Time Series Spring 2019 Final
1 May 2019
Connor Boone

Investigating Employment Inputs in the U.S. Telecommunication Sector

Introduction and Data:

Notes: Some Figures and Tables are at the end of the investigation.

This project focused on the impact of two explanatory variables, PPI (producer price index) of wireless communication services, and U.S. exports of services in the telecommunication sector on a dependent variable, total employees in the telecommunication sector in the U.S. The data was collected in time series format, on a non-seasonally adjusted monthly basis from 01/01/2000 to 12/01/2017, with a resulting 216 observations. The goal of this investigation is to determine a forecast estimator for the dependent variable, emp.

Dependent: Emp = All Employees: Information: Telecommunications
Explanatory: Ppi = Producer Price Index by Industry: Wireless Telecommunications Carriers: Wireless Telecommunications Services
Exp = U.S. Exports of Services: Telecommunications, Computer, and Information Services

It was hypothesized that PPI and service exports would affect change on the total number of employees due to acting on macroeconomic factors existing within this sector. It was assumed that PPI could affect total costs incurred by providers and that exports could serve as a measure of external industrial influence of US firms on the global telecommunications industry. Data was obtained from the Federal Reserve Economic Database (FRED). Graphical representation of the stationarity of the data is displayed in **Figures 1** and **2.**

Figure 1. Graphical Representation of Data.
Figure 2. Data post-differencing.

It was determined through ACF testing and visual inspection that the order of all variables were I(1). Before assessing the certainty of stationarity, all variables were logged and then after running the proper testing, were differenced to I(1). Following this, the variables were tested via Granger Causality for the presence of seasonality, which when left untreated can cause bias in results. This was achieved using a joint F-test on a restricted dynamic regression model without seasonal dummies versus another model which contained them. The p-value for this F-test was 0.027; therefore, the null was rejected at the 5% indicating the need for seasonal dummies in further regressions.

Methods:

Various methods were used to estimate the effects of the regressors on the dependent variable. Their models are shown below and will be discussed independently in the results section.

Model 1. VAR; **Model 2**. VECM; **Model 3**. ARIMA; **Model 4**. Dynamic Linear Model.

Models 1 and 2. VAR/VECM.

VAR

The VAR model was hypothesized to be able to capture and incorporate the effects of the independent variables in order to produce a more accurate forecast of employment in the telecommunications sector. The *vars* package uses VAR selection criteria in order to estimate the number of lags that should be used in the regression. The output from this are displayed here:

VAR selection criteria:

AIC(n)	HQ(n)	SC(n)	FPE(n)
3	**3**	**1**	**3**

VAR selection criteria suggested that 3 lags was the optimal number of autoregressive terms to add to the equation, even though SC(BIC) suggested one, BIC tends to underestimate lag length.

In order to estimate the degree to which the independent variables affected the dependent, the impulse response functions were calculated and visualized. These are displayed in **Figures 3** and **4.** The indicate the small degree to which the variables affected the moments of the impulse signals, with magnitudes for both functions <0.001 and > -0.003. The variance decomposition of the variables was also run to show the degree to ascertain the amount each variable contributes to others in the AR process. These results are displayed in **Table 1.**

Cointegration was tested for by using the Engle Granger method to run regressions on the lagged residuals of various combinations of the logged variable terms to see if their plots were stationary, of which there were. Cointegration was also tested for by using the Johansen procedure.

The results of this test are displayed here:

"eigen test"

	test	10pct	5pct	1pct
r <= 2 |	3.23	7.52	9.24	12.97
r <= 1 |	4.25	13.75	15.67	20.20
r = 0 |	68.47	19.77	22.00	26.81

"trace test"

	test	10pct	5pct	1pct
r <= 2 |	12.94	7.52	9.24	12.97
r <= 1 |	46.02	17.85	19.96	24.60
r = 0 |	88.99	32.00	34.91	41.07

These results show there is clear evidence of cointegration and therefore bias exists within the VAR model.

