Nonlinear Panel Data

Whitney Newey

Fall 2007

Panel data control for individual effects correlated with regressors.

Well known how to do this in linear models with additive effects.

Nonlinear model harder.

General set up:

Data: $Y_i = [Y_{i1}, ..., Y_{iT}]', X_i = [X_{i1}, ..., X_{iT}]', (i = 1, ..., n)$.

A linear model:

$$Y_{it} = X_{it}'\beta + \alpha_i + \eta_{it}, E[\eta_{it}|X_i, \alpha_i] = 0.$$

Alternative, equivalent formulation:

$$E[Y_{it}|X_i, \alpha_i] = X_{it}'\beta + \alpha_i.$$

Specifies the conditional mean of Y_i given X_i, α_i, and β.

Likelihood specifies conditional pdf $f(y|x, \alpha, \theta)$ of Y_i given X_i, α_i and parameter vector θ.

Example: Normal linear model: For e_T a $T \times 1$ vector of $1'$s,

$$Y_i|(X_i, \alpha_i) \sim N(X_i\beta + \alpha_i e_T, \sigma^2 I_T).$$

This is distributional version of a linear model.

Binary choice model: $Y_{it} \in \{0, 1\}$; e.g. labor force participation.

$$Y_{it}, (t = 1, ..., T) \text{ independent}, \text{Prob}(Y_{it} = 1|X_i, \alpha_i) = G(X_{it}'\beta + \alpha_i).$$

Count data: $Y_{i1}, ..., Y_{iT}$ indep, $Y_{it}|X_i, \alpha_i$ Poisson with mean $\exp(X_{it}'\beta + \alpha_i)$.

Linear model method is to transform data so α_i drops out. Differencing gives

$$E\left[Y_{it} - Y_{it-1}|X_i\right] = X_{it}'\beta + E[\alpha_i|X_i] - (X_{i,t-1}'\beta + E[\alpha_i|X_i]) = (X_{it} - X_{i,t-1})'\beta,$$

In nonlinear model, α_i does not drop out when we difference.

Binary choice example (What about linear probability model?):

$$E\left[Y_{it} - Y_{it-1}|X_i\right] = E[G(X_{it}'\beta + \alpha_i) - G(X_{it-1}'\beta + \alpha_i)|X_i].$$

Fixed Effects and the Incidental Parameters Problem

Fixed effects is maximizing the log-likelihood over each α_i as well as θ.

Fixed effects generally inconsistent in nonlinear model as n grows with T fixed.

In a linear model, least squares treating α_i as a parameter to be estimated is consistent.

Maximum likelihood treating α_i as a parameter to be estimated is generally not.

This is known as the *incidental parameters problem.*

It is caused by only having T observations to estimate each α_i, so that as n grows the estimate of α_i remains random.

In linear models this randomnes gets "averaged out." In nonlinear models it does not.

Limit of the fixed effects estimator as n grows with T fixed.

Estimator

$$\hat{\theta} = \arg \max_{\theta,\alpha_1,\ldots,\alpha_n} \frac{1}{n}\sum_{i=1}^{n} \ln f(Y_i|X_i,\theta,\alpha_i).$$

Concentrate out α_i: For a fixed θ each fixed effect is given by

$$\hat{\alpha}_i(\theta) = \max_{\alpha} \ln f(Y_i|X_i,\theta,\alpha_i).$$

Substituting in and maximize over θ to get $\hat{\theta}$,

$$\hat{\theta} = \arg\max_{\theta} \frac{1}{n}\sum_{i=1}^{n} \ln f(Y_i|X_i,\theta,\hat{\alpha}_i(\theta)).$$

By the usual extremum estimator, as n grows for fixed T the estimator $\hat{\theta}$ has plim

$$\theta_T = \arg\max_{\theta} E[\ln f(Y_i|X_i,\theta,\hat{\alpha}_i(\theta))].$$

$$\theta_T = \arg\max_{\theta} E[\ln f(Y_i|X_i, \theta, \hat{\alpha}_i(\theta))].$$

Randomness in $\hat{\alpha}_i(\theta)$ leads to inconsistnecy of $\hat{\theta}$.

$$\hat{\alpha}_i(\theta) = \max_{\alpha} \ln f(Y_i|X_i, \theta, \alpha_i).$$

If $\hat{\alpha}_i(\theta)$ were replaced by

$$\bar{\alpha}_i(\theta) = \arg\max_{\alpha} E[\ln f(Y|X, \theta, \alpha)],$$

would get consistency. Like measurement error in nonlinear model.

