UNIVERSITÄT BIELEFELD

Fakultät für Wirtschaftswissenschaften

Einführung Mikroökonometrie

310208 (VO)

Dietmar Bauer

Wiederholung

- Letztes Mal haben wir uns mit gepoolten Querschnittsdaten beschäftigt. Darunter versteht man gleichartige Daten aus zwei verschiedenen Stichproben, also etwa zu verschiedenen Zeitpunkten erhoben nicht notwendigerweise von den gleichen Individuen.
- Wir haben gesehen, dass Dummy-Variable verwendet werden können, um alle Daten mit einem Modell zu beschreiben.
- Dabei helfen Dummies und Interaktionsterme mit, unterschiedliche Parameter in verschiedenen Stichproben zu modellieren.
- Mittels t-Tests und F-Tests können wir Hypothesen prüfen, ob Koeffizienten in verschiedenen Stichproben ident sind oder statistische signifikant verschieden.

Wiederholung (II)

- Für Politikanalysen (Evaluierung der Auswirkung von Massnahmen) werden oft Paneldatensätze verwendet, bei denen die Stichproben von den gleichen Individuen zu mindestens zwei Zeitpunkten gezogen werden (vorher - nachher) Studien.
- Bei solchen Paneldatensätzen verwenden wir die Notation $\mathbf{y}_{i,t}$ (i ... Individuum, t ... Time), um kenntlich zu machen, dass wir mehr Beobachtungen der gleichen Individuen haben.
- Beim Vergleich von verschiedenen Individuen und verschiedenen Zeitpunkten besteht immer die Möglichkeit, dass sich die Individuen unterscheiden und dass es zeitliche Trends gibt, die auf alle Individuen im gleichen Maße zutreffen.
- Um diese zu modellieren, verwenden wir individuenspezifische Effekte und zeitliche Dummyvariable.

Wiederholung (III)

Beispiel für 2 Zeitpunkte:

$$\mathbf{y}_{i,t} = \beta_0 + \beta_1 \delta(t = 2) + \beta' \mathbf{X}_{i,t} + \mathbf{a}_i + \mathbf{u}_{i,t}, t = 1, 2, i = 1, ..., N.$$

- Darin enthalten sind $2 + K + N$ unbekannte Größen (plus die Fehlerterme).
- Für großes N sind das sehr viele.
- Betrachtet man aber die zeitlichen Differenzen (Veränderungen zwischen den 2 Zeitpunkten), dann fallen viele Größen heraus:

$$\Delta \mathbf{y}_{i,t} = \beta_1 + \beta' \Delta \mathbf{X}_{i,t} + \mathbf{u}_{i,t}, t = 2, i = 1, ..., N.$$

Paneldaten

Paneldaten

In den folgenden Stunden werden wir uns mit sogenannten **Paneldatensätzen** beschäftigen:

Beobachtung von mehreren Variablen einzelner Individuen zu verschiedenen aufeinanderfolgenden Zeitpunkten:

$$y_{i,t}, x_{i,t} : \qquad i = 1, ..., N, t = 1, ..., T$$

Unterscheidung:

- **balanzierte Panels:** für alle Individuen stehen alle Beobachtungen zur Verfügung.
- **unbalanzierte Panels:** sind nicht balanziert, es gibt also Individuen, die nicht zu allen Zeitpunkten beobachtet wurden.

Zunächst nur balanzierte Panels.

4. Paneldaten

$$\mathbf{y}_{i,t} = \beta' \mathbf{X}_{i,t} + \mathbf{a}_i + \mathbf{u}_{i,t}, \quad , t = 1, ..., T, i = 1, ..., N$$

- Auch für mehr als zwei Zeitpunkte kann man die ersten Differenzen betrachten, um die individuenspezifischen Effekte los zu werden:

$$\Delta \mathbf{y}_{i,t} = \mathbf{y}_{i,t} - \mathbf{y}_{i,t-1} = \beta' \Delta \mathbf{X}_{i,t} + \Delta \mathbf{u}_{i,t}, \quad , t = \mathbf{2}, ..., T, i = 1, ..., I$$

- Dabei 'verlieren' wir die ersten Beobachtung, weil wir zum ersten Zeitpunkt noch keine Änderung berechnen können.
- Und wir können keine Regressoren verwenden, die konstant über die Zeit sind: dann wäre die Änderung $\Delta \mathbf{x}_{i,t,k} = 0$ und wir hätten einen Regressor, der nicht variiert.

