

データ統合について (第3回)

株式会社ライトストーン

2016年1月12日

前回の一番最後で、移植先ワークファイルのみが時点情報を含んでいるケースを取扱いました。今回は前二回の内容の総仕上げとして、パネルデータからパネルデータに変数を移植する場合の操作例をご紹介します。

1 パネルデータからパネルデータへ元データを移植したい場合

1.1 概要

たとえば、市町村パネルデータで推定を行う場合に、都道府県レベルの変数の影響を考慮したい場合があります。そのような場合、時点を通じて一定の値を取る変数については固定効果推定を行うだけで良いわけですが、時点によって変動する観測値（たとえば都道府県人口）の影響は、固定効果推定を行うだけではコントロールできません。この時、次のような移動を行いたいというニーズが生じます。

Year	Prefecture	City	Z		Year	Prefecture	Z
2010	Tokyo	A	NA	←	2010	Tokyo	200
2010	Saitama	B	NA	←	2011	Tokyo	220
2010	Tokyo	C	NA	←	2010	Saitama	210
2010	Tokyo	D	NA	←	2011	Saitama	200
2010	Saitama	E	NA				
2011	Tokyo	A	NA	←			
2011	Saitama	B	NA	←			
2011	Tokyo	C	NA	←			
2011	Tokyo	D	NA	←			
2011	Saitama	E	NA	←			

図は、左側のパネルデータに、右側のパネルデータから Z を移動したい場合を表しています。分かりにくくなるため描画していませんが、右側の「2011 Tokyo」と「2010 Saitama」からも同様の矢印が左に伸びるということは容易にご理解いただけるかと思えます。すなわち、年度と都道府県の両方が一致している箇所に対し、データ（この例では Z）の移植を行いたいというケースです。

このような場合、たとえばエクセルの vlookup 関数等を使っても移動は可能でしょう。また、このくらいのデータ数でしたら、手で移動することもできなくはないでしょう。ただ、どちらの場合も少々手間がかかりますし、最終的に EViews で分析を行うのであれば、できるだけ EViews 上で一まとめに行えた方が、何かと便

利ではないかと考えられます。

幸いなことに、EViews でこのような移動を行いたい場合、前回までと同様に操作し、表示されるダイアログで prefecture と year の両方を、半角スペースで区切って入力すればよいだけです。前回、前々回と読まれてきた方はこれだけの説明でもすぐにでもお試しいただけるのではないかと思います。しかし念のため、パネルの場合に変わったポイントも含め、以下でもう一度おさらいしたいと思います。

1.2 操作説明

前々回と同様、都道府県ワークファイルから市町村ワークファイルにデータを移植したい場合を例に考えてみます。ここでの違いは、都道府県ワークファイル・市町村ワークファイルのいずれも、パネル構造を持っているということです。具体的には、今回利用する prefecdatapanel は東京都と埼玉県の、それぞれ 2010 年と 2011 年の観測値からなる都道府県パネルデータです。また、citydatapanel は東京と埼玉に属する 5 つの都市の、それぞれ 2010 年と 2011 年の観測値からなる市町村パネルデータです。まずは両方を一つのワークファイル内で開き、データをざっとご確認ください。パネルデータの場合の特別な点は、prefecdatapanel 側の prefecture と Z をグループ表示してみると分かります