Example: Binary logit, $Y_{it} \in \{0, 1\}$, $G(u) = e^u/(1 + e^u)$.

Known that the fixed effects estimator $\hat{\beta}_{FE}$ satisfies

$$\hat{\beta}_{FE} \xrightarrow{p} 2\beta_0$$

Bias in $\hat{\beta}$ can be severe. Not so severe in Tobit model.

Example: Gaussian linear model, FE estimator of σ^2 converges to

$$\sigma_T^2 = \frac{T-1}{T}\sigma^2.$$

Bias in estimates of marginal effects less severe.

In binary choice, marginal effect is

$$\int [G(\tilde{X}'\beta_0 + \alpha) - G(\bar{X}'\beta_0 + \alpha)] F_\alpha(d\alpha).$$

Fixed effects estimator is

$$\sum_{i=1}^{n} G(\tilde{X}'\hat{\beta} + \hat{\alpha}_i) - G(\bar{X}'\hat{\beta} + \hat{\alpha}_i)]/n$$

Hahn and Newey (2004) show quite small biases for probit.

Return to this below.

Discuss now how can get consistent estimators.

Conditional Maximum Likelihood

Occasionally there is statistic S_i such that α_i drops out of the conditional likelihood of Y_i given X_i and S_i.

That is,

$$f(Y_i|X_i, S_i, \theta, \alpha_i) = f(Y_i|X_i, S_i, \theta).$$

Conditional MLE (CMLE).

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} f(Y_i|X_i, S_i, \theta)$$

Consistent and asymptotically normal, and asymptotically efficient when the distribution of α_i conditional on X_i is unrestricted.

Problem is S_i only exists in a few cases, including Gaussian linear model, logit binary choice, oisson model for count data, and proportional hazards model.

In most other models there is no such S_i, so conditional MLE has limited usefulness.

Identification Issue:

θ may not be identified in the semiparametric model where the conditional pdf of Y_i given X_i, α_i is specified as $f(y|x, \alpha, \theta)$ and the conditional pdf of α_i given X_i is unspecified.

Chamberlain (1992): $T = 2$; $\Pr(Y_{it} = 1|X_i, \alpha_i) = G(\alpha_0 d_{it} + x_{it}'\beta_0 + \alpha_i)$, $d_{i1} = 0$, $d_{i2} = 1$, $G'(u) > 0$ everywhere, other regularity conditions. If X_i is bounded then β_0 is not identified if $G(u)$ is not logistic.

Also can show that θ_0 is not identified for $T = 2$, $\Pr(Y_{it} = 1|X_i, \alpha_i) = \Phi(\beta_0 X_{it} + \alpha_i)$, $X_{it} \in \{0, 1\}$. See following graph.

Extent of nonidentification (e.g. for censored models) is not clear.

No consistent estimator in nonidentified cases.

Could directly estimate identified set.

Recent progress, Honore and Tamer (2006) and other work.

Difficult when X_{it} takes on many values.

Other approaches are a) restrict distribution of α_i given X_i; b) find clever estimators for identified models; c) large T fixed effect bias corrections;

Correlated Random Effects:

Restricts conditional distribution of α_i given X_i.

Here consider parametric models; there are nonparametric and semiparametric versions.

Let $g(\alpha|X,\gamma)$ be conditional pdf of α given X.

Likelihood of Y given X is integrates out α, as in

$$f(Y|X,\beta,\gamma) = \int f(Y|X,\beta,\alpha)g(\alpha|X,\gamma)d\alpha.$$

The MLE is given by

$$\hat{\beta},\hat{\gamma} = \arg\max_{\beta,\alpha} \frac{1}{n}\sum_{i=1}^{n} \ln f(Y_i|X_i,\beta,\gamma) = \frac{1}{n}\sum_{i=1}^{n} \ln \int f(Y_i|X_i,\beta,\alpha)g(\alpha|X_i,\gamma)d\alpha$$

Consistency of $\hat{\beta}$ depends on the $g(\alpha|X,\gamma)$ being correctly specified.

May be difficult to calculate the integral.

Also, hard to form $g(\alpha|X,\gamma)$ in time consistent fashion.

Example: Correlated random effects probit.

