Using cointegrating relationships to capture the cyclical and long-run dynamics of revenue forecasts: A simple example from Massachusetts

Presented to the 2011 FTA Revenue Estimation & Tax Research Conference

October 18, 2011

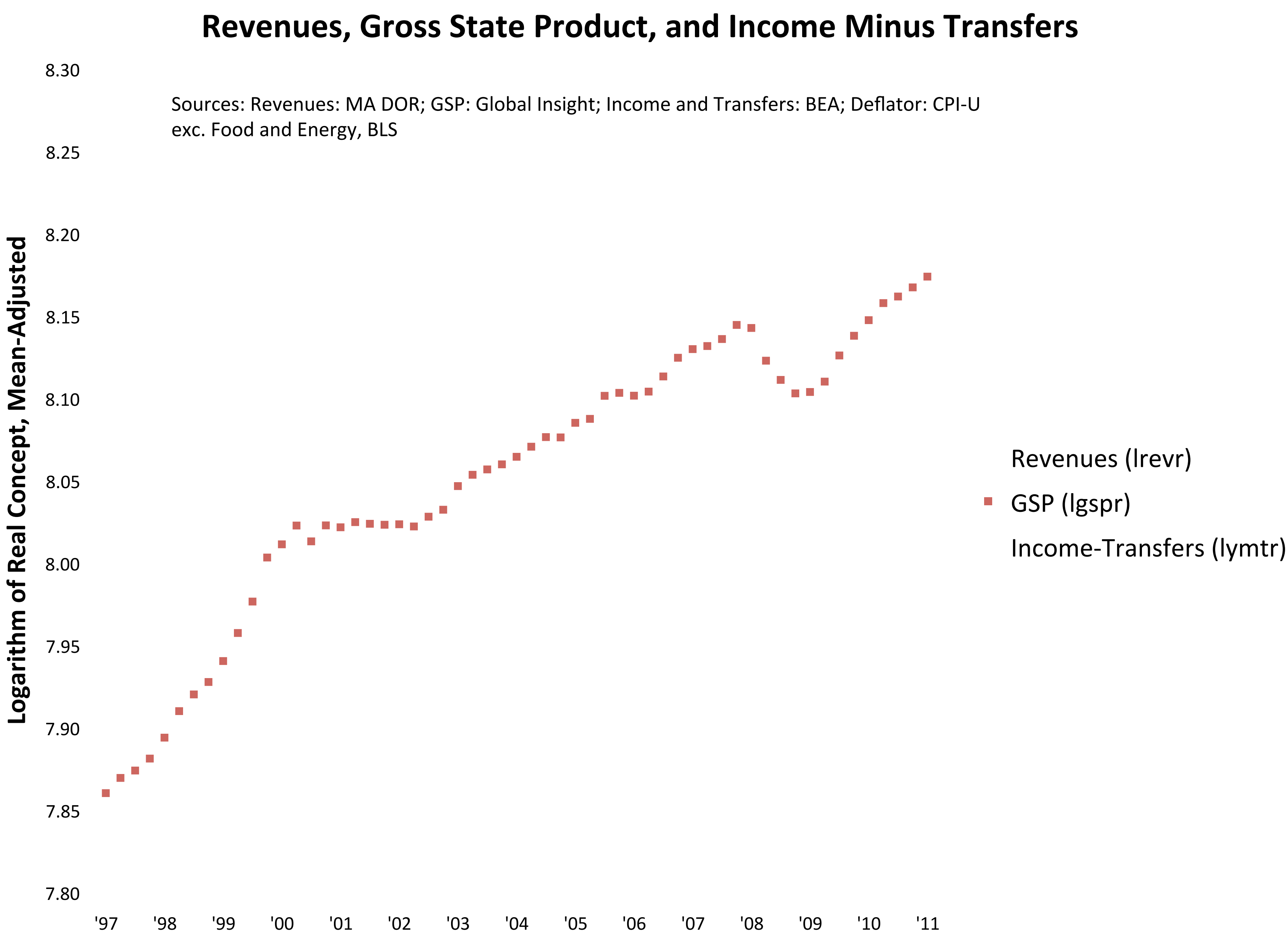
Alan Clayton-Matthews
Northeastern University

Advantages of Using Cointegrating Relationships

- Makes use of the dynamic adjustments towards a long-run equilibrium. Can capture cyclical dynamics better than a (misspecified) VAR model.
- Avoids the misspecification error of a VAR model in differences (when equilibrium relations exist).
- Can give more reliable (than VAR models in differences) estimates of long-term growth trends.
- The long-term trend can be used to estimate the cyclical deficit via a Beveridge-Nelson trend.
- It's easy to implement.
- Most state revenue systems are closely linked to output and/or incomes, so cointegrating relationships are likely to exist.

