is negative semi-definite, thus achieving higher efficiency.

3 Hausman et al., 1984

3.1 Random Effect

It's assumed that $\tilde{\lambda}_{it} = \lambda_{it}\tilde{\alpha}_i$, where λ_{it} is the parameter of Poisson distributions and $\tilde{\alpha}_i$ is a random firm specific effect. By assumption, $\tilde{\lambda}_{it} \not\perp \tilde{\lambda}_{i\prime t}$ for $t \neq \prime t$; $\tilde{\lambda}_{it} \perp \tilde{\lambda}_{jt}$.

Thus, rewrite $\tilde{\lambda}_{it}$ with respect to X_{it}: $\tilde{\lambda}_{it} = \lambda_{it}\alpha_i = \exp(X_{it}\beta + \mu_i)$, where μ_i is the firm specific effect. Then conditional pdf is:

$$
\begin{aligned}
P(n_{it}|X_{it},\mu_i) =& \frac{e^{-\exp(X_{it}\beta+\mu_i)}\left(e^{X_{it}\beta+\mu_i}\right)^{n_{it}}}{n_{it}!} \\
=& \frac{e^{-\lambda_{it}\exp(\mu_i)}\left(\lambda_{it}e^{\mu_i}\right)^{n_{it}}}{n_{it}!}
\end{aligned}
$$

Now, assume $g(\mu_i)$ is the pdf of μ, and the conditional density of μ_i on X_{it} is equal to the unconditional density of μ_i. That means μ_i is independent of X_{it} and thus μ is assumed to be randomly distributed across firms. Then, we have the joint pdf over all periods and μ_i:

$$
\begin{aligned}
P(n_{i1},\ldots,n_{iT},\mu_i|X_{i1},\ldots,_{iT}) =& P(n_{i1},\ldots,n_{iT}|X_{i1},\ldots,X_{iT},\mu_i)g(\mu_i) \\
=& \prod_t \left(\frac{\lambda_{it}^{n_{it}}}{n_{it}!}\right) e^{-exp(\mu_i)\sum\limits_t \lambda_{it}} \left(e^{\mu_i}\right)^{\sum\limits_t n_{it}} g(\mu_i)
\end{aligned}
$$

In order to integrate out μ_i from the joint pdf, we have to assume the distribution of μ_i, and follow the equation (2.3) to the joint pdf without any condition on μ_i and then get the log likelihood function. However, at most time, it's hard for us to know the distribution of μ_i.

3.2 Fixed Effects

In the derivation of random effects, we have assumptions like the independence of μ_i and X_{it} and what the distribution of μ_i is. These assumptions are strong, and will be relaxed in fixed effects setting.

In fixed effects, μ_i is considered as a covariate, and is not independent from X_{it}. But as we only care about β, μ_i is nuisance parameter, which may incur incidental parameter problem if we still use the pdf as shown in the random effects specification. Instead, Hausman et al., 1984 use the conditional MLE of Andersen, 1970 and condition on $\sum\limits_t n_{it}$.

For firm i, a sufficient statistic for $T\tilde{\lambda}_i = \sum\limits_t \tilde{\lambda}_{it}$ is $\sum_t n_{it}$, and $\sum_t n_{it}$ is distributed as Poisson with parameter $\sum\limits_t \tilde{\lambda}_{it} = \alpha_i \sum\limits_t \lambda_{it}$. Hence,

$$
\begin{aligned}
P(n_{i1},\cdots,n_{iT}|\sum_t n_{it}) =& \frac{P(n_{i1},\cdots,n_{i,T-1},\sum\limits_{t=1}^{T} n_{it}-\sum\limits_{t=1}^{T-1} n_{it})}{P(\sum\limits_t n_{it})} \\
=& \frac{\frac{e^{-\sum\limits_t \tilde{\lambda}_{it}}\prod\limits_t \tilde{\lambda}_{it}^{n_{it}}}{\prod\limits_t (n_{it}!)}}{\frac{e^{-\sum\limits_t \tilde{\lambda}_{it}}\left(\sum\limits_t \tilde{\lambda}_{it}\right)^{\sum\limits_t n_{it}}}{(\sum\limits_t n_{it})!}} \\
=& \frac{(\sum\limits_t n_{it})!}{\prod\limits_t (n_{it}!)}\prod_t \left(\frac{\tilde{\lambda}_{it}}{\sum\limits_t \tilde{\lambda}_{it}}\right)^{n_{it}}
\end{aligned}
$$

