

An emergence index

Sonia Mazzi

22/05/2019

The US patent data at hand can be best described as “pure trend” data, that is the main source of variation in the data is trend and noise.

We choose to model the data using a Local Linear Model (LLM).

Local linear trend (LLT) state-space model

- A signal plus noise model: Have $y = \{y_i\}_{i=1}^n$ with

$$y_i = \mu_i + \epsilon_i,$$

where μ_i is a stochastic process (with linear dynamic) and ϵ_i 's are iid with mean zero and constant variance.

- Note that y_i can be a vector.
- The stochastic specification of μ_i allows to model y in different ways.

Modeling the signal

- First assume $\mu_i = a + d i$, with a and d constant.
- Unsatisfactory for most time series. But, if we assume that $\mu = \{\mu_i\}$ is “smooth”, μ could be described LOCALLY by a linear model (Taylor's theorem).
- Let a and d vary with time to allow μ_i to adapt to the evolution of the series.
- If $\mu_i = a + d i$ then $\mu_{i+1} = \mu_i + d$.
- To allow for flexibility in the model

$$\begin{aligned}y_i &= \mu_i + \epsilon_i, \\ \mu_{i+1} &= \mu_i + d_i + \nu_i, \\ d_{i+1} &= \delta d_i + \eta_i.\end{aligned}$$

where $\nu = \{\nu_i\}$ and $\eta = \{\eta_i\}$ are independent white noise processes with mean zero and variance σ_ν^2 and σ_η^2 respectively. ϵ is independent of ν and η .

- The LLM is a state-space with a two-dimensional state vector with entries μ_i and d_i .
- μ_i is the signal
- d_i is the rate of change of the signal or “first derivative”.
- The signal can be extracted using a combination of Kalman and smoothing filtering.
- My preferred algorithms are the “diffuse” KF and SF.
- The same algorithms allow for (diffuse) prediction.
- For more details on the state-space model and signal-extraction algorithms see

de Jong, P.(1991). The diffuse Kalman filter. *The Annals of Statistics*. Vol 19, No.2, pp 1073 - 1083.

Generalisations

- Model easily generalised to handle unequally spaced data (e.g. missing data)
- Suppose y is observed at times t_1, \dots, t_n and let $h_i = t_{i+1} - t_i$. Then, if $\mu_i = \mu_{t_i}$

$$\begin{aligned} y_i &= \mu_i + \epsilon_i, \\ \mu_{i+1} &= \mu_i + h_i d_i + \nu_i, \\ d_{i+1} &= \delta d_i + \eta_i. \end{aligned}$$

- Other sources of variation such as seasonal effects, cycles and covariates, can be added to the state vector.

Application to patent data

$y = \{y_i\}_{i=1}^n$ represents one time series of quarterly counts. $\mu = \{\mu_i\}$ represents the signal and σ_ϵ is the noise standard deviation. Also, $d = \{d_i\}_{i=1}^n$ represents the rates of change, or first derivatives, of the signal at each time point.

Positive values of d_i indicate that the signal is increasing, negative values of d_i indicate that the signal is decreasing. If $d_i = 0$ then that indicates a point of maximum, minimum or perhaps an inflexion point.

The index

During a time period, say consecutive times j_1, \dots, j_m , we define

$$E = \sum_{k=1}^m d_{j_k}$$

as a measure of emergence. Positive values of E indicate that growth was dominant in the time period. Negative values of E indicate that decline was dominant in the time period. Values of E near zero, positive or negative, could arise in many ways. One way is if, for example, the d_{j_k} 's are mostly very small and so the series was stationary. Or, it could be that there are approximately an equal number of positive and negative d 's, and so there was growth and decline in the same period.

Now, what does “large” value of E mean? The derivative measures the rate of change of a function. Given a function f and two points x and y , the derivative of f at x , the instantaneous rate of change of f at x , is the limit as y tends to x of the ratio of the change of f from $f(y)$ to $f(x)$ divided by $y - x$,

$$f'(x) = \lim_{y \rightarrow x} \frac{f(y) - f(x)}{y - x}$$

So, we see that if the total variability of f is large, f' will be larger than the derivative of f if the total variability of f were small.

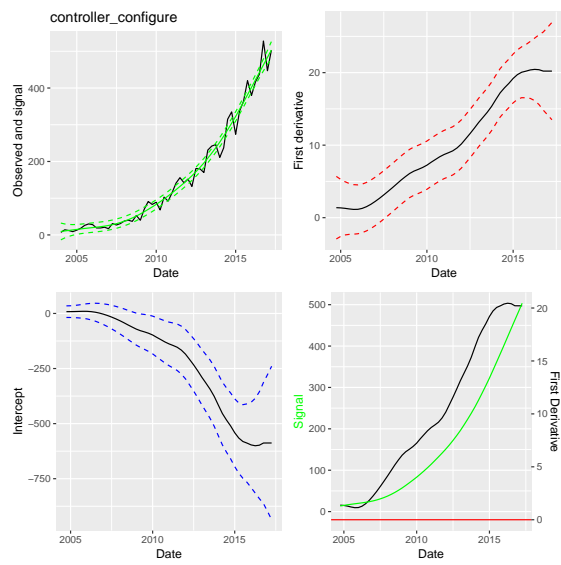
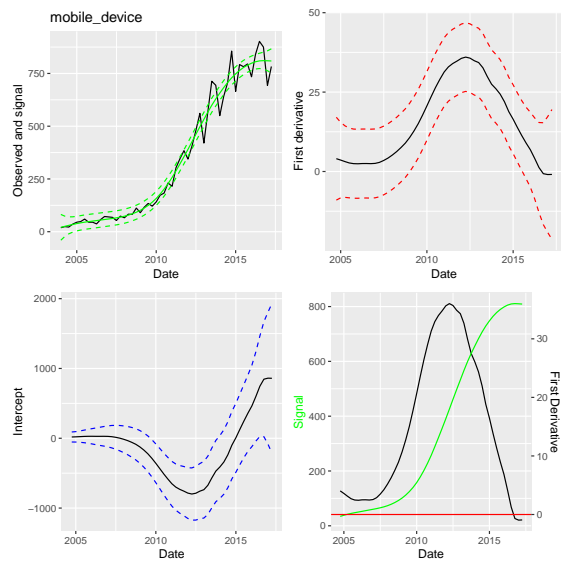
Therefore, we redefine E to be

$$E = \frac{\sum_{k=1}^m d_{j_k}}{\sigma_\epsilon}.$$

In this way we avoid large values of E , positive or negative, being driven by a large noise variance.

For example, consider the terms “mobile device” and “controller configure”. For the time period 1st quarter 2005 - 2nd quarter 2017 we have

Term	σ_ϵ	$\sum d_i$	$\sum d_i / \sigma_\epsilon$
mobile device	48.584	778.173	16.017
controller configure	20.266	511.429	25.235

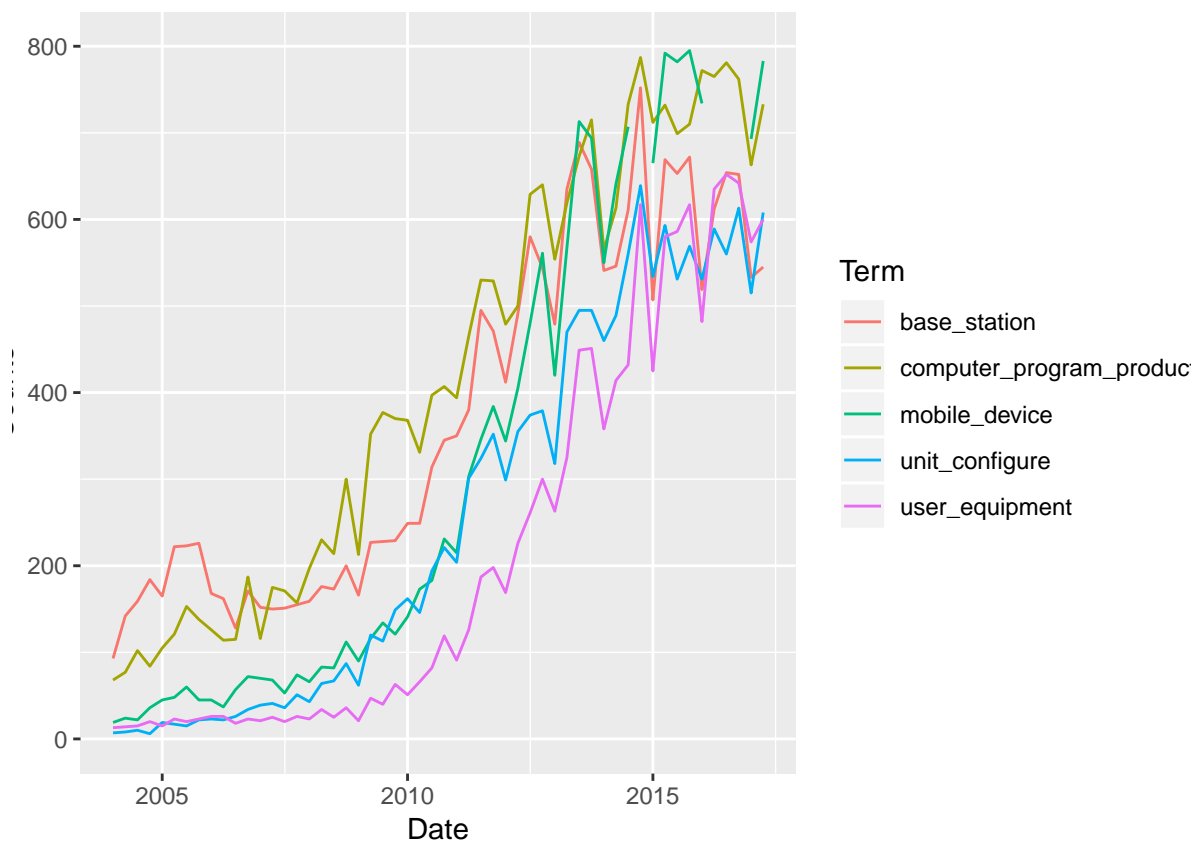
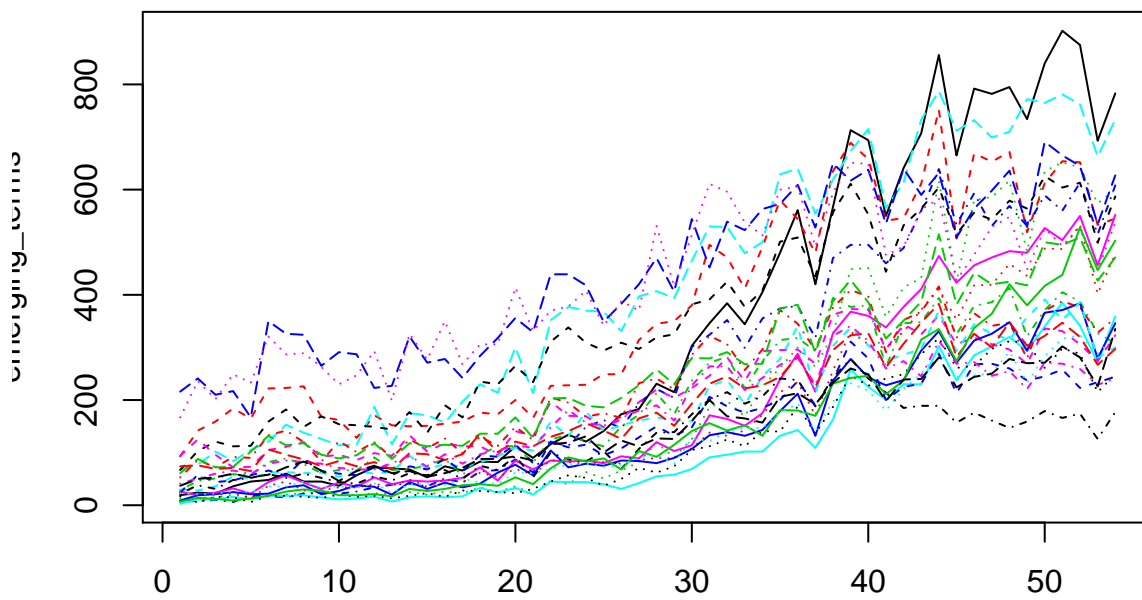


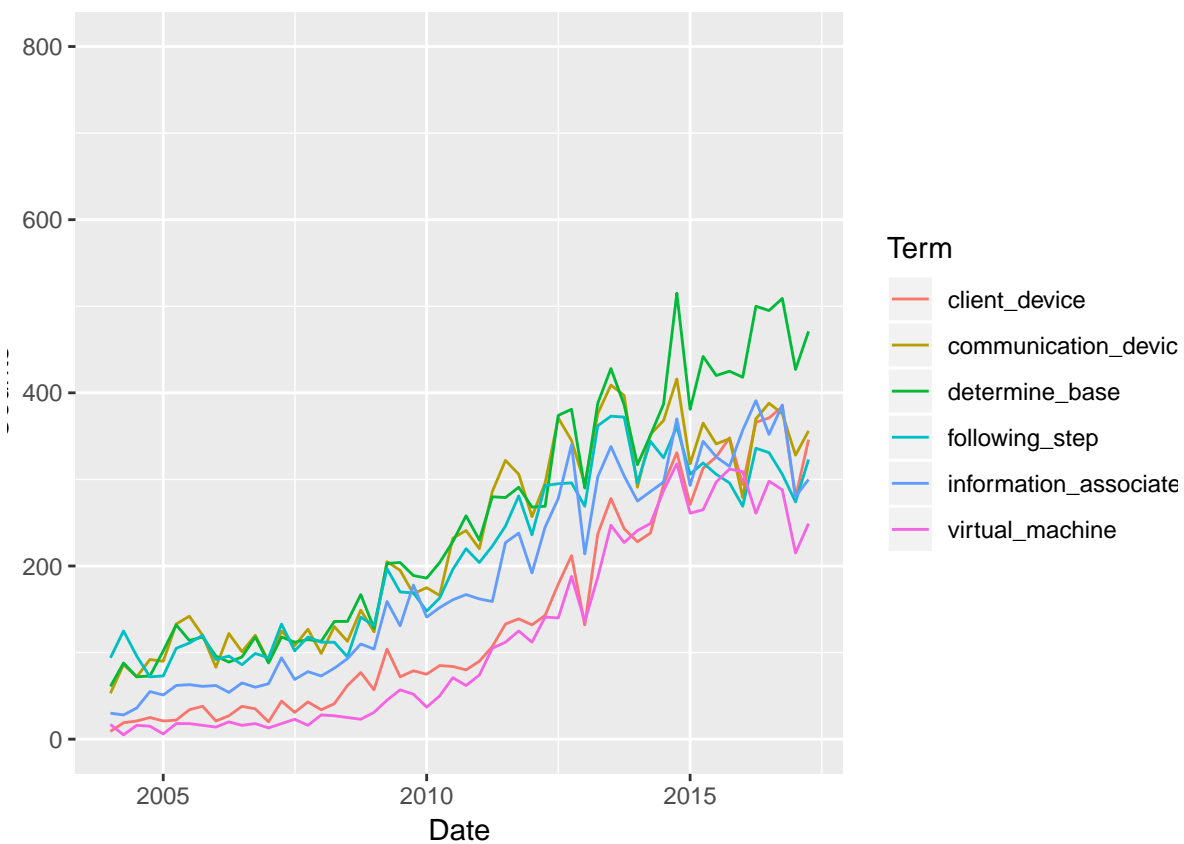
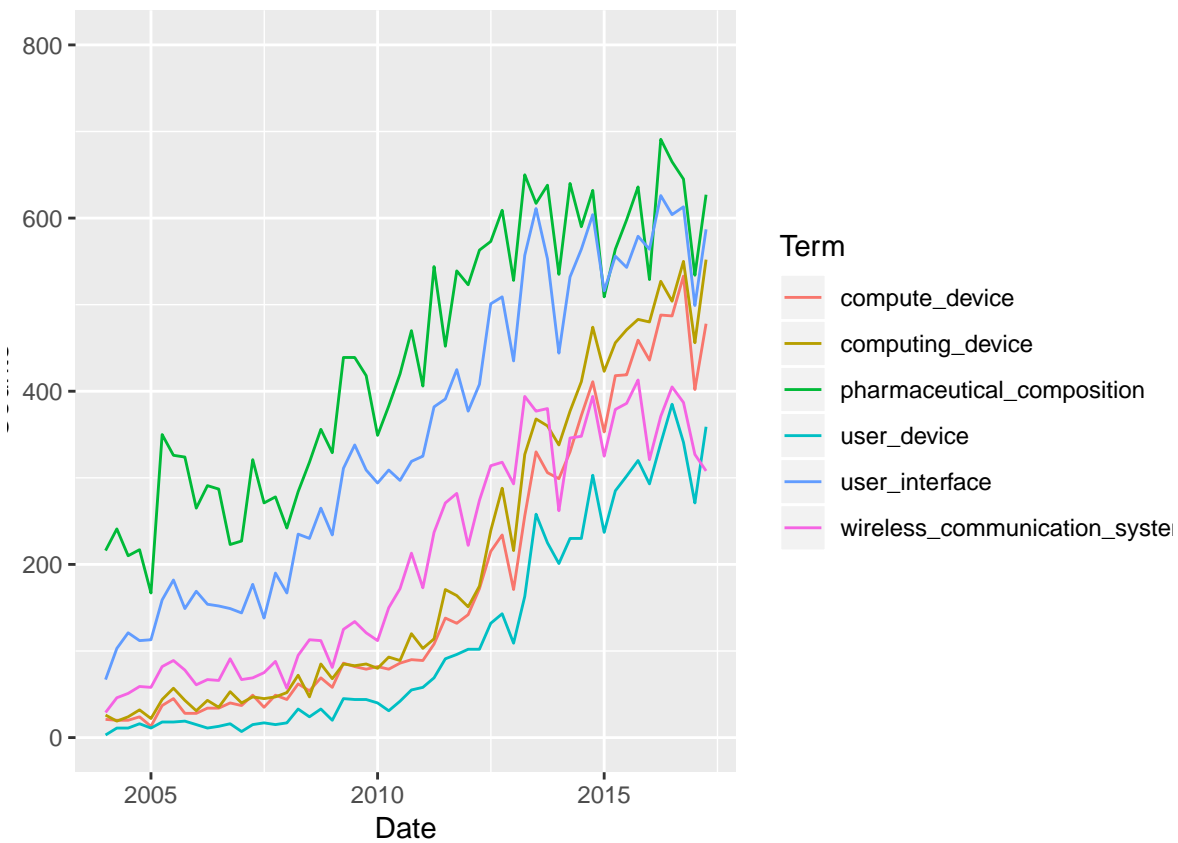
However, if we inspect the plots above, it is clear that although “mobile device” has a larger $\sum d_i$, “controller configure” has a steady grow over the time period, whereas “mobile device” is starting to decline after steady growth.

The data I used to test consisted of 74 series (for 74 terms) 25 of which were the top emergent terms, 25 were the top stationary terms and 24 top declining terms.

Terms classified as emergent by Porter

The terms are all clearly emergent. I am not sure if Porter's ranking is the same as mine as I don't have Porter's scores.





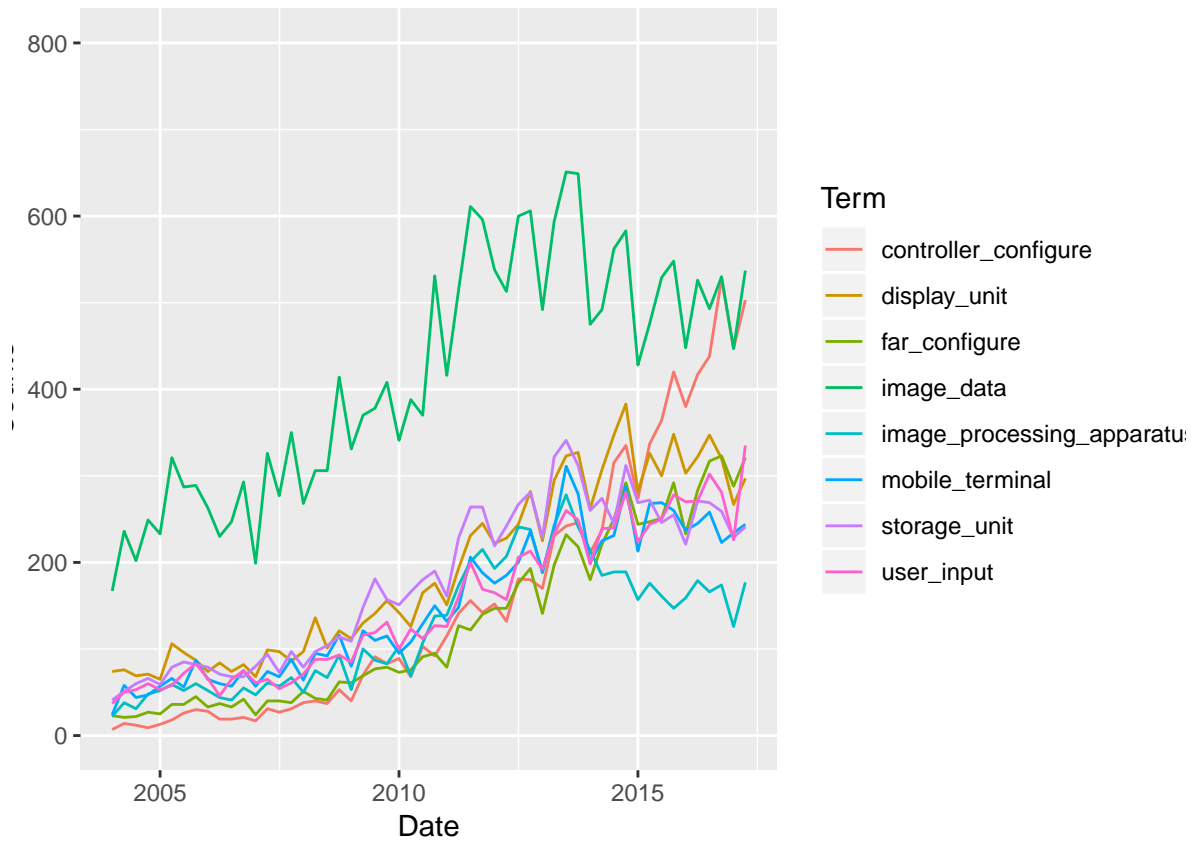


Table 1: Emerging Terms

Term	σ_ϵ	σ_ν	σ_η	δ	E(sum 1st der)	E(sum 1st der/sd)
controller configure	20.26621	1.4538156	0.0000002	1.0000000	511.4290	25.235554
computing device	22.46527	3.0752379	0.0000095	0.9737273	509.1460	22.663692
compute device	22.77279	2.9092781	0.0000001	0.9706548	464.9410	20.416514
unit configure	30.61585	3.4494549	0.0000044	0.9707211	572.5834	18.702186
far configure	16.86448	0.8969558	0.0000147	1.0000000	300.0263	17.790421
user equipment	38.21802	3.6276779	0.0000034	0.9767679	620.3227	16.231156
mobile device	48.58455	5.1622083	0.0000534	0.9675438	778.1726	16.016875
user device	21.44104	1.8749874	0.0000002	0.9813091	343.3377	16.013114
computer program product	41.39129	2.8362622	0.0000024	0.9798947	658.9539	15.920109
user interface	33.97358	0.0000004	14.7416176	0.9945274	492.6912	14.502186
client device	23.50435	1.8719932	0.0000008	0.9742848	336.8822	14.332760
virtual machine	16.62287	3.1675641	0.0000001	0.9420660	226.5827	13.630783
determine base	32.57034	1.2181121	0.0000011	0.9958347	413.7848	12.704346
user input	19.74913	0.8249383	0.0000002	0.9945268	246.9970	12.506727
information associate	27.75152	1.6463558	0.0000013	0.9706949	297.8188	10.731620
wireless communication system	28.47157	2.9646210	0.0000026	0.9442072	289.4167	10.165115
display unit	22.43741	2.0637547	0.0000496	0.9562891	224.0652	9.986231
communication device	29.40785	0.0000000	13.3705902	0.9903924	287.5876	9.779279
mobile terminal	21.44692	1.7022585	0.0000076	0.9464027	191.8905	8.947228
following step	23.30157	2.7713324	0.0001550	0.9267517	205.9476	8.838356
storage unit	21.48599	2.4570829	0.0000122	0.9290738	181.7098	8.457130
base station	48.62254	6.0875813	0.0000171	0.9170816	410.4200	8.440942
pharmaceutical composition	48.86269	3.0236779	0.0001709	0.9472690	381.2242	7.801949
image processing apparatus	16.46404	4.8710888	0.0000033	0.8095802	121.1640	7.359309
image data	50.01925	5.6798145	0.0000001	0.8585361	274.1318	5.480526

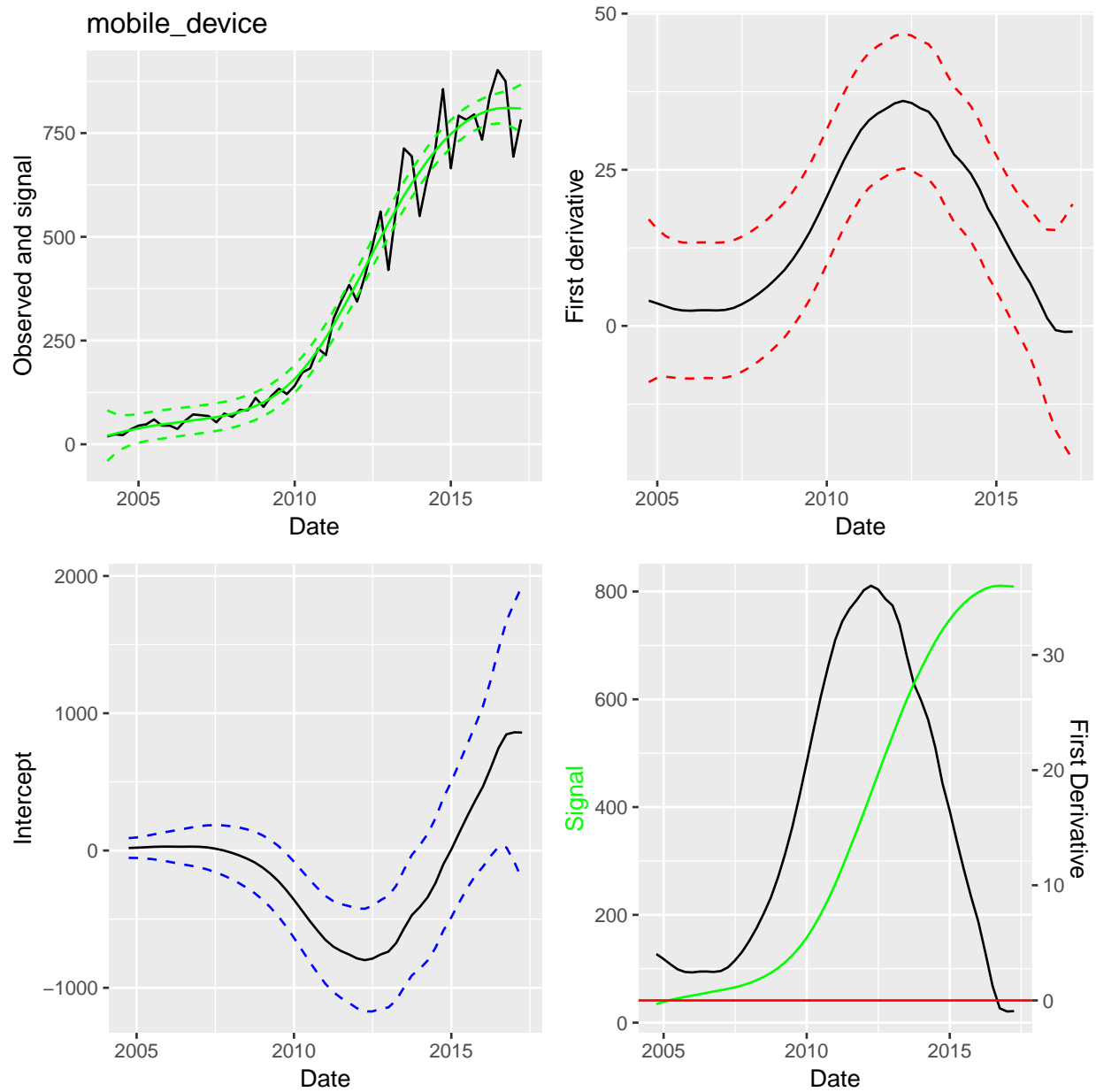


Figure 1:

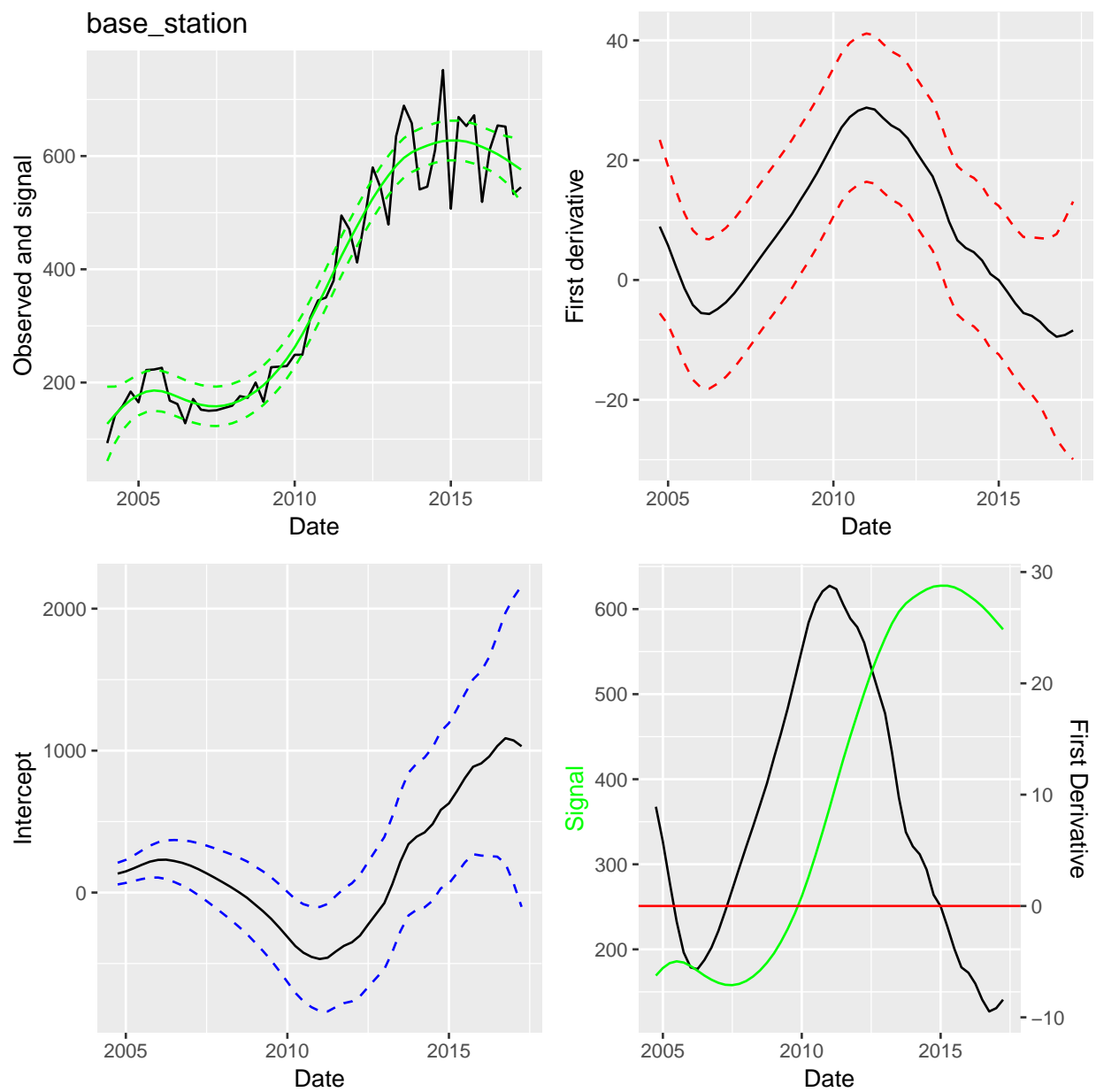


Figure 2:

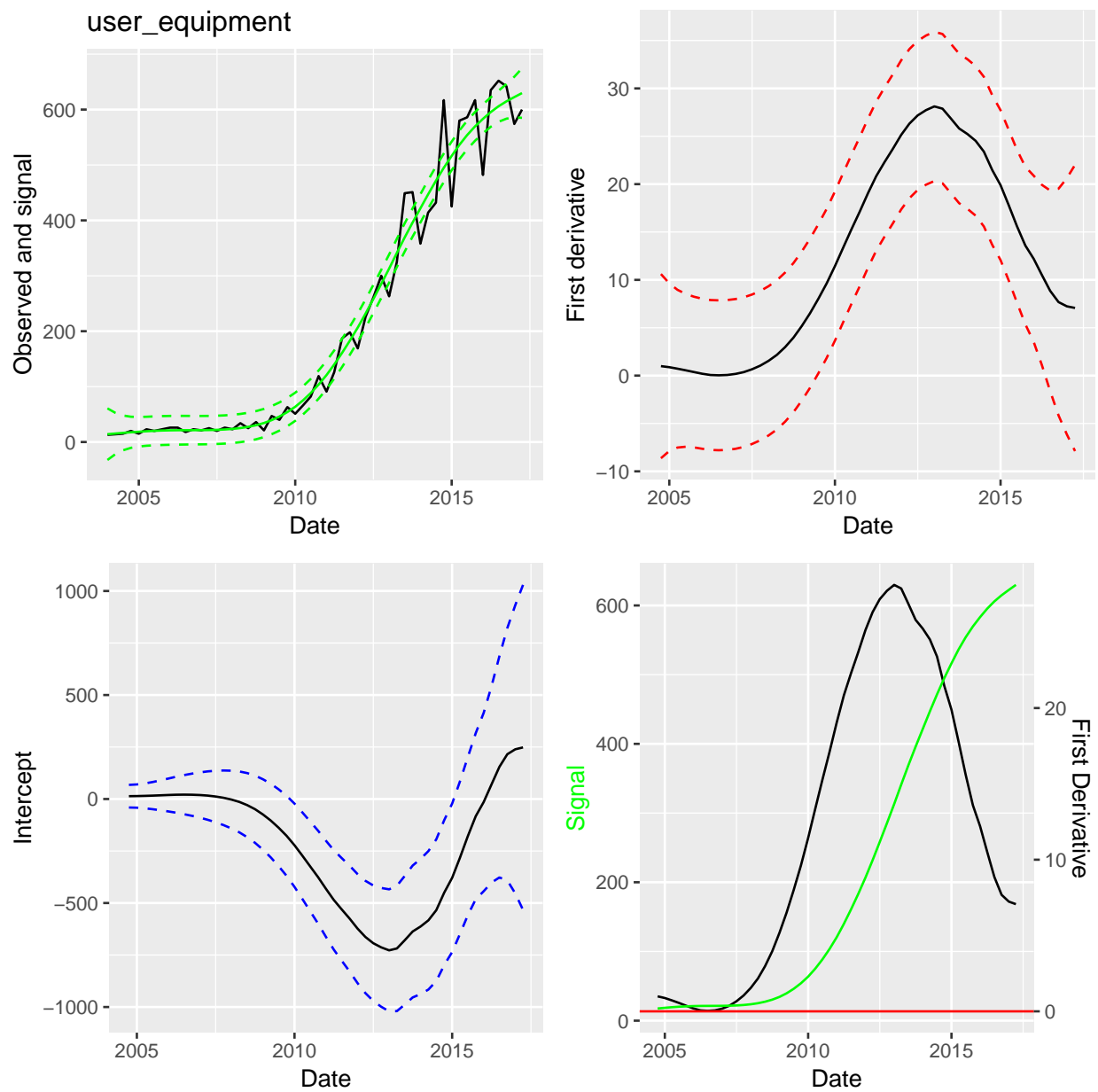


Figure 3:

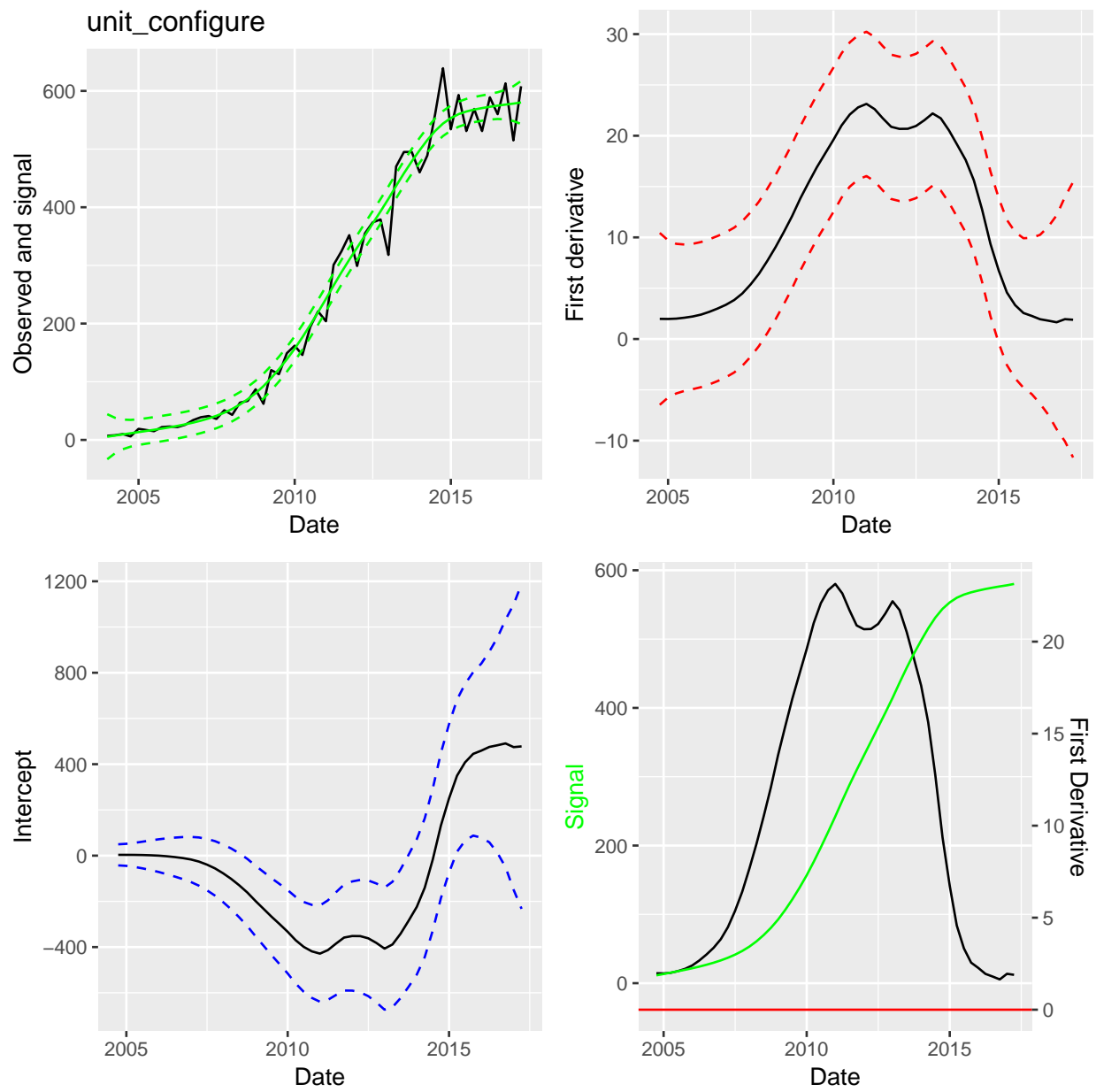


Figure 4:

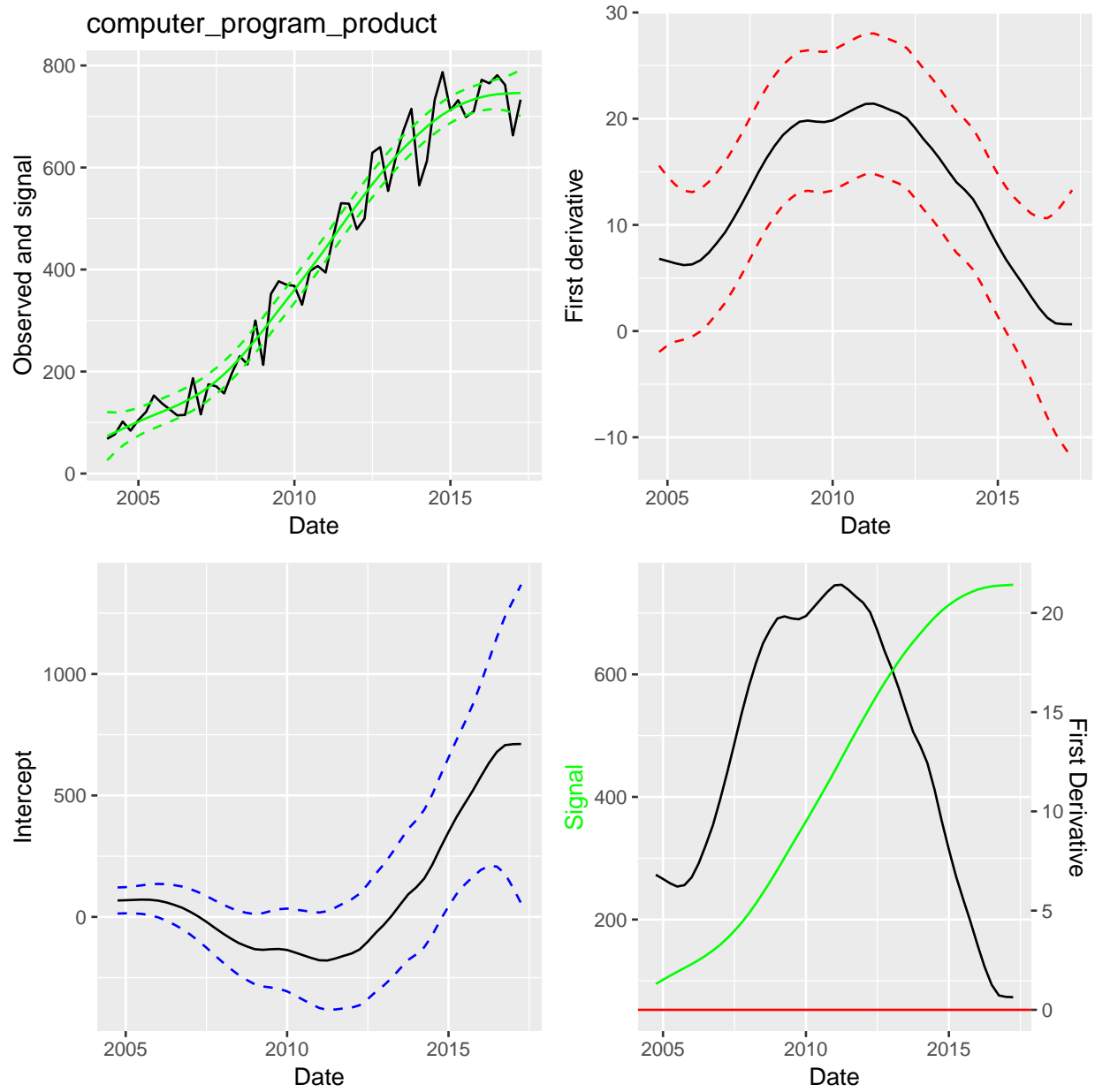


Figure 5:

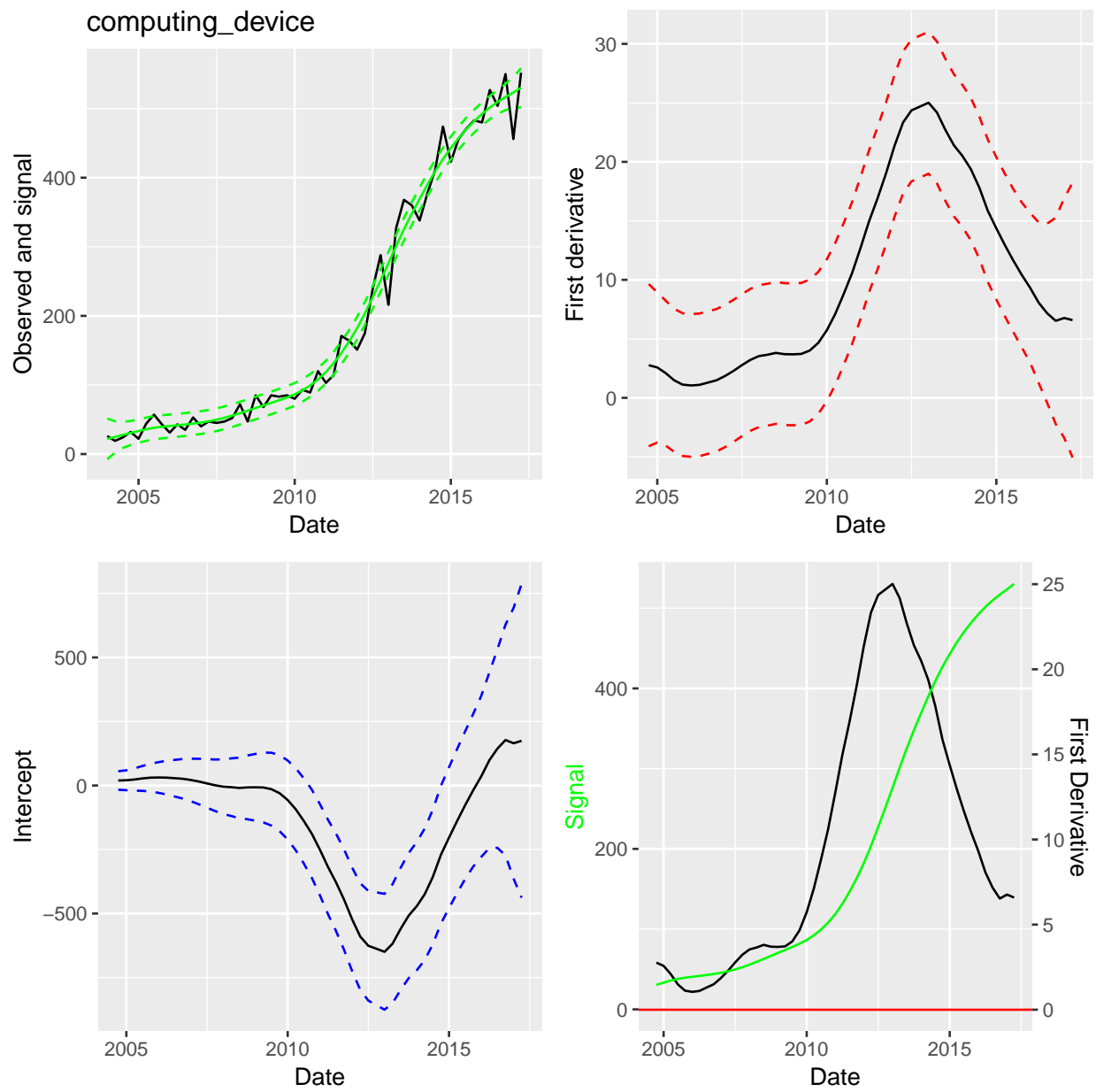


Figure 6:

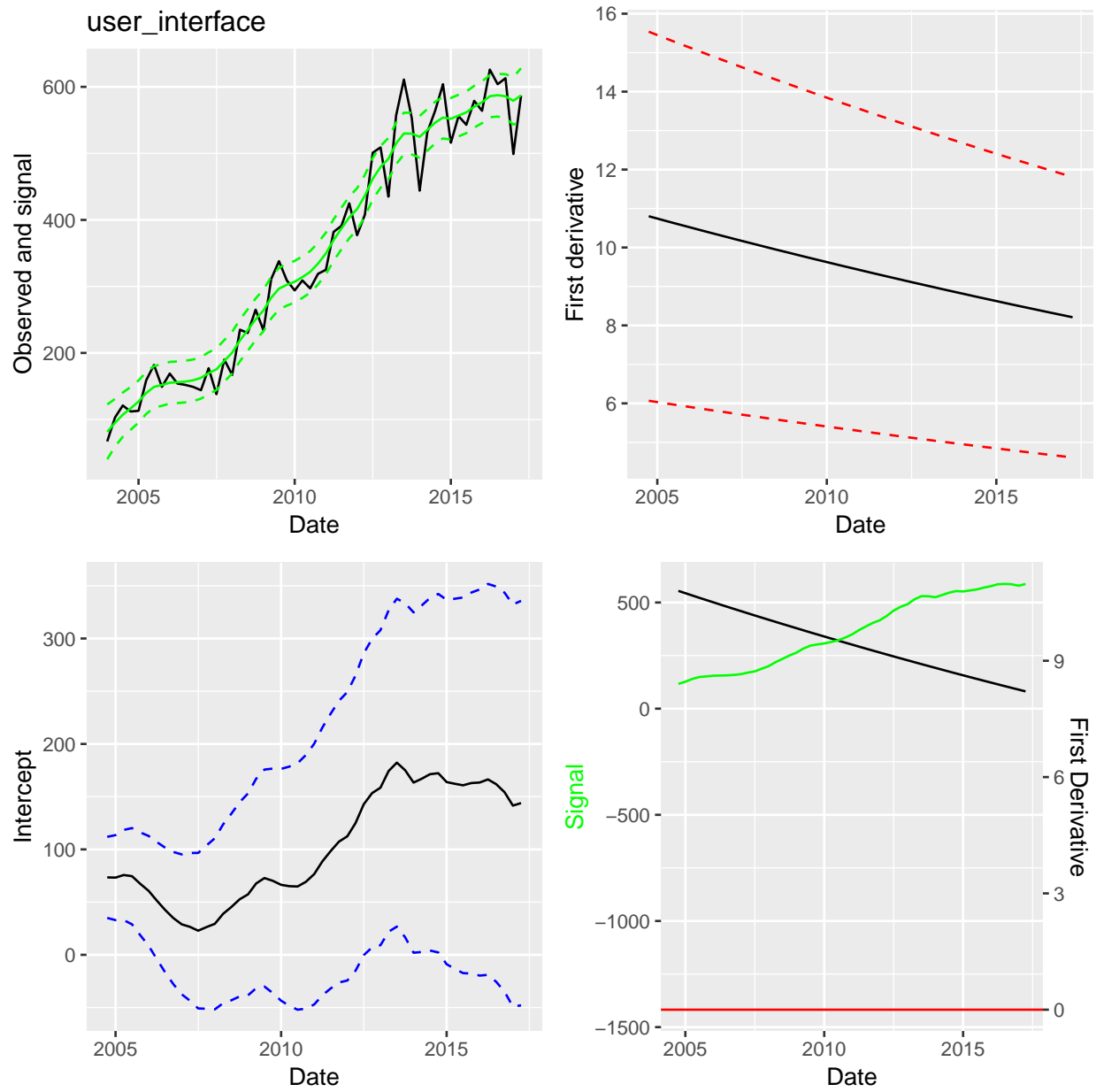


Figure 7:

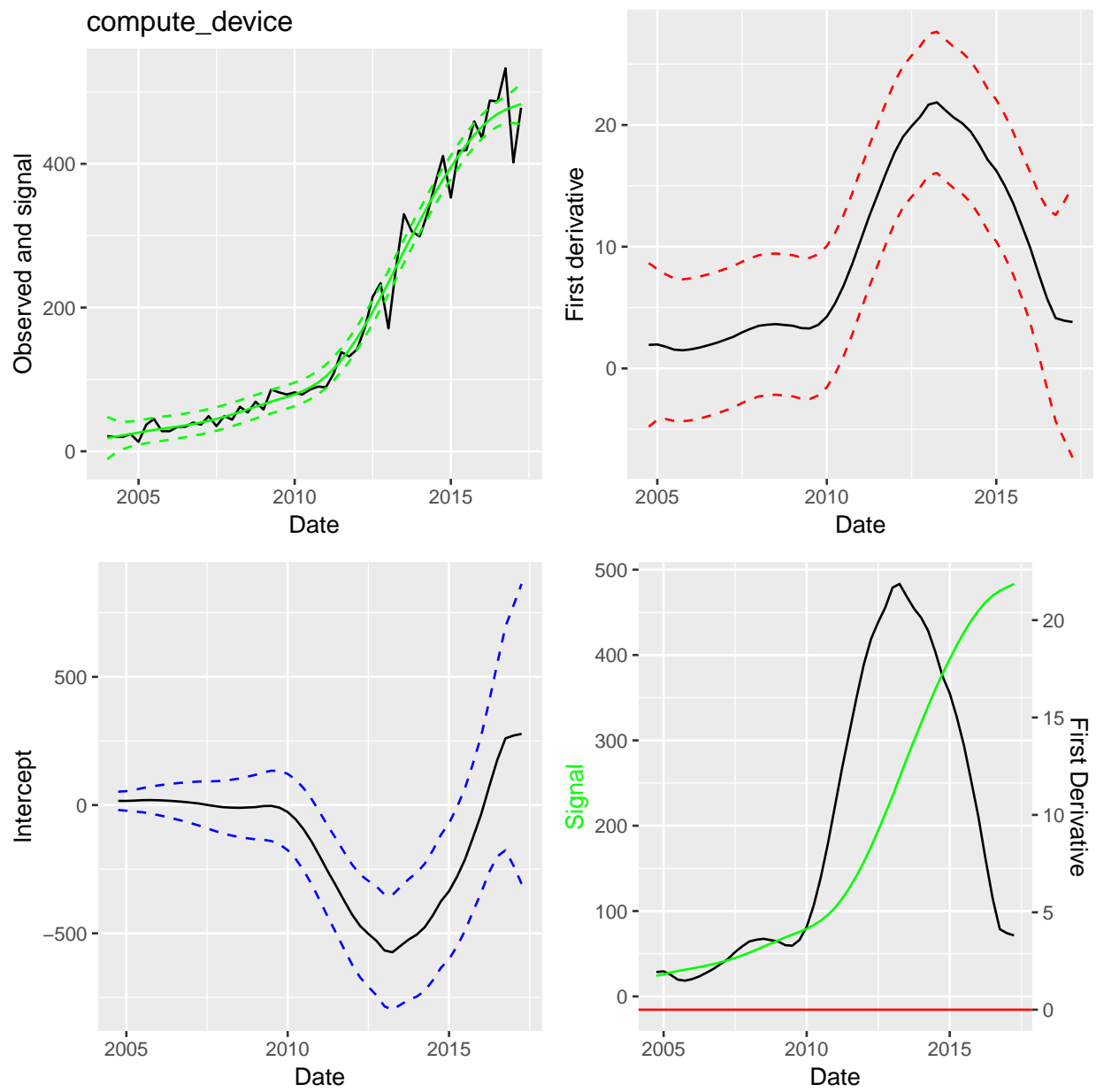


Figure 8:

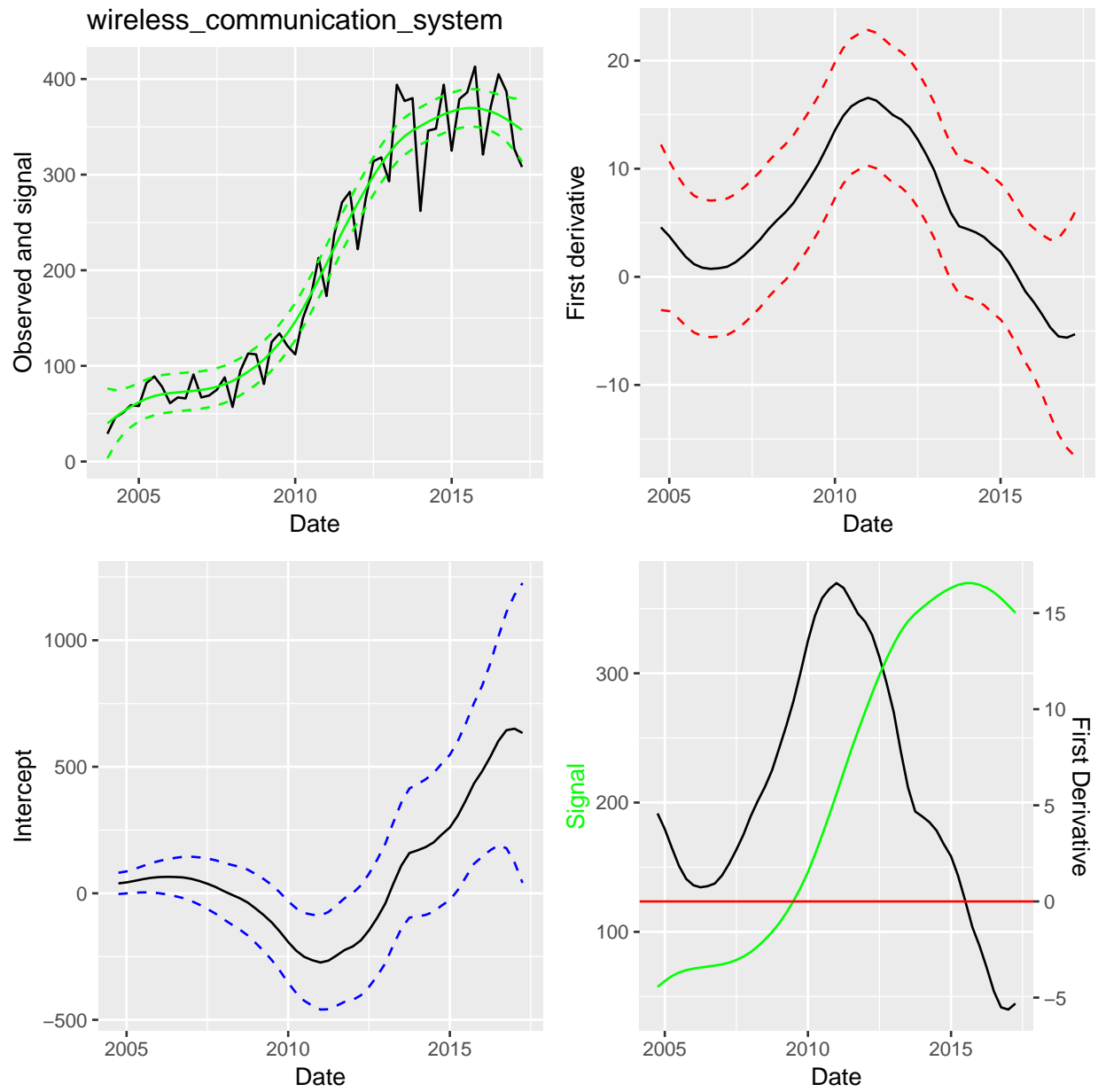


Figure 9:

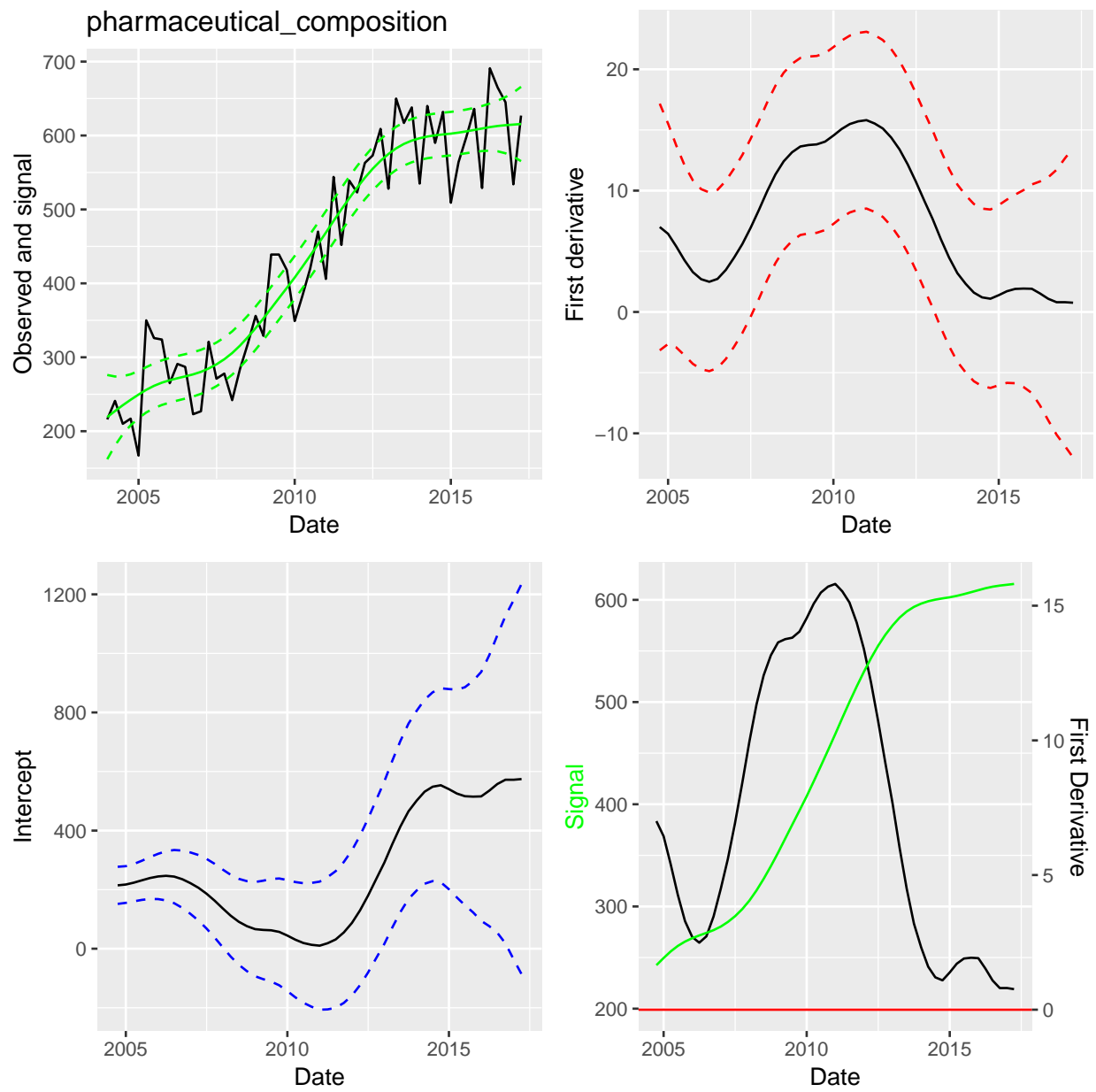


Figure 10:

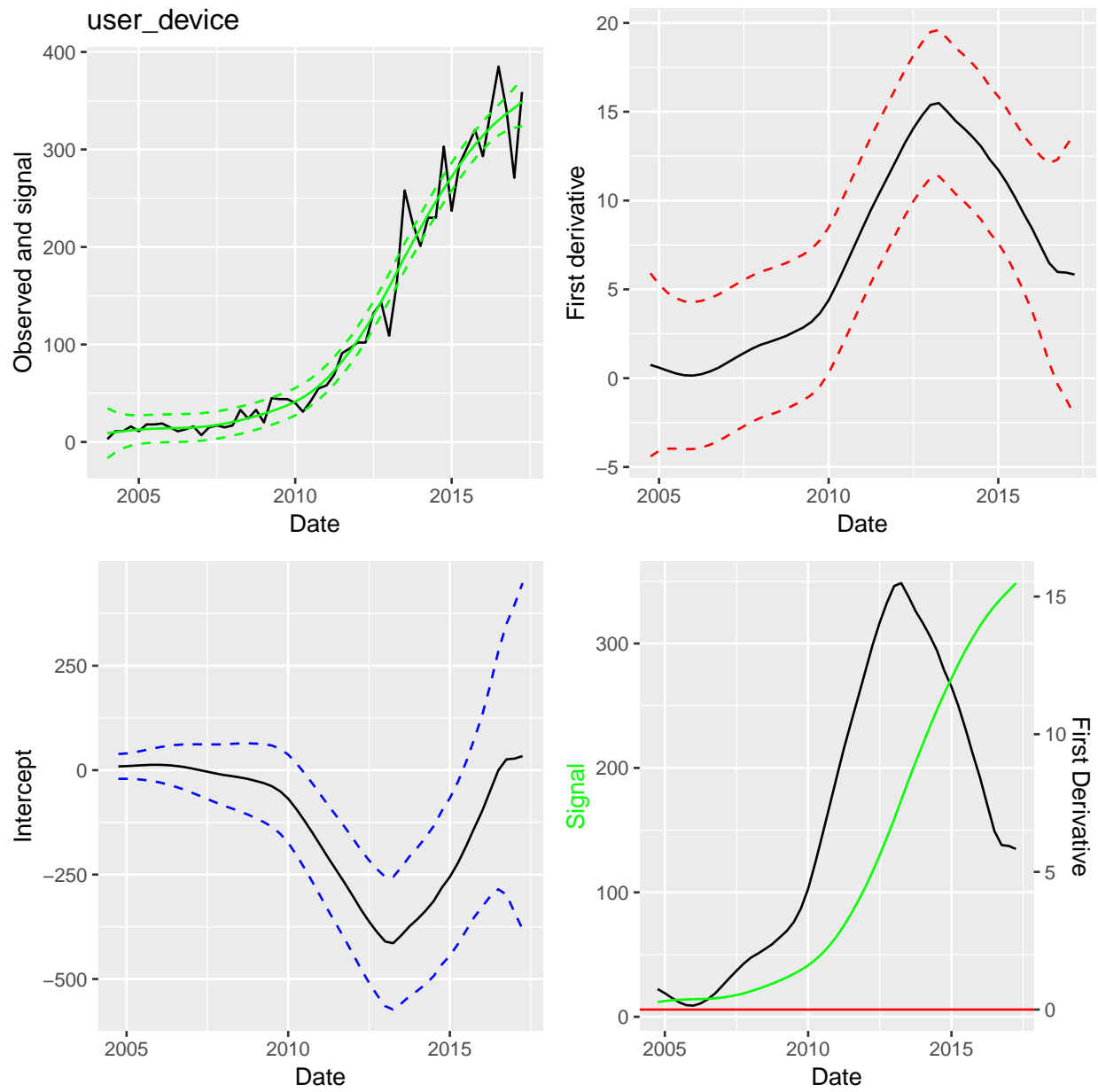


Figure 11:

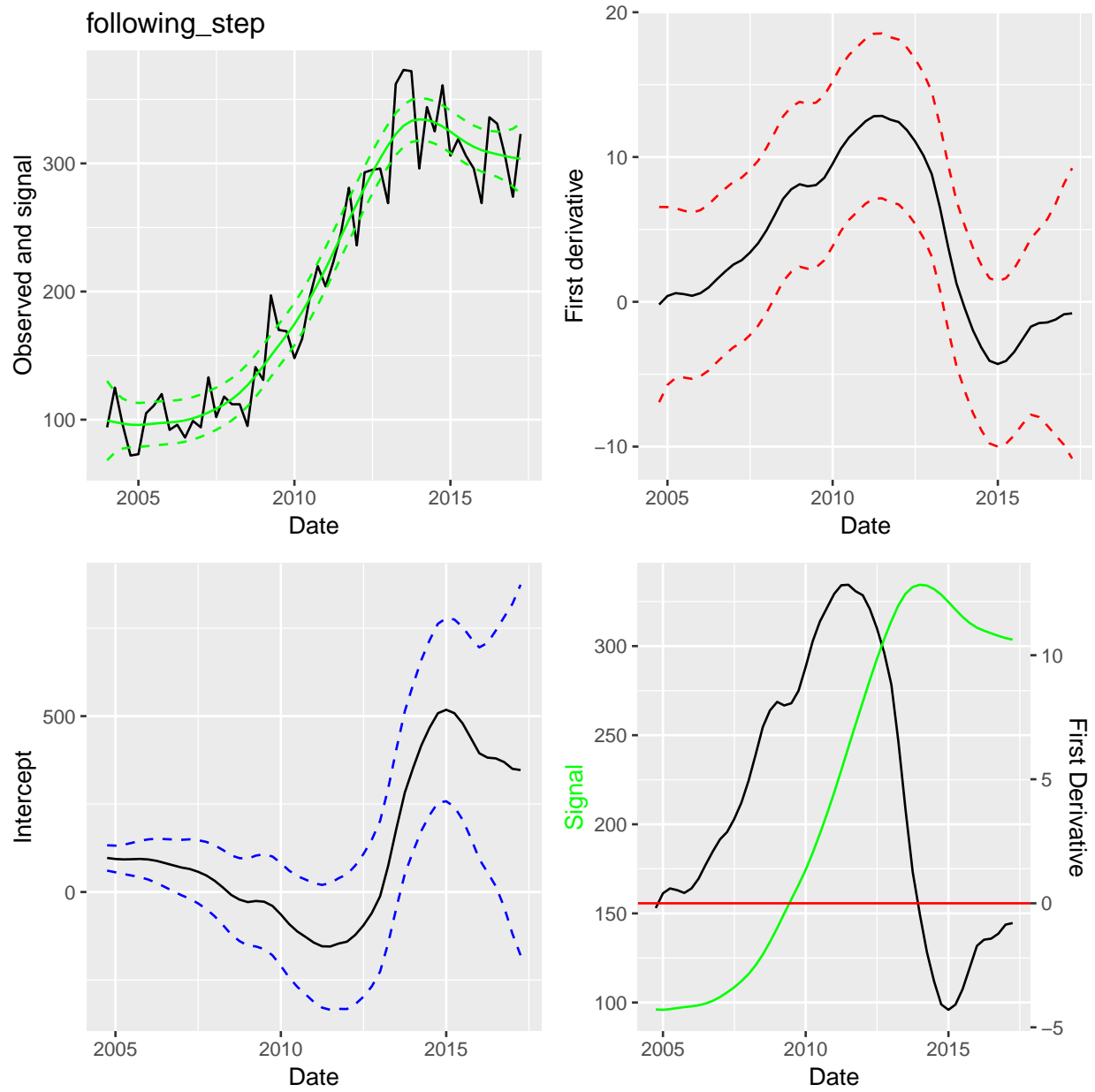


Figure 12:

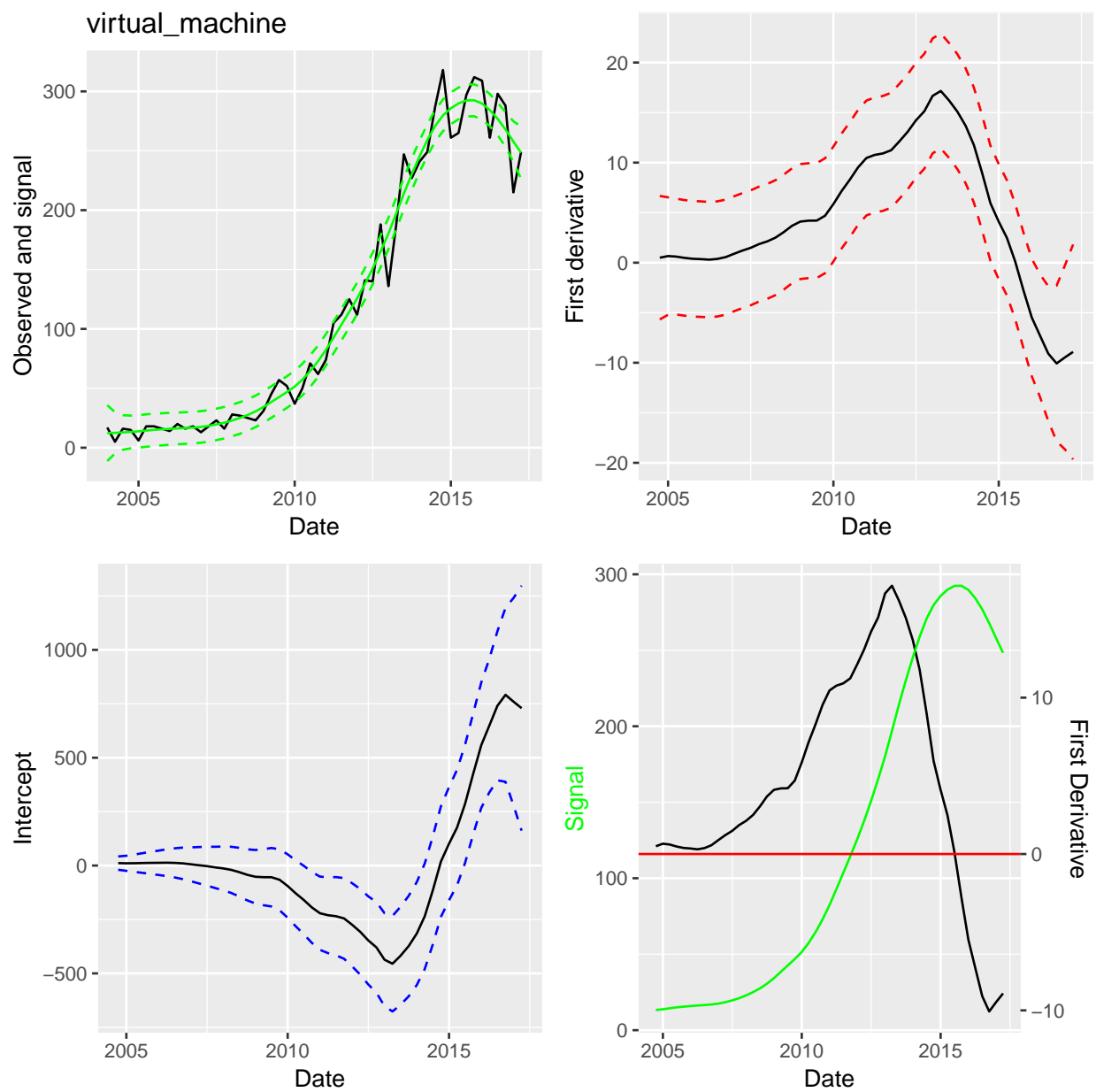


Figure 13:

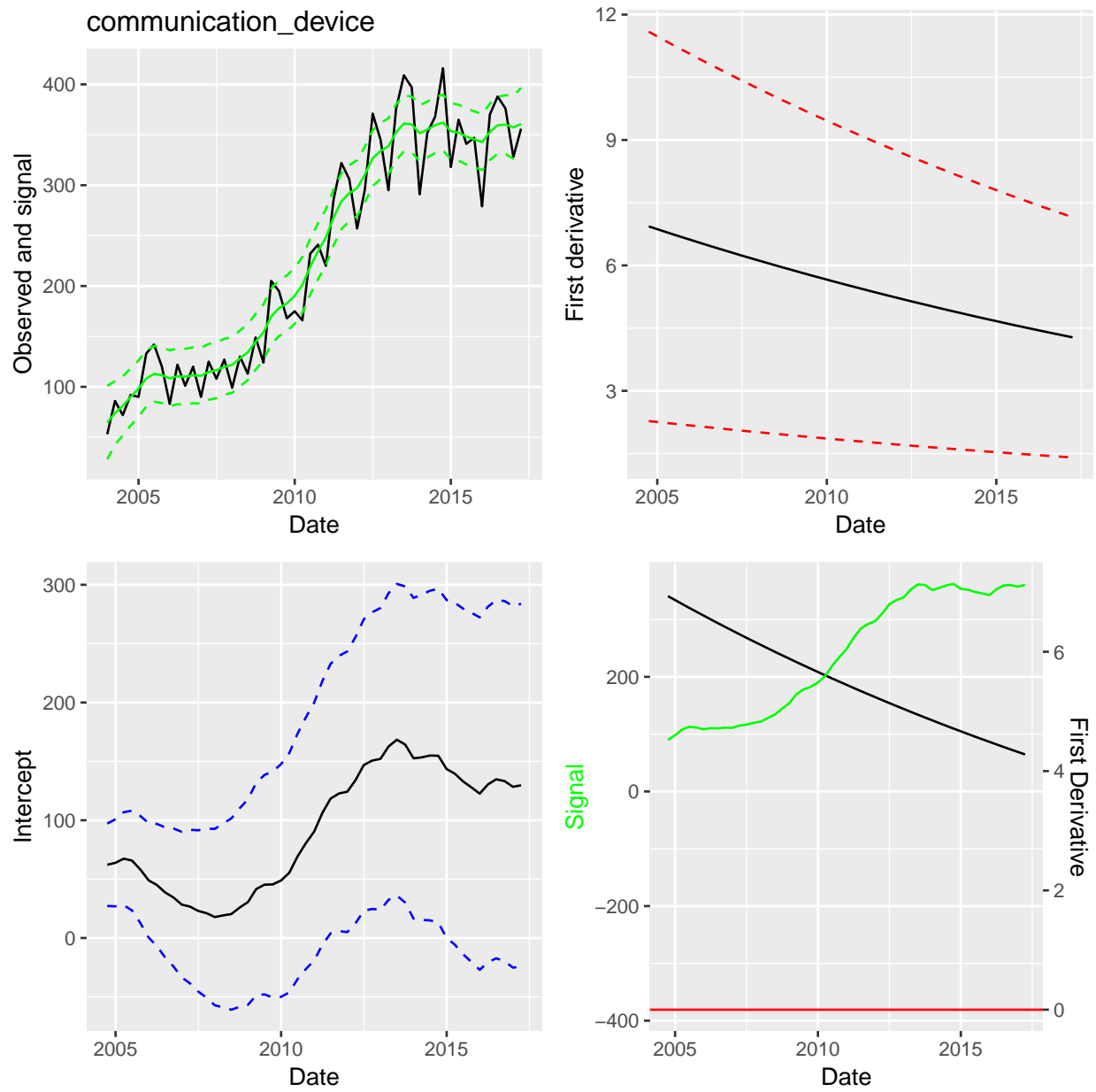


Figure 14:

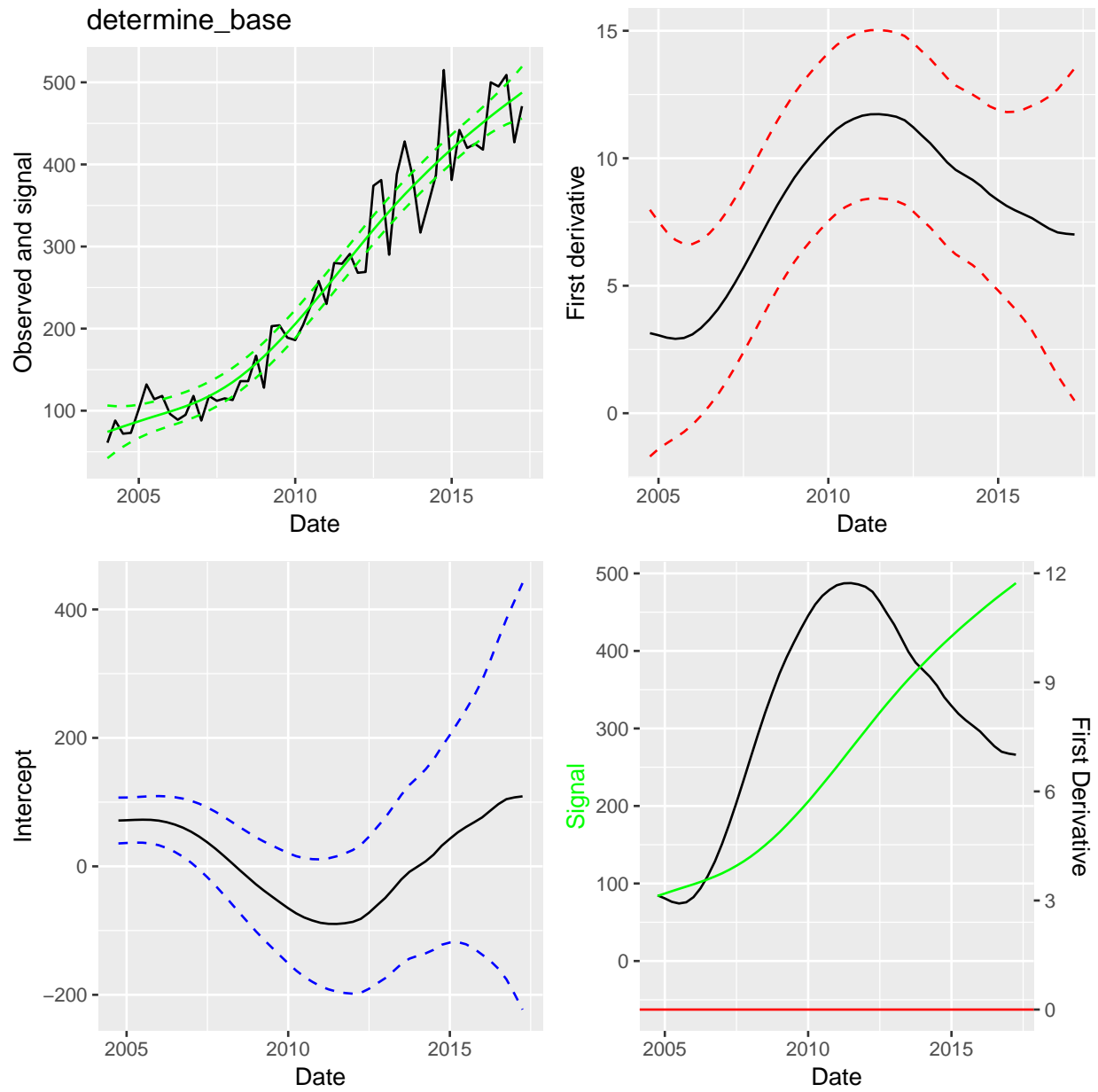


Figure 15:

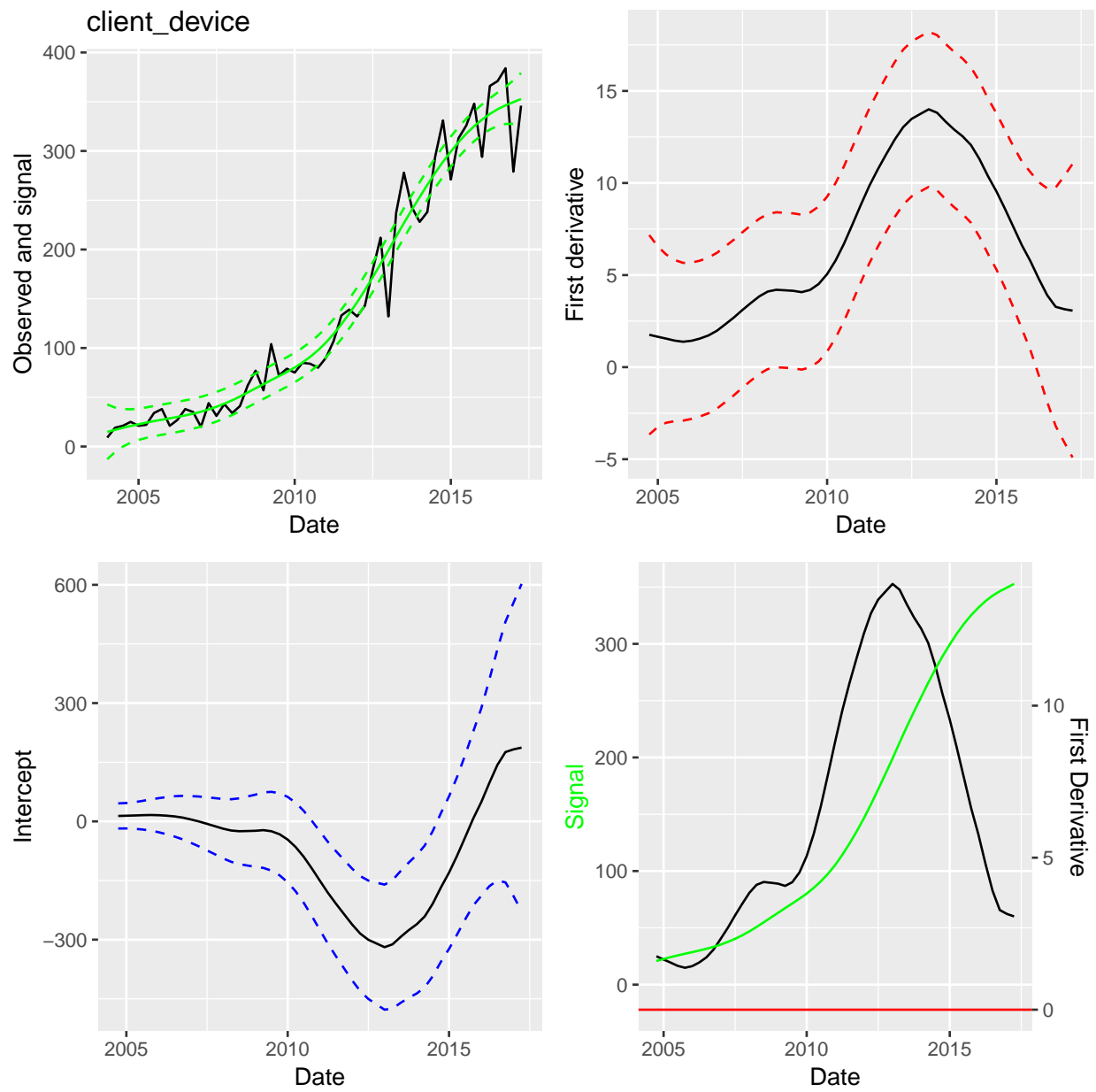


Figure 16:

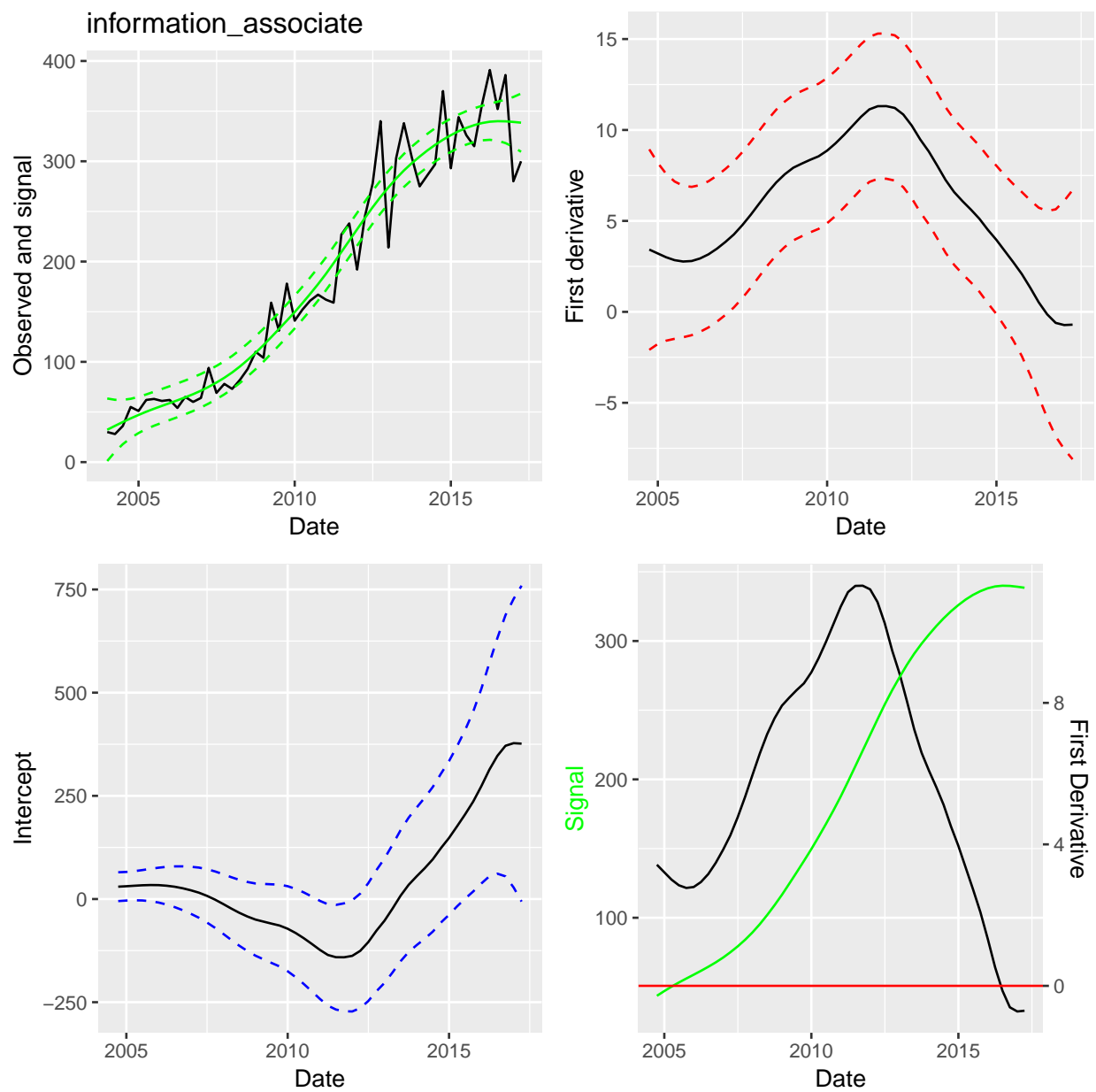


Figure 17:

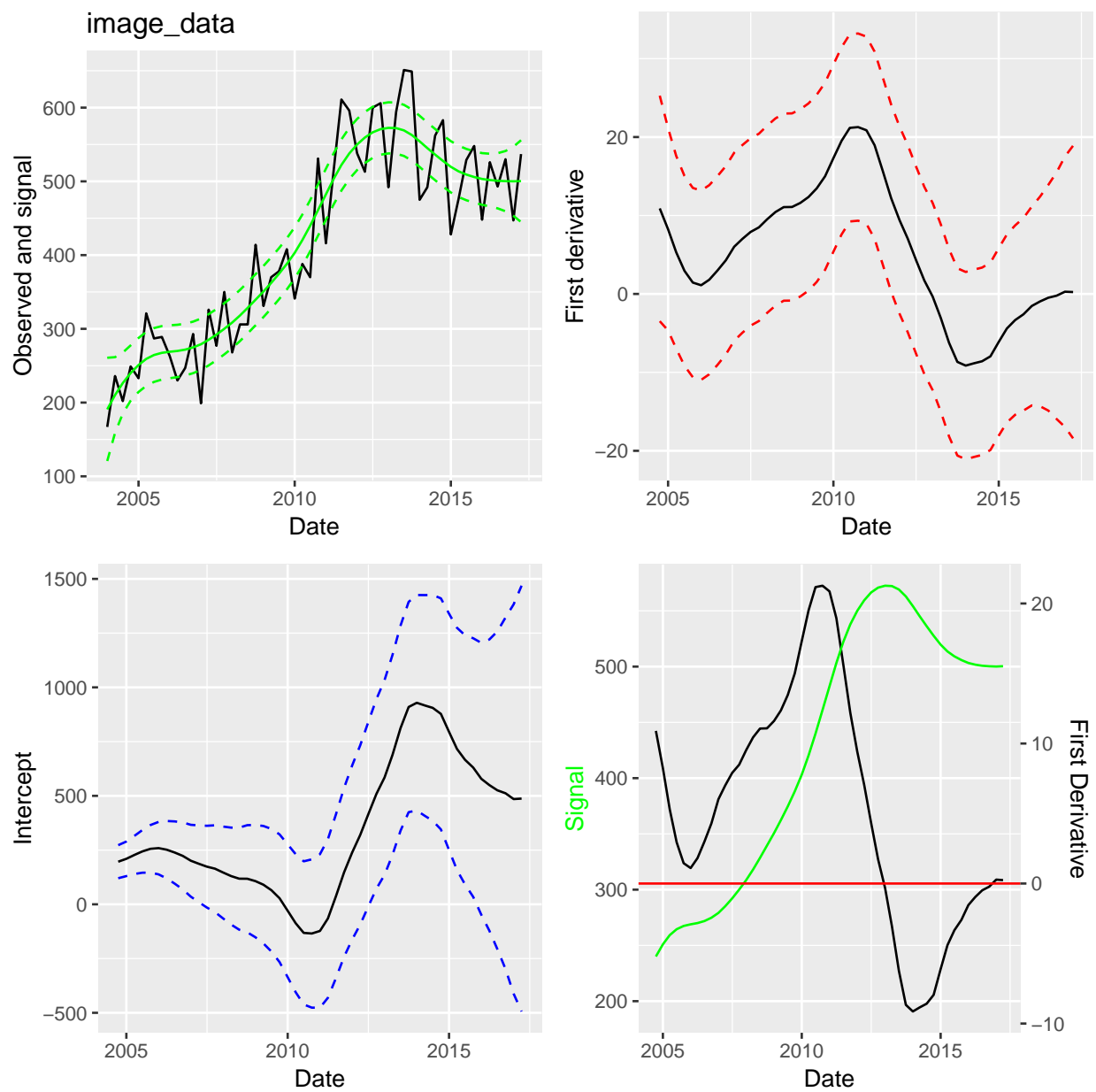


Figure 18:

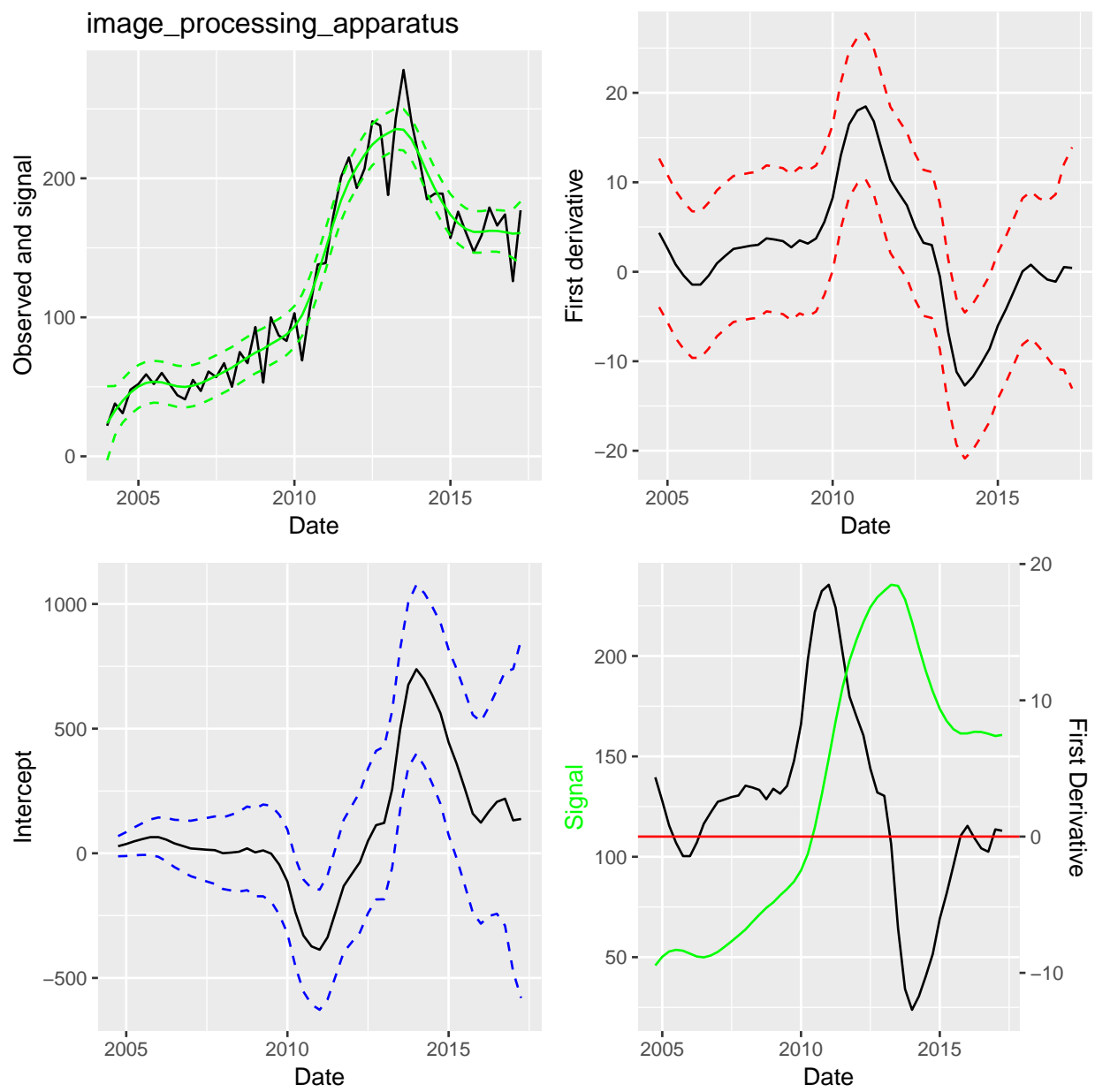


Figure 19:

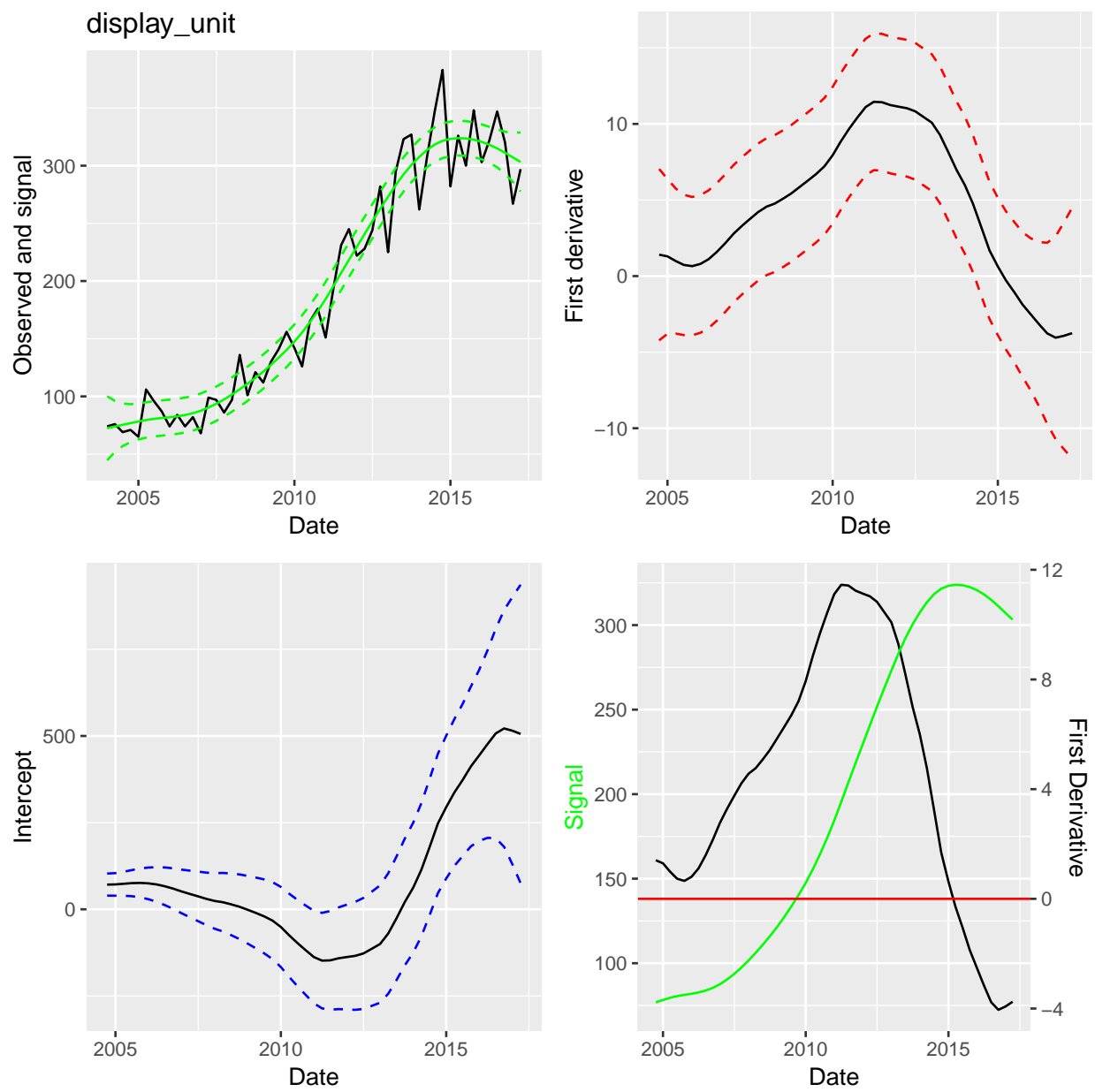


Figure 20:

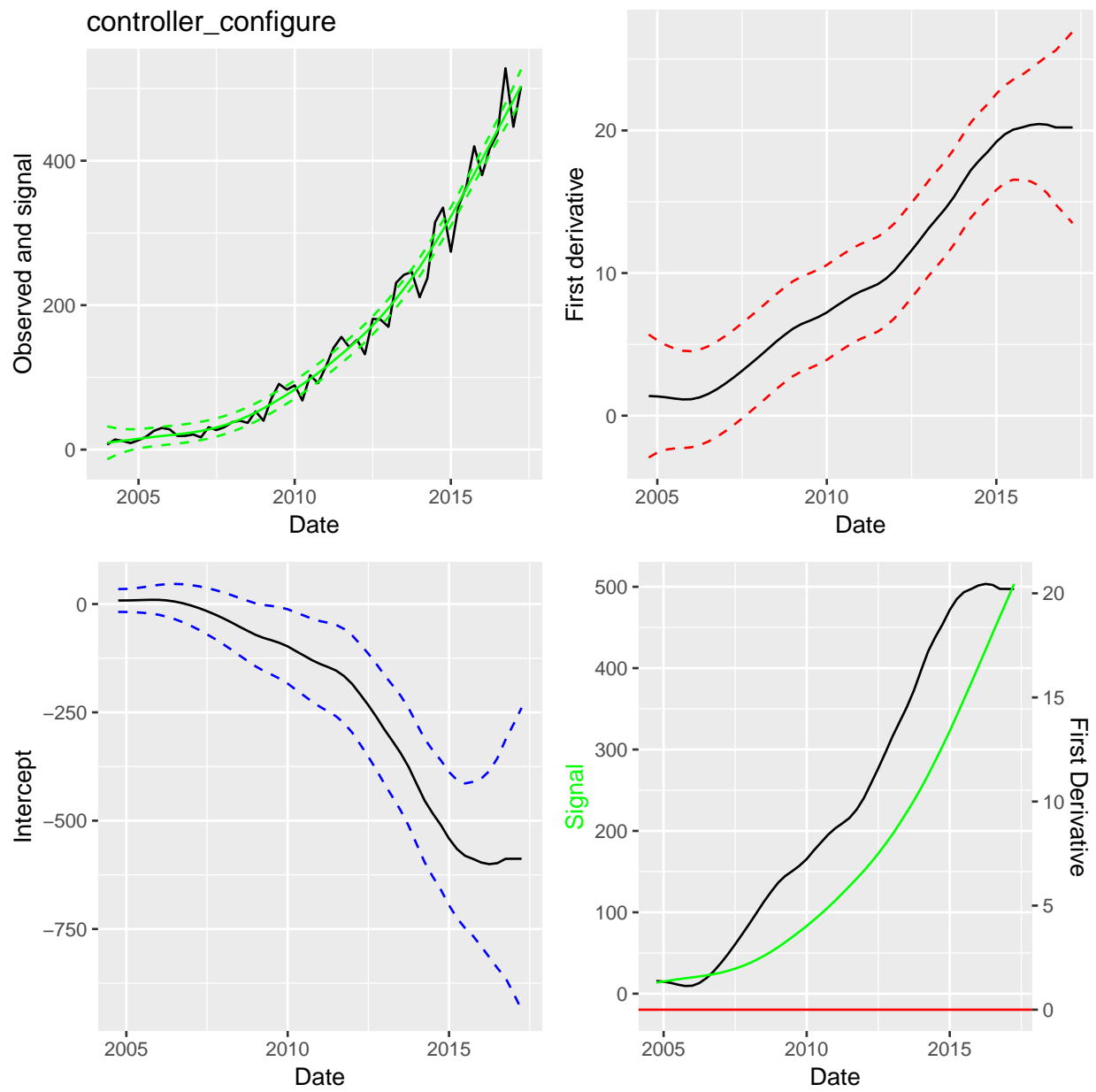


Figure 21:

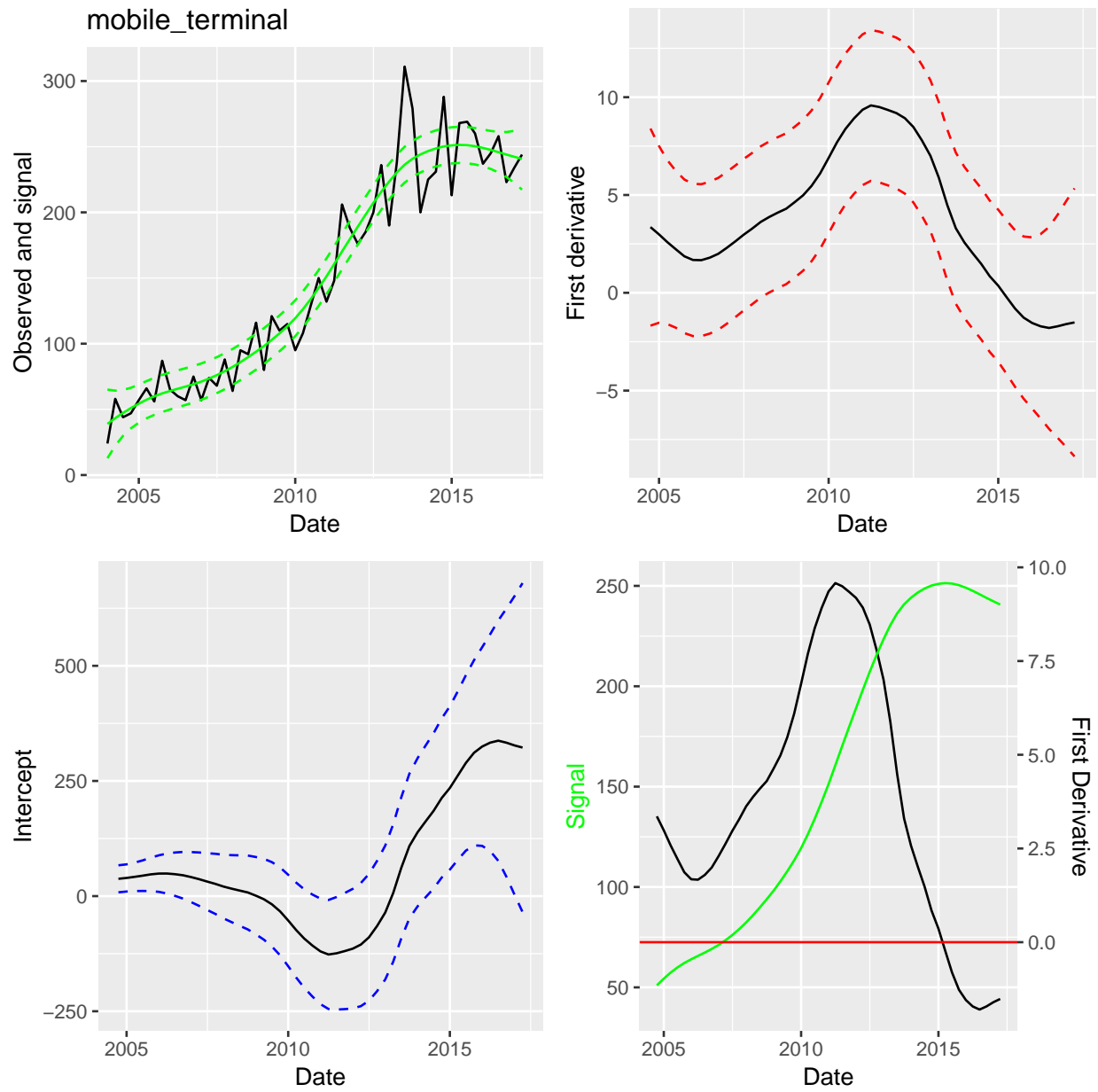


Figure 22:

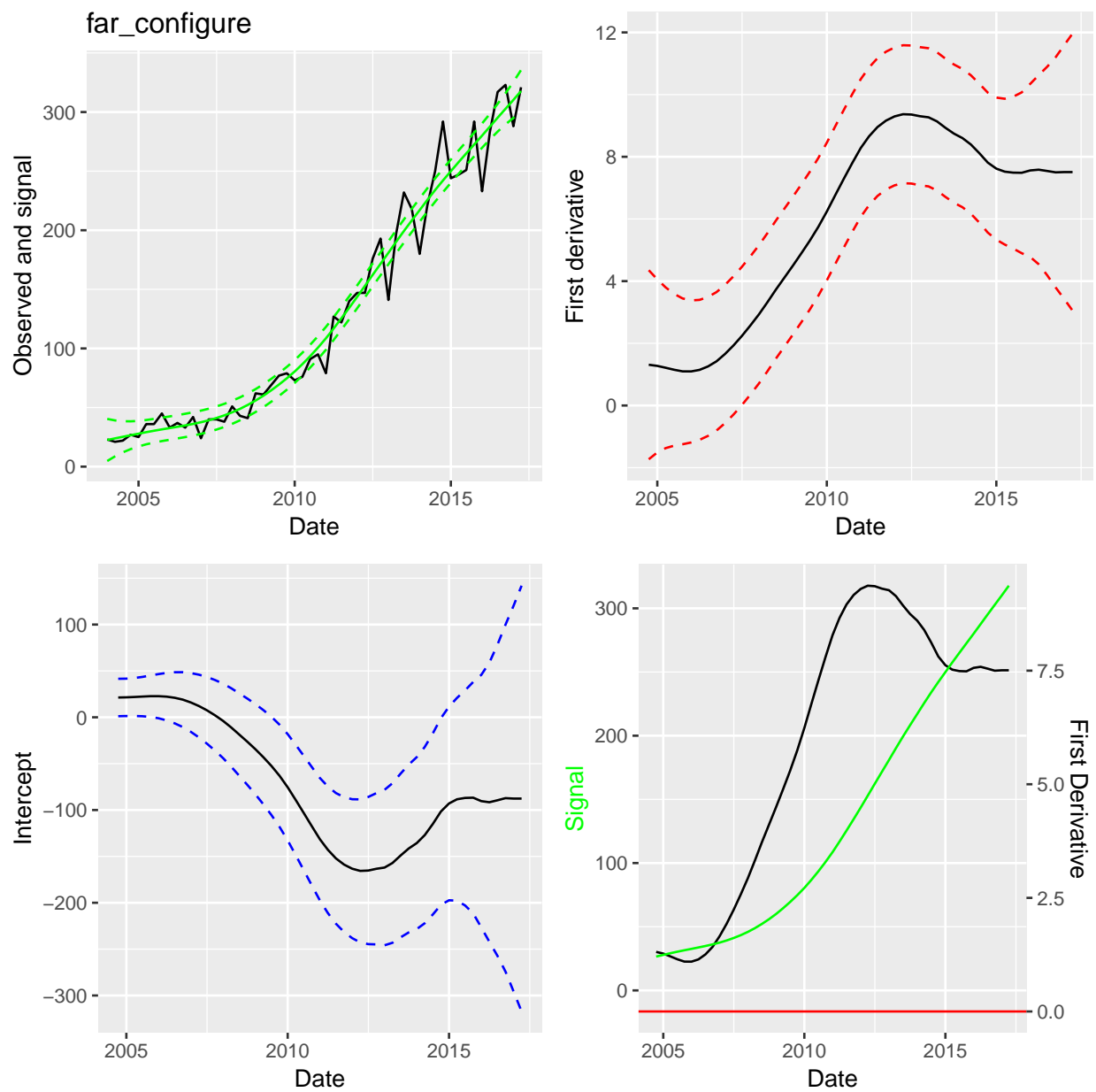


Figure 23:

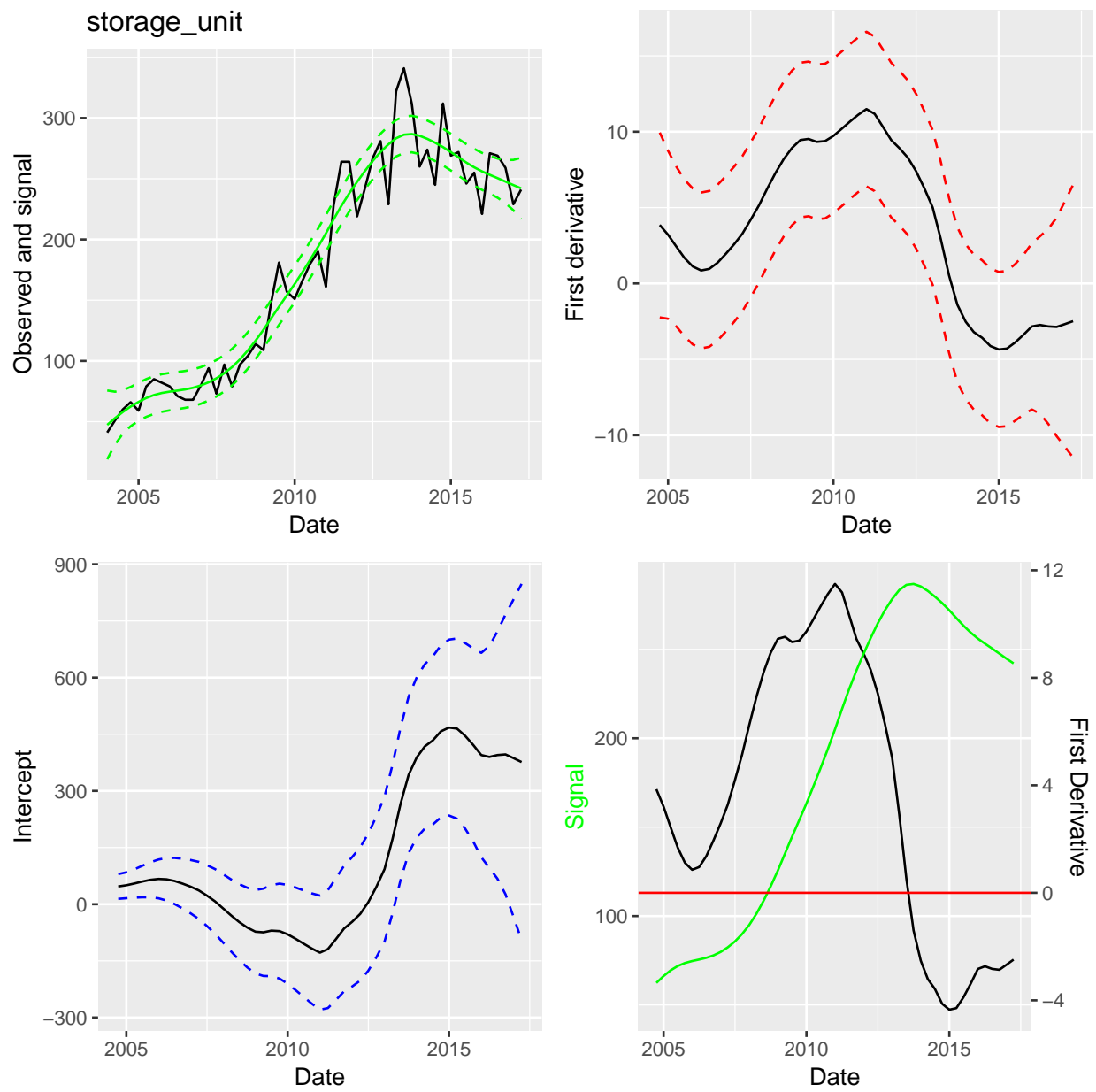


Figure 24:

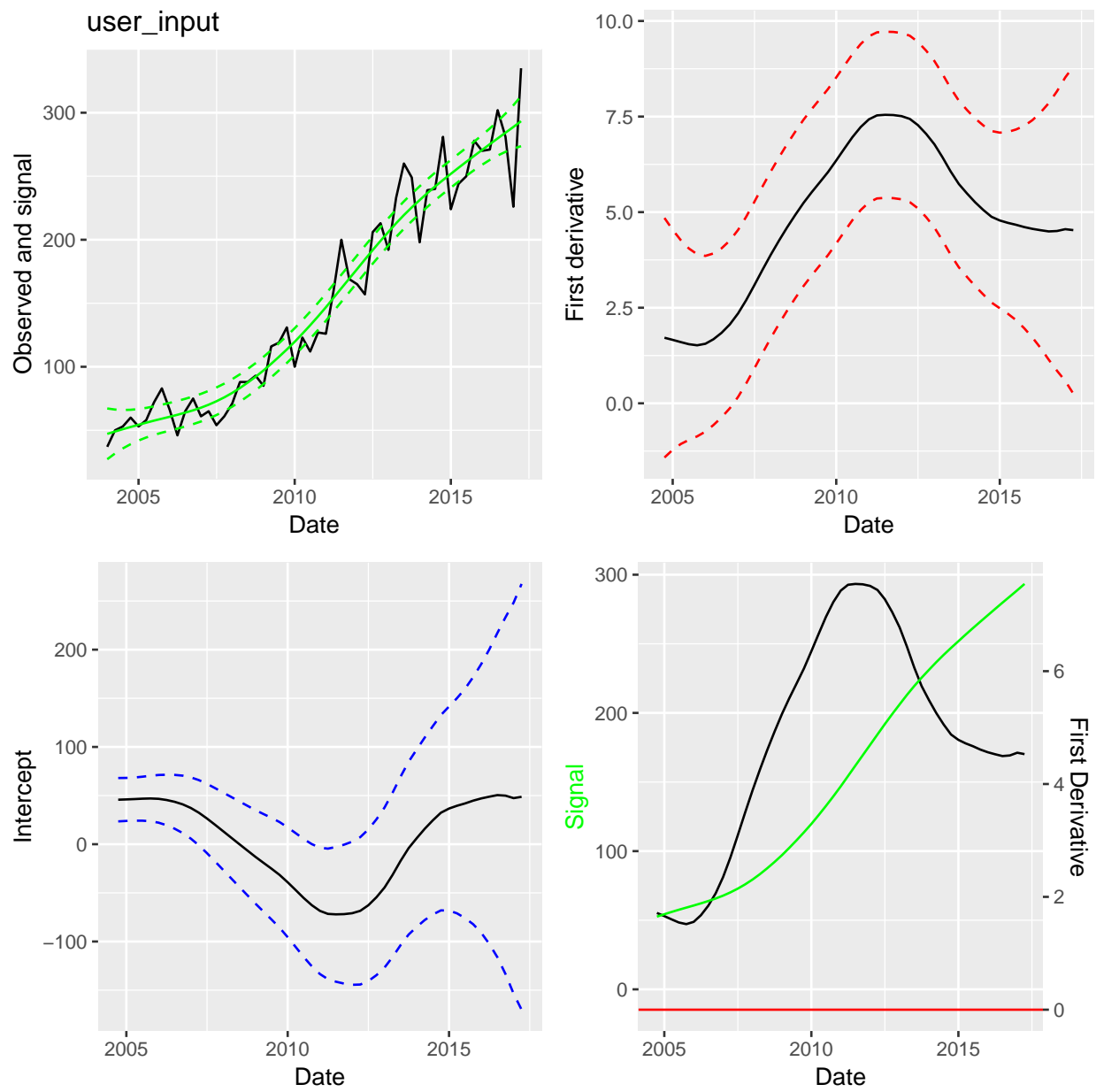


Figure 25:

Terms classified as stationary by Porter

The trends of some terms in this group are clearly increasing functions.

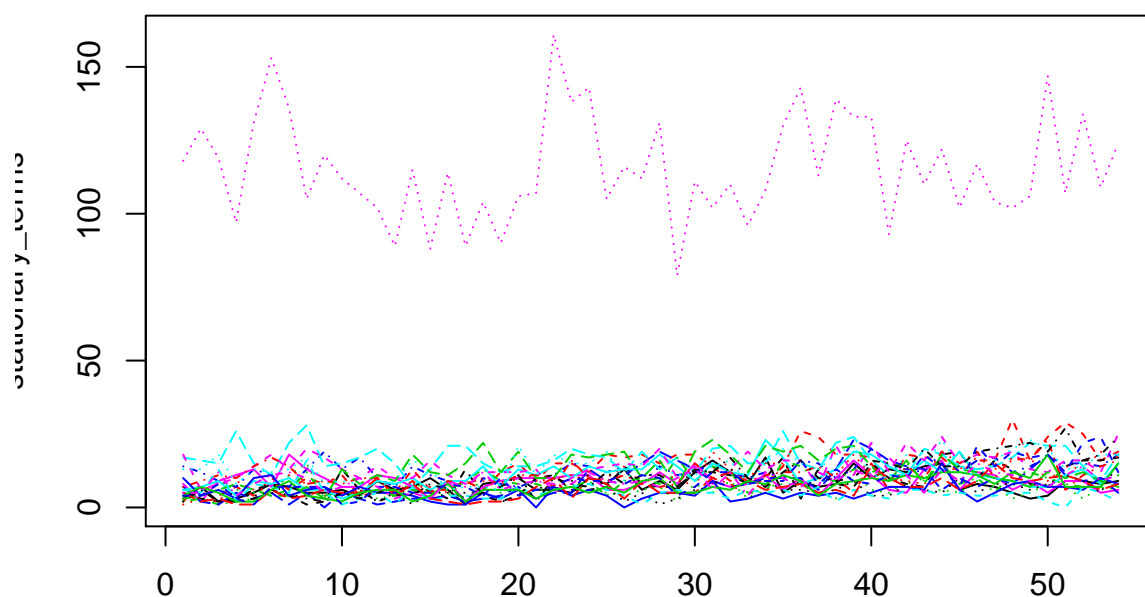
The terms “expression level”, “axially displaceable”, “ovarian cancer”, “wavelength conversion”, “virtual memory”, “cylindrical housing”, “damp element”, “mass flow”, “structure arrange”, “processor module”, “sequentially apply” are clearly emergent terms. The trends are linear with positive slope.

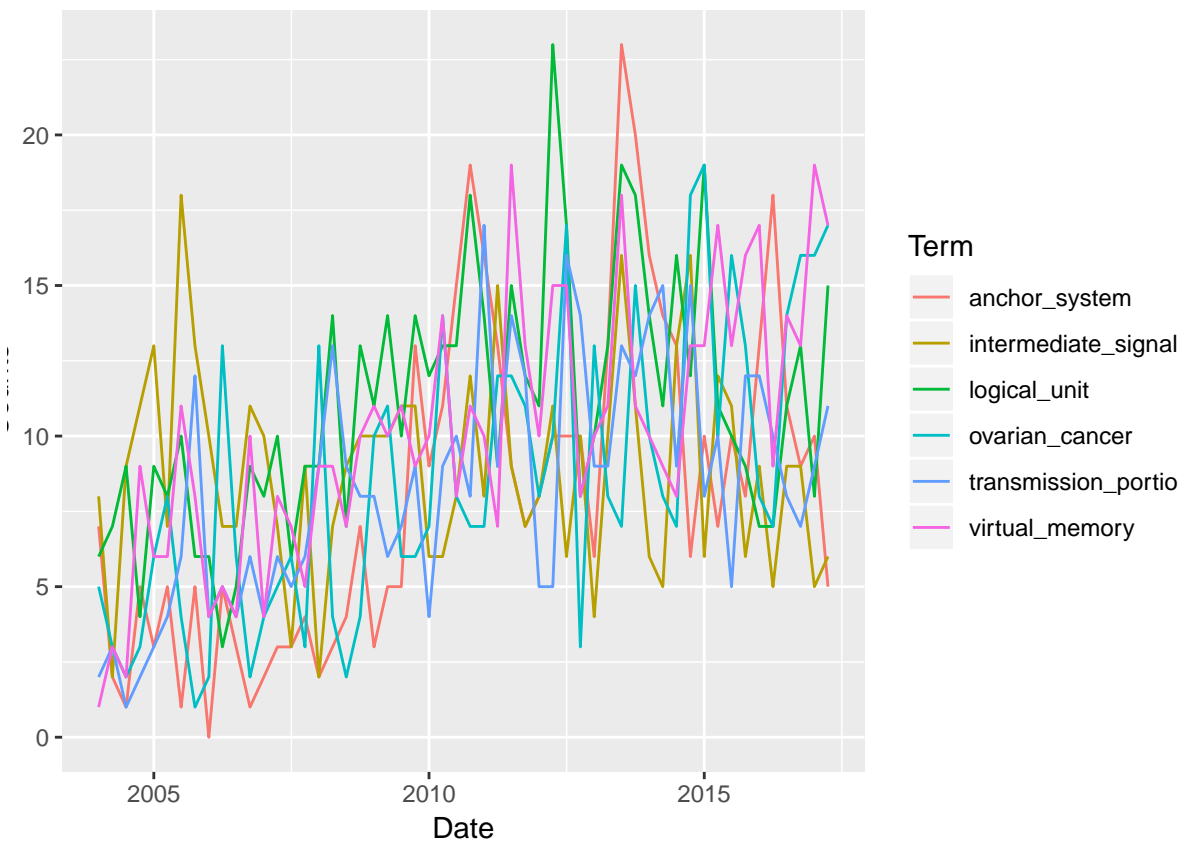
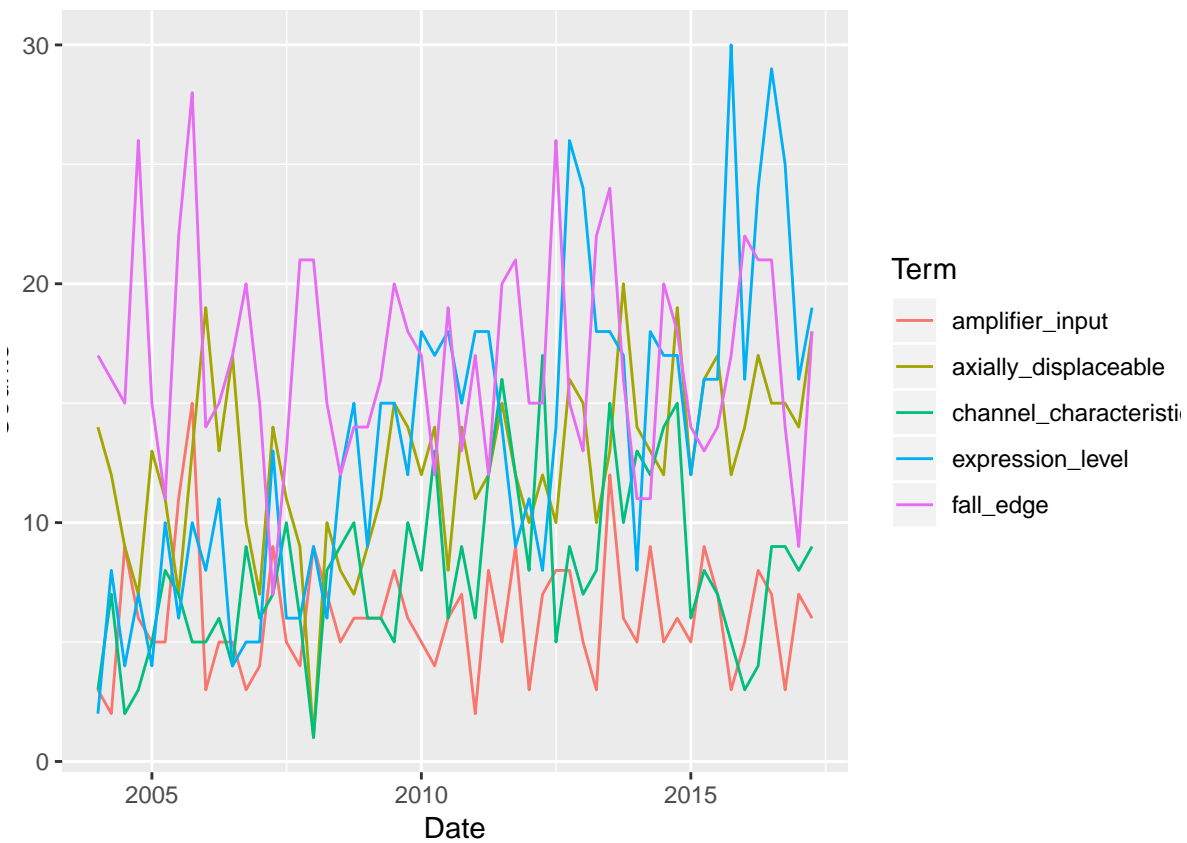
Some terms in this group have a U- or inverted-U-shaped trend. Therefore it’s difficult to classify the term. To avoid this, I suggest that only, say, 5 years are used for the index. In this case we are using 12 years.

The only terms we can definitely categorise as stationary are “amplifier input”, “fall-edge”, “intermediate signal” and “prior art”.

Some terms were tricky to fit (expression level, axially displaceable and anchor system). I still can’t find the optimal parameters of “anchor system”. This is in my TO-DO list.

Generally the issue of fitting the parameters when the counts are low is difficult.





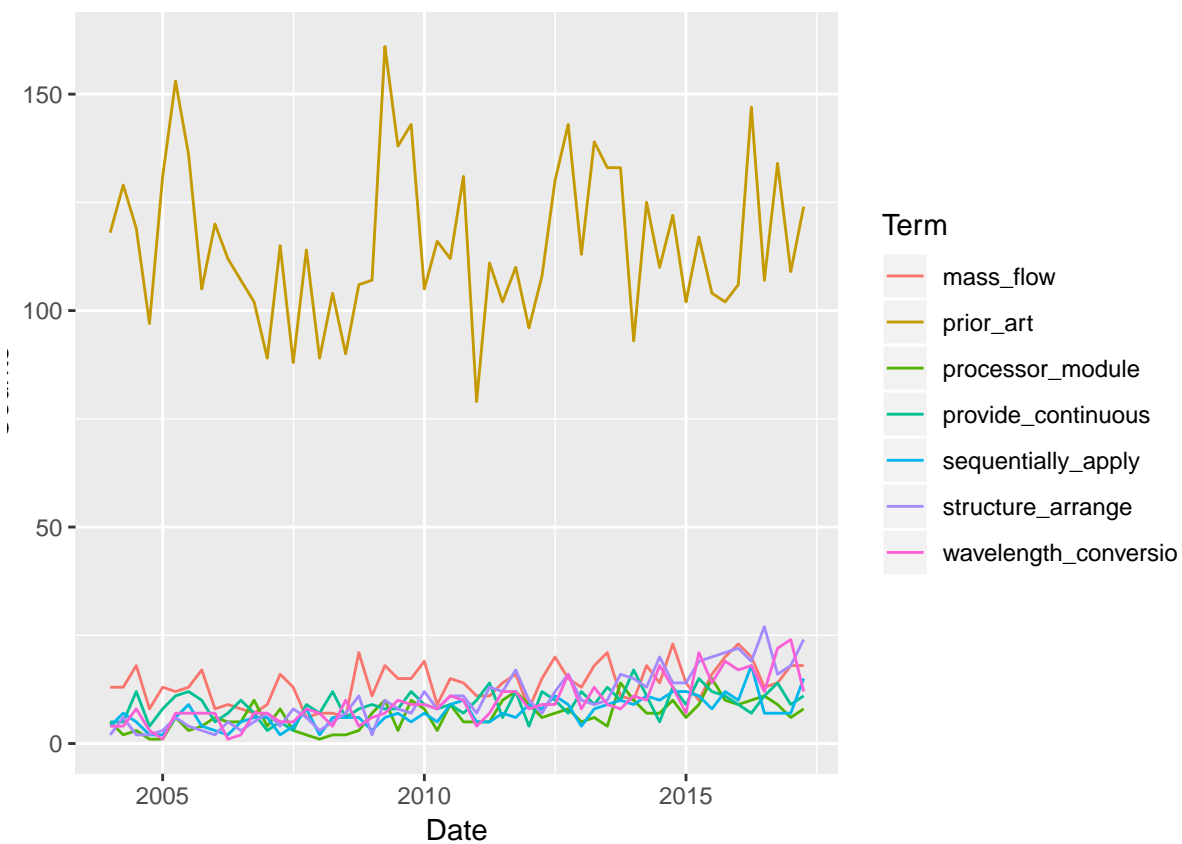
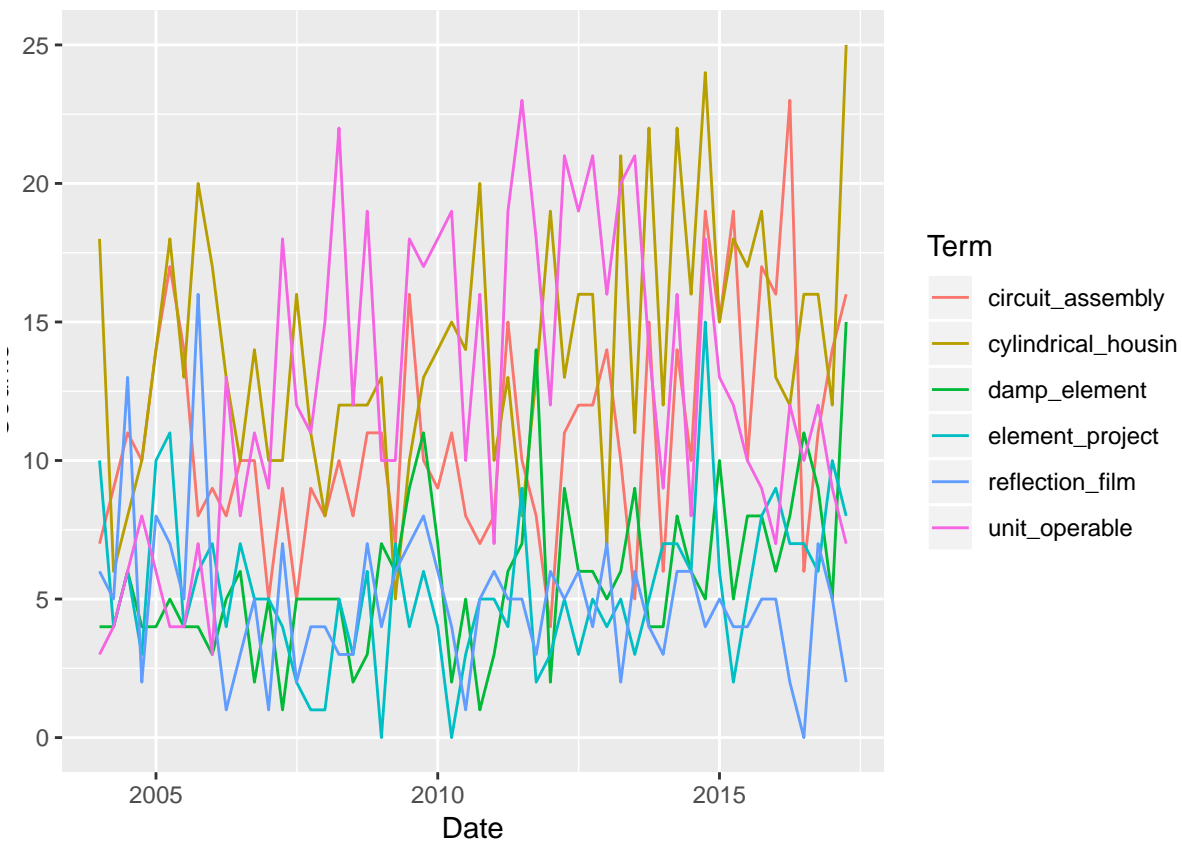