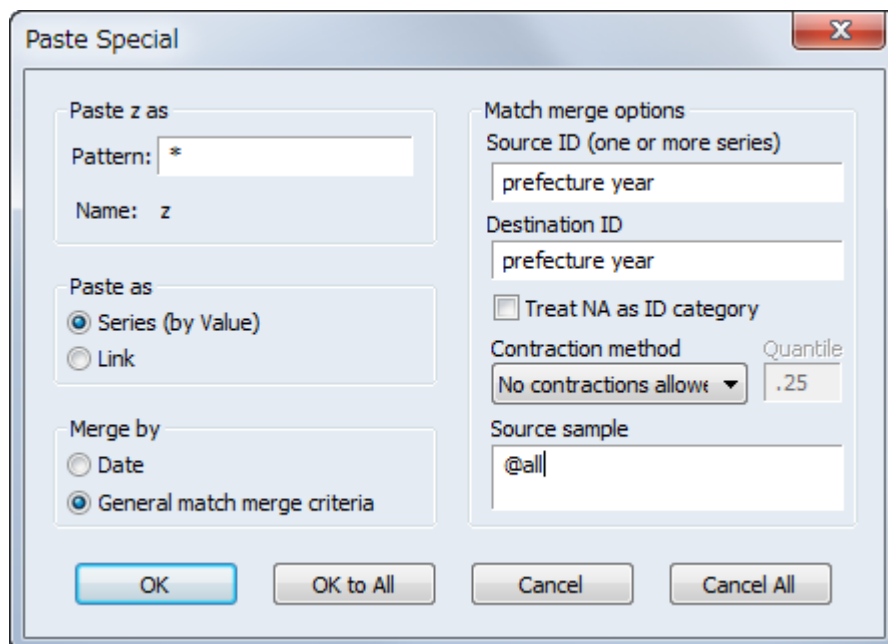
```
group temp prefecture z
show z
```

Tokyo - 10	Tokyo	200
Tokyo - 11	Tokyo	220
Saitama - 10	Saitama	210
Saitama - 11	Saitama	200

パネルであるため、Tokyo に対し 200 と 220 という二つの値が入力されています。Saitama に対しても、210 と 200 という二つの値が入力されています。言うまでもないことですが、これらはそれぞれ 2010 年と 2011 年のデータに対応します。つまり、prefecture だけをキーとして使う場合、元データを移植する場合に必要な、One to Many の構造になっていないということです*1。

ただし、先に概要で示したように、パネルデータからパネルデータに元データを移動したい場合、それぞれの時点別にデータを移動したいことが多いでしょう。prefecture だけでなく year も ID として使えば、このようなニーズに対応可能です。具体的には、prefecdatapanel の側の Z をドラッグアンドドロップか、Copy し、citydatapanel 側に張り付けます。ダイアログが自動で表示されますので、次のように入力してください。

*1 言い換えれば、prefecture の情報だけでは、どの数字を用いていいのかが分からないということです。



prefecture と year の間には半角スペースが入ります。このように入力して複数の ID で条件づけると、すべての ID 情報を使ってマッチさせて、データの移植が行われます。実際に citydata 側にコピーされた Z を開いて、この点を確認してください。

	PREFECTURE	YEAR	Z
A - 10	Tokyo	2010	200
A - 11	Tokyo	2011	220
B - 10	Saitama	2010	210
B - 11	Saitama	2011	200
C - 10	Tokyo	2010	200
C - 11	Tokyo	2011	220
D - 10	Tokyo	2010	200
D - 11	Tokyo	2011	220
E - 10	Saitama	2010	210
E - 11	Saitama	2011	200