Annahme (FD: first differences)

FD.1: Der datengenerierende Prozess erfüllt die Gleichung

$$\mathbf{y}_{i,t} = \beta' \mathbf{X}_{i,t} + \mathbf{a}_i + \mathbf{u}_{i,t}, \quad , t = 1, ..., T, i = 1, ..., I$$

wobei $T \geq 2$.

FD.2: Wir beobachten eine Zufallsstichprobe im Querschnitt (die Daten sind unabhängig über die Individuen hinweg).

FD.3: Kein Regressor ist konstant über die Zeit und über Individuen hinweg und es existieren keine perfekten linearen Beziehungen zwischen den Regressoren.

FD.4 (strikte Exogenität): Wenn $\mathbf{X}_i \in \mathbb{R}^{T \times K}$ die Regressormatrix für Individuum i bezeichnet dann gelte:

$$\mathbb{E}(\mathbf{u}_{i,t} | \mathbf{X}_i, \mathbf{a}_i) = 0$$

FD.5: $\mathrm{Var}(\Delta \mathbf{u}_{i,t} | \mathbf{X}_i) = \sigma^2, t = 2, ..., T.$

FD.6: Für alle $t \neq s$ sind die differenzierten Fehler unkorreliert in der Zeit:

$$\mathrm{Cov}(\Delta \mathbf{u}_{i,t}, \Delta \mathbf{u}_{i,s} | \mathbf{X}_i) = 0$$

- Unter den Annahmen FD ist der FD-Schätzer der BLUS. Besonders FD.6 ist sehr speziell.
- Wenn FD.6 nicht zutrifft aber FD.1-FD.5, dann ist der FD Schätzer immer noch unverzerrt, aber nicht mehr optimal.
- Dann gibt es andere Schätzer, die kleinere Varianz besitzen.
- Die Formulierung der Annahmen auf die ersten Differenzen $\Delta \mathbf{u}_{i,t}$ ist ungewohnt, wir sind Annahmen gewohnt, die die Fehler $\mathbf{u}_{i,t}$ direkt betreffen.
- Wenn $\mathbf{u}_{i,t}$ zum Beispiel Messfehler sind, dann ist die Annahme von Unabhängigkeit der Messfehler oft aus der Theorie her motiviert.
- Dann gilt aber:

$$\mathrm{Cov}(\Delta \mathbf{u}_{i,t}, \Delta \mathbf{u}_{i,t-1}) = \mathrm{Cov}(\mathbf{u}_{i,t} - \mathbf{u}_{i,t-1}, \mathbf{u}_{i,t-1} - \mathbf{u}_{i,t-2}) = -\mathrm{Var}(\mathbf{u}_{i,t-1}) \neq 0.$$

Daher suchen wir nach anderen Ideen!

4. 1. Beispieldatensatz

Datenbeispiel:

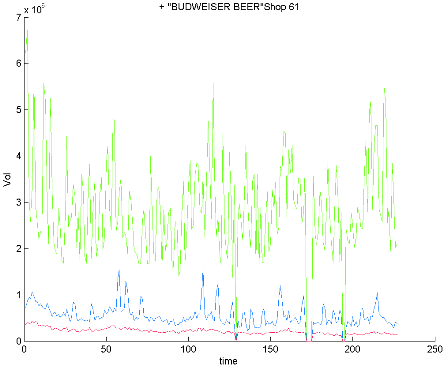
- Dominick's Fine Foods Trade-Panel Datensatz.
- Supermarktkette in Chicago, stellte einen großen Datensatz der Öffentlichkeit zur Verfügung: Website.
- über vier Jahre an Daten in 25 verschiedenen Produktkategorien in allen 100 Geschäften der Kette.

4. 1. Beispieldatensatz

Wir verwenden Bierverkäufe als Beispiel. Die Daten haben vier Dimensionen:

- Zeit: wöchentliche Verkaufszahlen über insgesamt mehr als 4 Jahre, $T = 227$.
- Ort: wir betrachten die 27 umsatzstärksten Geschäfte, also $N = 27$.
- Biersorten: Die Produkte weisen eine Hierarchie auf:
  - **Produzent:** Anheuser Busch, Miller, Pabst, ...
  - **Marke:** Budweiser, Miller, ...
  - **Submarke:** Bud Light, Miller Ice, ...
  - **Verpackung:** Sixpack, Kiste, ...
- Variablen:
  - Umsatz in Volumen (Unzen)
  - Preis (in USD pro Unze)
  - Marketingmassnahmen: Display, Folder, Preisnachlass (nicht verlässlich kodiert).

