A glance at the **price elasticity of demand** (~~marginal product of energy~~) from the lens of a time-varying panel data **demand** (~~production~~) function with latent 'type-heterogeneity'

David C. Broadstock

Presented at IAEE Euroopean Meeting, August, 2019–Slovenia
Slides last updated: August 28, 2019

IDEAS Innovation-driven Education and Scholarship | Faculty of Business 工商管理學院
Discover · Design · Deliver

School of Accounting & Finance 會計及金融學院

Outline of today's talk

We have a fair bit of ground to cover in 40 minutes...

The price elasticity of energy demand

- The idea has is not to challenge the theory of demand, or extend it in any way, in fact I am very simplistic in how I 'attack' this aspect of the project:

$$\ln(Q_{it}) = \mu_{it} + \beta_{Pit} ln(P_{it}) + \beta_{Yit} ln(Y_{it}) + u_{it}$$

Type-herterogeneity

- What do I mean by type heterogeneity there are two (sequential) aspects to how I approach this:
  - ***Heterogeneity*** In a first pass 'simple' estimation round I do not want to impose common coefficients e.g. to allow: $\beta_{Pi} \neq \beta_{Pj}, \forall\{i,j\}, \{i,j\} \in I, i \neq j$ (ignoring the t subscript for simplicity).
  - ***Type identification (reduction)*** in my world is about isolating panel 'synchronicity' in marginal products e.g. the idea that $\beta_{Pi} \approx \beta_{Pj}, \forall\{t\} \in T, \{i,j\} \in I, i \neq j$.

The econometric specification and its ‘challenges’

A production function with time-varying (in-)efficiency

$$\ln(Q_{it}) = \mu_{it} + \beta_{Pit} ln(P_{it}) + \beta_{Yit} ln(Y_{it}) + u_{it}, \qquad u \sim \mathbf{NID}(0, \sigma_u^2) \tag{1a}$$

$$\mu_{it} = \mu_{it-1} + e_{it}, \qquad e_{1it} \sim \mathbf{NID}(0, \sigma_{e_{1i}}^2) \tag{1b}$$

$$\beta_{Pit} = \beta_{Pit-1} + e_{2t}, \qquad e_{2it} \sim \mathbf{NID}(0, \sigma_{e_{2i}}^2) \tag{1c}$$

$$\beta_{Yit} = \beta_{Yit-1} + e_{3t}, \qquad e_{3it} \sim \mathbf{NID}(0, \sigma_{e_{3i}}^2) \tag{1d}$$

- Panel formed of OECD 17 countries, taken from Adeyemi et al. (2010) - a little dated, but a valid test case nonetheless.

Challenges...

- Can we estimate a time-varying coefficent accurately in modest panel dimensions?
- If there is an *i* dimension to address (i) can it be handled with accuracy, and (ii) can a well performing dimension reduction strategy be devised?

A preview of the main results

Since we have too much ground to cover in 15 minutes…

...this is clearly a thought in progress..., but...

- Show that a panel modified STSM performs well under 'normal' conditions.
- Clarify that OLS-FE lacks precision compared with panel STSM
- I further show that a-priori unknown, complex, 'clubbing' patterns can be uncovered ***without*** a high computational overhead, and with respectable levels of accuracy

Orientation: Defining and interpreting 'accuracy'

Coverage, significance and relative accuracy scores.

'Coverage' of the true parameter in the confidence set:

- Shows if the true parameter contained within the 95% confidence interval* of the estimated coefficient e.g. $\hat{\beta}_{LOW} < \beta < \hat{\beta}_{UP}$.

'Significance' of estimated coefficients:

- Shows if the estimated coefficients are deemed significant at the 95% level, noting that (by design) all terms are significant e.g. $sgn(\hat{\beta}_{LOW}) = sgn(\hat{\beta}_{UP})$.
- **Coverage and significance should be considered simultaneously, since accuracy in one without the other implies erroneous policy implications.**

Relative accuracy scores (RAS):

- RAS scores are based on averages of dummy variables that take the value one for the estimator when it provides the most accurate point estimate of the true coefficient, and zero if some other estimator was more accurate *i.e. it is defined relative to the other models it is competing against*.
- RAS can be defined for individual coefficients, as well as overall model fit.