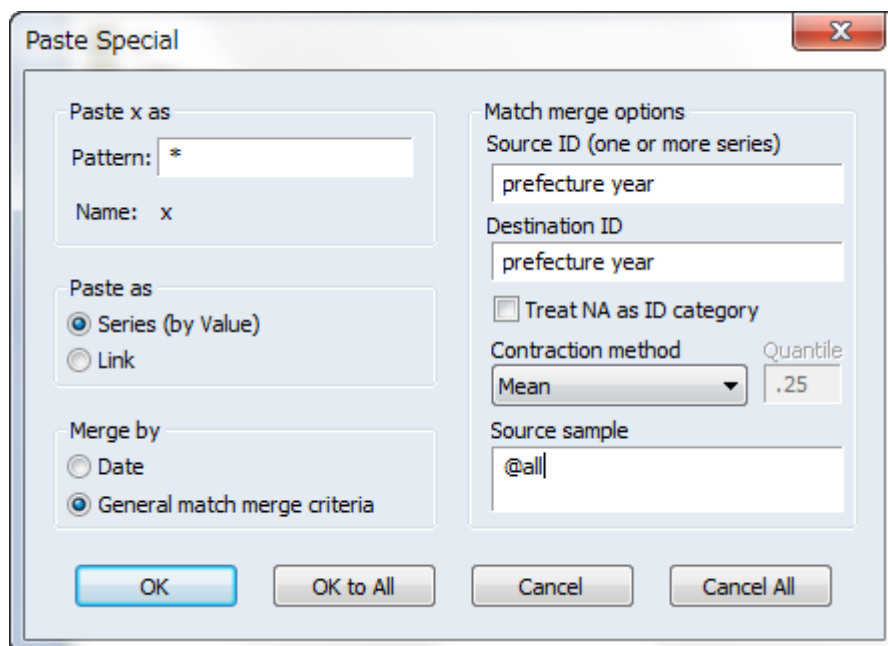
データは意図した通りに移植されています。

2 パネルデータからパネルデータへ Contraction を行いながらデータを移植したい場合

また、前回のケースのように、市町村ワークファイルから都道府県ワークファイルにデータを移動することも可能です。ただしその場合は、前回同様に Contraction が必要になります。具体的な方法については前回の内容をご確認いただき、上記のように ID を複数設定してください。パネルデータのように複数の次元がある場

合、どの次元を使って Contraction が行われるのかは必ずしも明瞭ではありません。そこで、以下では代表的なケース別に、具体的にどのような Contraction が行われるかについて分かりやすく図示いたします。

2.1 prefecture と year の両方を ID として使うケース



と入力したとします。この場合、prefecture と year の両方が一致している市町村において、グループ平均を計算します。同じ都道府県でも年度が違ったり、年度が同じでも異なる都道府県は別のグループになります。具体的には以下の色別グループ内で計算が行われます。試しに右側の表の同じ色の箇所の平均を取り、左の同じ色の箇所と比較してみてください。

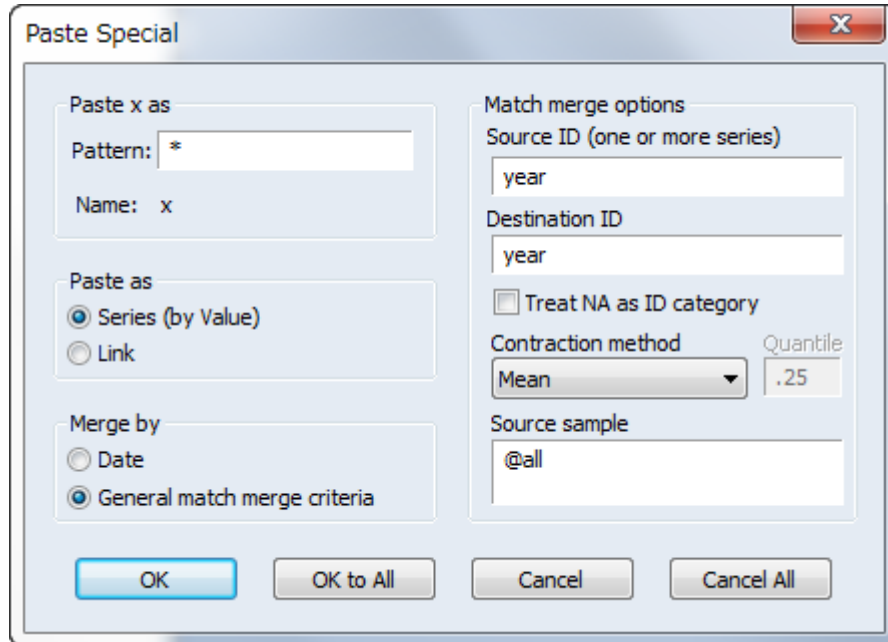
PREFECTURE	YEAR	X
Tokyo	2010	90
Tokyo	2011	110
Saitama	2010	105
Saitama	2011	125