4. 1. Beispieldatensatz



4. 1. Beispieldatensatz

Was heißt Preis eines Produktes genau?

- Produkt kann eine ganze Produktkategorie sein: zum Beispiel alle Bud Light Produkte.
- In dem Fall ist der Preis der Mittelwert aller Preise der Subprodukte.
- Gemittelt wird gewichtet mit dem Volumenumsatzanteil.
- Das Gleiche gilt für andere Variablen.

4. 1. Beispieldatensatz

Marketingtheorie sagt uns, dass ...

- Kunden nach dem Basispreis und der Preisreduktion getrennt entscheiden:
  - Basispreis: langfristig, geringe Variation
  - Preisreduktion: kurzfristig, durch Aktionen bestimmt
- wir mit verzögerten Effekten rechnen sollten: Leads und Lags von Preisnachlässen.
  - Vorziehkäufe verlagern Umsätze nach vorne
  - Stockpiling nach hinten: Warten auf den nächsten Preisnachlass.
- Aktionen bei Konkurrenten den Umsatz beeinflussen können. Aktionen bei verwandten Produkten heben möglicherweise den Umsatz.
- es unbeobachtete Faktoren gibt: Events und TV-Werbung sind nicht in den Daten enthalten. Ebenso Aktionen bei Waren, die oft im gleichen Warenkorb landen.

4. 1. Beispieldatensatz

- Wir wollen den Umsatz (in *log*(Volumen), nicht in Dollar) für ein Produkt in den Geschäften modellieren.
- Als Regressoren stehen uns die Marketingmassnahmen und der Preis zur Verfügung.
- zusätzlich wird der Umsatz von Store-Variablen abhängen: Größe, Einzugsbereich, ...

Modell:

$$\log(\mathbf{Y}_{i,t}) = \mathbf{y}_{i,t} = \beta' \mathbf{X}_{i,t} + \mathbf{a}_i + \mathbf{u}_{i,t}$$

In $\mathbf{a}_i$ steckt ein multiplikativer Faktor:

$$\mathbf{Y}_{i,t} = \exp(\beta' \mathbf{X}_{i,t} + \mathbf{a}_i + \mathbf{u}_{i,t}) = \exp(\mathbf{a}_i) \left(\exp(\beta' \mathbf{X}_{i,t}) \exp(\mathbf{u}_{i,t}) \right)$$

4. 1. Beispieldatensatz

Was steckt in den individuenspezifischen Effekten $\mathbf{a}_i$?

- Größe der Geschäfte
- Einzugsbereich
- Vorlieben der Kunden eines Geschäfts
- alles, was sich nicht über die Zeit ändert, aber von Shop zu Shop unterschiedlich sein kann.

Loswerden der fixen Effekte

Within Transformation: (innerhalb der Gruppe)

$$\begin{aligned}
\mathbf{y}_{i,t} &= \beta'\mathbf{X}_{i,t} + \mathbf{a}_i + \mathbf{u}_{i,t}, \\
\overline{\mathbf{y}_{i,:}} &= \beta'\overline{\mathbf{X}_{i,:}} + \mathbf{a}_i + \overline{\mathbf{u}_{i,:}} \\
\mathbf{y}_{i,t} - \overline{\mathbf{y}_{i,:}} &= \beta'(\mathbf{X}_{i,t} - \overline{\mathbf{X}_{i,:}}) + \mathbf{u}_{i,t} - \overline{\mathbf{u}_{i,:}} \\
\dot{\mathbf{y}}_{i,t} &= \beta'\dot{\mathbf{X}}_{i,t} + \dot{\mathbf{u}}_{i,t}.
\end{aligned}$$

In der letzten Gleichung kann der KQ-Schätzer berechnet werden, da hier die individuenspezifischen Effekte nicht vorkommen.

Bemerkung:

- In der Gleichung taucht keine Konstante auf.
- Sie würde durch die Within Transformation auch wegfallen.