Steps in Estimating and Forecasting Cointegrated Models

- Identify the probable cointegrating relationships.
- Test for cointegration: Augmented Dickey-Fuller tests for a unit root and Engle-Granger Augmented Dickey-Fuller tests for stationary linear combinations are simple to do.
- Dynamic OLS (DOLS) estimates of the cointegrating relationships are easy to do and are better than OLS for short periods of history.
- Estimate a Vector Error-Correction (VECM) model – simply a VAR model with error correction terms.
- Use the usual time-series techniques, e.g., AIC and BIC criteria, for finding the best lag structures in the DOLS and VECM models, as well as for the ADF and EG-ADF tests.
- Carry out the forecast.
- Calculate the Beveridge-Nelson trend if it is useful for you.


Revenues, Gross State Product, and Income Minus Transfers
Sources: Revenues: MA DOR; GSP: Global Insight; Income and Transfers: BEA; Deflator: CPI-U exc. Food and Energy, BLS
Logarithm of Real Concept, Mean-Adjusted
8.30
8.25
8.20
8.15
8.10
8.05
8.00
7.95
7.90
7.85
7.80
'97
'98
'99
'00
'01
'02
'03
'04
'05
'06
'07
'08
'09
'10
'11
Revenues (lrevr)
GSP (lgspr)
Income-Transfers (lymtr)

The Cointegrating Relationships and Error-Correction Terms

- $$lrevr_t = \theta_{g0} + \theta_{g1} lgspr_t + \eta_{gt}$$
  $$lrevr_t = \theta_{y0} + \theta_{y1} lymtr_t + \eta_{yt}$$

Where: $lrevr, lgspr,$ and $lymtr$ are $\mathrm{I}(1)$, and η_{gt}, η_{yt} are stationary.

$$ec\_g_t = lrevr_t - \hat{\theta}_{g0} - \hat{\theta}_{g1} lgspr_t$$
$$ec\_y_t = lrevr_t - \hat{\theta}_{y0} - \hat{\theta}_{y1} lymtr_t$$

Dynamic OLS (DOLS) Estimates of the Cointegrating Coefficients

- $$lrevr_t = \gamma_{g0} + \theta_{g1} lgspr_t + \sum_{i=-p}^{p} \delta_{gi} \Delta lgspr_{t-i} + e_{gt}$$

$$lrevr_t = \gamma_{y0} + \theta_{y1} lymtr_t + \sum_{i=-q}^{q} \delta_{yi} \Delta lymtr_{t-i} + e_{yt}$$

Note: use HAC (Newey-West) standard errors for hypothesis tests.