The initial data generating process

A simulated panel with unobserved common time trend

Here I outline the data gnerating process that I will use in establishing the efficacy of the panel STSM model for the purpose of recovering time varying latent trends.

To ensure generality we will for now denote the two exogenous variables by x_{1it} and x_{2it} rather then p_{it} and y_{it}, similarly we will denote the left hand side variable by y_{it}, rather than q_{it}.

1. **Generate exogenous variables:** $\{x_{1it}, x_{2it}\} \sim N(0, 1)$
2. **Specify coefficient values:** $\{\beta_1, \beta_2\} = 1$
3. **Generate unobserved trend:** $\alpha_t = \phi\alpha_{t-1} + \nu_{it}$, with $\phi = 1$ and $\nu \sim N(0, 1)$
4. **Construct systematic component of observed data:** $y^*_{it} = \alpha_t + \beta_1 x_{1it} + \beta_2 x_{2it}$
5. **Generate non-systematic component of observed data:** $u_{it} \sim N(0, 1)$
6. **Construct observed data:** $y_{it} = y^*_{it} + u_{it}$

Moving to a world with varying coefficients

A simulated panel with both unobserved common trend & time varying coefficients

Now I shall move towards a more demanding data generating process in which the coefficients are varying over time

1. **Generate exogenous variables:** $\{x_{1it}, x_{2it}\} \sim N(0,1)$
2. **Specify coefficient values:** $\beta_{1t} = \phi\beta_{1t-1} + v_{1it}$; $\beta_{2t} = \phi\beta_{2t-1} + v_{2it}$ with $\{v_{1it}, v_{2it}\} \sim N(0,1)$
3. **Generate unobserved trend:** $\alpha_t = \phi\alpha_{t-1} + \nu_{it}$, with $\phi = 1$ and $\nu_{it} \sim N(0,1)$
4. **Construct systematic component of observed data:** $y^*_{it} = \alpha_t + \beta_{1t}x_{1it} + \beta_{2t}x_{2it}$
5. **Generate non-systematic component of observed data:** $u_{it} \sim N(0,1)$
6. **Construct observed data:** $y_{it} = y^*_{it} + u_{it}$

The above steps will be repeated for combinations in $\{N, T\} = \{5, 10, 15, 20, 25, 30\}$, and for $M = 1000$ monte-carlo replications.

Two approaches to estimation

Here the OLS-FE model becomes visibly limited in its potential

Traditional fixed effects estimation:

$$y_{it} = \alpha_t + \beta_1 x_{1it} + \beta_2 x_{2it} + u_{it} \tag{2}$$

We could in theory interact the x variables with time trends here also, though this may become cumbersome quite quickly

Panel models in state space form:

$$y_{it} = \alpha_t + \beta_{1t} x_{1it} + \beta_{2t} x_{2it} + u_{it} \tag{3a}$$

$$\alpha_t = \phi \alpha_{t-1} + v_{it} \tag{3b}$$

$$\beta_{1t} = \phi \beta_{1t-1} + v_{1it} \tag{3c}$$

$$\beta_{2t} = \phi \beta_{2t-1} + v_{2it} \tag{3d}$$

That is one case only - let's multiply!!

By now we have some sense that in one type of scenario (data generating process or d.g.p.), and with one set of random data, it is not inconceivable that the TVP model might be 'at least no worse' than FE models. I do not give the nonparametric model further consideration in this study.

We are now going to do a more thorough and fairer comparison with multiple replications and random draws on the data.

- The d.g.p. contiunes to reflect the world we have explored so far, in which key model parameters are in fact constant over time (favoring the FE model), but in which there is a time-varying intercept:

$$Y^*_{it} = \alpha_t + \beta_1 X_{1it} + \beta_2 X_{2it}; \quad Y_{it} = Y^*_{it} + u_{it}; \quad u_{it} \sim N(0,1)$$

$$\beta_1 = 1; \quad \beta_2 = 1; \quad X_{1it} \sim N(0,1); \quad X_{2it} \sim N(0,1)$$

$$\alpha_t = \alpha_{t-1} + e_{0t}; \quad M = 1000$$

$$N = 5, 10, 15, 20, 25, 30; \quad T = 5, 10, 15, 20, 25, 30$$

Simulation results: $Y_{it}^{*} = \alpha_t + \beta_1 X_{1it} + \beta_2 X_{2it}$

TVP model is generally no worse in inference than OLS-FE, irrespective of sample sizes, and much butter in terms of relative accuracy in all cases.

