Discussion of "A Generalized Value-Added Model with Conditional Random Effects and Multivariate Shrinkage"

Christopher Taber

Department of Economics
University of Wisconsin-Madison

April 23, 2008

Outline

Outline

1. My Interpretation of the Approach
2. One comment

Basic Model

Here is my basic impression of the model

We start with the 3 period value added model defined at the school level k

$$Y_{1i} - Y_{0i} = S'_{1i}\alpha_{1k} + \gamma_i + \eta_{1i}$$
$$Y_{2i} - Y_{1i} = S'_{2i}\alpha_{2k} + \gamma_i + \eta_{2i}$$

Our main goal is estimation of α_{1k} and α_{2k}.

Problem is that we are worried that S_{1i} and S_{2i} are correlated with γ_i

Basic Model

Here is my basic impression of the model

We start with the 3 period value added model defined at the school level k

$$Y_{1i} - Y_{0i} = S'_{1i}\alpha_{1k} + \gamma_i + \eta_{1i}$$
$$Y_{2i} - Y_{1i} = S'_{2i}\alpha_{2k} + \gamma_i + \eta_{2i}$$

Our main goal is estimation of α_{1k} and α_{2k}.

Problem is that we are worried that S_{1i} and S_{2i} are correlated with γ_i

Basic Model

Here is my basic impression of the model

We start with the 3 period value added model defined at the school level k

$$Y_{1i} - Y_{0i} = S'_{1i}\alpha_{1k} + \gamma_i + \eta_{1i}$$
$$Y_{2i} - Y_{1i} = S'_{2i}\alpha_{2k} + \gamma_i + \eta_{2i}$$

Our main goal is estimation of α_{1k} and α_{2k}.

Problem is that we are worried that S_{1i} and S_{2i} are correlated with γ_i

Conditional Random Effect (CRE)

The idea of the conditional random effect is to allow for this type of correlation.

In particular assume that

$$\gamma_i = \gamma_{0k} + S'_{1i} g_{1k} + S'_{2i} g_{2k} + u_i$$

Note that this is better than random effects because γ_i can be correlated with the other stuff

It is generally not as good as fixed effects because we have restricted the form of the relationship

Conditional Random Effect (CRE)

The idea of the conditional random effect is to allow for this type of correlation.

In particular assume that

$$\gamma_i = \gamma_{0k} + S_{1i}' g_{1k} + S_{2i}' g_{2k} + u_i$$

Note that this is better than random effects because γ_i can be correlated with the other stuff

It is generally not as good as fixed effects because we have restricted the form of the relationship

Conditional Random Effect (CRE)

The idea of the conditional random effect is to allow for this type of correlation.

In particular assume that

$$\gamma_i = \gamma_{0k} + S_{1i}' g_{1k} + S_{2i}' g_{2k} + u_i$$

Note that this is better than random effects because γ_i can be correlated with the other stuff

It is generally not as good as fixed effects because we have restricted the form of the relationship

Under this restriction we can rewrite the model as

$$Y_{1i} - Y_{0i} = \gamma_{0k} + S'_{1i}\alpha_{1k} + S'_{1i}g_{1k} + S'_{2i}g_{2k} + u_i + \eta_{1i}$$
$$Y_{2i} - Y_{1i} = \gamma_{0k} + S'_{2i}\alpha_{2k} + S'_{1i}g_{1k} + S'_{2i}g_{2k} + u_i + \eta_{2i}$$

or

$$Y_{1i} - Y_{0i} = \gamma_{0k} + S'_{1i}\left(\alpha_{1k} + g_{1k}\right) + S'_{2i}g_{2k} + u_i + \eta_{1i}$$
$$Y_{2i} - Y_{1i} = \gamma_{0k} + S'_{2i}\left(\alpha_{2k} + g_{2k}\right) + S'_{1i}g_{1k} + u_i + \eta_{2i}$$

Pretty clearly you can identify $(\alpha_{1k}, \alpha_{2k}, g_{1k}, g_{2k})$ from this model.