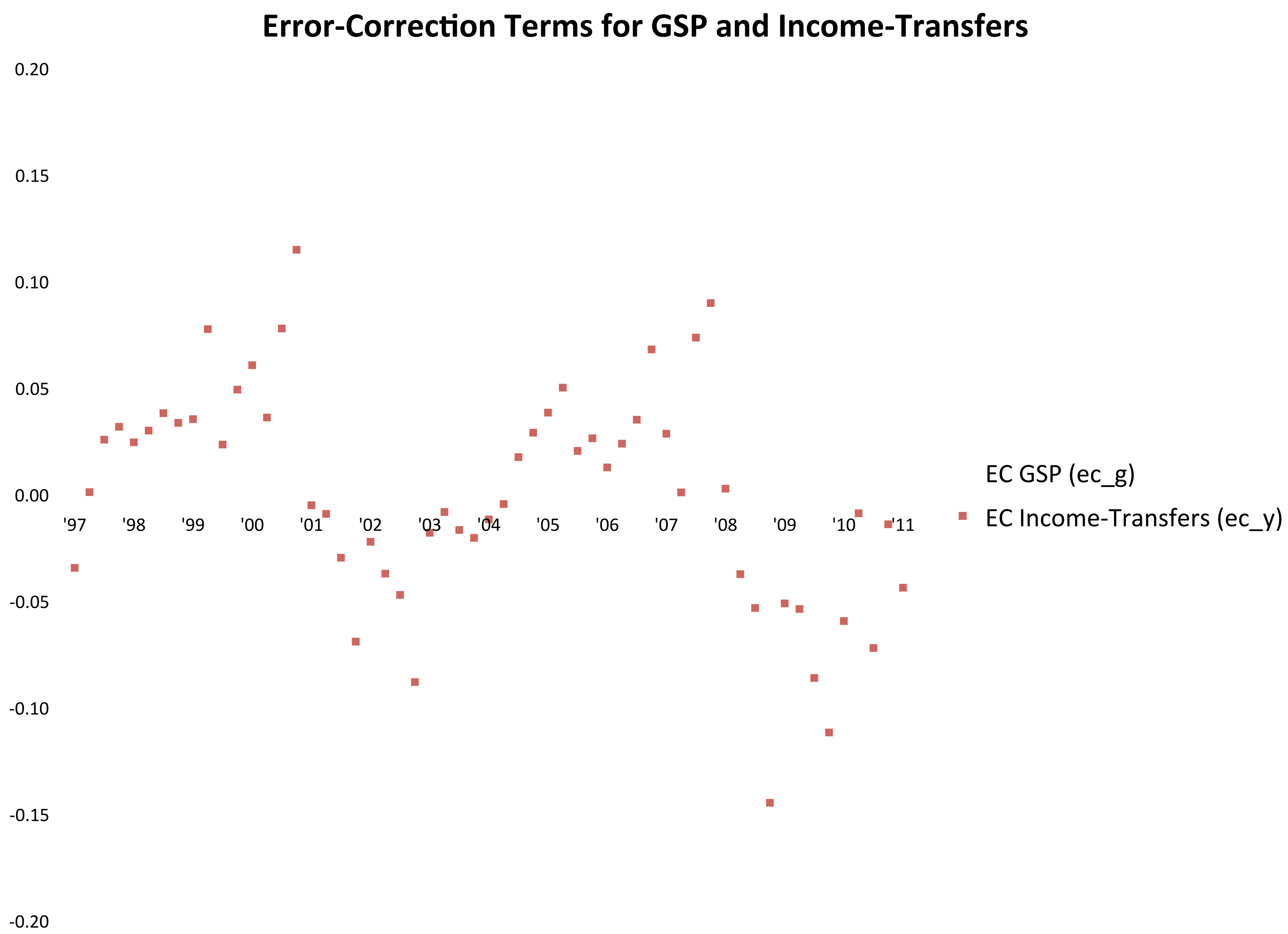
$\hat{\theta}_{g0} =$ mean value of: $lrevr_t - \hat{\theta}_{g1} lgspr_t$
$\hat{\theta}_{y0} =$ mean value of: $lrevr_t - \hat{\theta}_{y1} lymtr_t$

DOLS Estimates of the Cointegration Coefficient on GSP (lgspr)