Set $p_{it} = \frac{\tilde{\lambda}_{it}}{\sum\limits_t \tilde{\lambda}_{it}}$ and then we have multinomial distribution since $\sum\limits_t p_{it} = 1$. And by using $\tilde{\lambda}_{it} = \exp(X_{it}\beta + \mu_i)$, we have:

$$
\begin{aligned}
p_{it} =& \frac{\exp(X_{it}\beta + \mu_i)}{\sum\limits_t \exp(X_{it}\beta + \mu_i)} \\
=& \frac{\exp(X_{it}\beta)}{\sum\limits_t \exp(X_{it}\beta)}
\end{aligned}
$$

This is the so-called multinomial logit model used by McFadden, 1974. Define the share of patents for firm i in a given year by $s_{it} = \frac{n_{it}}{\sum\limits_t n_{it}}$. The logit model then explains the share of total patents in each year given the firms' total number of patents in T years.

The log likelihood function is then

$$
L(\beta) = C_3 - \sum_{i=1}^{N}\sum_{t=1}^{T} n_{it}\log\sum_{s=1}^{T} e^{-(X_{it}-X_{is})\beta}
$$

See Appendix 6.2 for the derivation of the non-constant part of log likelihood function. FOC and SOC can thus be calculated. This MLE is slightly different from discrete choice likelihood function in McFadden, 1974, as the dependent variable is ratio instead of dummies. The existing logit model algorithm needs to be revised to fit into this specification.

3.3 Simulation

In this section, simulation on the fixed effects specification above is performed based on an available R package poisFErobust. In my setup, I observe iid data $\{(Y_{it}, X_{it}) : i = 1, \ldots, n; t = 1, \ldots, 10\}$ with $Y_{it} \in \mathbb{R}$ and $X_{it} = (X_{it}^{(1)}, X_{it}^{(2)}, X_{it}^{(3)}, X_{it}^{(4)})^T \in \mathbb{R}^4$, following the model:

$$Y_{it} = \text{Poisson}(\exp(X_{it}^T \beta + u_i)), \quad i = 1, \ldots, n; t = 1, \ldots, 10$$

where $\beta = (0.5, 1, 1.5, 2)^T$ and $u_i \sim \mathcal{N}(0, 0.0025)$ is fixed effect. This setup satisfies the conditional mean assumption. Then, I use the proposed Fixed effect Poisson regression to estimate β. Figure 1 are histograms of each coefficients when $n = 1500$ and when I simulate the data for 100 times.