4. 1. Beispieldatensatz

Annahme (Fixe Effekte)

FE.1: Für jedes i ist der datengenerierende Prozess

$$\mathbf{y}_{i,t} = \beta' \mathbf{X}_{i,t} + \mathbf{a}_i + \mathbf{u}_{i,t}$$

FE.2: Die Daten sind eine Zufallsstichprobe aus dem Querschnitt.

FE.3: Jeder Regressor variiert über die Zeit (zumindest für ein i) und es bestehen keine exakten linearen Zusammenhänge zwischen den Regressoren.

FE.4: Für jedes t ist der bedingte Erwartungswert bezüglich den Regressoren aller Perioden und der fixen Effekte gleich null:

$$\mathbb{E}\left(\mathbf{u}_{i,t} | \mathbf{X}_i, \mathbf{a}_i\right) = 0$$

Unter diesen Annahmen ist der KQ-Schätzer nach der Within- Transformation unverzerrt.

4. 1. Beispieldatensatz

Annahme (Fixe Effekte (cont'ed))

FE.5:

$$\mathrm{Var}(\mathbf{u}_{i,t}|\mathbf{X}_i, \mathbf{a}_i) = \sigma^2, \quad i = 1, ..., N, t = 1, ..., T.$$

FE.6: Für alle $t \neq s$ gilt:

$$\mathrm{Cov}(\mathbf{u}_{i,t}, \mathbf{u}_{i,s}|\mathbf{X}_i, \mathbf{a}_i) = 0, t \neq s, \quad i = 1, ..., N$$

FE.7: Bedingt auf $\mathbf{X}_i$ und $\mathbf{a}_i$ sind die Fehler $\mathbf{u}_{i,t}$ unabhängig und identisch normalverteilt mit Mittel null und Varianz σ^2.

Unter Annahmen FE.1-FE.6 ist der KQ-Schätzer der BLUS.
Unter den Annahmen FE.1-FE.4, FE.7 haben die t-Tests und F-Tests die namensgebenden Verteilungen.

Bemerkung: Wir haben keine Annahme eingeführt, die die Beziehung zwischen den fixen Effekten $\mathbf{a}_i$ und den Regressoren $\mathbf{X}_i$ beschränkt.

Warum ist das so?

Zurück zum Querschnittsmodell $\mathbf{y}_i = \beta_0 + \beta' \mathbf{X}_i + \mathbf{u}_i$.

Zwei Möglichkeiten der Schätzung:

- Gemeinsame Schätzung von β_0, β, in Matrixschreibweise
  $Y = [\iota_N, X] \begin{bmatrix} \beta_0 \\ \beta \end{bmatrix} + U = X_+ \beta_+ + U \Rightarrow \hat{\beta}_+ = (X_+' X_+)^{-1} X_+' Y$.
- Zuerst β_0 schätzen bedingt auf β, dann β:
  $\overline{Y} = \hat{\beta}_0 + \overline{X}\beta + \overline{U}$, $\dot{Y} = Y - \iota_N \overline{Y} = (X - \iota \overline{X})\beta + U - \iota_N \overline{U} = \dot{X}\beta + \dot{U} \Rightarrow$
  $\hat{\beta} = (\dot{X}'\dot{X})^{-1}\dot{X}'\dot{Y} \Rightarrow \hat{\beta}_0 = \overline{Y} - \overline{X}(\dot{X}'\dot{X})^{-1}\dot{X}'\dot{Y}$.

Man kann zeigen, dass beide Wege zum gleichen Ergebnis führen.

Das FWL Theorem

Theorem (Frisch-Waugh-Lovell)

Bei der Schätzung der Gleichung $\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\gamma + \mathbf{u}_i$ kann man wie folgt vorgehen:

1. Regressiere $\mathbf{y}_i$ auf $\mathbf{Z}_i$, d.h. schätze die Gleichung $\mathbf{y}_i = \mathbf{Z}_i\gamma + \mathbf{v}_i$. Das ergibt die Residuen $\mathbf{y}_{i,-Z}$.
2. Regressiere $\mathbf{X}_i$ auf $\mathbf{Z}_i$, d.h. schätze die Gleichung $\mathbf{X}_i = \mathbf{Z}_i\Pi + \mathbf{V}_i$. Das ergibt die Residuen $\mathbf{X}_{i,-Z}$.
3. Schätze die Gleichung $\mathbf{y}_{i,-Z} = \mathbf{X}_{i,-Z} + \mathbf{u}_{i,-Z}$. Das ergibt die Schätzer $\tilde{\boldsymbol{\beta}}$.