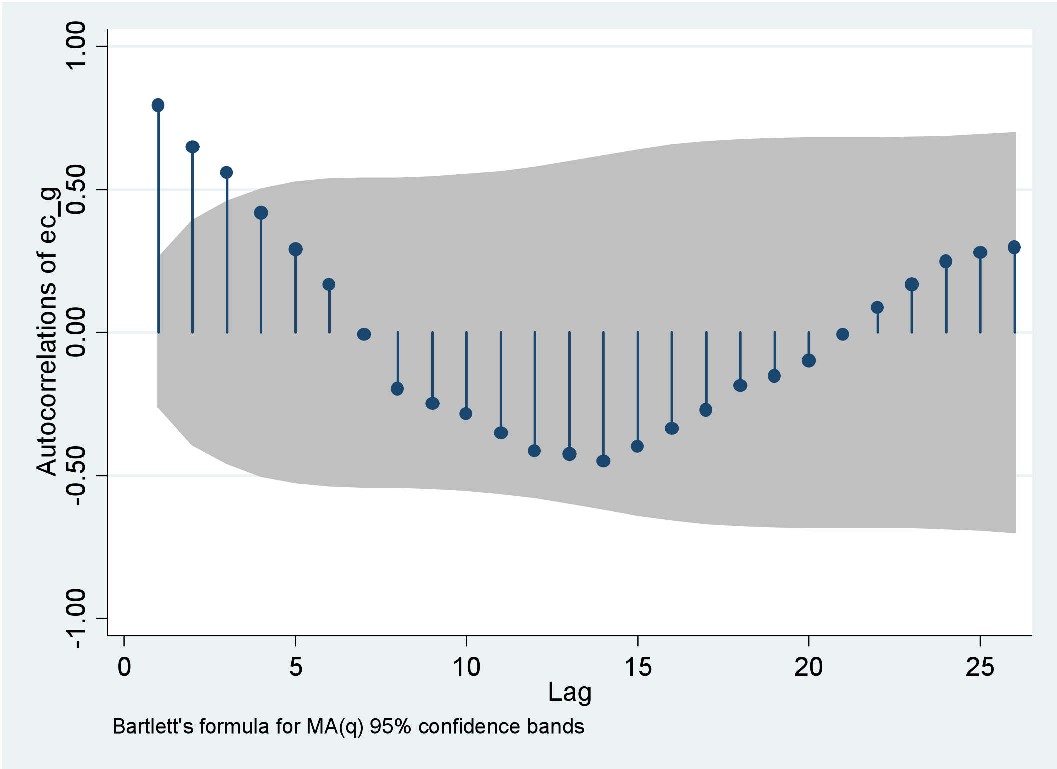
```
*
*  DOLS with newy-west std err : GSP
*
. newey lrevr lgspr L(-4/4).dlgspr, lag(3)

Regression with Newey-West standard errors          Number of obs  =        48
maximum lag: 3                                      F( 10,    37)  =     11.82
                                                    Prob > F       =    0.0000

------------------------------------------------------------------------------
             |             Newey-West
       lrevr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lgspr |   .8728016   .3089833     2.82   0.008     .2467419    1.498861
             |
      dlgspr |
         F4. |  -.5481994    .923495    -0.59   0.556    -2.419378    1.322979
         F3. |  -1.386603   1.336892    -1.04   0.306    -4.095404    1.322198
         F2. |  -.6752612   .7453761    -0.91   0.371    -2.185537    .8350143
         F1. |   .4675015   1.270931     0.37   0.715    -2.107649    3.042652
         --. |   1.021733    1.05158     0.97   0.338    -1.108971    3.152436
         L1. |   .8863019   1.040486     0.85   0.400    -1.221922    2.994526
         L2. |   .5469576   1.595779     0.34   0.734    -2.686398    3.780313
         L3. |    1.94619   1.232476     1.58   0.123    -.5510433    4.443424
         L4. |   3.798816   1.210382     3.14   0.003     1.346349    6.251282
             |
       _cons |   3.658146   2.502795     1.46   0.152    -1.412999    8.729291
------------------------------------------------------------------------------
```

DOLS Estimates of the Cointegration Coefficient on Income-Transfers (lymtr)

```
. *
. *  DOLS with newy-west std err: Income-transfers
. *
. newey lrevr lymtr L(-4/4).dlymtr, lag(3)

Regression with Newey-West standard errors          Number of obs  =        48
maximum lag: 3                                      F( 10,    37)  =     53.24
                                                    Prob > F       =    0.0000

------------------------------------------------------------------------------
             |             Newey-West
       lrevr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lymtr |   .8541221   .0929745     9.19   0.000     .6657378    1.042506
             |
      dlymtr |
         F4. |  -.3384281   .3536237    -0.96   0.345    -1.054938    .3780816
         F3. |  -.5674194   .4812396    -1.18   0.246    -1.542503    .4076647
         F2. |   .1285599    .412168     0.31   0.757    -.7065717    .9636916
         F1. |   .4563609   .5019408     0.91   0.369    -.5606678     1.47339
         --. |    .113619   .4269499     0.27   0.792    -.7514637    .9787016
         L1. |   1.225325   .5346357     2.29   0.028     .1420501      2.3086
         L2. |   .3031446     .36933     0.82   0.417     -.445189    1.051478
         L3. |   .5551846   .3782327     1.47   0.151    -.2111876    1.321557
         L4. |   1.379641   .3823295     3.61   0.001     .6049674    2.154314
             |
       _cons |   4.047577   .7257703     5.58   0.000     2.577027    5.518128
------------------------------------------------------------------------------
```