FE Coverage and significance: α_t

	Length of time series T					
N	5	10	15	20	25	30
5	0.93	0.91	0.88	0.85	0.72	0.77
10	0.87	0.94	0.92	0.91	0.88	0.86
15	0.71	0.85	0.90	0.92	0.92	0.90
20	0.76	0.88	0.92	0.94	0.94	0.93
25	0.80	0.90	0.94	0.95	0.95	0.94
30	0.83	0.93	0.95	0.96	0.89	0.96

TVP Coverage and significance: α_t

	Length of time series T					
N	5	10	15	20	25	30
5	0.55	0.57	0.58	0.60	0.60	0.62
10	0.79	0.85	0.86	0.87	0.86	0.87
15	0.89	0.91	0.90	0.90	0.91	0.91
20	0.92	0.93	0.93	0.93	0.92	0.92
25	0.93	0.94	0.94	0.94	0.93	0.94
30	0.94	0.95	0.94	0.94	0.94	0.94

FE Relative accuracy score: α_t

	Length of time series T					
N	5	10	15	20	25	30
5	0.32	0.30	0.28	0.25	0.25	0.23
10	0.36	0.36	0.33	0.32	0.30	0.30
15	0.33	0.33	0.34	0.32	0.32	0.31
20	0.34	0.35	0.34	0.34	0.33	0.32
25	0.33	0.35	0.34	0.34	0.34	0.33
30	0.33	0.34	0.35	0.34	0.33	0.33

TVP Relative accuracy score: α_t

	Length of time series T					
N	5	10	15	20	25	30
5	0.68	0.70	0.72	0.75	0.75	0.77
10	0.64	0.64	0.67	0.68	0.70	0.70
15	0.67	0.67	0.66	0.68	0.68	0.69
20	0.66	0.65	0.66	0.66	0.67	0.68
25	0.67	0.65	0.66	0.66	0.66	0.67
30	0.67	0.66	0.65	0.66	0.67	0.67

Now lets complicate things...

Let us introduce more complex d.g.p. with full TVP's

With some reliable comparisons that we can trust the TVP framework at least as much as we can trust the 'usual' panel techniques we apply, we now turn attention towards some more interesting cases that can only be considered using a TVP approach.

- The d.g.p. reflects a more complilcated world in which we have key model parameters that are themselves varying over time (thereby favoring the FE model), and in which we continue to include a time-varying intercept:

$$Y_{it}^{*} = \alpha_t + \beta_{1t} X_{1it} + \beta_{2t} X_{2it}; \quad Y_{it} = Y_{it}^{*} + u_{it}; \quad u_{it} \sim N(0,1)$$

$$\beta_{1t} = \beta_{1t-1} + e_{1t}; \quad \beta_{2t} = \beta_{2t-1} + e_{2t}; \quad X_{1it} \sim N(0,1); \quad X_{2it} \sim N(0,1)$$

$$\alpha_t = \alpha_{t-1} + e_{0t}; \quad M = 1000$$

$$N = 5, 10, 15, 20, 25, 30; \quad T = 5, 10, 15, 20, 25, 30$$

In the next slide we will not show relative accuracy scores, but those results should be implicit.

Simulation results: $Y_{it}^{*} = \alpha_t + \beta_{1t}X_{1it} + \beta_{2t}X_{2it}$

By this stage we should be expecting the FE approach to falter. What is remarkable however, is the striking performance of TVP models in modest panel dimensions. With $N = 15$ and $T = 15$ we observe 75% 'accuracy' in our estimates for all parts of the model. By the time we hit still modest dimensions of $N = 30$ and $T = 30$ we are pushing close to 90% accuracy!!