VECM

The VECM used the Johansen procedure to correct for the cointegration that existed within the dataset. The code for this was as follows:

```
vecm1 <- ca.jo(y, ecdet = "const", type="eigen", K=14, spec="longrun",
          season=12)
```

This new VECM was tested for serial correlation via the BG test. Even though VARselect had suggested that K=3 lags was the optimal length for the regression, this caused serial correlation in the data. Additional lengths were added until it was determined that K=14 was the optimal length to control for serial correlation and allow the results to remain relatively unaffected. AIC/BIC values were negatively affected by changing lag length from 3 to 14. This is shown here:

	K=3	K=14
AIC:	-4084.6	-3886.7
BIC:	-3893	-3351.6

Model 3. ARIMA.

The fitARIMA program suggested that an optimal AR length was 3 lags.

Code:

```
fitARIMA(log(emp), order=c(1,1,1))
fitARIMA(log(emp), order=c(2,1,1))
fitARIMA(log(emp), order=c(3,1,1))
smp <- demp[0:(length(demp)-11)]
arima1 <- arima(smp), order=c(3,0,1) seasonal = list(order = c(0,0,0), period =
12),method="ML"))
AIC(arima1); BIC(arima1)
> AIC(arima1); BIC(arima1)
[1] -1392
[1] -1375.2
```

Testing was performed on the ARIMA model to assess whether there was autocorrelation in the results. The Ljung-Box test was used to test for this, and the null failed to be rejected at the 5% level, indicating there is evidence of some serial correlation in the model.

Model 4. Dynamic Linear Model.
Several models with varying lags were chosen to estimate Granger Causality of the variables and to estimate the goodness of fit of the model. The best fit model was:

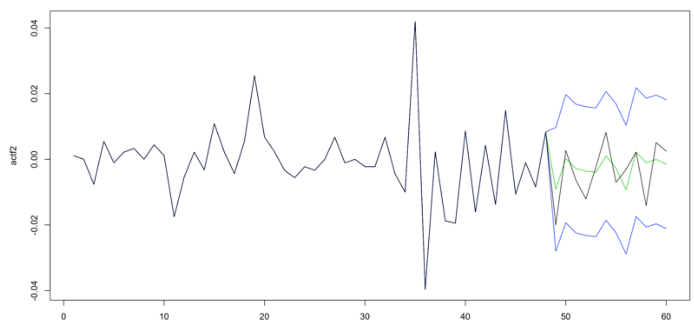
```
rb2 = dynlm(demp~L(demp,1:3)+season(demp)+L(dppi,0:6)+ L(dexp,0:12),
start=c(2000,01), end = c(2017,12), frequency = 12)
> AIC(rb2); BIC(rb2)
[1] -1377.4
[1] -1258.2
```

An estimate of the Residuals versus their fitted values for this equation are shown in **Figure 5,** and **Table 2** shows the output results.

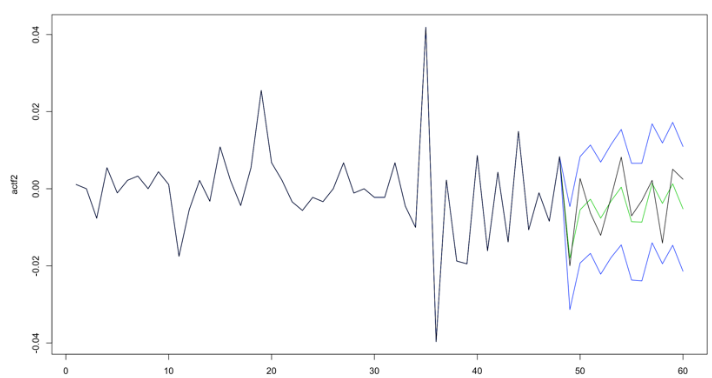
The dynamic regression results were interesting because they really call into question the utility of using d(exports) in the equation at all. At no point on the results were any of the lags or the nonlagged value found to be significant even at the 10% level (**Table 5**).

As it stands considering all of the above information, at this point in the investigation, the best estimator is the VECM model. It has the lowest AIC/BIC values, and has had several testing conditions satisfied via its error correction term.