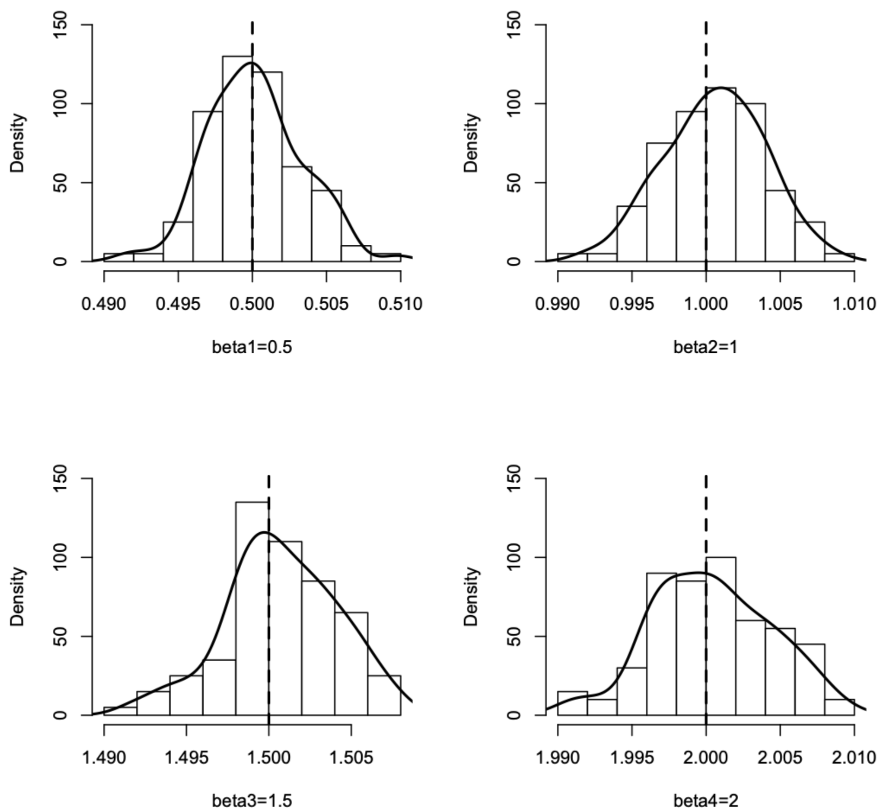

Figure 1: *Unbiasedness*

From Figure 1, it can be generally seen that the estimator is unbiased, and the variance is not large. Also, to observe the consistency of the estimator, I run simulations for $n = 10, 50, 100, 200, \ldots, 1500$, and 2 depicts how the mean of estimated coefficients over 100 simulations changes with n.

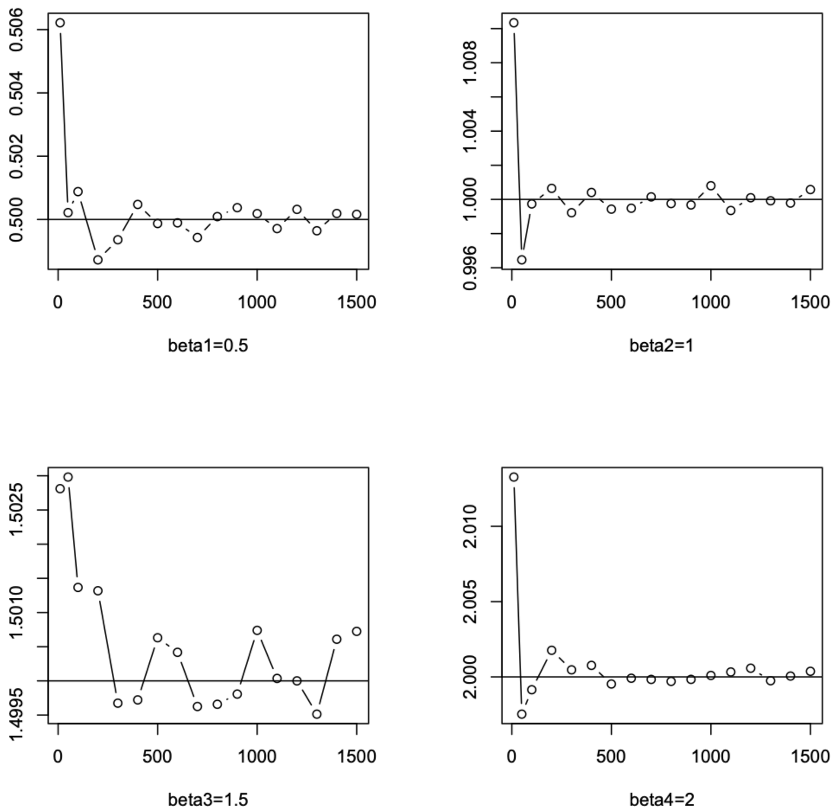

Figure 2: *Consistency*

Generally, with the increase of n, the estimator converges to true parameter in probability.