Der Schätzer $\tilde{\boldsymbol{\beta}}$ ist dann ident mit dem Schätzer $\hat{\boldsymbol{\beta}}$, der aus der gemeinsamen Gleichung geschätzt wird.

Man kann also die Schätzung aufteilen, und zuerst einen Teil der Regressoren 'herausrechnen' und dann eine Gleichung schätzen, die die jeweiligen Residuen einbezieht.

4. 1. Beispieldatensatz

Der Beweis folgt aus einer Anwendung der sogenannten Block-Matrizen-Inversionsformel:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & D^{-1} \end{bmatrix} + \begin{bmatrix} I \\ -D^{-1}C \end{bmatrix} (A - BD^{-1}C)^{-1} \begin{bmatrix} I & -BD^{-1} \end{bmatrix}.$$

Diese wird angewandt auf die gemeinsame Regression:

$$\begin{aligned} \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} &= \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix}^{-1} \begin{bmatrix} X'Y \\ Z'Y \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & (Z'Z)^{-1} \end{bmatrix} \begin{bmatrix} X'Y \\ Z'Y \end{bmatrix} \\ &+ \begin{bmatrix} I \\ -(Z'Z)^{-1}Z'X \end{bmatrix} (X'X - X'Z(Z'Z)^{-1}Z'X)^{-1} \begin{bmatrix} I & -X'Z(Z'Z)^{-1} \end{bmatrix} \begin{bmatrix} X'Y \\ Z'Y \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ (Z'Z)^{-1}Z'Y \end{bmatrix} + \begin{bmatrix} I \\ -(Z'Z)^{-1}Z'X \end{bmatrix} (X'P_Z X)^{-1} X'P_Z Y \end{aligned}$$

wobei $P_Z = I - Z(Z'Z)^{-1}Z'$ die Projektion auf das orthogonale Komplemtent des Spaltenraums der Matrix Z bezeichnet. Das Theorem folgt dann aus der Tatsache, dass $P_Z X$ die Residuen einer Regression von X auf Z ergibt (Hier hilft die Idee, dass die KQ-Schätzung eine Projektion im $\mathbb{R}^N$ ist).

4. 1. Beispieldatensatz

Warum ist das so?

Betrachte man nun $\mathbf{a}_i$ als Parameter durch die Einführung von einer Dummyvariablen pro Individuum:

$$\mathbf{y}_{i,t} = \beta'\mathbf{X}_{i,t} + \sum_{r=1}^{N} a_r \delta(r = i) + \mathbf{u}_{i,t}$$

Dann kann man das FWL-Theorem anwenden, wobei Z durch die Spalten der Dummies gebildet wird.

Paneldaten: Fixed Effects

In Matrixschreibeweise:

$$
Y = \left(\begin{array}{c}
y_{1,1} \\ y_{1,2} \\ \vdots \\ y_{1,T} \\ \hline y_{2,1} \\ \vdots \\ y_{2,T} \\ \hline \vdots \\ \hline y_{N,1} \\ \vdots \\ y_{N,T}
\end{array}\right), \quad
[X, Z] = \left(\begin{array}{cccc|cccc}
x_{1,1,1} & x_{1,1,2} & \dots & x_{1,1,K} & 1 & 0 & \dots & 0 \\
x_{1,2,1} & x_{1,2,2} & \dots & x_{1,2,K} & 1 & 0 & \dots & 0 \\
\vdots & \vdots & & \vdots & 1 & 0 & \dots & 0 \\
x_{1,T,1} & x_{1,T,2} & \dots & x_{1,T,K} & 1 & 0 & \dots & 0 \\ \hline
x_{2,1,1} & x_{2,1,2} & \dots & x_{2,1,K} & 0 & 1 & \dots & 0 \\
\vdots & \vdots & & \vdots & 0 & 1 & \dots & 0 \\
x_{2,T,1} & x_{2,T,2} & \dots & x_{2,T,K} & 0 & 1 & \dots & 0 \\ \hline
\vdots & \vdots & & \vdots & \vdots & & & \vdots \\ \hline
x_{N,1,1} & x_{N,1,2} & \dots & x_{N,1,K} & 0 & 0 & \dots & 1 \\
\vdots & \vdots & & \vdots & 0 & 0 & \dots & 1 \\
x_{N,T,1} & x_{N,T,2} & \dots & x_{N,T,K} & 0 & 0 & \dots & 1
\end{array}\right)
$$

Daraus kann obiges gezeigt werden.