移植先ワークファイル

PREFECTURE	YEAR	X
Tokyo	2010	100
Tokyo	2011	120
Saitama	2010	90
Saitama	2011	110
Tokyo	2010	80
Tokyo	2011	100
Tokyo	2010	90
Tokyo	2011	110
Saitama	2010	120
Saitama	2011	140

ソースワークファイル

2.2 year のみを ID として使うケース



PREFECTURE	YEAR	X
Tokyo	2010	96
Tokyo	2011	116
Saitama	2010	96
Saitama	2011	116

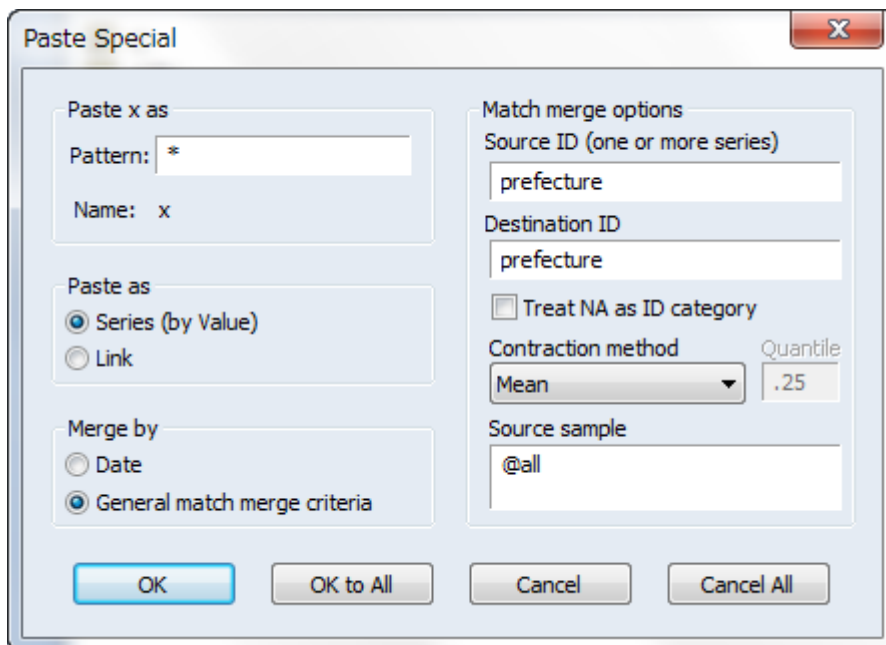
移植先ワークファイル

PREFECTURE	YEAR	X
Tokyo	2010	100
Tokyo	2011	120
Saitama	2010	90
Saitama	2011	110
Tokyo	2010	80
Tokyo	2011	100
Tokyo	2010	90
Tokyo	2011	110
Saitama	2010	120
Saitama	2011	140

ソースワークファイル

都道府県に関わらず、同じ年のすべてのデータで平均を計算し、移植を行います。

2.3 prefecture のみを ID として使うケース



PREFECTURE	YEAR	X
Tokyo	2010	100
Tokyo	2011	100
Saitama	2010	115
Saitama	2011	115

移植先ワークファイル

PREFECTURE	YEAR	X
Tokyo	2010	100
Tokyo	2011	120
Saitama	2010	90
Saitama	2011	110
Tokyo	2010	80
Tokyo	2011	100
Tokyo	2010	90
Tokyo	2011	110
Saitama	2010	120
Saitama	2011	140

ソースワークファイル

年に関係なく、各都道府県内で平均を計算し、移植を行います。

今回は複数の ID 情報を持つデータの代表例である、パネルデータ間のデータの移動について解説しました。しかしながら、別にパネルデータでなくとも、複数 ID を使った移動は可能です。意図した通りに移植が行われるように気を付けながら、必要に応じて ID を組み合わせ、データの統合に挑戦してみてください。

以上で、3 回に渡って説明してきたデータマッチングの解説は終了です。次回は EViews9 で追加されたコマンドキャプチャ機能を紹介します。