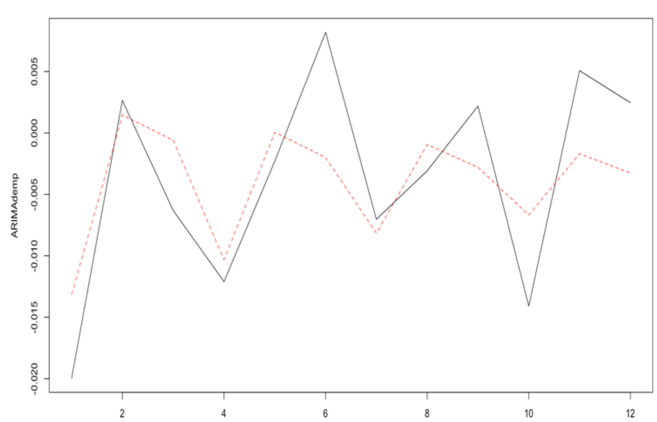
Results (Forecasting):
<u>I used the last 12 periods instead of 8 for an out of sample forecast, then predicted for the next 12 periods because I misread the problem, so I just rolled with it.</u>



VAR (demp): RMPSE: 0.0066444



VECM (demp): RMPSE: 0.0056068



ARIMA (demp): RMPSE: 0.0054567

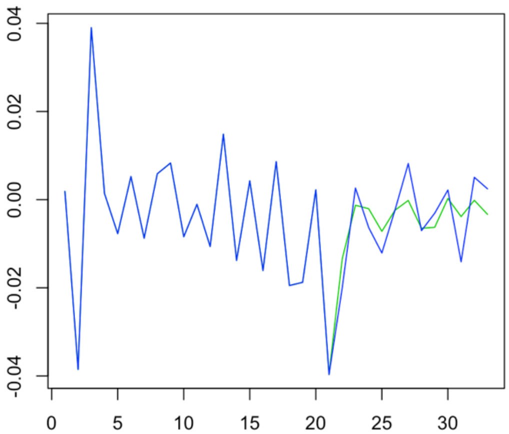
The following models were used to forcast the results of the last 12 observations and then compare them to the actual values to determine the integrity of the model. The models estimated values are in green in the VAR and VECM models and dotted red in the ARIMA model. The confidence intervals are marked in blue in VAR and VECM. In all three models, the solid black lines indicate the actual logged values.

According to the RMPSE, the model with the closest estimate to the actual values is the ARIMA model.

The ARIMA model is likely the most accurate forecasting tool because the other two models are using a VAR that includes a value that is likely degrading the confidence of the results (exports). The AR model probably should have had more lags to better control for serial correlation.

The average of all estimates versus the real values and their text readouts are listed on the next page for further comparison of the estimates.

Final values (blue) vs averaged forecasted values (green):



green (avg) blue

	varfct	vecmfct	arma	actual
13	-0.009171018	-0.01798444	-0.013154235	-0.0199832
14	0.000196709	-0.00546051	0.001461014	0.0026457
15	-0.002826337	-0.00271496	-0.000599464	-0.0063108
16	-0.003611746	-0.00762905	-0.010321866	-0.0121013
17	-0.003950450	-0.00321813	0.000043197	-0.0023095
18	0.001049214	0.00040561	-0.001999500	0.0081873
19	-0.002733474	-0.00854513	-0.008162938	-0.0070319
20	-0.009244190	-0.00865049	-0.000947602	-0.0030840
21	0.002194975	0.00140956	-0.002800679	0.0021855
22	-0.001032476	-0.00380161	-0.006680629	-0.0140966
23	-0.000048735	0.00127694	-0.001694236	0.0050666
24	-0.001509655	-0.00522172	-0.003265285	0.0024591

Discussion and Conclusion:

This project investigated the effect of producer price index and US exports of telecommunications services on the number of people employed in the telecommunications sector. The results show that a large portion of this work is influenced by factors such as time, seasonality, and PPI. They also showed that US exports of telecommunications services likely doesn't have a large impact on employment in this sector.