Error-Correction Terms for GSP and Income-Transfers
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.20
'97
'98
'99
'00
'01
'02
'03
'04
'05
'06
'07
'08
'09
'10
'11
EC GSP (ec_g)
EC Income-Transfers (ec_y)

Autocorrelations for GSP Error Correction Term (ec_g)



The VECM Model

1. $\Delta lrevr_t = \sum_{i=1}^{p} \beta_{1r}(i)\, \Delta lrevr_{t-i} + \sum_{i=1}^{p} \beta_{1g}(i)\, \Delta lgspr_{t-i} + \sum_{i=1}^{p} \beta_{1y}(i)\, \Delta lymtr_{t-i} + \propto_{1g} ec\_g_{t-1} + \propto_{1y} ec\_y_{t-1} + \beta_{10} + u_{1t}$
2. $\Delta lgspr_t = \sum_{i=1}^{p} \beta_{2r}(i)\, \Delta lrevr_{t-i} + \sum_{i=1}^{p} \beta_{2g}(i)\, \Delta lgspr_{t-i} + \sum_{i=1}^{p} \beta_{2y}(i)\, \Delta lymtr_{t-i} + \propto_{2g} ec\_g_{t-1} + \propto_{2y} ec\_y_{t-1} + \beta_{20} + u_{2t}$
3. $\Delta lymtr_t = \sum_{i=1}^{p} \beta_{3r}(i)\, \Delta lrevr_{t-i} + \sum_{i=1}^{p} \beta_{3g}(i)\, \Delta lgspr_{t-i} + \sum_{i=1}^{p} \beta_{3y}(i)\, \Delta lymtr_{t-i} + \propto_{3g} ec\_g_{t-1} + \propto_{3y} ec\_y_{t-1} + \beta_{30} + u_{3t}$

VECM Model Estimates

```
. var dlrevr dlgspr dlymtr, lags(1/1) exog(L.ec_g L.ec_y)

Vector autoregression

Sample:  1998q1 - 2011q3                          No. of obs      =        55
Log likelihood =  500.2322                        AIC             = -17.53572
FPE            =  4.88e-12                        HQIC            = -17.28167
Det(Sigma_ml)  =  2.53e-12                        SBIC            = -16.87877

Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
dlrevr               6     .035701   0.3895   35.08475   0.0000
dlgspr               6     .005611   0.5275   61.40625   0.0000
dlymtr               6     .010567   0.5185   59.23084   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
dlrevr       |
      dlrevr |
         L1. |  -.0137055   .1415883    -0.10   0.923    -.2912136    .2638025
             |
      dlgspr |
         L1. |   .4725516   .7533629     0.63   0.530    -1.004013    1.949116
             |
      dlymtr |
         L1. |   1.580362   .3952457     4.00   0.000     .8056952     2.35503
             |
        ec_g |
         L1. |   .2256739    .350384     0.64   0.520    -.4610661    .9124139
             |
        ec_y |
         L1. |  -.6713756   .4753648    -1.41   0.158    -1.603074    .2603224
             |
       _cons |  -.0073395   .0058534    -1.25   0.210     -.018812    .0041329
-------------+----------------------------------------------------------------
```

VECM Model Estimates, Continued

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
dlgspr						
dlrevr L1.	.0391867	.0222508	1.76	0.078	-.0044241	.0827975
dlgspr L1.	.2614355	.1183922	2.21	0.027	.0293911	.4934798
dlymtr L1.	.2683569	.0621135	4.32	0.000	.1466167	.390097
ec_g L1.	.0241871	.0550634	0.44	0.660	-.0837352	.1321093
ec_y L1.	-.0618759	.0747043	-0.83	0.408	-.2082937	.0845419
_cons	.0023456	.0009199	2.55	0.011	.0005426	.0041485
dlymtr						
dlrevr L1.	-.0322506	.0419082	-0.77	0.442	-.1143891	.0498879
dlgspr L1.	.6404445	.2229849	2.87	0.004	.2034022	1.077487
dlymtr L1.	.3081897	.1169872	2.63	0.008	.0788991	.5374804
ec_g L1.	-.3594172	.1037088	-3.47	0.001	-.5626826	-.1561518
ec_y L1.	.4654486	.1407013	3.31	0.001	.189679	.7412181
_cons	.0005072	.0017325	0.29	0.770	-.0028885	.0039029