3.4 Negative Binomial Model

With fixed effects Poisson model, we have restriction $En_{it} = \text{var}(n_{it}) = \lambda_{it}$, while with random effects Poisson has a variance to mean ratio of $\frac{\lambda_{it}+\delta}{\delta}$ which increases with λ_{it}. Hausman et al., 1984 wants to somehow combine the properties of these two models to permit the variance to grow with mean and have a conditional fixed effects α_i. He begins with a usual negative binomial regression without fixed effects.

It is assumed that the parameter λ_{it} follows a gamma distribution with paramters (γ, δ) and it is specified that $\gamma = \exp(X_{it}\beta)$ and δ is constant. Then, according to the properties of gamma distribution, $E(\lambda_{it}) = \frac{\exp(X_{it}\beta)}{\delta}$ and $\mathrm{var}(\lambda_{it}) = \frac{\exp(X_{it}\beta)}{\delta^2}$. Because $\mathrm{var}(\lambda_{it}) \neq 0$, λ_{it} varies even when X_{it} remains the same for a firm i over all time periods. Here, we haven't allowed for firm specific effects, so λ_{it} are independent for a given firm over time. Hausman et al., 1984 now takes the gamma distribution of λ_{it} and integrate over it to find the pdf of n_{it}.

$$
\begin{aligned}
P(n_{it}) &= \int_0^\infty \frac{1}{n_{it}!} e^{-\lambda_{it}} \lambda_{it}^{n_{it}} f(\lambda_{it}) d\lambda_{it} \\
&= \frac{\Gamma(\gamma + n_{it})}{\Gamma(\gamma)\Gamma(n_{it}+1)} \left(\frac{\delta}{1+\delta} \right)^{\gamma} (1+\delta)^{-n_{it}}
\end{aligned}
$$

which is NB distribution with (γ, δ). See Appendix 6.3 for why gamma mixture of Poisson distribution is negative binomial distribution. So, we have $E(n_{it}) = \frac{\exp(X_{it}\beta)}{\delta}$ and $\mathrm{var}(n_{it}) = \frac{\exp(X_{it}\beta)(1+\delta)}{\delta^2}$. Therefore, the variance to mean ratio is $\frac{1+\delta}{\delta} > 1$, meaning the negative binomial regression allows for overdispersion, and the overdispersion is the same for all i. Then, by following similar arguments in Poisson regression, we can have the maximum likelihoof function, and thus FOC and SOC.

3.5 NB model with firm specific effects

It is easier to describe the fixed effects model and then add the random interpretation to it. Similar to what he has done in FE Poisson regression, he needs to find the distribution of $\sum\limits_t n_{it}$ for a given firm i. The Moment generating function (MGF) of negative binomial distribution is $m(t) = \left(\frac{1+\delta-e^t}{\delta} \right)^{-\gamma}$. Because the MGF of the sum of independent random variables(RV) is the product of MFG of each RV. So, we can see that if δ is common for two independent negative binomial RV w_1 and w_2, then $w_1 + w_2 = z$ is distributed as a negative binomial distribution with parameters $(\gamma_1 + \gamma_2, \delta)$. So, the conditional distribution of w_1 given z is:

$$P(w_1|z = w_1 + w_2) = \frac{P(w_1)P(z - w_1)}{P(z)}$$
$$= \frac{\Gamma(\gamma_1 + w_1)\Gamma(\gamma_2 + w_2)\Gamma(\gamma_1 + \gamma_2)\Gamma(w_1 + w_2 + 1)}{\Gamma(\gamma_1 + \gamma_2 + z)\Gamma(\gamma_1)\Gamma(\gamma_2)\Gamma(w_1 + 1)\Gamma(w_2 + 1)}$$