The best model to estimate the effect of these factors on employment is VECM. This is evident by it having the low AIC/BIC values by far, and the goodness of fit for the out of sample model. The ARIMA model is a good estimate of future movements in employment, but is likely skewed by bias from excluded variables, such as market indicators, government actions, financial environment, and consumer sentiment.

Tables and figures.

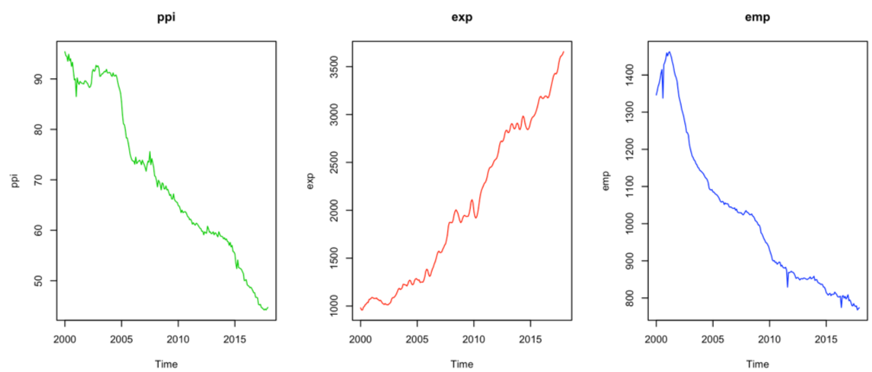
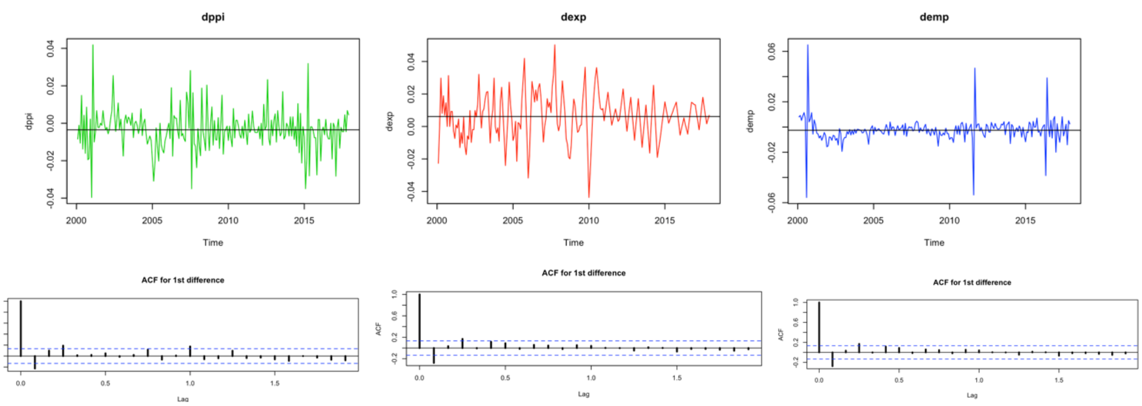


Figure 1.
Nonstationary variables



dppi dexp demp

Figure 2.
Variables that have been differenced to order I(1).

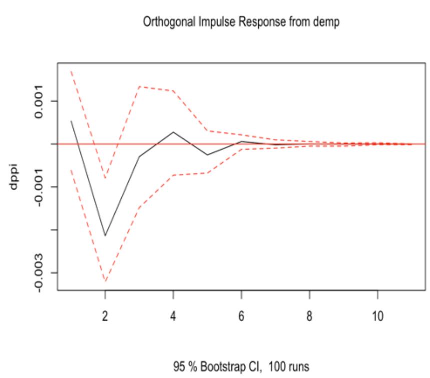


Figure 3. Impuse Response Function on demp from dppi.

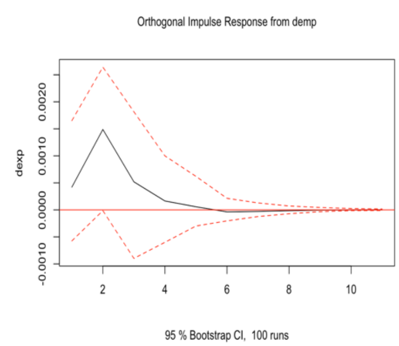


Figure 4. Impuse Response Function on demp from dexp.