Steps to Produce the n-period Forecast from t+1 to t+n

1. Use the VECM to predict $\Delta lrevr_{t+1|t}$, $\Delta lgspr_{t+1|t}$, $\Delta lymtr_{t+1|t}$
2. $lrevr_{t+1|t} = lrevr_t + \Delta lrevr_{t+1|t}$, etc. for $lgspr_{t+1|t}, lymtr_{t+1|t}$.
3. $ec\_g_{t+1|t} = lrevr_{t+1|t} - \hat{\theta}_{g0} - \hat{\theta}_{g1} lgspr_{t+1|t}$
   $ec\_y_{t+1|t} = lrevr_{t+1|t} - \hat{\theta}_{y0} - \hat{\theta}_{y1} lymtr_{t+1|t}$
4. Iterate through steps 1-3 replacing "$t+1|t$" by "$t+2|t$", "t" by "$t+1|t$", until "$t+n|t$".

Prediction Code for Stata After "var" Command

```
*  prediction loop
*    obs 58  is 2011q4, first quarter in forecast
*    obs 178 is 2041q4, far in the future
*
local i=58
while `i'<=178 {
                capture drop pdlrevr
                capture drop pdlgspr
                capture drop pdlymtr

                predict pdlrevr in `i', xb equation(dlrevr)
                predict pdlgspr in `i', xb equation(dlgspr)
                predict pdlymtr in `i', xb equation(dlymtr)

                replace lrevr=l.lrevr + pdlrevr in `i'
                replace lgspr=l.lgspr + pdlgspr in `i'
                replace lymtr=l.lymtr + pdlymtr in `i'

                replace dlrevr=d.lrevr in `i'
                replace dlgspr=d.lgspr in `i'
                replace dlymtr=d.lymtr in `i'

                replace ec_g = lrevr - bconsg - bg*lgspr in `i'
                replace ec_y = lrevr - bconsy - bymt*lymtr in `i'

                local i=`i'+1
}
```

Estimation of the Beveridge-Nelson (BN) Trend

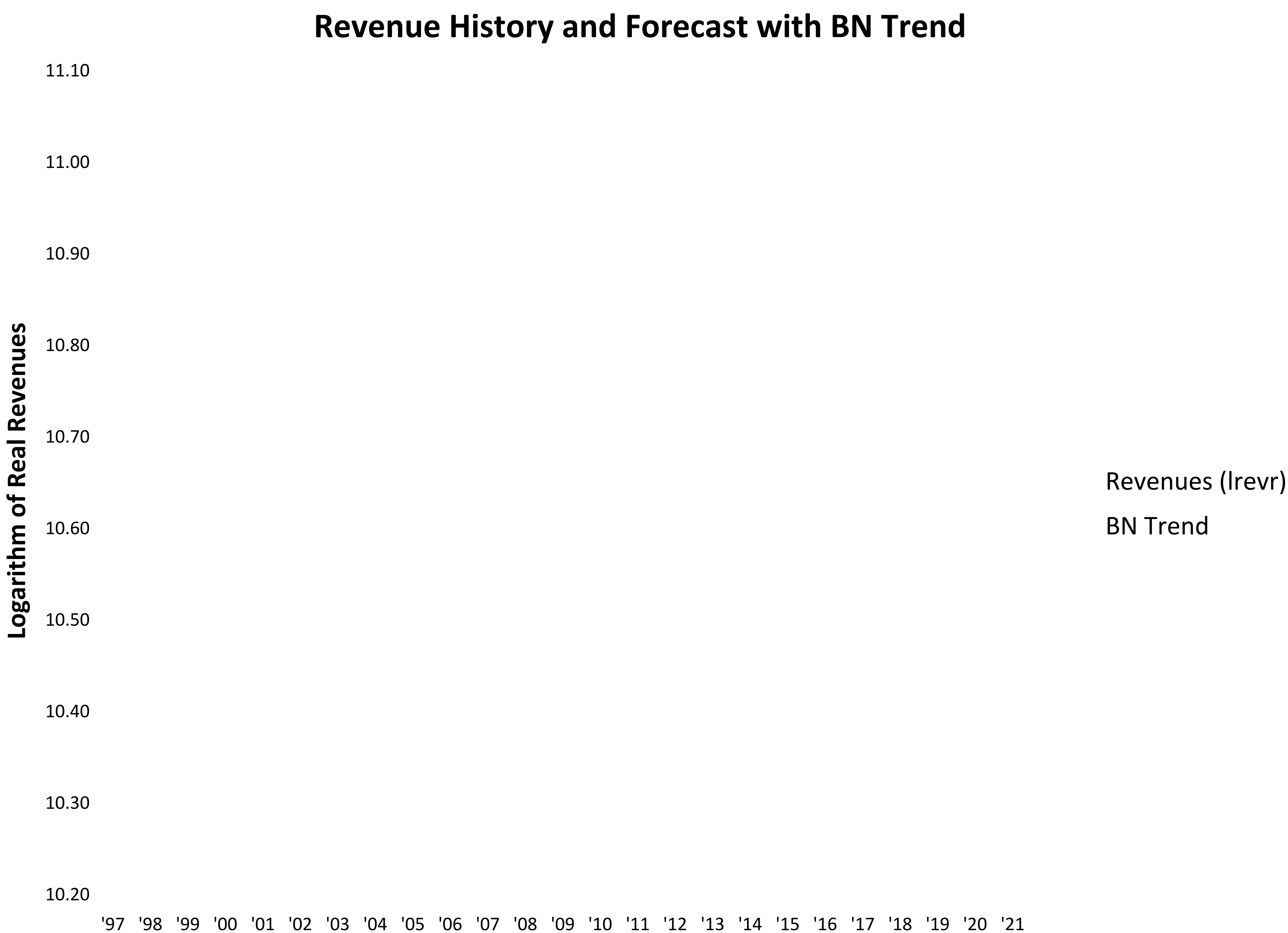
To estimate the Beveridge-Nelson trend, calculate the slope of the long-term trend in the far future and extrapolate this linear trend backward to the end of history.