4. 1. Beispieldatensatz

$$Z'Y = \begin{bmatrix} \sum_{t=1}^{T} y_{1,t} \\ \sum_{t=1}^{T} y_{2,t} \\ \vdots \\ \sum_{t=1}^{T} y_{N,t} \end{bmatrix}, Z'Z = \begin{bmatrix} T & 0 & \dots & 0 \\ 0 & T & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & T \end{bmatrix} \Rightarrow (Z'Z)^{-1} Z'Y = \begin{bmatrix} \overline{y_{1,:}} \\ \overline{y_{2,:}} \\ \vdots \\ \overline{y_{N,:}} \end{bmatrix}$$

Damit erhält man als Residuen einer Regression auf die Dummies genau die transformierten Größen $\dot{y}_{i,t}$ und $\dot{X}_{i,t}$.

Der Within-Schätzer heisst daher auch *Least-Squares Dummy Variable (LSDV) Schätzer*.

Paneldaten: Fixed Effects

Berechnung der Freiheitsgrade bei den Tests:

Freiheitsgrade = Stichprobengröße - Zahl der Regressoren (inklusive Konstante)

Hier:

- Stichprobengröße: NT
- Zahl der Regressoren: $K + N$ (X und die Dummies für $\mathbf{a}_i$)
- Freiheitsgrade: $NT - K - N = N(T - 1) - K$

Paneldaten: Fixed Effects

Within-Schätzer oder erste Differenzenschätzer?

- Die individuellen Effekte würden auch beim erste Differenzenschätzer wegfallen.
- Annahmen FD.1-FD.4 sind ident mit FE.1-FE.4.
- FD.5-FD.6 und FE.5-FE.6 sind verschieden, aber die FE Annahmen sind oft einfacher zu motivieren.
- Within-Schätzer als KQ-Schätzer ist der BLUS im ursprünglichen Modell.
- Der FD-Schätzer ist der BLUS im Modell für die ersten Differenzen (wenn diese unabhängig über die Zeit sind) aber nicht im ursprünglichen.

Paneldaten: gepoolter Querschnitt, Produkt Budweiser Light

```
Pooling Model

Call:
plm(formula = log(VOL) ~ PR_int + PR_BASE, data = pdat, subset = (VOL >
0), model = "pooling")

Unbalanced Panel: n = 27, T = 127-220, N = 5441

Residuals:
Min.  1st Qu.   Median  3rd Qu.     Max.
-3.02342 -0.55018  0.11157  0.61044  2.11688

Coefficients:
Estimate Std. Error t-value  Pr(>|t|)
(Intercept)  0.907751   0.308755   2.940  0.003295 **
PR_int      -0.443745   0.016763 -26.471 < 2.2e-16 ***
PR_BASE      0.492961   0.029503  16.709 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    4161.1
Residual Sum of Squares: 3676.8
R-Squared:      0.11638
Adj. R-Squared: 0.11605
```

Paneldaten: Fixed Effects

```
Oneway (individual) effect Within Model

Call:
plm(formula = log(VOL) ~ PR_int + PR_BASE, data = pdat, subset = (VOL >
0), effect = "individual", model = "within")

Unbalanced Panel: n = 27, T = 127-220, N = 5441

Residuals:
Min.    1st Qu.    Median   3rd Qu.      Max.
-3.633338 -0.325439  0.058116  0.392258  1.810246

Coefficients:
          Estimate Std. Error t-value  Pr(>|t|)
PR_int  -0.441860   0.011817 -37.393 < 2.2e-16 ***
PR_BASE  0.506770   0.020751  24.422 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    2285.7
Residual Sum of Squares: 1805.3
R-Squared:      0.21016
Adj. R-Squared: 0.20607
F-statistic: 720.001 on 2 and 5412 DF, p-value: < 2.22e-16
```