So, we see that the resulting conditional pdf doesn't depend on δ. This δ can vary across firms as long as it does not vary over time. This δ is considered fixed effects. This distribution is usually called negative hypergeometric distribution. More generally, we have:

$$P(n_{i1}, \ldots, n_{iT}| \sum_t n_{it}) = \left(\prod_t \frac{\Gamma(\gamma_{it} + n_{it})}{\Gamma(\gamma_{it})\Gamma(n_{it} + 1)} \right) \left(\frac{\Gamma(\sum_t \gamma_{it})\Gamma(\sum_t n_{it} + 1)}{\Gamma(\sum_t \gamma_{it} + \sum_t it)} \right)$$

This distribution is sometimes called negative multivariate hypergeometric distribution. Hausman et al., 1984 also shows an alternative way to get this conditional pdf, which leads to the same result. The log likelihood function follows once γ_{it} is specified.

It is assumed that the parameters of the underlying model is $(\gamma_{it}, \delta_i) = \left(\exp(X_{it}\beta), \frac{\Phi_i}{\exp(\mu_i)}\right)$ where Φ_i and μ_i are allowed to vary across firms. Thus, $E(n_{it}) = \frac{\exp(X_{it}\beta + \mu_i)}{\Phi_i}$ and $\text{var}(n_{it}) = \frac{\exp(X_{it}\beta) + \frac{\mu_i}{\Phi_i^2}}{1 + \frac{\Phi_i}{\exp(\mu_i)}}$. So, the variance to mean ratio is $\left(\frac{\exp(\mu_i)}{\Phi_i}\right)\left(1 + \frac{\Phi_i}{\exp(\mu_i)}\right)$.

By doing so, this model allows for both overdispersion and a firm specific variance to mean ratio. Log likelihood can thus be computed, and FOC, SOC are thus found.

4 Is it possible to combine SUR Poisson and FE Poisson?

KING, 1989 talks about Seemingly Unrelated Poisson, while Hausman et al., 1984 discusses panel data count models including fixed effects Poisson regression. One may think about if it is possible to combine these two models by taking advantage of both? My answer is No.

Generally speaking, seemingly unrelated regression needs the estimated covariance between residuals of each regression. However, we are unable to estimate the error term in Poisson regression like how we do in OLS, unless we define the error term as how KING, 1989 does. Let's first explain why we don't have error term in Poisson regression.

In reference to aphe, n.d., suppose we have generalized linear model $g(\mu_i) = \alpha + x_i^T\beta$, where $g(\mu_i)$ is considered to be link function and u_i is the conditional expectation. Hence, for linear regression, $g(\mu_i) = \mu_i$; for logistic regression, $g(\mu_i) = \log(\frac{\mu_i}{1-\mu_i})$; and for Poisson regression, $g(\mu_i) = \log(\mu_i)$. A straight forward idea of writing error term is that the observed outcome is: $y_i = g^{-1}(\alpha + x_i^T\beta) + e_i$ where $E(e_i) = 0$ or a constant and $\mathrm{var}(e_i) = \sigma^2(\mu_i)$. For Poisson regression, $\sigma^2(\mu_i) = \exp(\alpha + x_i^T\beta)$. But, in general, we can not explicitly state that e_i has a Poisson distribution unless we specify that like KING, 1989.

Another way of thinking is that the mean and variance relationship is already captured in the generalized linear model, hereby Poisson model. So, if we have two Poisson models (of course regressors need to be exogenous, otherwise we have bipoisson), these models are more than just seemingly unrelated, but completely unrelated.

It is just mentioned that if the error term in Poisson regression is defined to follow a Poisson distribution, then we could estimate error term. But, this type of specification is not in line with individual specific effect. When we say individual specific effect, we mean this effect would be the same for each individual over all time periods. However, if we define the individual specific effect as an additive variable with Poisson distribution, then the effect to the main model would never be the same due to the variance of Poisson distribution itself.

Therefore, there is no way to combine SUR Poisson in KING, 1989 and FE Poisson in Hausman et al., 1984. I also found another paper about SUR Negative Binomial in Winkelmann, 2000, and the error term is defined in a similar way as KING, 1989, so by following the same arguments, there is no way to combine SUR Negative Binomial and FE NB.