FE Coverage and significance: α_t

	Length of time series T					
N	5	10	15	20	25	30
5	0.94	0.91	0.88	0.85	0.80	0.77
10	0.87	0.94	0.92	0.91	0.88	0.86
15	0.71	0.85	0.90	0.92	0.92	0.90
20	0.75	0.88	0.92	0.94	0.94	0.92
25	0.80	0.90	0.94	0.95	0.95	0.94
30	0.83	0.92	0.95	0.95	0.96	0.96

TVP Coverage and significance: α_t

	Length of time series T					
N	5	10	15	20	25	30
5	0.55	0.57	0.58	0.59	0.60	0.62
10	0.79	0.85	0.86	0.87	0.87	0.87
15	0.89	0.91	0.90	0.90	0.91	0.91
20	0.92	0.93	0.93	0.93	0.92	0.92
25	0.93	0.94	0.94	0.94	0.93	0.94
30	0.94	0.95	0.95	0.94	0.94	0.94

FE Coverage and significance: β_{2t}

	Length of time series T					
N	5	10	15	20	25	30
5	0.89	0.94	0.94	0.94	0.94	0.93
10	0.93	0.91	0.89	0.86	0.85	0.83
15	0.89	0.84	0.78	0.72	0.67	0.63
20	0.81	0.67	0.57	0.51	0.47	0.43
25	0.75	0.57	0.48	0.42	0.37	0.34
30	0.73	0.57	0.48	0.42	0.38	0.35

TVP Coverage and significance: β_{2t}

	Length of time series T					
N	5	10	15	20	25	30
5	0.57	0.63	0.65	0.66	0.66	0.68
10	0.81	0.84	0.83	0.82	0.80	0.81
15	0.87	0.83	0.80	0.78	0.76	0.76
20	0.86	0.81	0.83	0.84	0.86	0.87
25	0.85	0.83	0.86	0.87	0.89	0.90
30	0.81	0.79	0.81	0.84	0.85	0.86

The simulation is all well and good but...

What about the differences between panel members

In practice there will be considerable differences between certain panel members, enough to require handling during the estimation stage.

- The simulation exercises presented above were somewhat dismissive of this important heterogeneity.
- This was a simplifying assumption that facilitated meaningful and objective evaluation of the estimators, yet this might reasonably be considered a strong restriction nonetheless.
- The maintained assumption in much empirical research, is that coefficients are common to all panel members. This assumption can be relaxed, and a yet more general panel representation might be given by:

$$y_{it} = \alpha_t^{(\kappa_j)} + \beta_{1t}^{(\kappa_k)} x_{1it} + \beta_{2t}^{(\kappa_m)} x_{2it} + \eta_{it} \tag{4}$$

- κ_j, κ_k and κ_m are identifier functions used to denote membership/clustering of coefficients into clubs with $j \in J$, $k \in K$ and $m \in M$, and $\{J, K, M\} \leq N$.
- It is possible to identify club membership using a relatively simple detection mechanism

A D.G.P. for a panel with 'synchronous types'

Common time-varying coefficients for subsets of panel members

Now we wish to allocate the N panel members into coefficient clubs. We do this by first creating K_1 and K_2 which are (in practice unobserved) club membership indicators:

$$K_1 = \begin{cases} 1 & \forall\, N < 0.5N \\ 0 & \text{Otherwise} \end{cases}$$

$$K_2 = \begin{cases} 1 & \forall\, N \in 0.25N, \ldots, 0.75N \\ 0 & \text{Otherwise} \end{cases}$$

We can then simulate a dataset with clubbed coefficients through the following relationship:

$$Y_{it}^* = K_1\alpha_{1t} + (1 - K_1)\alpha_{2t} + K_2\beta_{1t}X_{1it} + (1 - K_2)\beta_{2t}X_{1it}; \quad Y_{it} = Y_{it}^* + u_{it}; \quad u_{it} \sim N(0, 1)$$

With all other assumptions remaining similar to the previous D.G.P.'s presented. In this manner, we arrive at four latent club assignments to be 'detected' during the estimation process.

Objectives of the second stage of the study:

- To correctly assign each of the N panel members into their true coefficient club(s)
- To obtain precise estimates of the true coefficients

MC results: $Y_{it}^* = \hat{K}_1\alpha_{1t} + (1-\hat{K}_1)\alpha_{2t} + \hat{K}_2\beta_{1t}X_{1it} + (1-\hat{K}_2)\beta_{2t}X_{1it}$

Monte-Carlo simulation results with $M = 200^*$

TVP Coverage and significance: α_{1t}

	Length of time series T				
N	5	10	15	20	25
20	0.422	0.756	0.831	0.867	0.853
40	0.548	0.829	0.854	0.854	0.840
60	0.588	0.835	0.829	0.835	0.850
80	0.622	0.848	0.880	0.859	0.869
100	0.641	0.843	0.858	0.866	0.850