$Y_{it} = 1(Y_{it}^* > 0)$ where conditional on (X_i, α_i), $Y_{i1}^*, ..., Y_{iT}^*$ are independent and Y_{it}^* has distribution $N(X_{it}'\beta_0 + \alpha_i, \sigma_t^2)$. Let $x_i = vec(X_i')$ be the vector of all observations across t on the regressors. Suppose also that the conditional distribution of α_i given X_i is $N(x_i'\lambda, \sigma_\alpha^2)$. Note that conditional on X_i,

$$Y_{it}^* \sim N(X_{it}'\beta_0 + x_i'\lambda, \sigma_t^2 + \sigma_\alpha^2).$$

Then for $\theta = (\beta', \lambda', \sigma_1^2, ..., \sigma_T^2, \sigma_\alpha^2)'$ and e_t the t^{th} $T \times 1$ unit vector,

$$\begin{aligned} \Pr(Y_{it} &= 1|X_i, \theta) = -x_i'\lambda)/\sigma_\alpha^2)d\alpha = \Phi\left(\frac{X_{it}'\beta + x_i'\lambda}{\sqrt{\sigma_t^2 + \sigma_\alpha^2}}\right) \\ &= \Phi\left(x_i'\pi_t\right), \pi_t = \frac{e_t \otimes \beta + \lambda}{\sqrt{\sigma_t^2 + \sigma_\alpha^2}}. \end{aligned}$$

This is a marginal likelihood for Y_{it}.

Joint likelihood is very complicated. $Y_{i1}, ..., Y_{iT}$ not independent conditional on X_i. This is generally true in models where integrate out α_i.

Estimation: Do marginal likelihood (probit) to get $\hat{\pi}_1, ..., \hat{\pi}_T$. Normalize $\delta_1 = 1$ and let $\delta_t = 1/\sqrt{\sigma_t^2 + \sigma_\alpha^2}$, $(t = 1, ..., T)$, where we normalize $\delta_1 = 1$. Reparameterize so that $\theta = (\beta', \lambda', \delta_2, ..., \delta_T)'$ and for $\pi = (\pi_1', ..., \pi_T')'$ let

$$h(\pi, \theta) = \begin{pmatrix} \delta_1 \pi_1 - e_1 \otimes \beta - \lambda \\ \vdots \\ \delta_T \pi_T - e_T \otimes \beta - \lambda \end{pmatrix}.$$

We can then do minimum distance, using $\hat{\pi} = (\hat{\pi}_1', ..., \hat{\pi}_T')'$ mentioned above.

$$\hat{\theta} = \arg\min_\theta h(\hat{\pi}, \theta)' \hat{W} h(\hat{\pi}, \theta).$$

$h(\hat{\pi}, \theta)$ is linear in θ so easy to do.

Efficient two-step estimator. For $\hat{V}$ an estimator of the joint asymptotic variance of $\hat{\pi}$, let $\tilde{\theta} = \arg\min_\theta h(\hat{\pi}, \theta)' \hat{V}^{-1} h(\hat{\pi}, \theta)$. Then let $\hat{D} = diag(I, \tilde{\delta}_2 I, ..., \tilde{\delta}_T I)$ where I is an identity matrix with the same dimension as π. Then $\hat{D}\hat{V}\hat{D}$ is estimator of the variance of $\sqrt{n}(\hat{\pi} - \pi_0)$, so optimal minimum distance is

$$\hat{\theta} = \arg\min_\theta h(\hat{\pi}, \theta)' \left(\hat{D}\hat{V}\hat{D}\right)^{-1} h(\hat{\pi}, \theta).$$

Empirical example from Chamberlain (1984).

Labor force participation, with $n = 924$ and $T = 4$, four years. 1968, 70, 72, 74.

Two X_{it} number of children under 6 and number of children. Here are the results:

Probit	-.121	-.058	Logit	.-573	-.336
	(.046)	(.029)		(.115)	(.120)

Quite different estimates; ratios are similar.

Correlated random effects depends on T in an essential way.

Many coefficients. A more parsimonius model is $\alpha_i \sim N(\bar{X}_i'\lambda, \sigma_\alpha^2)$, $\bar{X}_i = \sum_{t=1}^{T} X_{it}/T$.