5 Wooldridge, 1999

I will not go through the proof of Wooldridge, 1999, but only introduce the main results.

In section 3 of Wooldridge, 1999, it is shown that multinomial Quasi-conditional Maximum Likelihood Estimator (QCMLE), also known as fixed effects Poisson[1], is consistent with only the conditional mean assumption (i.e., $E(y_{it}|X_{it}, \phi_i) = \phi_i \mu(X_{it}, \beta)$). Wooldridge, 1999 also proposes a specification testing to test this assumption. This consistency is proved to be robust to any failure of Poisson assumptions.

It is also said QCMLE based on multinomial distribution provides consistent estimates for β under the assumption that (1) conditional pdf is negative binomial and (2) no serial correlation. So, FE Negative Binomial estimator is also consistent under certain conditions, which are stronger than FE Poisson. In section 5 of Wooldridge, 1999, it is mentioned that any failure of assumptions (5.1) to (5.3) would lead to misspecification of FENB, and the consistency may not hold under such a misspecification. Moreover, FENB assumes overdispersion for all i, which is also too restrictive. Consistency of FE Poisson allows both underdispersion and overdispersion.

Prof. Wooldridge clearly states the advantage of using FE Poisson over FENB at a forum. The support of his claim at the forum could all be found in Wooldridge, 1999.

[1] The QCMLE is defined in a more generalized way with ϕ_i and $\mu(X_{it}, \beta)$. For FE Poisson in Hausman et al., 1984, $\mu(X_{it}, \beta) = \exp(X_{it}\beta)$ and $\phi_i = \exp(u_i)$;

6 Appendix

6.1 Sum of Variables with Poisson Distribution

Proposition 1. *If* $X \sim \text{Poisson}(\lambda)$ *and* $Y \sim \text{Poisson}(\mu)$*, then* $X + Y \sim \text{Poisson}(\lambda + \mu)$

Proof.

$$\begin{aligned}
P(X+Y=k) &= \sum_{i=0}^{k} P(X+Y=k, X=i) \\
&= \sum_{i=0}^{k} P(X=i)P(Y=k-i) \\
&= \sum_{i=0}^{k} \exp(-\lambda)\frac{\lambda^i}{i!}\exp(-\mu)\frac{\mu^{k-i}}{(k-i)!} \\
&= \exp(\lambda+\mu)\sum_{i=0}^{k}\frac{\lambda^i\mu^{k-i}}{i!(k-i)!} \\
&= \exp(\lambda+\mu)\frac{1}{k!}\sum_{i=0}^{k}\frac{k!}{i!(k-i)!}\lambda^i\mu^{k-i} \\
&= \exp(\lambda+\mu)\frac{1}{k!}(\lambda+\mu)^k \\
&\sim \text{Poisson}(\lambda+\mu)
\end{aligned}$$

∎

6.2 Log likelihood function of non-constant part of fixed effects specification

Proof.

$$
\begin{aligned}
\log\left(\prod_i\prod_t p_{it}^{n_{it}}\right) &= \log\left(\prod_i\prod_t\left(\frac{\exp(X_{it}\beta)}{\sum_t \exp(X_{it}\beta)}\right)^{n_{it}}\right)\\
&= \log\left(\prod_i\prod_t\left(\frac{\exp(X_{it}\beta)}{\sum_s \exp(X_{is}\beta)}\right)^{n_{it}}\right)\\
&= \log\left(\prod_i\prod_t\left(\frac{1}{\sum_s \exp\left(\left(X_{is}-X_{it}\right)\beta\right)}\right)^{n_{it}}\right)\\
&= \sum_i\sum_t n_{it}\log\left(\frac{1}{\sum_s \exp\left(-\left(X_{it}-X_{is}\right)\beta\right)}\right)\\
&= -\sum_i\sum_t n_{it}\log\left(\sum_s \exp\left(-\left(X_{it}-X_{is}\right)\beta\right)\right)
\end{aligned}
$$