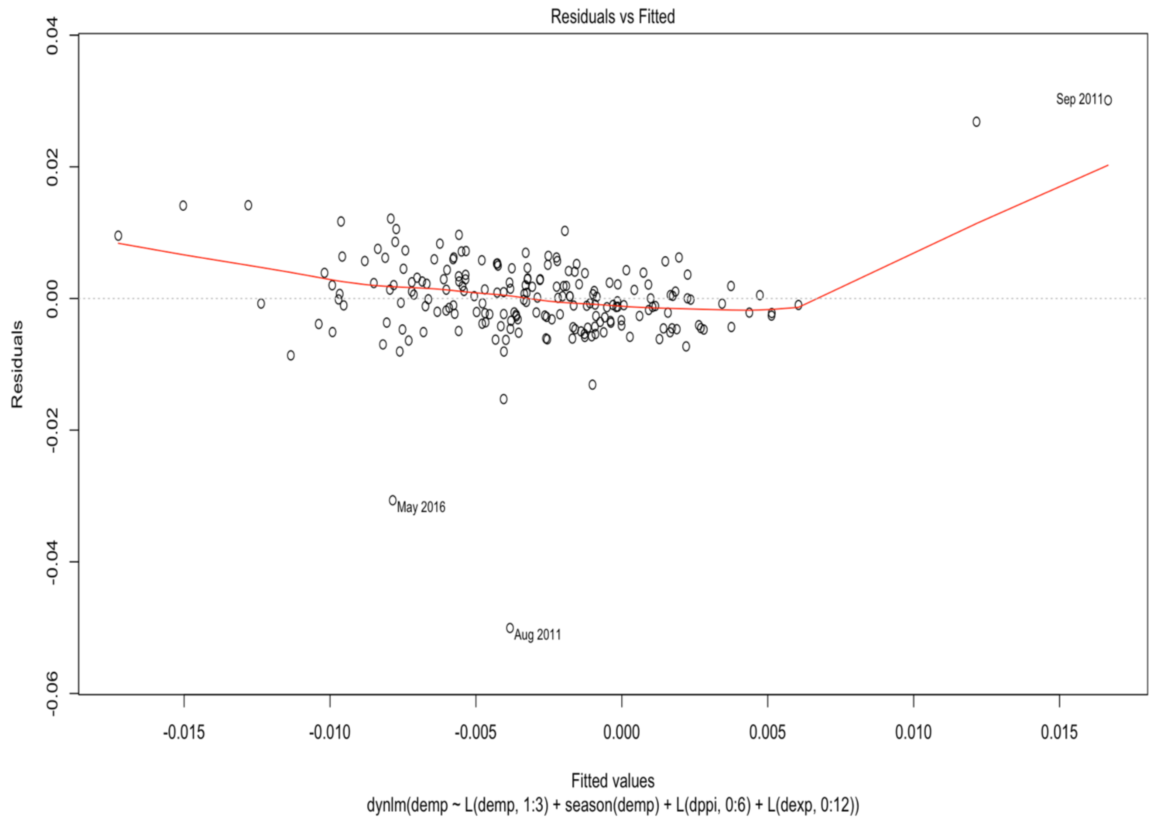


Figure 5.
Dynamic Linear Regression, fitted versus residual values.