Marginal Effects

Marginal effect for change in X is, for $F(\alpha)$ the CDF of α,

$$\mu_t(\tilde{X}) - \mu_t(X), \mu(X) = \int \Phi((X'\beta_0 + \alpha)/\sigma_t)F(d\alpha)$$

By iterated expectations, holding X fixed,

$$\begin{aligned} \mu_t(X) &= E[1(X'\beta_0 + \alpha_i + \eta_{it} > 0)] = E[E[1(X'\beta_0 + \alpha_i + \eta_{it} > 0)|X_i]] \\ &= E[\Phi(\delta_t(X'\beta_0 + x_i'\lambda_0))] \end{aligned}$$

This object can be estimated by

$$\hat{\mu}_t(X) = \sum_{i=1}^{n} \Phi(\hat{\delta}_t(X'\hat{\beta} + x_i'\hat{\lambda}))/n$$

Would be interesting to compare this estimator with fixed effects marginal effect in the empirical example.

Some Semiparametric Results

Some distribution free results that are useful.

Poisson model: Conditional on (X_i, α_i), Y_{it} is independent over time and Poisson with mean $e^{X_{it}'\beta+\alpha_i}$. Good model for patents; see Hausman, Hall, Griliches (1984). Wooldridge showed that consistency of CMLE only requires

$$E\left[Y_{it}|X_i, \alpha_i\right] = e^{X_{it}\beta+\alpha_i}$$

Binary choice: Manski maximum score estimator; Conditions for consistency include infinite support.

Tobit: Honore

Manski and Honore require homoskedasticity over time.

Does not hold in linear model applications.

Large T Fixed Effects Bias Correction

Let θ_T denote plim of fixed effects estimator.

As T grows $\lim_{T \longrightarrow \infty} \theta_T = \theta_0$.

Under smoothness,

$$\theta_T = \theta_0 + \frac{B}{T} + O(\frac{1}{T^2}).$$

Example: Gaussian linear model

$$\sigma_T^2 = \frac{T-1}{T}\sigma^2 = \sigma^2 - \frac{\sigma^2}{T} = \sigma^2 + \frac{B}{T}, B = -\sigma^2.$$

Also n and T grow, we should have

$$(nT)^{1/2}\left(\widehat{\theta} - \theta_T\right) \xrightarrow{d} N(0, \Omega).$$

$$\theta_T = \theta_0 + \frac{B}{T} + O(\frac{1}{T^2}), (nT)^{1/2}\left(\widehat{\theta} - \theta_T\right) \xrightarrow{d} N(0, \Omega).$$

As a way to think about how bad fixed effects bias can be, consider $n/T \longrightarrow \rho$.

$$\begin{aligned}
(nT)^{1/2}\left(\widehat{\theta} - \theta_0\right) &= (nT)^{1/2}\left(\widehat{\theta} - \theta_T\right) + (nT)^{1/2}(\theta_T - \theta_0) \\
&= (nT)^{1/2}\left(\widehat{\theta} - \theta_T\right) + (nT)^{1/2}\frac{B}{T} + O((nT)^{1/2}/T^2) \\
&\xrightarrow{d} N\left(B\rho^{1/2}, \Omega\right).
\end{aligned}$$

Here there is asymptotic bias.

Consequently, usual asymptotic confidence intervals incorrect.

Asymptotic normality of $\widehat{\theta}$, centered at its probability limit, like misspecification result (e.g. White, 1982).

Analytical Bias Correction

Find formula for B, construct estimator $\hat{B}$. Bias corrected estimator is

$$\hat{\theta}_1 = \hat{\theta} - \hat{B}/T.$$

To show when this works, suppose

$$(nT)^{1/2}(\hat{B} - B)/T \xrightarrow{p} 0.$$

For example, if $\hat{B}$ itself has $(nT)^{1/2}(\hat{B} - B)$ asymptotically normal then holds.

Plugging in as before we get,

$$\begin{aligned}
(nT)^{1/2}\left(\widehat{\theta}_1 - \theta_0\right) &= (nT)^{1/2}\left(\widehat{\theta} - \theta_T\right) \\
&\quad +(nT)^{1/2}(\theta_T - \theta_0 - \hat{B}/T) \\
&= (nT)^{1/2}\left(\widehat{\theta} - \theta_T\right) + (nT)^{1/2}(B - \hat{B})/T \\
&\quad +O((nT)^{1/2}/T^2) \\
&\xrightarrow{d} N(0, \Omega).
\end{aligned}$$

Iterated Analytical Correction

Often the bias formula will depend on θ, so that $\hat{B} = \tilde{B}(\hat{\theta})$.