Only real problem is g_{1k}-you need to take the model pretty seriously

Under this restriction we can rewrite the model as

$$Y_{1i} - Y_{0i} = \gamma_{0k} + S'_{1i}\alpha_{1k} + S'_{1i}g_{1k} + S'_{2i}g_{2k} + u_i + \eta_{1i}$$
$$Y_{2i} - Y_{1i} = \gamma_{0k} + S'_{2i}\alpha_{2k} + S'_{1i}g_{1k} + S'_{2i}g_{2k} + u_i + \eta_{2i}$$

or

$$Y_{1i} - Y_{0i} = \gamma_{0k} + S'_{1i}\left(\alpha_{1k} + g_{1k}\right) + S'_{2i}g_{2k} + u_i + \eta_{1i}$$
$$Y_{2i} - Y_{1i} = \gamma_{0k} + S'_{2i}\left(\alpha_{2k} + g_{2k}\right) + S'_{1i}g_{1k} + u_i + \eta_{2i}$$

Pretty clearly you can identify $(\alpha_{1k}, \alpha_{2k}, g_{1k}, g_{2k})$ from this model.

Only real problem is g_{1k}-you need to take the model pretty seriously

Under this restriction we can rewrite the model as

$$Y_{1i} - Y_{0i} = \gamma_{0k} + S'_{1i}\alpha_{1k} + S'_{1i}g_{1k} + S'_{2i}g_{2k} + u_i + \eta_{1i}$$
$$Y_{2i} - Y_{1i} = \gamma_{0k} + S'_{2i}\alpha_{2k} + S'_{1i}g_{1k} + S'_{2i}g_{2k} + u_i + \eta_{2i}$$

or

$$Y_{1i} - Y_{0i} = \gamma_{0k} + S'_{1i}\left(\alpha_{1k} + g_{1k}\right) + S'_{2i}g_{2k} + u_i + \eta_{1i}$$
$$Y_{2i} - Y_{1i} = \gamma_{0k} + S'_{2i}\left(\alpha_{2k} + g_{2k}\right) + S'_{1i}g_{1k} + u_i + \eta_{2i}$$

Pretty clearly you can identify $(\alpha_{1k}, \alpha_{2k}, g_{1k}, g_{2k})$ from this model.

Only real problem is g_{1k}-you need to take the model pretty seriously

Under this restriction we can rewrite the model as

$$Y_{1i} - Y_{0i} = \gamma_{0k} + S'_{1i}\alpha_{1k} + S'_{1i}g_{1k} + S'_{2i}g_{2k} + u_i + \eta_{1i}$$
$$Y_{2i} - Y_{1i} = \gamma_{0k} + S'_{2i}\alpha_{2k} + S'_{1i}g_{1k} + S'_{2i}g_{2k} + u_i + \eta_{2i}$$

or

$$Y_{1i} - Y_{0i} = \gamma_{0k} + S'_{1i}\left(\alpha_{1k} + g_{1k}\right) + S'_{2i}g_{2k} + u_i + \eta_{1i}$$
$$Y_{2i} - Y_{1i} = \gamma_{0k} + S'_{2i}\left(\alpha_{2k} + g_{2k}\right) + S'_{1i}g_{1k} + u_i + \eta_{2i}$$

Pretty clearly you can identify $(\alpha_{1k}, \alpha_{2k}, g_{1k}, g_{2k})$ from this model.

Only real problem is g_{1k}-you need to take the model pretty seriously

Shrinkage

Now what do we do with this?

We have a consistent estimate so we could be done, but Rob wants to do more.

If k were picked at random and I wanted to get a good estimate of α_{1k} what would I do?

1. Just use the data from school k and use the estimate $\hat{\alpha}_k$
2. Use the data from all schools taking the average of the $\hat{\alpha}_k$ across all schools

There are advantages to each, thinking about the second relative to the first

Shrinkage

Now what do we do with this?

We have a consistent estimate so we could be done, but Rob wants to do more.

If k were picked at random and I wanted to get a good estimate of α_{1k} what would I do?

1. Just use the data from school k and use the estimate $\hat{\alpha}_k$
2. Use the data from all schools taking the average of the $\hat{\alpha}_k$ across all schools

There are advantages to each, thinking about the second relative to the first

Shrinkage

Now what do we do with this?