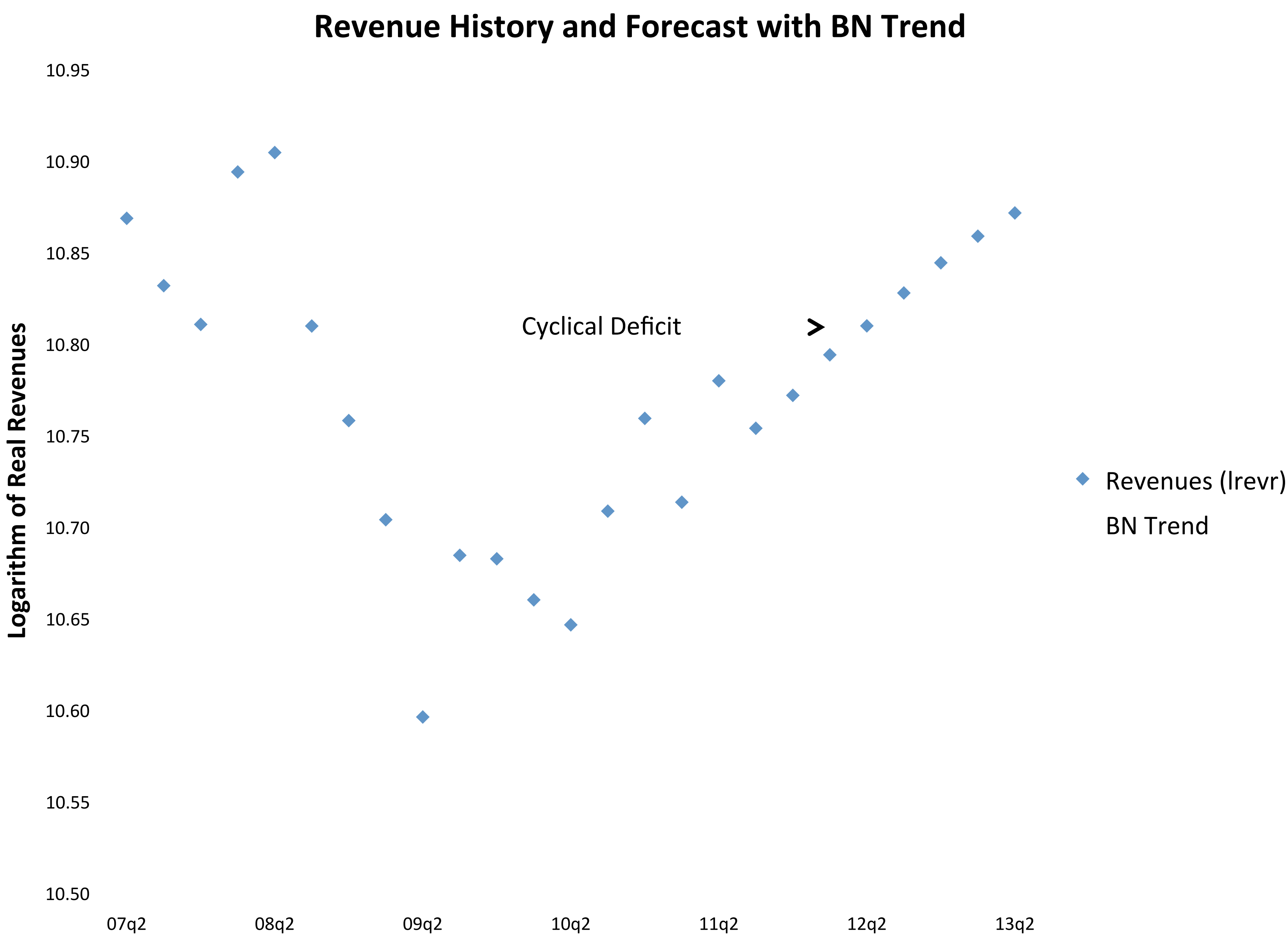
This is easily done by estimating an OLS regression of *lrevr* on time near the end of the long-term forecast and “backcasting” the fitted line back to the end of history.