Can iterate the bias correction:

$$\hat{\theta}_j = \hat{\theta} - \tilde{B}(\hat{\theta}_{j-1})/T.$$

Iterating to convergence would give

$$\hat{\theta}_\infty = \hat{\theta} - \tilde{B}(\hat{\theta}_\infty)/T.$$

Does not improve asymptotic properties.

Can improve small sample properties.

Jackknife Bias Correction

Use how $\hat{\theta}$ changes with T to form implicit bias correction.

Does not require formula for B.

Let $\hat{\theta}_{(t)}$ denote fixed effects estimator not using t^{th} time period.

Jackknife estimator is

$$\widetilde{\theta} \equiv T\widehat{\theta} - (T-1)\sum_{t=1}^{T}\widehat{\theta}_{(t)}/T.$$

Explain with expansion,

$$\theta_T = \theta_0 + \frac{B}{T} + \frac{D}{T^2} + O\left(\frac{1}{T^3}\right).$$

imit of $\widetilde{\theta}$ for fixed T and how it changes with T shows bias correction.

$$\begin{aligned} \tilde{\theta} &\xrightarrow{p} T\theta_T - (T-1)\,\theta_{T-1} = \theta_0 + \left(\frac{1}{T} - \frac{1}{T-1}\right)D + O\left(\frac{1}{T^2}\right) \\ &= \theta_0 + O\left(\frac{1}{T^2}\right). \end{aligned}$$

Example: Variance estimation in Gaussian I model (Neyman and Scott, 1948):

z_{it} is i.i.d. with distribution $N(\alpha_i, \theta_0)$.

Here

$$\theta_T = \frac{T-1}{T}\theta_0 = \theta_0 - \frac{\theta_0}{T}.$$

Thus $B = -\theta_0$. Analytical correction:

$$\hat{\theta}_1 = \hat{\theta} + \hat{\theta}/T \xrightarrow{p} \left(\frac{T-1}{T} + \frac{T-1}{T^2}\right)\theta_0$$

Is not consistent for fixed T. Iterating analytical correction is

$$\begin{aligned} \hat{\theta}_\infty &= \hat{\theta} + \hat{\theta}_\infty/T, \\ \hat{\theta}_\infty &= \frac{T}{T-1}\hat{\theta}. \end{aligned}$$

Can also show that this is jackknife. Here is consistent for fixed T.

Monte Carlo Example: Like Heckman (1981). Design is:

$$\begin{aligned} y_{it} &= 1(x_{it}\theta_0 + \alpha_i + \varepsilon_{it} > 0), \\ \alpha_i &\sim N(0,1), \varepsilon_{it} \sim N(0,1), \\ x_{it} &= t/10 + x_{i,t-1}/2 + u_{it}, \\ x_{i0} &= u_{i0},\, u_{it} = U(-1/2, 1/2). \\ N &= 100, T = 8; \beta = 1, -1. \end{aligned}$$

Marginal effect is average derivative of $\Phi(x'\theta + \alpha)$,

$$\mu = \theta_0 \bar{E}[\phi(x'\theta_0 + \alpha_i)].$$

The fixed effects estimator of this object is

$$\hat{\mu} = \hat{\theta} \sum_{i=1}^{n} \phi\left(x'\hat{\theta} + \hat{\alpha}_i\right)/n.$$

Consider analytical and jacknife bias corrections.

Table Three: Properties of $\hat{\theta}$, $T = 8$.					
Estimator of θ_0	Mean	Med.	SD	$\hat{p}$; .05	$\hat{p}$; .10
MLE	1.18	1.17	.151	.267	.370
Jackknife	.953	.950	.119	.056	.102
Analytic	1.05	1.05	.134	.062	.135
Analytic-M	1.05	1.05	.132	.060	.126

Table Five: Properties of $\hat{\theta}$, $T = 4$					
Estimator of θ_0	Mean	Med.	SD	$\hat{p}$; .05	$\hat{p}$; .10
MLE	1.42	1.41	.397	.269	.373
Jackknife	.752	.743	.262	.100	.177
Analytic	1.12	1.11	.306	.055	.101
Analytic-M	1.21	1.20	.335	.102	.172