We have a consistent estimate so we could be done, but Rob wants to do more.

If k were picked at random and I wanted to get a good estimate of α_{1k} what would I do?

1. Just use the data from school k and use the estimate $\hat{\alpha}_k$
2. Use the data from all schools taking the average of the $\hat{\alpha}_k$ across all schools

There are advantages to each, thinking about the second relative to the first

Shrinkage

Now what do we do with this?

We have a consistent estimate so we could be done, but Rob wants to do more.

If k were picked at random and I wanted to get a good estimate of α_{1k} what would I do?

1. Just use the data from school k and use the estimate $\hat{\alpha}_k$
2. Use the data from all schools taking the average of the $\hat{\alpha}_k$ across all schools

There are advantages to each, thinking about the second relative to the first

Shrinkage

Now what do we do with this?

We have a consistent estimate so we could be done, but Rob wants to do more.

If k were picked at random and I wanted to get a good estimate of α_{1k} what would I do?

1. Just use the data from school k and use the estimate $\hat{\alpha}_k$
2. Use the data from all schools taking the average of the $\hat{\alpha}_k$ across all schools

There are advantages to each, thinking about the second relative to the first

Shrinkage

Now what do we do with this?

We have a consistent estimate so we could be done, but Rob wants to do more.

If k were picked at random and I wanted to get a good estimate of α_{1k} what would I do?

1. Just use the data from school k and use the estimate $\hat{\alpha}_k$
2. Use the data from all schools taking the average of the $\hat{\alpha}_k$ across all schools

There are advantages to each, thinking about the second relative to the first

Advangtage: It will give a more precise estimate because it uses all of the data

Disadvantage: It will be biased. It is a consistent estimate of the unconditional mean, not α_k

Optimally there is a tradeoff between the two

What Shrinkage does is to weight the estimates in a way to minimize the Mean Squared Error

Advangtage: It will give a more precise estimate because it uses all of the data

Disadvantage: It will be biased. It is a consistent estimate of the unconditional mean, not α_k

Optimally there is a tradeoff between the two

What Shrinkage does is to weight the estimates in a way to minimize the Mean Squared Error

Advangtage: It will give a more precise estimate because it uses all of the data

Disadvantage: It will be biased. It is a consistent estimate of the unconditional mean, not α_k

Optimally there is a tradeoff between the two

What Shrinkage does is to weight the estimates in a way to minimize the Mean Squared Error

Advangtage: It will give a more precise estimate because it uses all of the data

Disadvantage: It will be biased. It is a consistent estimate of the unconditional mean, not α_k

Optimally there is a tradeoff between the two

What Shrinkage does is to weight the estimates in a way to minimize the Mean Squared Error

Outline

One comment

I think this is a cool idea and should be really useful in practice

My only issue is, what is special about mean squared error?

In general we are not interested in a particular value of α_k, but rather features of the distribution of $\alpha_1,...,\alpha_K$

One comment

I think this is a cool idea and should be really useful in practice

My only issue is, what is special about mean squared error?

In general we are not interested in a particular value of α_k, but rather features of the distribution of $\alpha_1,...,\alpha_K$

One comment

I think this is a cool idea and should be really useful in practice

My only issue is, what is special about mean squared error?

In general we are not interested in a particular value of α_k, but rather features of the distribution of $\alpha_1,...,\alpha_K$

Examples:

- Unconditional Mean
- Variance
- Median
- Maximum value
- Quartiles
- Full distribution

If our goal is to estimate one of these objects above it is not obvious (to me) that minimizing mean squared error of the individual coefficients is the right thing to do

I don't view this as a problem, but rather straight forward extension of the idea

Whatever our estimate is, we can choose a criterion that we want and choose the estimate to minimize that thing

Thus the general idea here is broader than just mean squared error.

I don't view this as a problem, but rather straight forward extension of the idea

Whatever our estimate is, we can choose a criterion that we want and choose the estimate to minimize that thing

Thus the general idea here is broader than just mean squared error.

I don't view this as a problem, but rather straight forward extension of the idea

Whatever our estimate is, we can choose a criterion that we want and choose the estimate to minimize that thing

Thus the general idea here is broader than just mean squared error.