TVP Coverage and significance: α_{2t}

	Length of time series T				
N	5	10	15	20	25
20	0.430	0.768	0.826	0.854	0.846
40	0.565	0.820	0.873	0.852	0.854
60	0.614	0.828	0.836	0.837	0.851
80	0.630	0.846	0.879	0.852	0.873
100	0.651	0.856	0.864	0.875	0.849

TVP Coverage and significance: β_{1t}

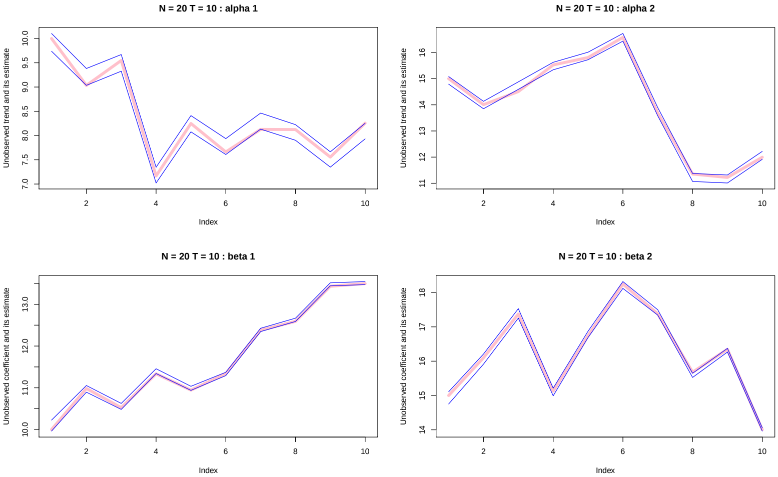
	Length of time series T				
N	5	10	15	20	25
20	0.509	0.782	0.862	0.896	0.901
40	0.560	0.851	0.877	0.912	0.926
60	0.645	0.850	0.900	0.921	0.933
80	0.653	0.852	0.918	0.926	0.937
100	0.653	0.852	0.912	0.937	0.934

TVP Coverage and significance: β_{2t}

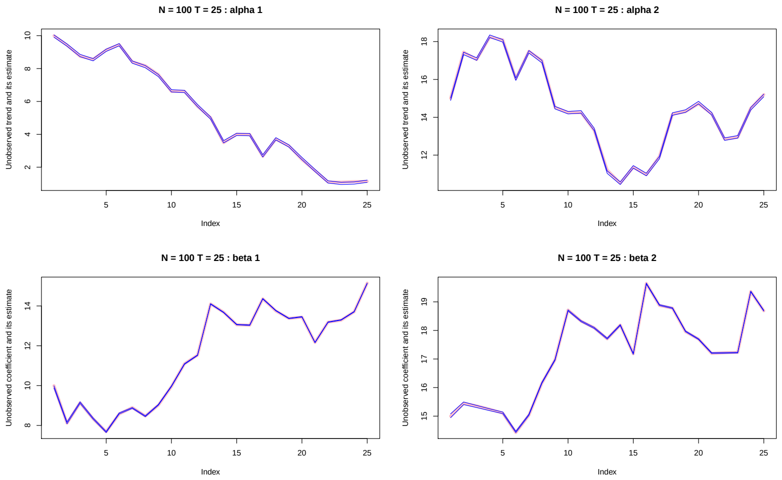
	Length of time series T				
N	5	10	15	20	25
20	0.513	0.785	0.867	0.910	0.912
40	0.540	0.826	0.901	0.892	0.922
60	0.618	0.835	0.904	0.923	0.928
80	0.621	0.850	0.913	0.918	0.939
100	0.618	0.856	0.912	0.925	0.936

* Note, in the tables above, the coverage and signficance scores for α_{1t} and α_{2t} underestimate their true values due to a minor programing typo.