Table Four: Properties of $\hat{\mu}$, $T = 8$.					
Estimator of μ/μ_0	Mean	Med.	SD	$\hat{p}$;.05	$\hat{p}$;.10
MLE	1.02	1.02	.131	.078	.140
Jackknife	1.00	.992	.130	.086	.159
Analytic	1.02	1.02	.133	.090	.153
Analytic-M	1.02	1.02	.131	.087	.154

Table Six: Properties of $\hat{\mu}$, $T = 4$.					
Estimator of μ/μ_0	Mean	Med.	SD	$\hat{p}$; .05	$\hat{p}$; .10
MLE	1.00	1.00	.257	.103	.168
Jackknife	1.06	1.05	.307	.159	.224
Analytic	.996	.994	.265	.113	.178
Analytic-M	1.05	1.05	.266	.117	.185

Bounds for Marginal Effects:

Assume $X_{it} \in \{0,1\}$. $\Pr(Y_{it} = 1|X_i, \alpha_i) = \Phi(\theta_0 X_{it} + \alpha_i)$.

Object of interest

$$\mu_0 = \int [\Phi(\theta_0 + \alpha) - \Phi(\alpha)] F_0(d\alpha)$$

Average change in the probability of $Y_{it} = 1$.

Let $\overrightarrow{0}$ and $\overrightarrow{1}$ denote $T \times 1$ vectors of $0's$ and $1's$ respectively.

Define

$$\mu^* = \int [\Phi(\theta_0 + \alpha) - \Phi(\alpha)] F_0(d\alpha | X_i \notin \{\overrightarrow{0}, \overrightarrow{1}\}).$$

Then μ^* is identified.

$$\mu^* = \int [\Phi(\theta_0 + \alpha) - \Phi(\alpha)] F_0(d\alpha | X_i \notin \{\overrightarrow{0}, \overrightarrow{1}\}).$$

μ^* is identified.

Proof: Consider $X \notin \{\overrightarrow{0}, \overrightarrow{1}\}$. Then there is $t(X)$ such that $x_{t(X)} = 1$ and $s(X)$ such that $x_{s(X)} = 1$. Then we have

$$\begin{aligned} E[y_{i,t(X)} - y_{i,s(X)} | X_i &= X] = E[y_{i,t(X)} - y_{i,s(X)} | X_i = X, \alpha_i] | X_i = X] \\ &= \int [\Phi(\theta_0 + \alpha) - \Phi(\alpha)] F_0(d\alpha | X_i = X). \end{aligned}$$

Let $P(X) = \Pr(X_i = X)$. Then

$$\mu^* = \sum_{X \notin \{\overrightarrow{0}, \overrightarrow{1}\}} P(X) E[y_{i,t(X)} - y_{i,s(X)} | X_i = X].$$

$$\mu^* = \sum_{X \notin \{\overrightarrow{0}, \overrightarrow{1}\}} P(X) E[y_{i,t(X)} - y_{i,s(X)} | X_i = X].$$

Cannot identify $\int [\Phi(\theta_0 + \alpha) - \Phi(\alpha)] F_0(d\alpha | \overrightarrow{x})$ for $x \in \{\overrightarrow{0}, \overrightarrow{1}\}$.

μ^* is over identified for $T > 2$.

Simple estimator:

Let $n^* = \#\{i : X_i \notin \{\overrightarrow{0}, \overrightarrow{1}\}\}$.

$$\hat{\mu}^* = \frac{1}{n^*} \sum_{X \notin \{\overrightarrow{0}, \overrightarrow{1}\}} \sum_{\{i | X_i = X\}} \left(y_{i,t(X)} - y_{i,s(X)} \right).$$

Bounds for μ_0.

Let $D = 1(\mu_* > 0)$. Let $\bar{P} = P(\overrightarrow{0}) + P(\overrightarrow{1})$

$$(1 - \bar{P})\mu^* - (1 - D)\bar{P} \leq \mu_0 \leq \mu^*(1 - \bar{P}) + D\bar{P}$$

Tight bounds use the form $\Phi(\theta_0 + \alpha) - \Phi(\alpha)$.

Bounds shrink to a point exponentially fast at T grows.

There are 2^T possible X so $P(\overrightarrow{0}) + P(\overrightarrow{1})$ will shrink like $C2^{-T}$ for some constant C.

This fast shrinkage rate might be conjectured fom the bias corrections.

In smooth models (all derivative existing) one can form a bias correction that approaches the truth at T^{-J} for any integer J.