Calculation of the BN Trend

- . *
- . * Calculate the BN Trend
- . *
- .
- . regress lrevr time if tin(2025q1,2041q4)
-
- Source | SS df MS Number of obs = 68
- -------------+------------------------------ F(1, 66) = .
- Model | .619436925 1 .619436925 Prob > F = 0.0000
- Residual | 5.8334e-09 66 8.8385e-11 R-squared = 1.0000
- -------------+------------------------------ Adj R-squared = 1.0000
- Total | .61943693 67 .009245327 Root MSE = 9.4e-06
-
- --
- lrevr | Coef. Std. Err. t P>|t| [95% Conf. Interval]
- -------------+--
- time | .0048626 5.81e-08 8.4e+04 0.000 .0048625 .0048628
- _cons | 9.831054 .0000171 5.8e+05 0.000 9.83102 9.831088
- --
-
- . display (1+_b[time])^4-1
- .01959292
-
- . predict trend_lrevr


Revenue History and Forecast with BN Trend
Logarithm of Real Revenues
11.60
11.40
11.20
11.00
10.80
10.60
10.40
10.20
10.00
'97 '98 '99 '00 '01 '02 '03 '04 '05 '06 '07 '08 '09 '10 '11 '12 '13 '14 '15 '16 '17 '18 '19 '20 '21 '22 '23 '24 '25 '26 '27 '28 '29 '30 '31 '32 '33 '34 '35 '36 '37 '38 '39 '40 '41
Revenues (lrevr)
BN Trend


Revenue History and Forecast with BN Trend
Logarithm of Real Revenues
11.10
11.00
10.90
10.80
10.70
10.60
10.50
10.40
10.30
10.20
'97 '98 '99 '00 '01 '02 '03 '04 '05 '06 '07 '08 '09 '10 '11 '12 '13 '14 '15 '16 '17 '18 '19 '20 '21
Revenues (lrevr)
BN Trend


Revenue History and Forecast with BN Trend
Logarithm of Real Revenues
10.95
10.90
10.85
10.80
10.75
10.70
10.65
10.60
10.55
10.50
07q2
08q2
09q2
10q2
11q2
12q2
13q2
Cyclical Deficit
Revenues (lrevr)
BN Trend

References

- Stock, James H., and Watson, Mark W. (2011). *Introduction to Econometics, 3rd Edition*, Pearson Education.
- Enders, Walter. (2010). *Applied Econometric Time Series, 3e*, John Wiley & Sons.
- Beveridge, Stephen, and Nelson, Charles R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle', *Journal of Monetary Economics*, 7 (1981), pp. 151-174.