Table 1.
Variance Decompositions (continued below on pg. 8)

	$demp			$dppi		
	demp	dppi	dexp	demp	dppi	dexp
[1,]	1.00000	0.0000000	0.0000000	0.011477	0.98852	0.00000000
[2,]	0.98635	0.0095422	0.0041039	0.013529	0.98607	0.00040463
[3,]	0.95789	0.0346529	0.0074543	0.013021	0.97811	0.00886901
[4,]	0.95663	0.0326519	0.0107211	0.012349	0.97774	0.00991018
[5,]	0.95131	0.0321904	0.0164956	0.014946	0.97474	0.01031627
[6,]	0.94450	0.0353175	0.0201796	0.018215	0.97171	0.01007480
[7,]	0.94562	0.0345378	0.0198408	0.017236	0.97316	0.00960026
[8,]	0.93293	0.0353207	0.0317483	0.017857	0.97169	0.01045355
[9,]	0.92972	0.0365818	0.0336954	0.018178	0.97031	0.01150966
[10,]	0.93013	0.0354949	0.0343765	0.017813	0.97077	0.01141296
[11,]	0.91922	0.0351056	0.0456772	0.017820	0.97095	0.01122816
[12,]	0.92114	0.0342808	0.0445761	0.018658	0.96955	0.01179390

$dexp

	demp	dppi	dexp
[1,]	0.011241	0.00026562	0.98849
[2,]	0.012953	0.00130856	0.98574
[3,]	0.026214	0.00130103	0.97248
[4,]	0.039855	0.00580239	0.95434
[5,]	0.052540	0.00614618	0.94131
[6,]	0.071089	0.00825664	0.92065
[7,]	0.087666	0.01059717	0.90174
[8,]	0.112233	0.01043871	0.87733
[9,]	0.115253	0.01479925	0.86995
[10,]	0.113666	0.03011525	0.85622
[11,]	0.110008	0.05436318	0.83563
[12,]	0.106795	0.08003973	0.81317

Table 5.
Results from dynlm regression.

DynLM Results	Dependent variable:
	demp
L(demp, 1:3)1	-0.300***
	(0.076)
L(demp, 1:3)2	-0.037
	(0.079)
L(demp, 1:3)3	0.165**
	(0.075)
season(demp)Feb	0.006*
	(0.003)
season(demp)Mar	0.005
	(0.003)
season(demp)Apr	0.005
	(0.004)
season(demp)May	0.002
	(0.003)
season(demp)Jun	0.010***
	(0.003)
season(demp)Jul	0.007**
	(0.003)
season(demp)Aug	0.002
	(0.003)
season(demp)Sep	0.008**
	(0.003)
season(demp)Oct	0.009**
	(0.004)
season(demp)Nov	0.012***
	(0.003)
season(demp)Dec	0.008***
	(0.003)
L(dppi, 0:6)0	0.056
	(0.058)
L(dppi, 0:6)1	0.047
	(0.060)
L(dppi, 0:6)2	-0.095
	(0.060)
L(dppi, 0:6)3	-0.123**
	(0.061)
L(dppi, 0:6)4	-0.096
	(0.060)
L(dppi, 0:6)5	-0.045
	(0.061)
L(dppi, 0:6)6	0.005
	(0.058)
L(dexp, 0:12)0	0.065
	(0.068)
L(dexp, 0:12)1	0.001
	(0.090)
L(dexp, 0:12)2	0.020
	(0.090)
L(dexp, 0:12)3	0.055
	(0.091)
L(dexp, 0:12)4	0.039
	(0.090)
L(dexp, 0:12)5	0.009
	(0.088)
L(dexp, 0:12)6	-0.009
	(0.087)
L(dexp, 0:12)7	0.117
	(0.087)
L(dexp, 0:12)8	-0.085
	(0.088)
L(dexp, 0:12)9	0.061
	(0.086)
L(dexp, 0:12)10	0.120
	(0.081)
L(dexp, 0:12)11	-0.082
	(0.082)
L(dexp, 0:12)12	0.015
	(0.065)
Constant	-0.013***
	(0.002)
Observations:203;R2:0.281	
Residual Std. Error:0.007 (df = 168)	
F Statistic	1.927*** (df = 34; 168)

Sources:
Emp - https://fred.stlouisfed.org/series/CES5051700001
Ppi - https://fred.stlouisfed.org/series/PCU5172105172101
Exp - https://fred.stlouisfed.org/series/ITXTCIM133S

List of packages used:
lmtest
dynlm
season
vars
urca
robust
stargazer
"intord.R"
"serialcorr.R"
"fitarima(1).R"
"RMSPE.R"