Sample simulation run, $N = 20$ and $T = 10$

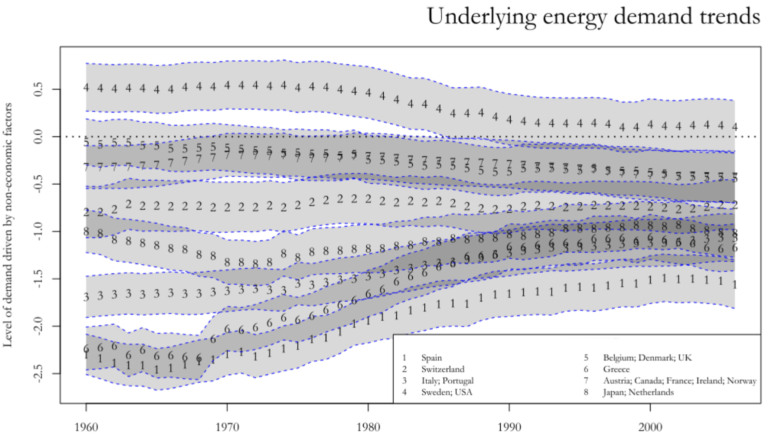


Sample simulation run, $N = 100$ and $T = 25$



How many alpha's are there? Answer: $< N$

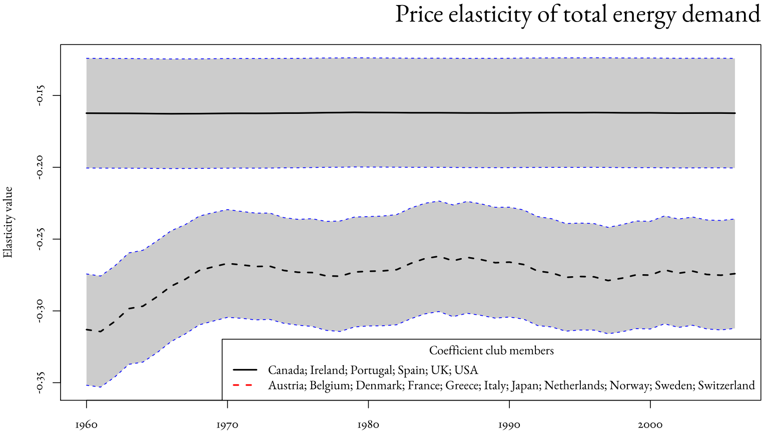
A lot of heterogeneity, and incidental pattern of slow convergence between 1970-1990



(95% Confidence intervals shown by the shaded grey regions)

How many price elasticities? Answer:>1

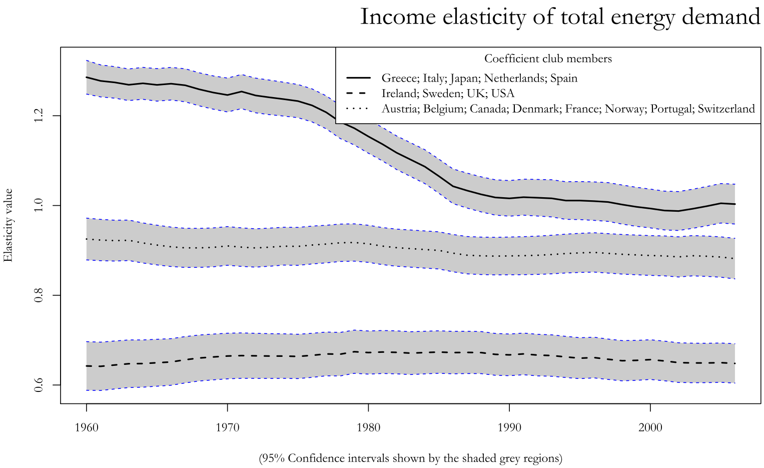
Perhaps most interesting here is that one varies over time and the other does not.



(95% Confidence intervals shown by the shaded grey regions)

How many income elasticities? Answer:>price

Price policy groups ≠ income policy groups



Closing remarks (1/2)

Lessons, limitations and next steps

Lessons learned:

- A panel modified variant of a multivariate STSM is quite effective at estimating time varying model features and at least as good as OLS-FE for 'simple' cases of unobserved trend estimation
- Standard inference techniques hold good 'power', even in small samples
- Can accurately assign panel members into latent coefficient clubs, leaving open the possibility of more intricate policy coordination related insights

Limitations of the work done so far:

- The application considered needs to more more closely aligned to the original studiy
- Still need to vary simulation parameters to alleviate some of the key concerns e.g. starting parameters for convergence etc.

Next steps for the work:

- Consider different and more challenging (though still realistic) d.g.p.'s

Closing remarks (1/2)

Lessons, limitations and next steps

... and get more firmly into the production function based applications ...

Thanks for listening!

Any questions/comments are warmly welcomed.
david.broadstock@polyu.edu.hk