■

6.3 Gamma mixture of Poisson distributions is a negative binomial distribution

Proof. In reference to Hardy, n.d., suppose the distribution of λ_{it} is Gamma distribution with (γ, δ):

$$\frac{1}{\Gamma(\alpha)}\left(\delta\lambda\right)^{\gamma-1}e^{-\delta\lambda}\left(d\delta\lambda\right)$$

Also, suppose the conditional distribution of n_{it} given λ_{it} is the Poisson distribution:

$$\frac{e^{-\lambda_{it}}\lambda_{it}^{n_{it}}}{n_{it}!}$$

So, we thus have:

$$
\begin{aligned}
P(n_{it}) =& E(P(n_{it}|\lambda_{it})) \\
=& E(\frac{e^{-\lambda_{it}}\lambda_{it}^{n_{it}}}{n_{it}!}) \\
=& \int_0^\infty \left(\frac{e^{-\lambda_{it}}\lambda_{it}^{n_{it}}}{n_{it}!}\right) \frac{1}{\Gamma(\alpha)} \left(\delta\lambda_{it}\right)^{\gamma-1} e^{-\delta\lambda_{it}} \left(d\delta\lambda_{it}\right) \\
=& \frac{\delta^\gamma}{n_{it}!\Gamma(\gamma)} \int_0^\infty \lambda_{it}^{n_{it}+\gamma-1} e^{-\lambda_{it}(1+\delta)} d\lambda_{it} \\
=& \frac{\delta^\gamma}{n_{it}!\Gamma(\gamma)} \left(\frac{1}{1+\delta}\right)^{n_{it}+\gamma} \int_0^\infty \left(\lambda_{it}\left(1+\delta\right)\right)^{n_{it}+\gamma-1} e^{-\lambda_{it}(1+\delta)} d\left(\lambda_{it}\left(1+\delta\right)\right) \\
=& \frac{\delta^\gamma}{n_{it}!\Gamma(\gamma)} \left(\frac{1}{1+\delta}\right)^{n_{it}+\gamma} \int_0^\infty u^{n_{it}+\gamma-1} e^{-u} du \\
=& \frac{\delta^\gamma \Gamma(n_{it}+\gamma)}{n_{it}!\Gamma(\gamma)} \left(\frac{1}{1+\delta}\right)^{n_{it}+\gamma} \\
=& \frac{\Gamma(n_{it}+\gamma)}{n_{it}!\Gamma(\gamma)} \left(\frac{1}{1+\delta}\right)^{n_{it}} \left(\frac{\delta}{1+\delta}\right)^{\gamma} \\
=& \frac{\Gamma(n_{it}+\gamma)}{\Gamma(n_{it}+1)\Gamma(\gamma)} \left(\frac{1}{1+\delta}\right)^{n_{it}} \left(\frac{\delta}{1+\delta}\right)^{\gamma}
\end{aligned}
$$

where the first step is by Law of Iterated Expectation.

■

References

Andersen, E. B. (1970). Asymptotic Properties of Conditional Maximum-Likelihood Estimators. *Journal of the Royal Statistical Society: Series B (Methodological)*, *32*(2).

aphe. (n.d.). Logistic regression - error term and its distribution [URL:https://stats.stackexchange.com/q/124847 (version: 2014-11-20)].

Hardy, M. (n.d.). Relationship between negative binomial distribution and Bayesian Poisson with Gamma priors [URL:https://stats.stackexchange.com/q/263101 (version: 2017-02-21)].

Hausman, J. A., Hall, B. H., & Griliches, Z. (1984). *Econometric models for count data with an application to the patents-rd relationship* (Working Paper No. 17). National Bureau of Economic Research.

KING, G. (1989). A seemingly unrelated poisson regression model. *Sociological Methods & Research*, *17*(3), 235–255.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. *Economic Theory and Mathematical Economics*, (105-142).

Winkelmann, R. (2000). Seemingly Unrelated Negative Binomial Regression. *Oxford Bulletin of Economics and Statistics*, (553-560).

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, *90*(1), 77–97.