Table 2: Stationary Terms

Term	σ_ϵ	σ_ν	σ_η	δ	E(1stder)	E(sum1stder/sd)
structure arrange	2.815327	0.0227443	0.0000001	1.0000000	19.2133521	6.8245554
wavelength conversion	3.238696	0.0000000	0.0000001	1.0000000	13.3533067	4.1230506
expression level	4.281186	0.0000000	0.0000000	0.9908559	17.1364327	4.0027297
virtual memory	2.832714	0.0000000	0.0000000	0.9816148	10.3607917	3.6575498
sequentially apply	2.377261	0.0000000	0.0000000	1.0000000	7.7360778	3.2541981
ovarian cancer	3.527516	0.0000000	0.0000000	1.0000000	10.8241662	3.0684951
transmission portion	2.921351	0.0000000	0.0000000	0.9354029	8.2599000	2.8274251
processor module	2.636422	0.0000000	0.0000000	0.9936448	6.7360768	2.5550071
anchor system	1.851920	2.5902237	0.0000000	0.0440073	4.2671578	2.3041809
logical unit	3.155938	0.0000000	0.5510841	0.9499389	6.5023744	2.0603617
damp element	2.523229	0.0000000	0.0000000	1.0000000	4.6678102	1.8499350
channel characteristic	3.150927	0.0000000	0.0000000	0.9253172	5.2756506	1.6743169
axially displaceable	3.193998	0.0000000	0.0000004	1.0000000	5.3454545	1.6735937
mass flow	4.019558	0.0000000	0.0000001	1.0000000	6.6241281	1.6479744
cylindrical housing	4.105870	0.0000000	0.0000000	1.0000000	5.5419097	1.3497527
unit operable	3.754142	0.1217633	0.0000002	0.9440418	4.4542629	1.1864930
provide continuous	2.825754	0.1732193	0.0000001	0.4240296	2.8933171	1.0239099
circuit assembly	3.493271	0.3933979	0.0000001	0.3576427	3.3692957	0.9645103
amplifier input	2.453408	0.0000000	0.0000000	0.3973656	0.5979061	0.2437043
element project	2.483974	0.0535886	0.0000001	0.9128285	0.5532855	0.2227421
intermediate signal	3.340594	0.0000000	0.0000000	0.4867379	0.6964338	0.2084760
fall edge	4.343637	0.0000000	0.0000000	0.8638916	-1.1027240	-0.2538711
prior art	17.489111	0.0000001	0.0000006	0.7796256	-6.0176466	-0.3440796
reflection film	2.547699	0.0000000	0.0000000	0.8827550	-2.4671565	-0.9683861

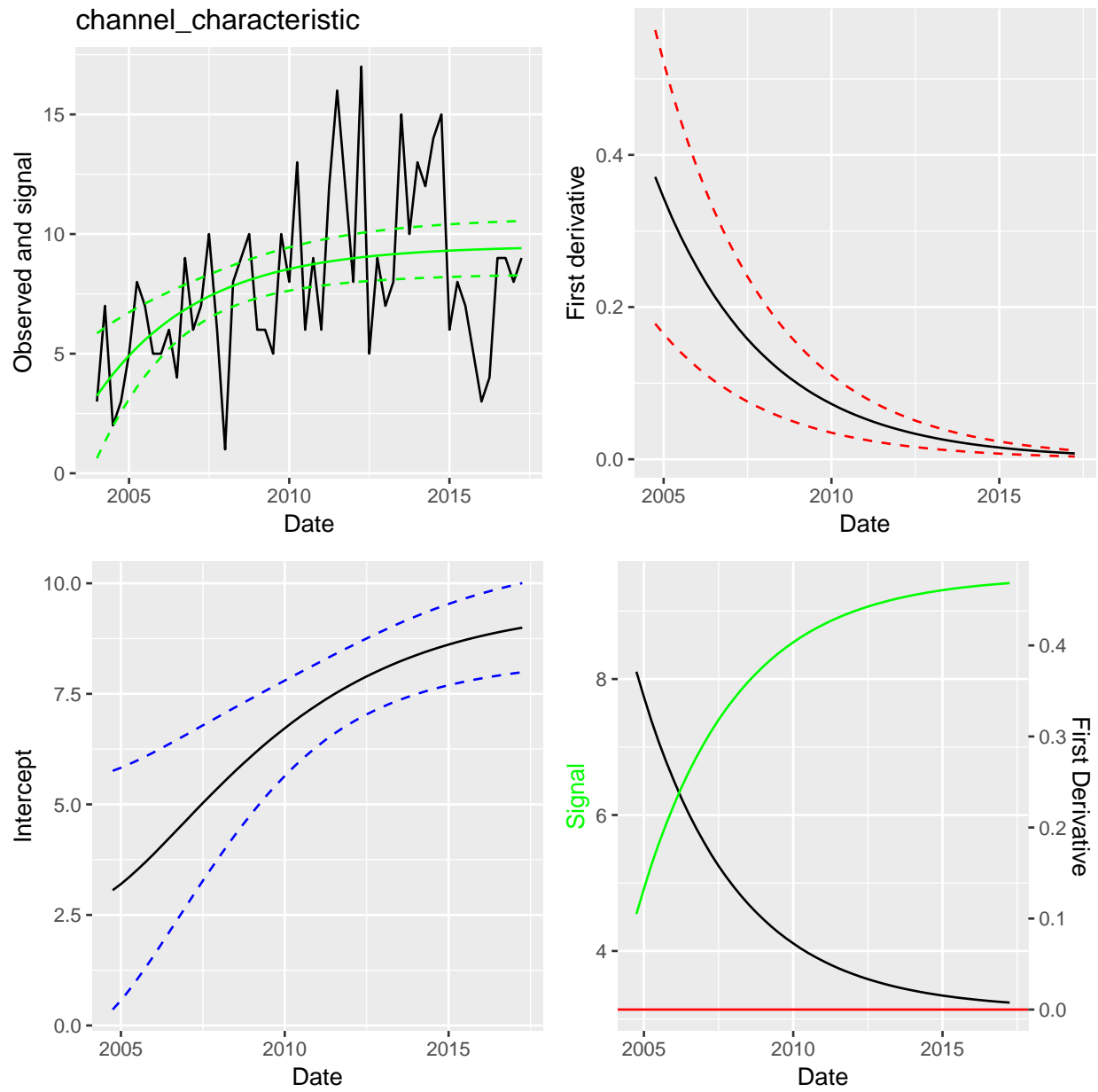


Figure 26:

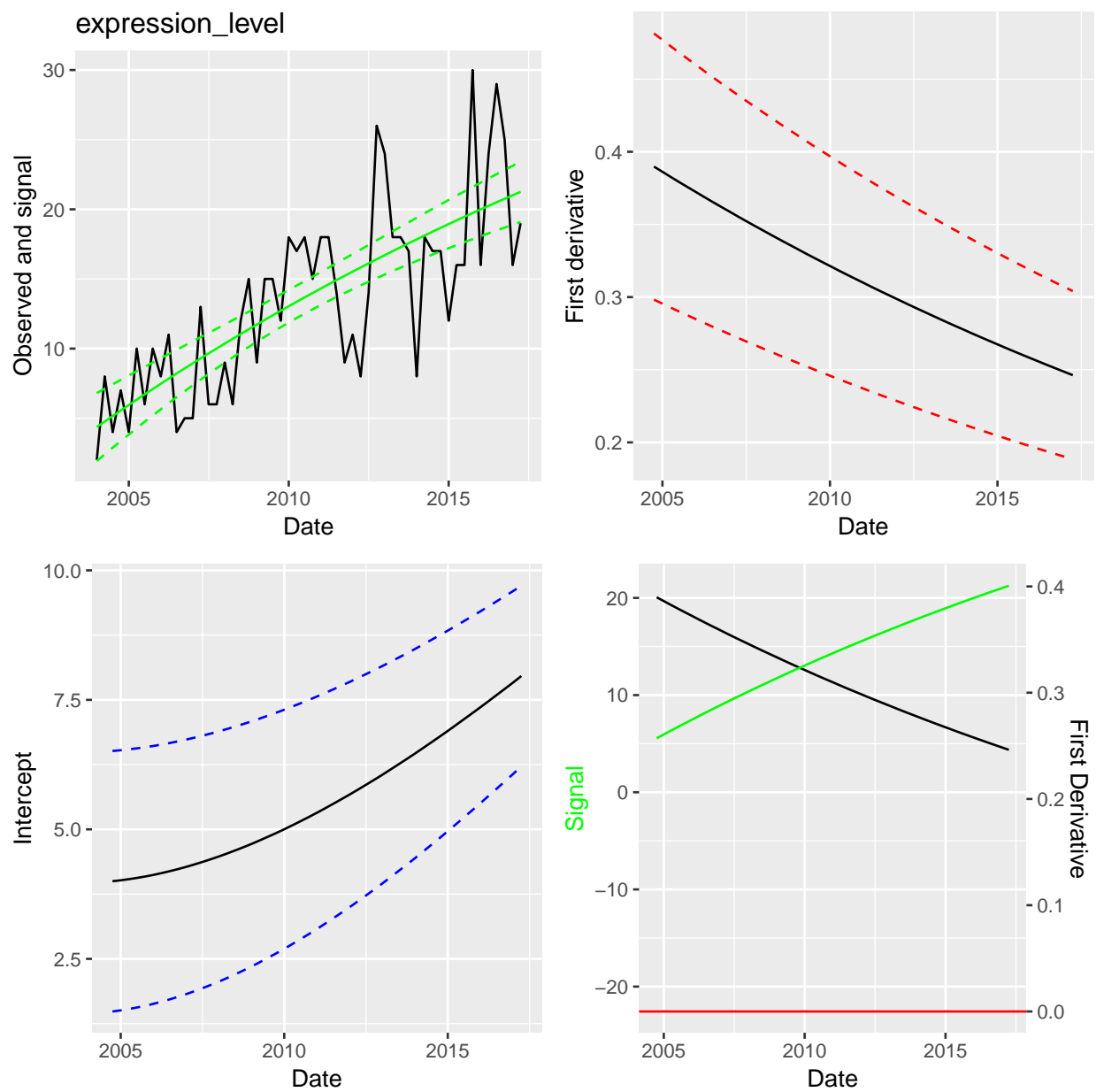


Figure 27:

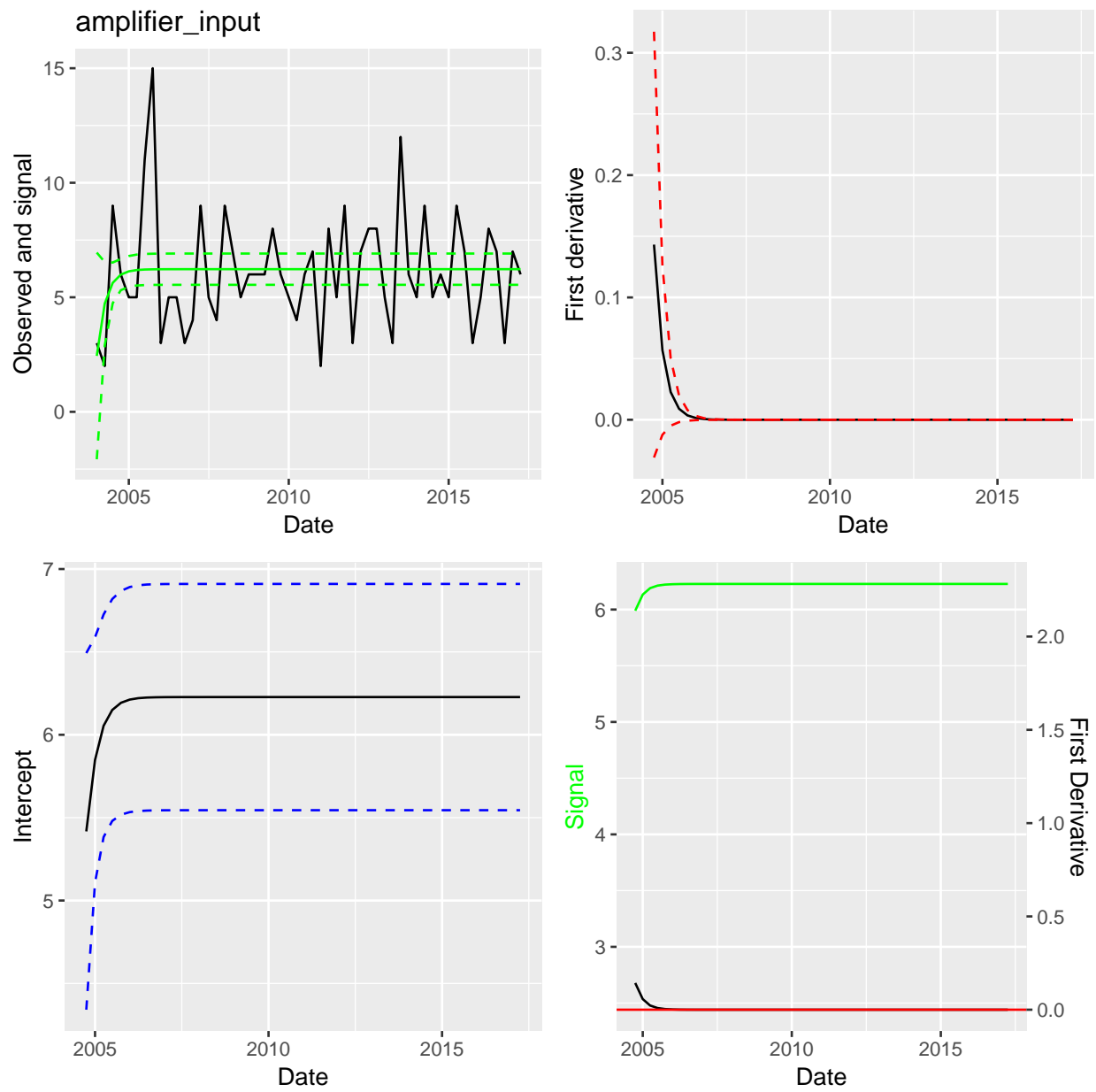


Figure 28:

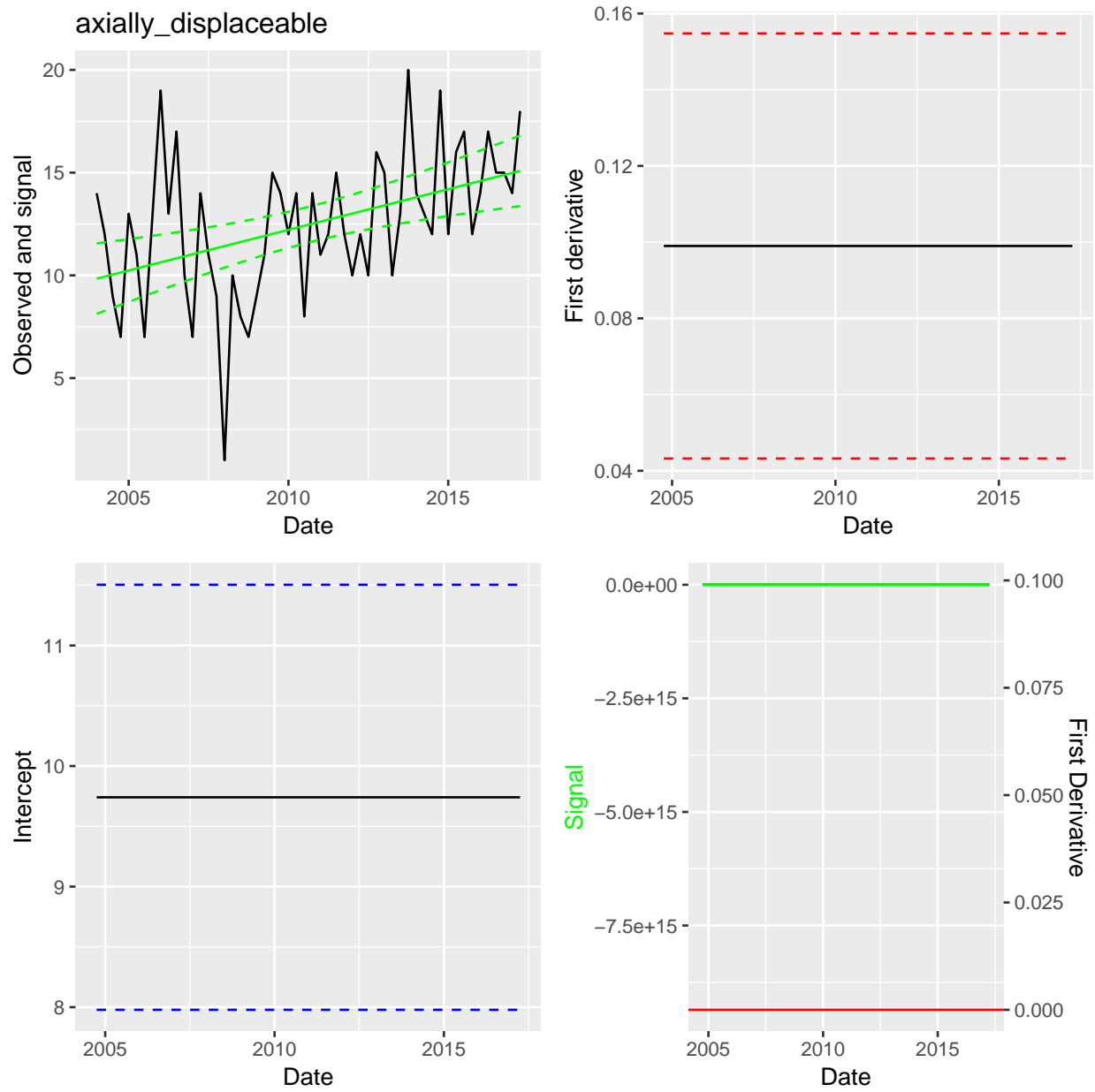


Figure 29:

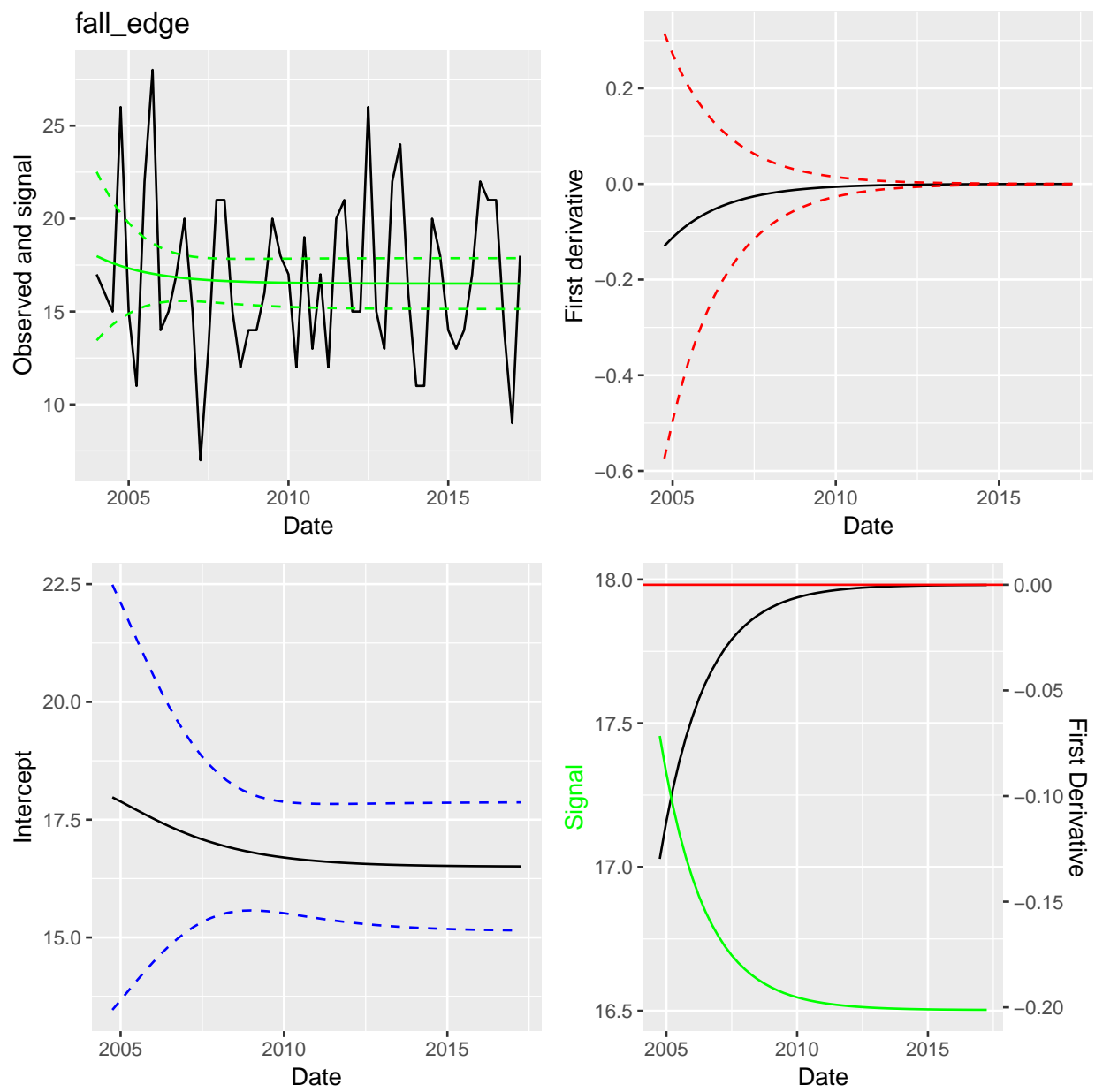


Figure 30:

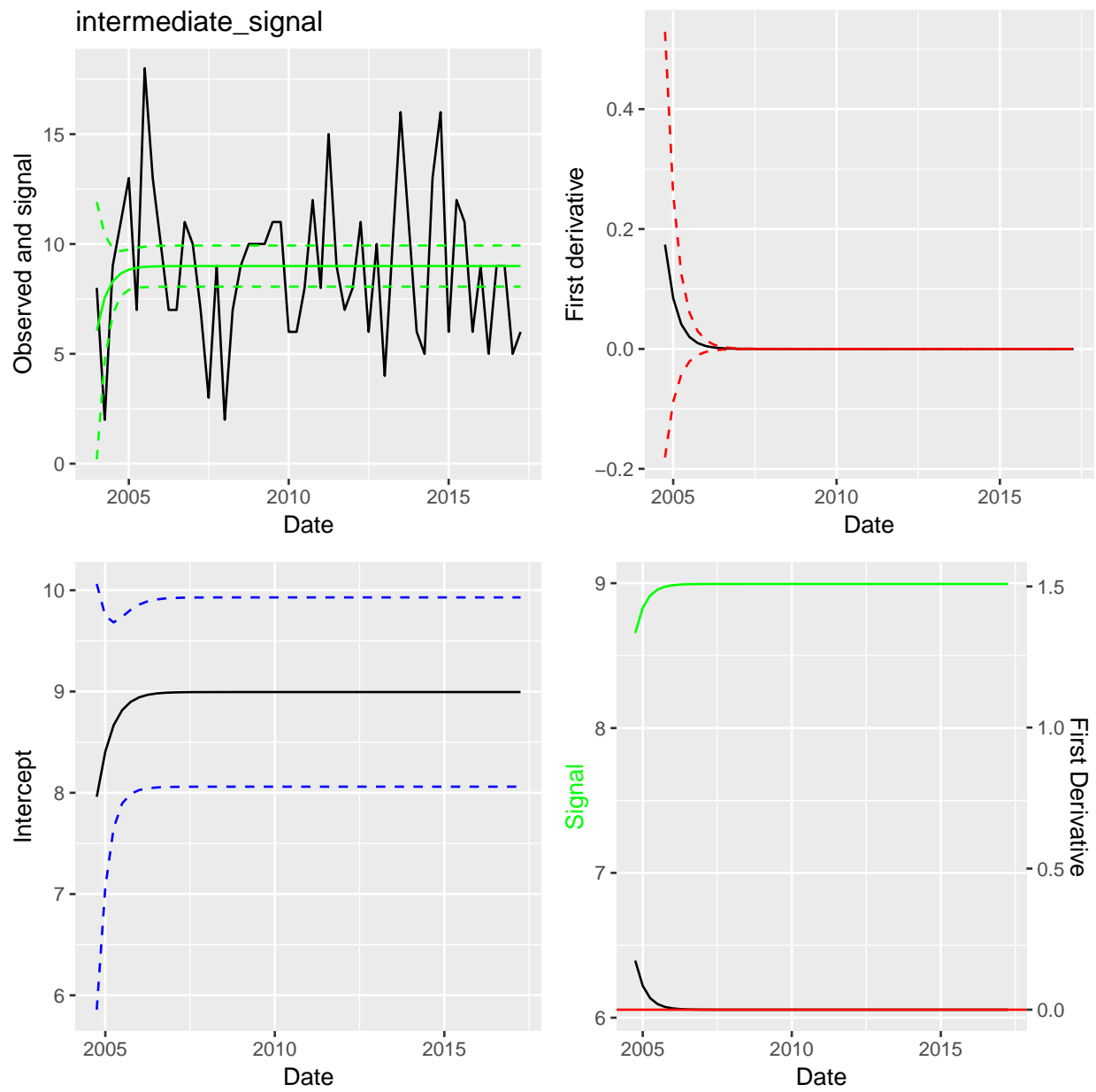


Figure 31:

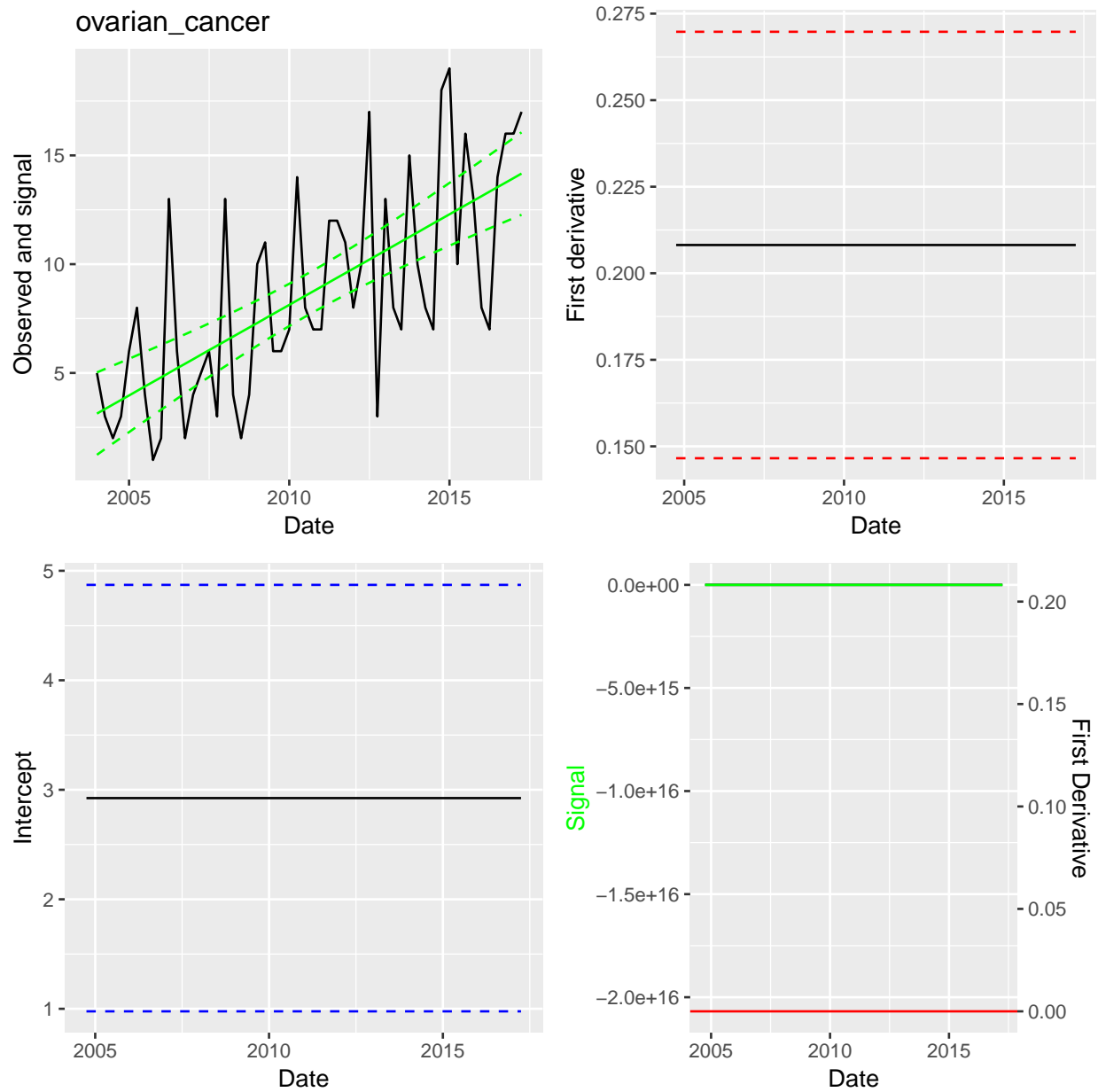


Figure 32:

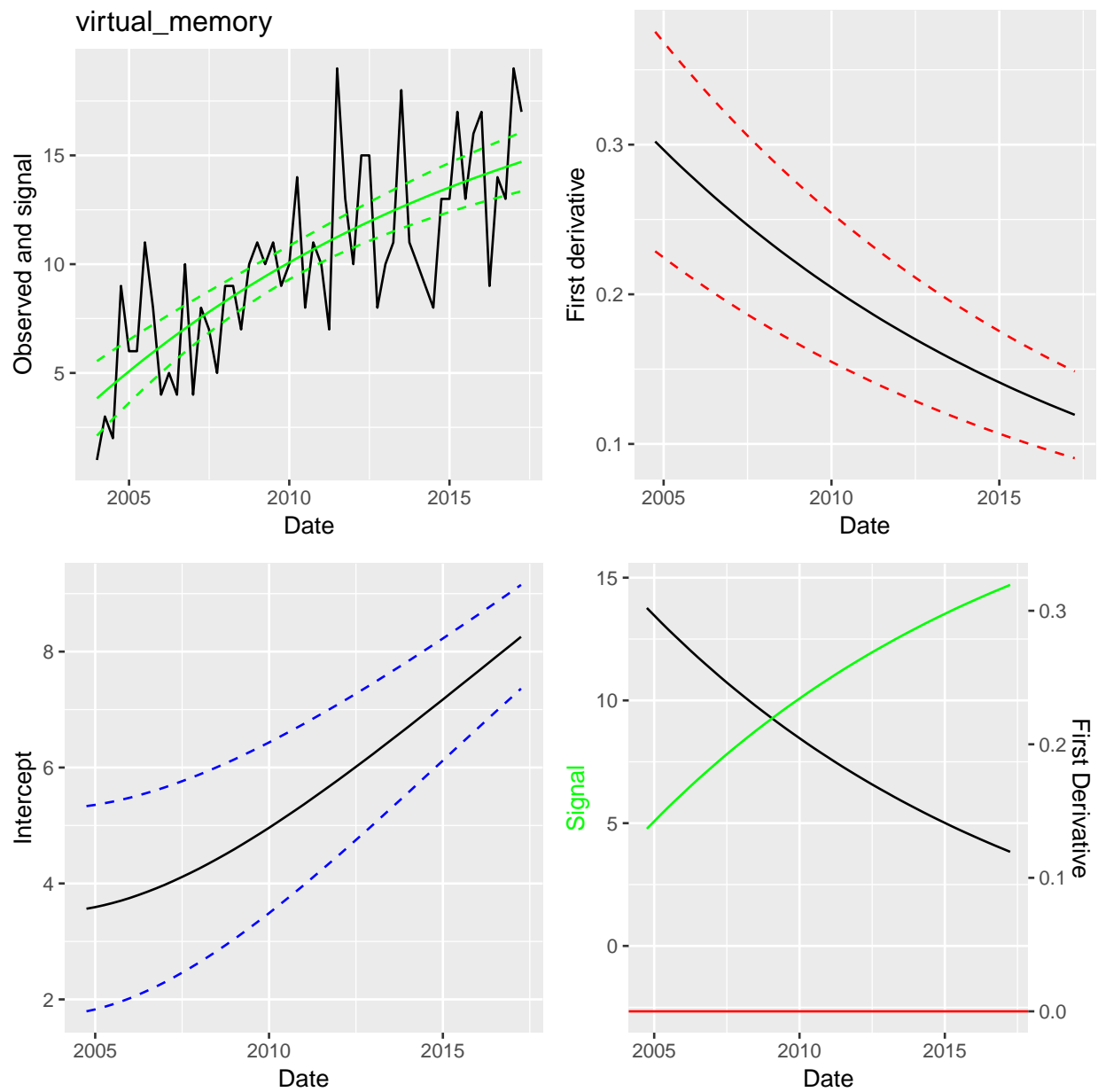


Figure 33:

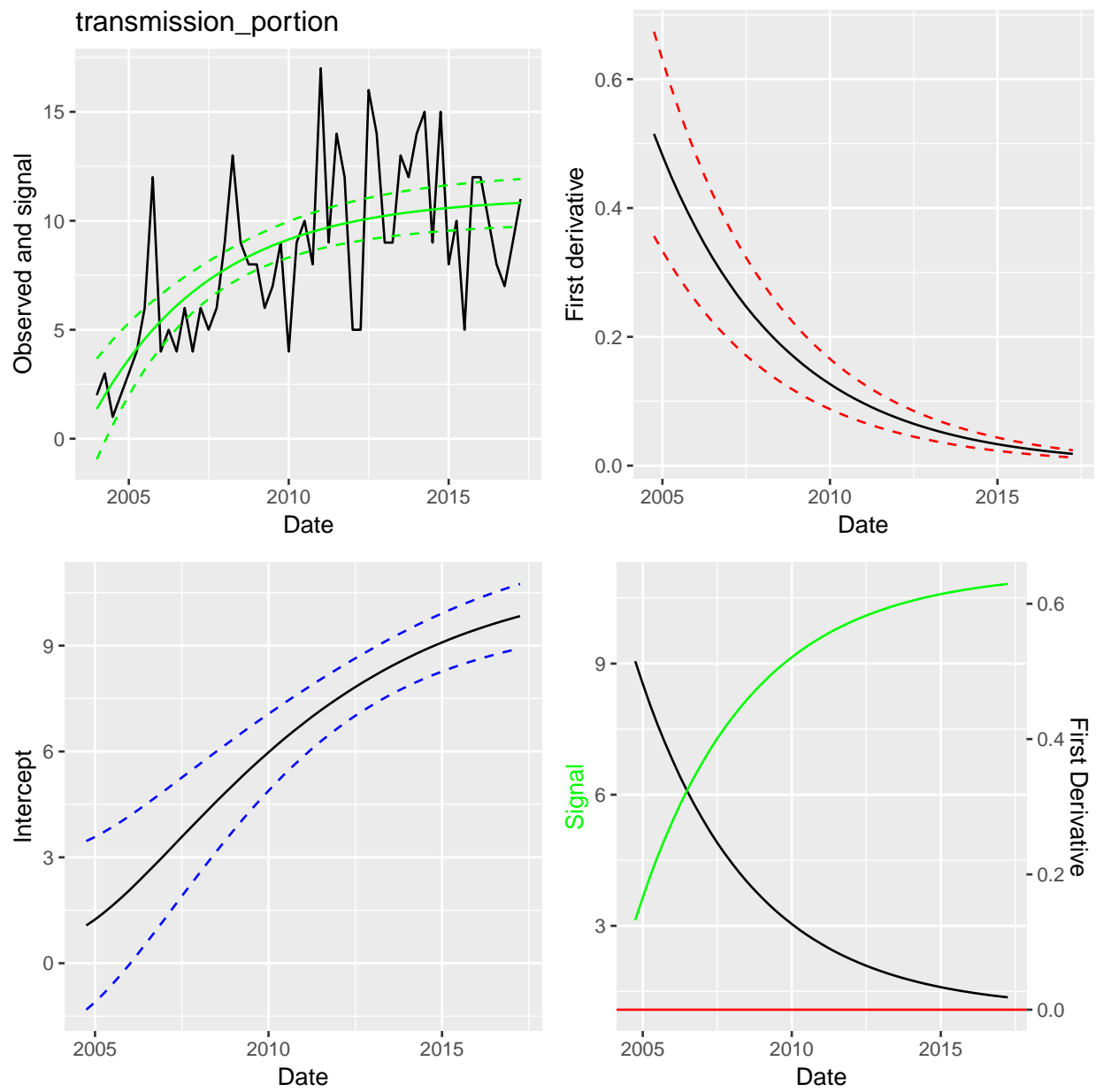


Figure 34:

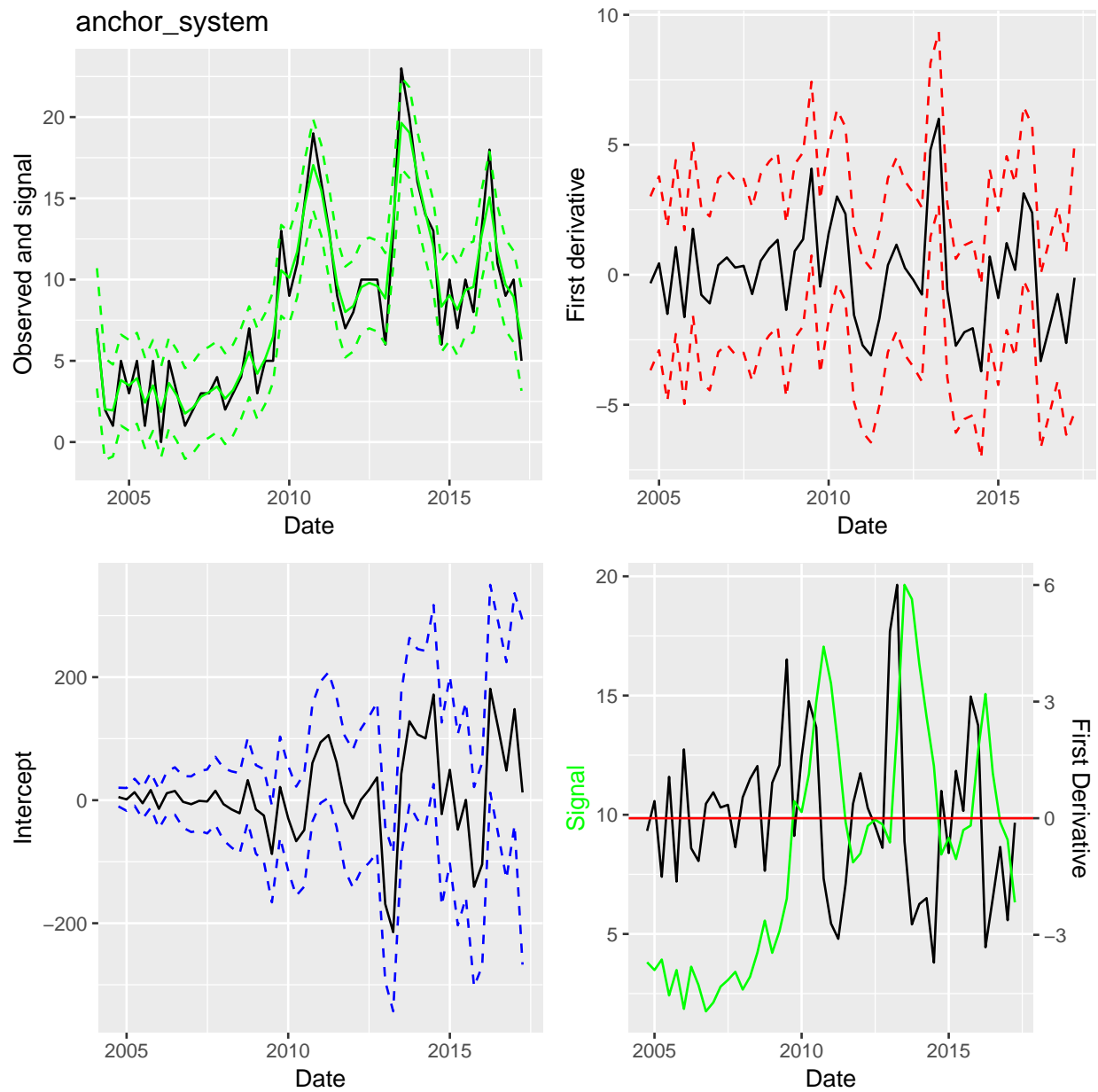


Figure 35:

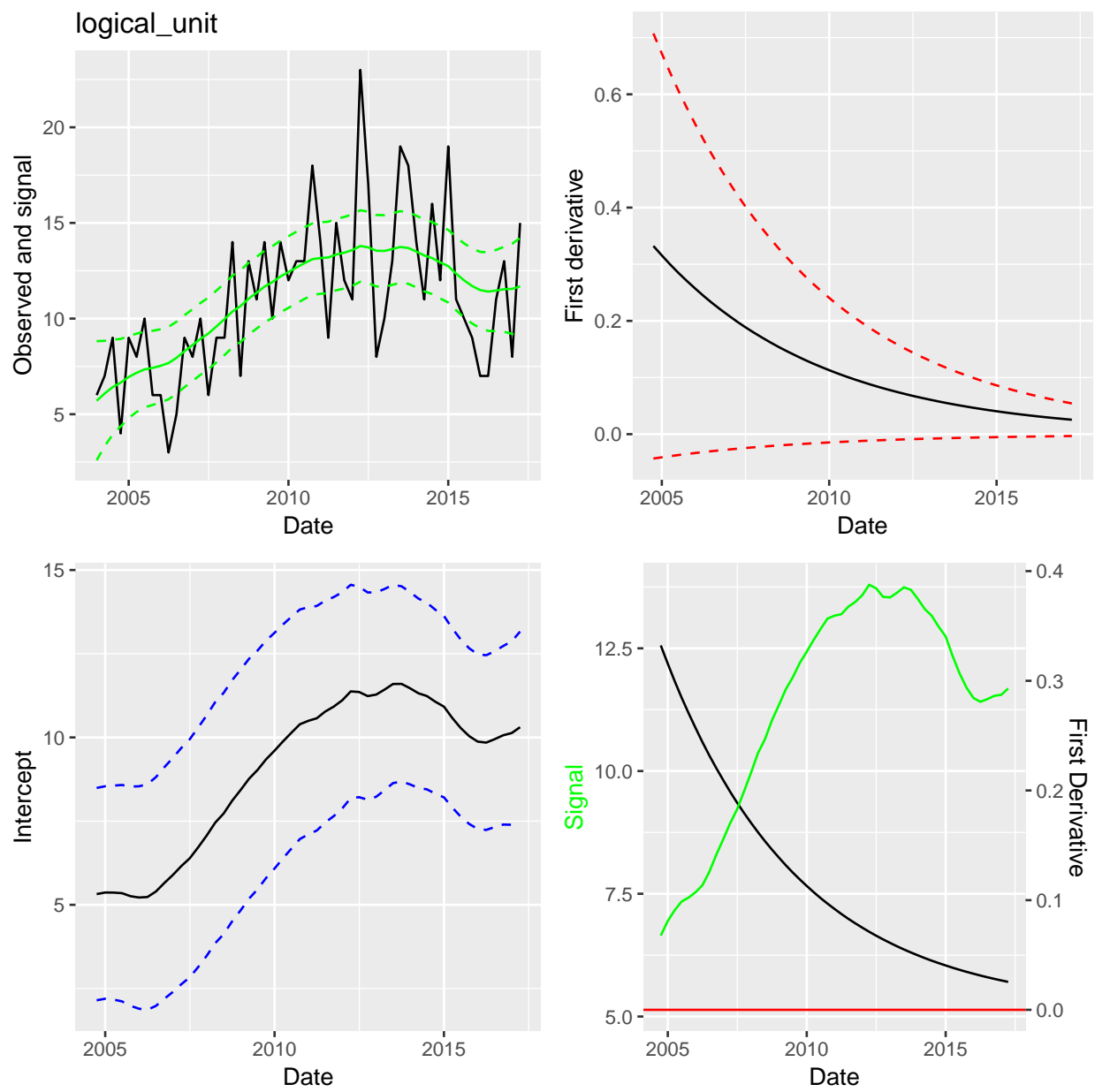


Figure 36:

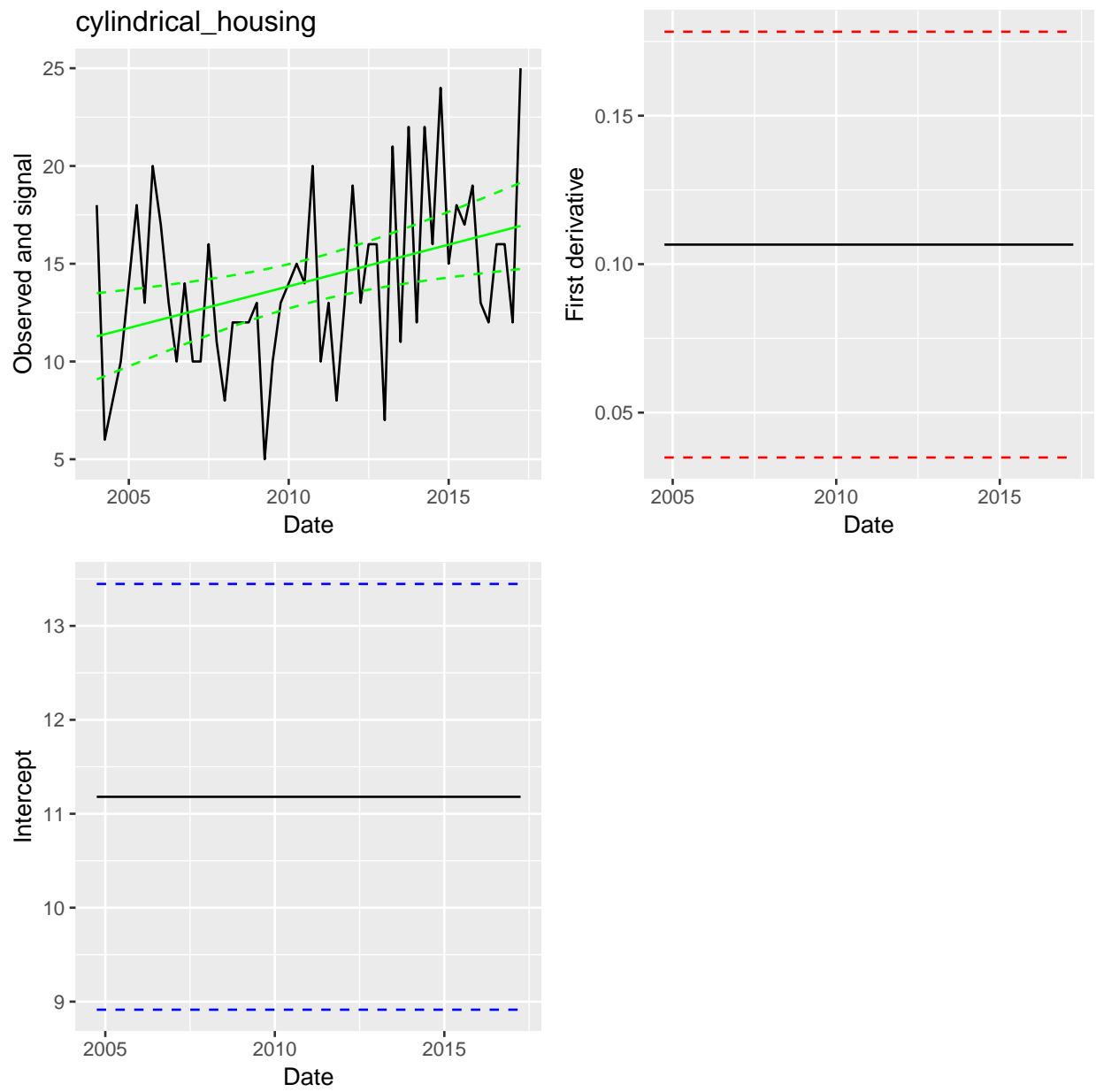


Figure 37:

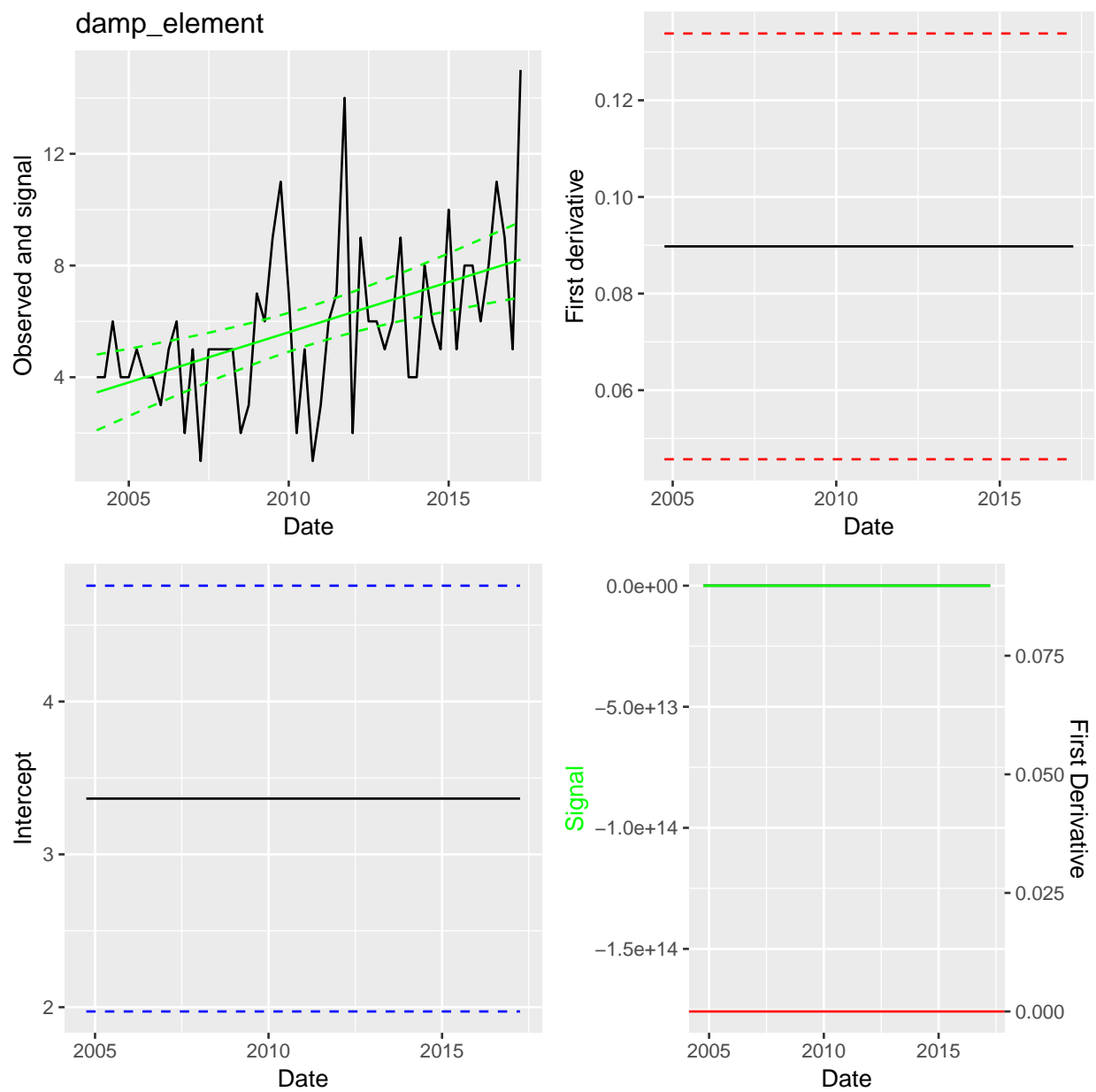


Figure 38:

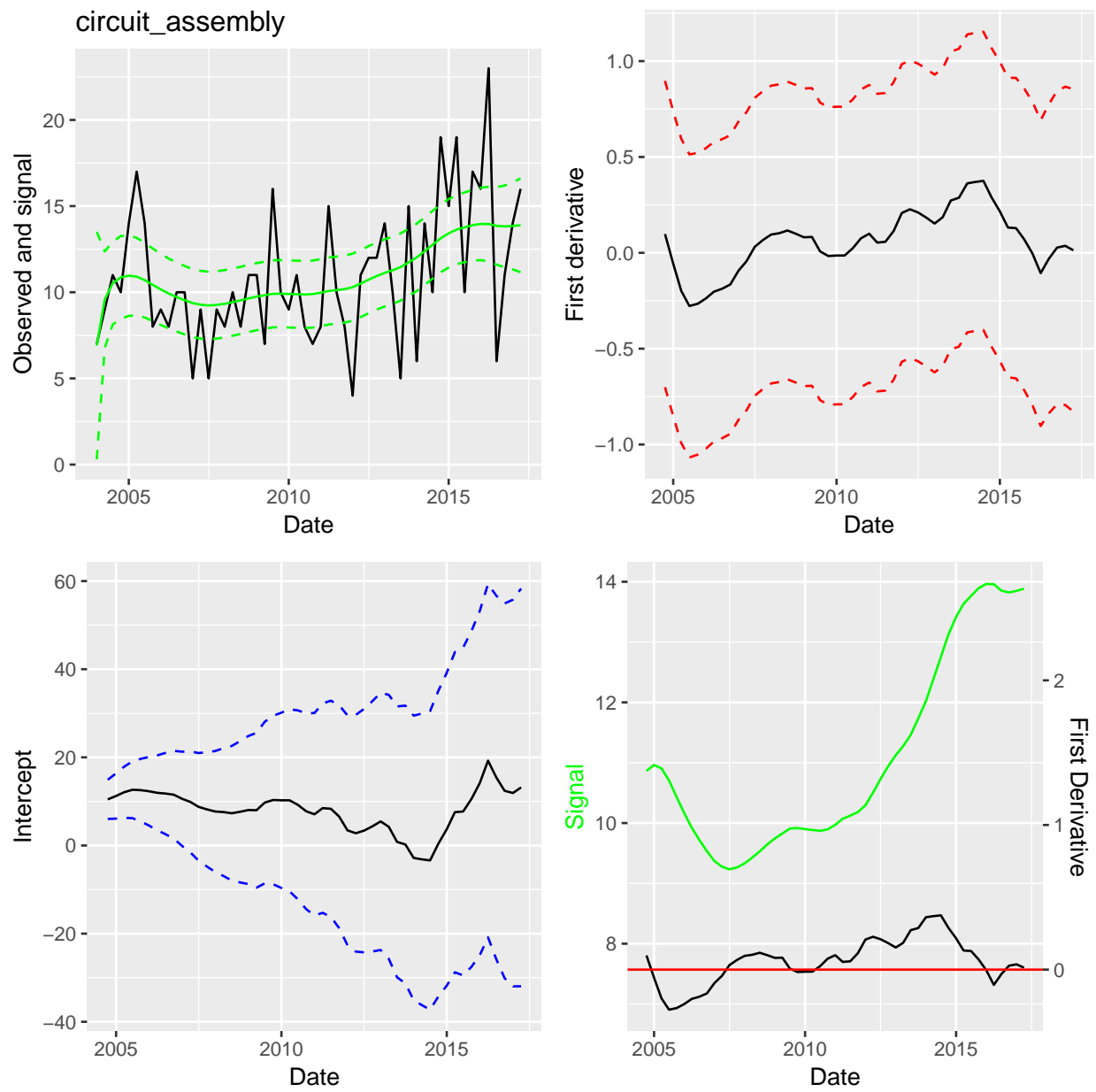


Figure 39:

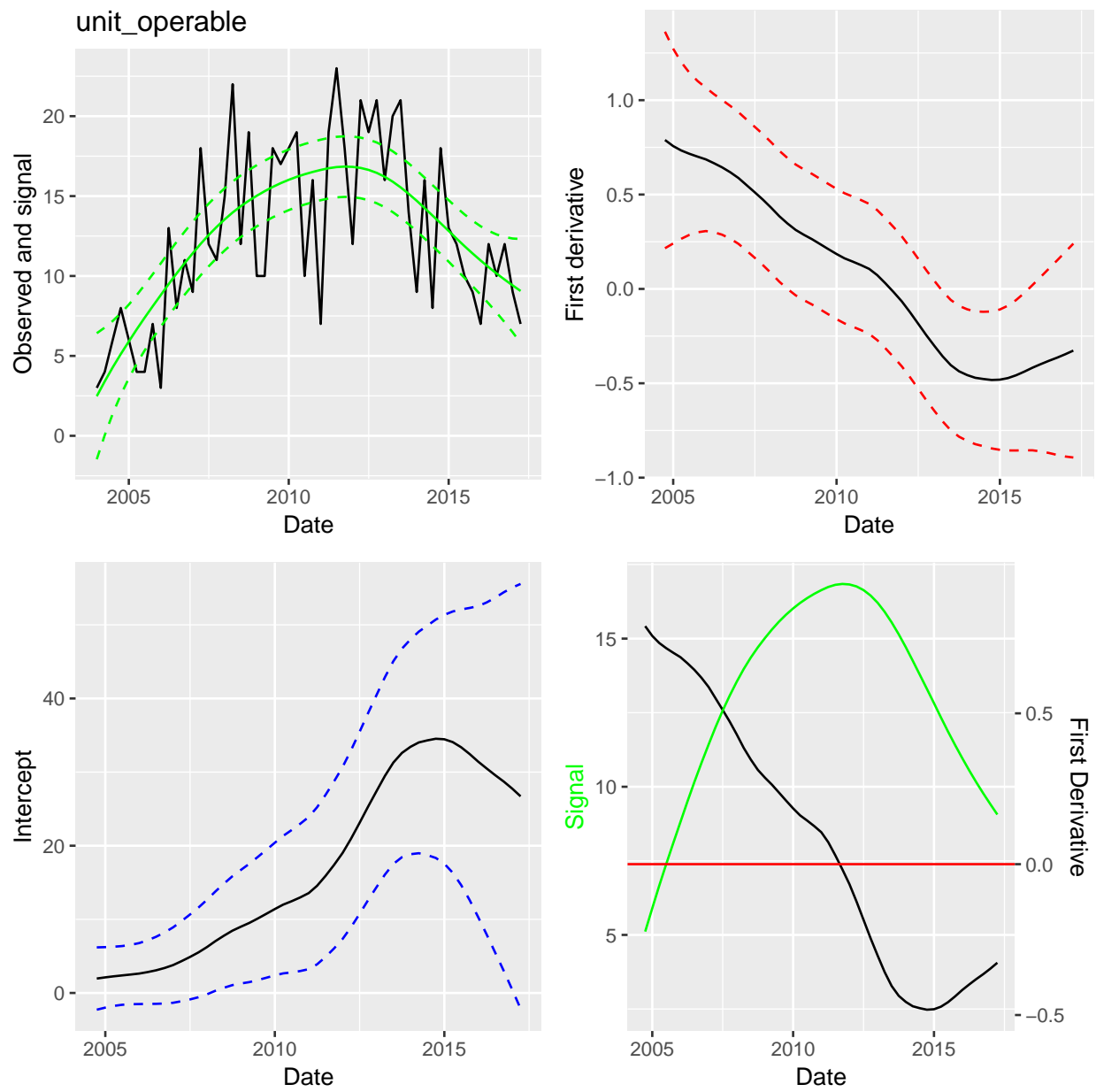


Figure 40:

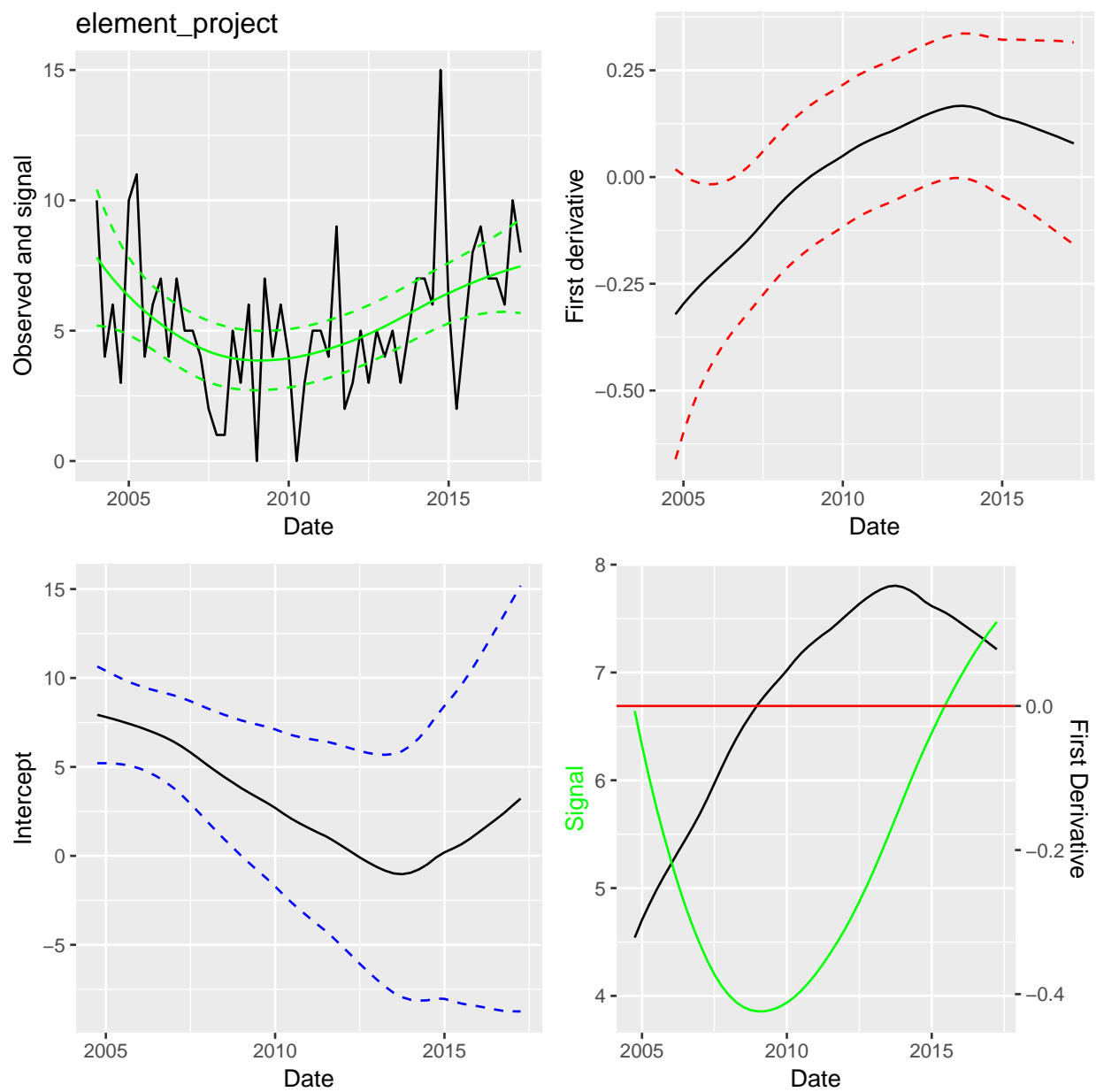


Figure 41:

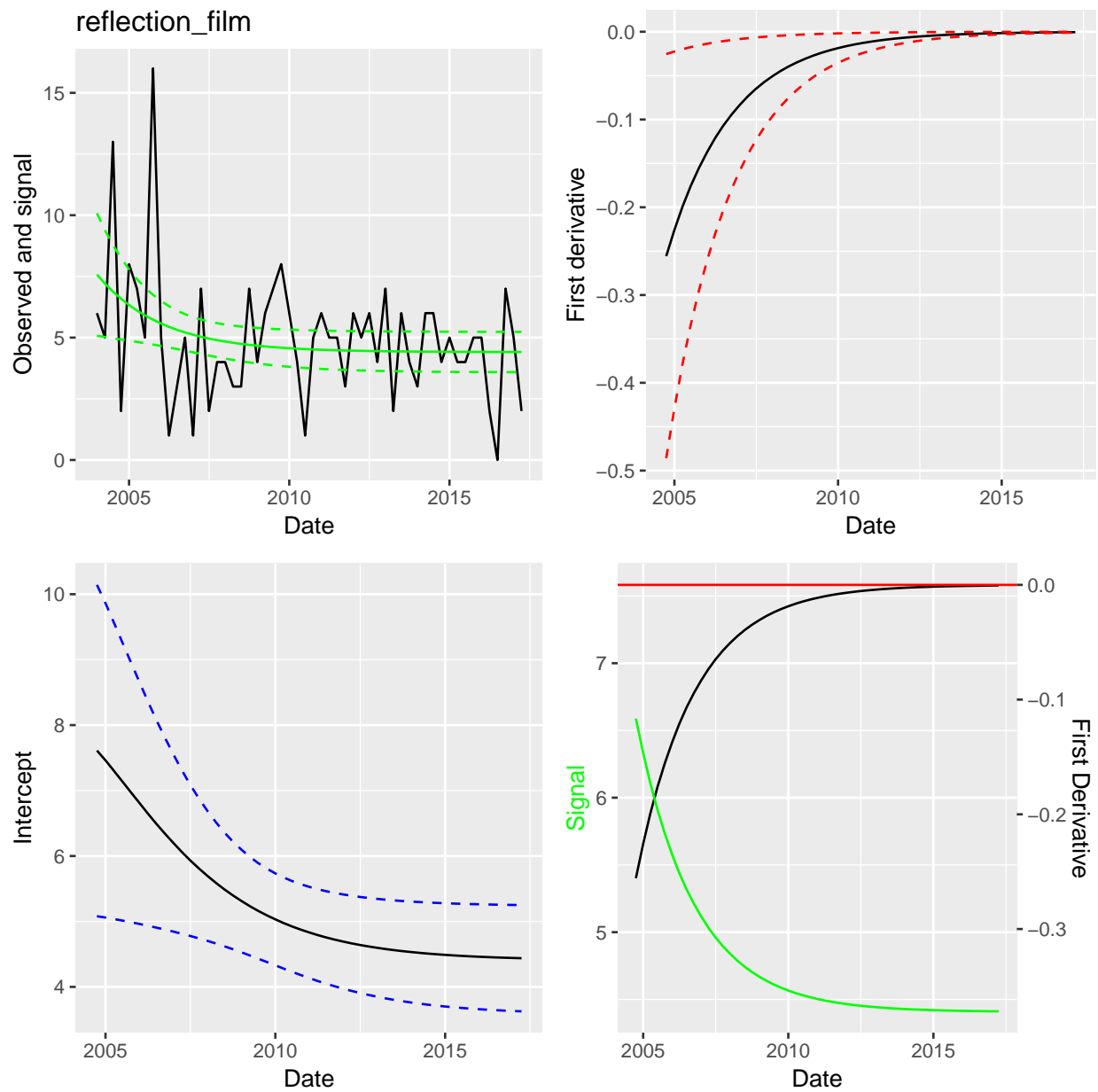


Figure 42:

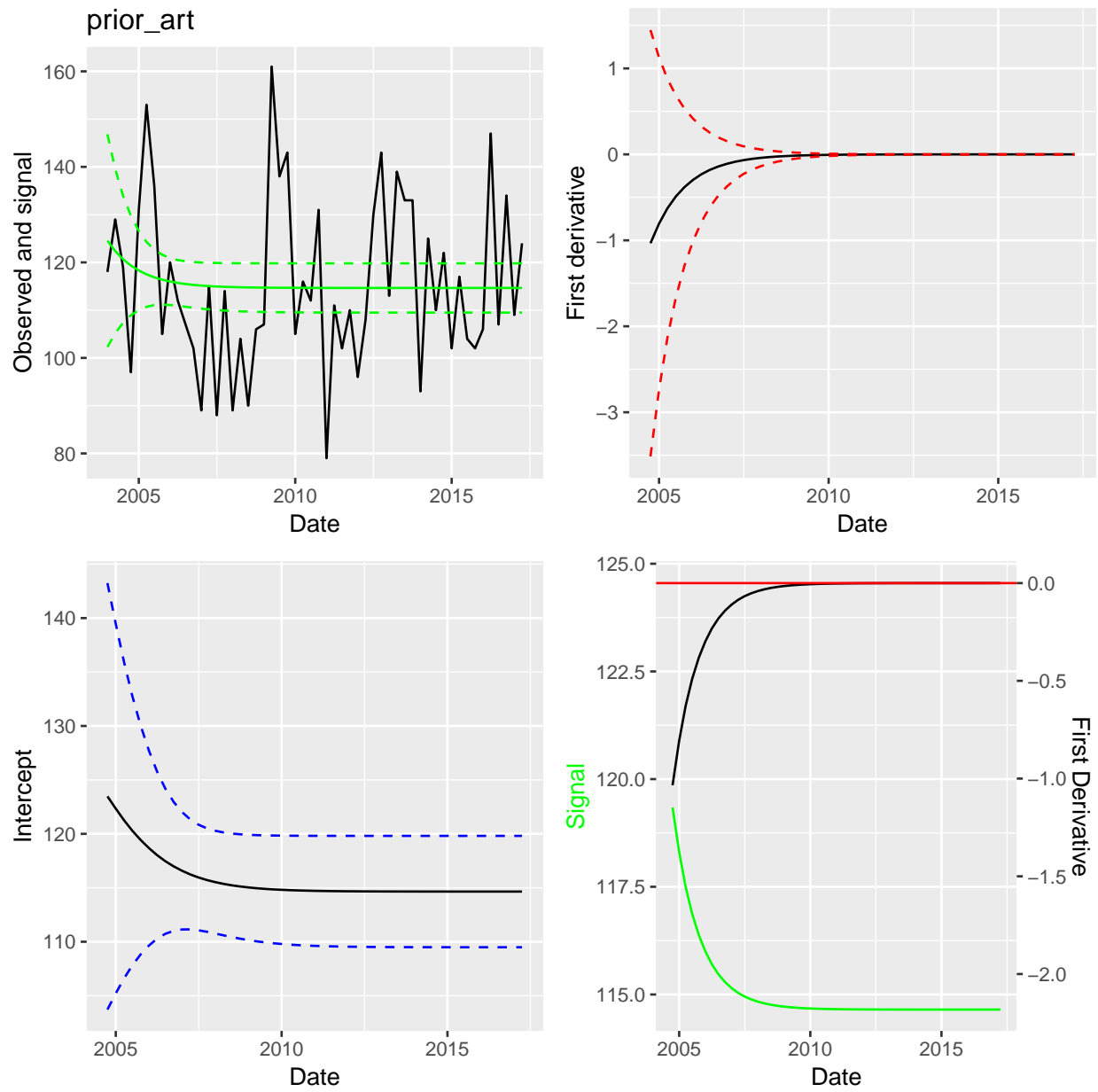


Figure 43:

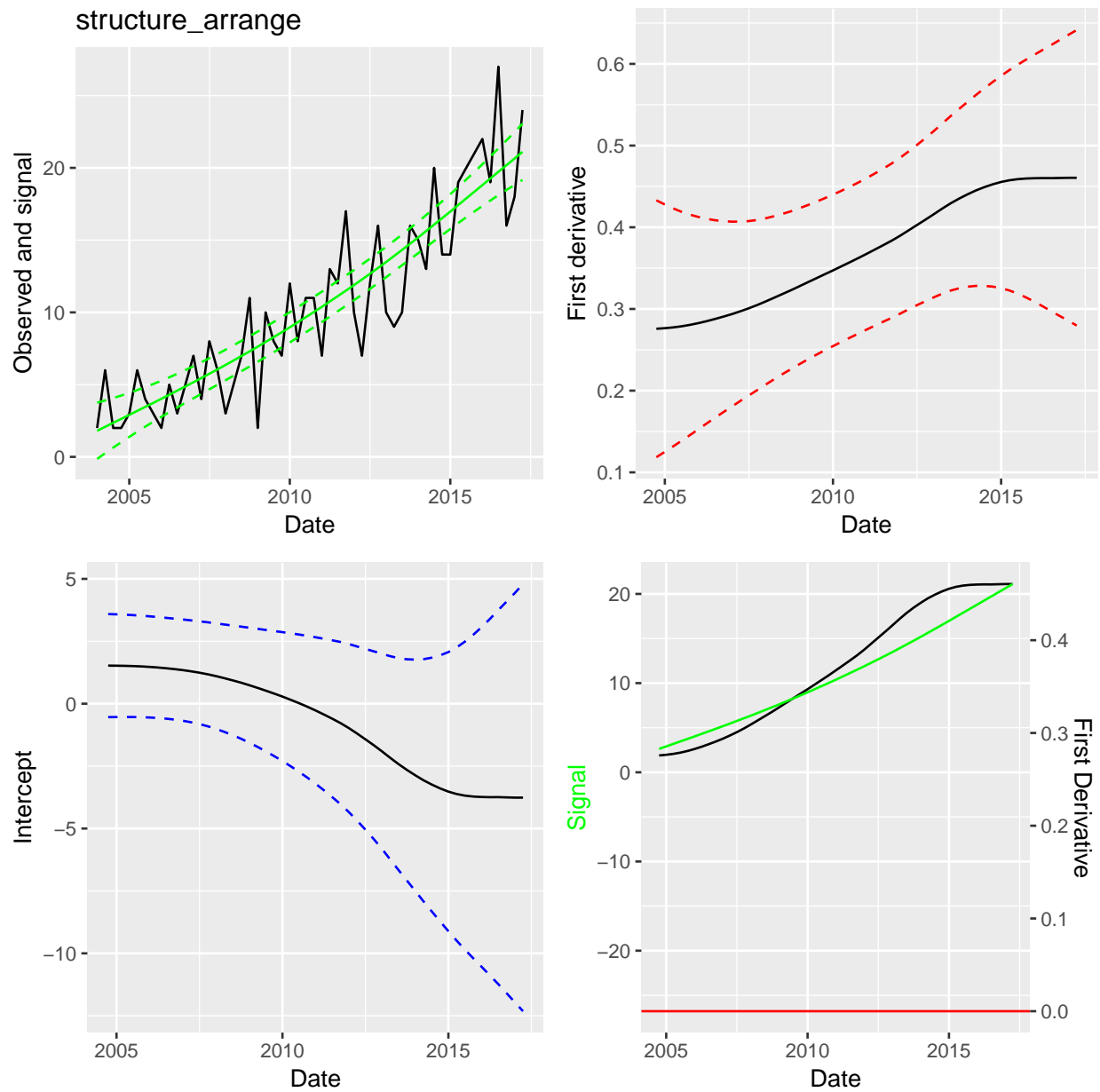


Figure 44:

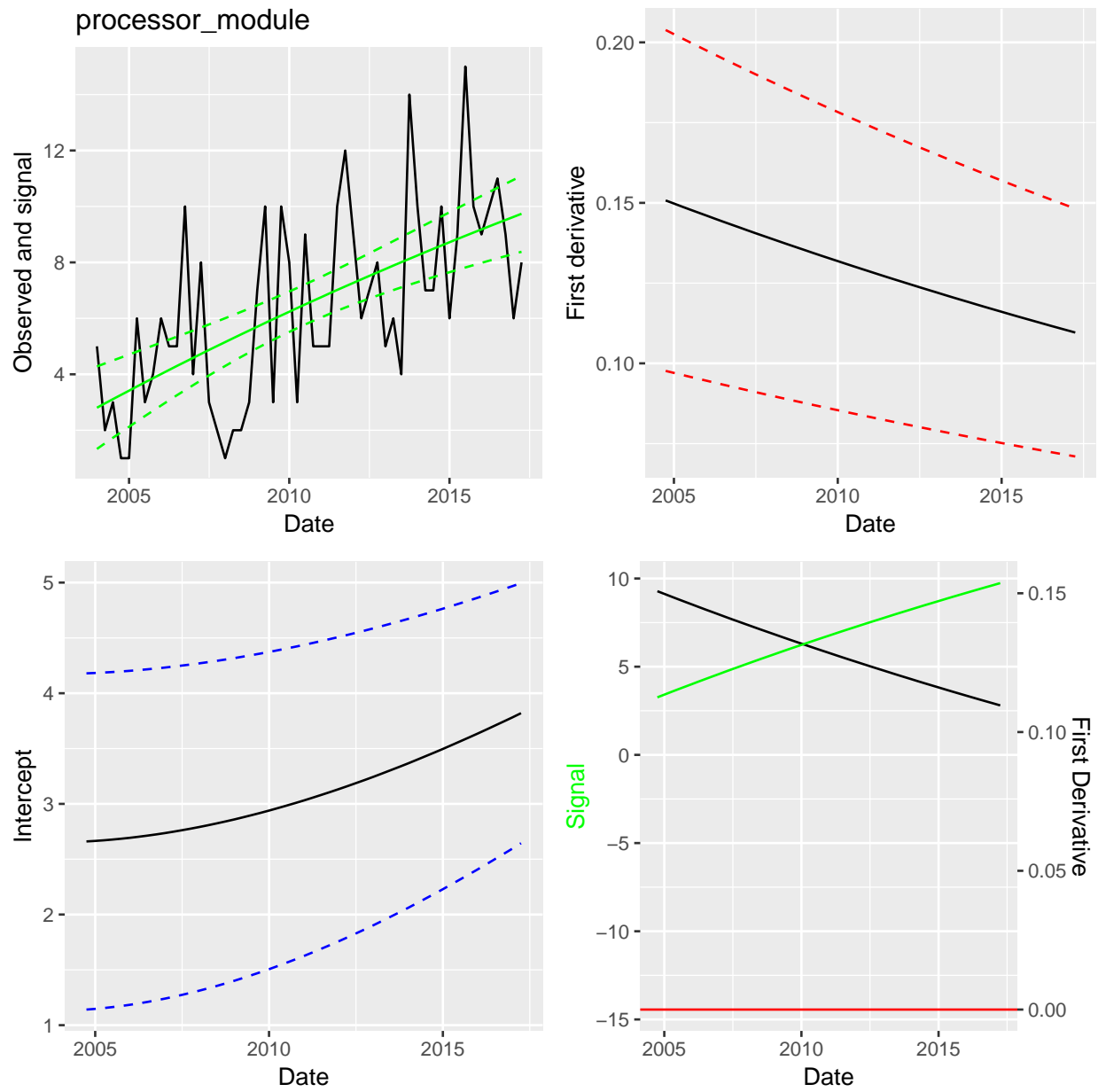


Figure 45:

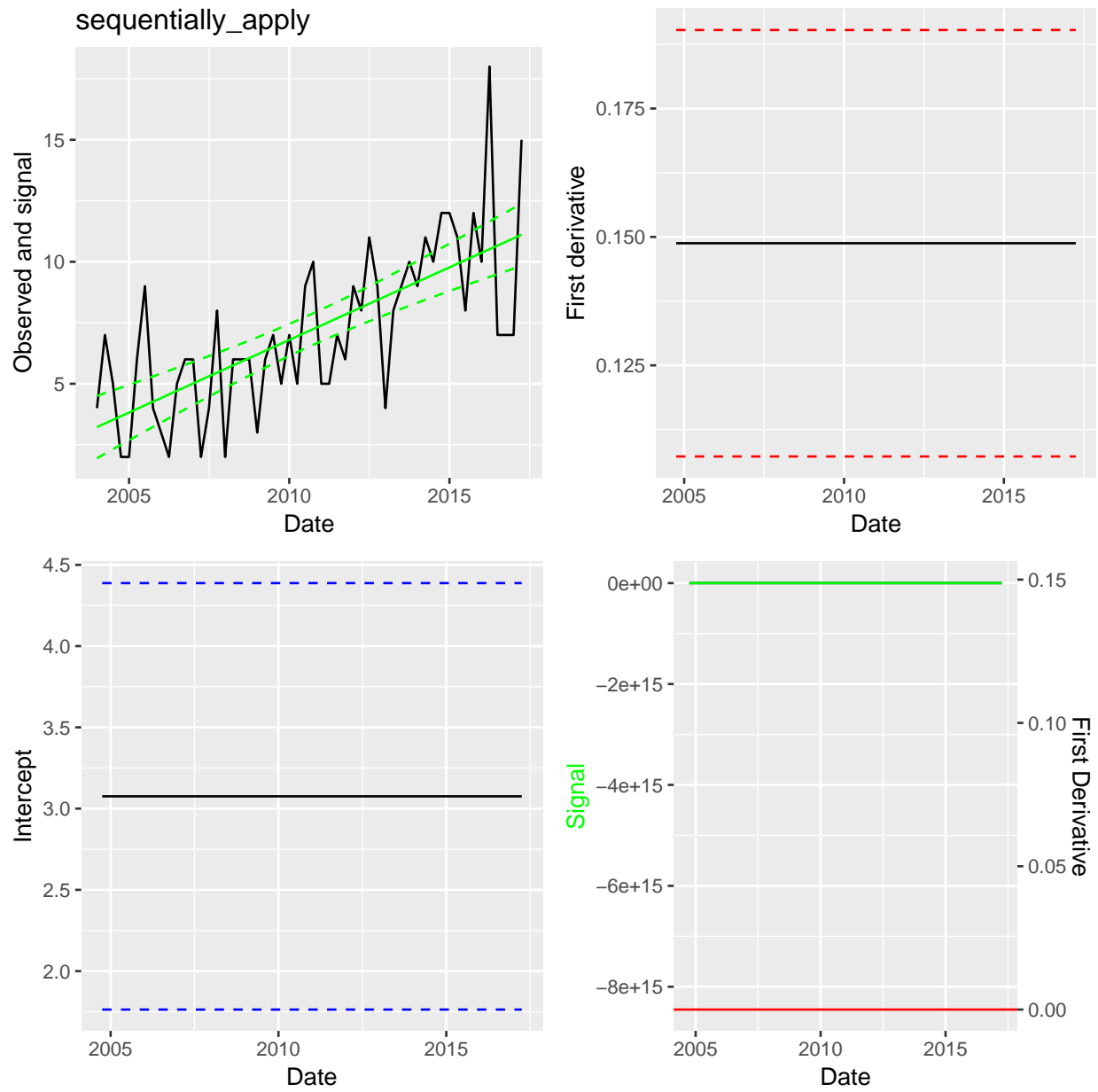


Figure 46:

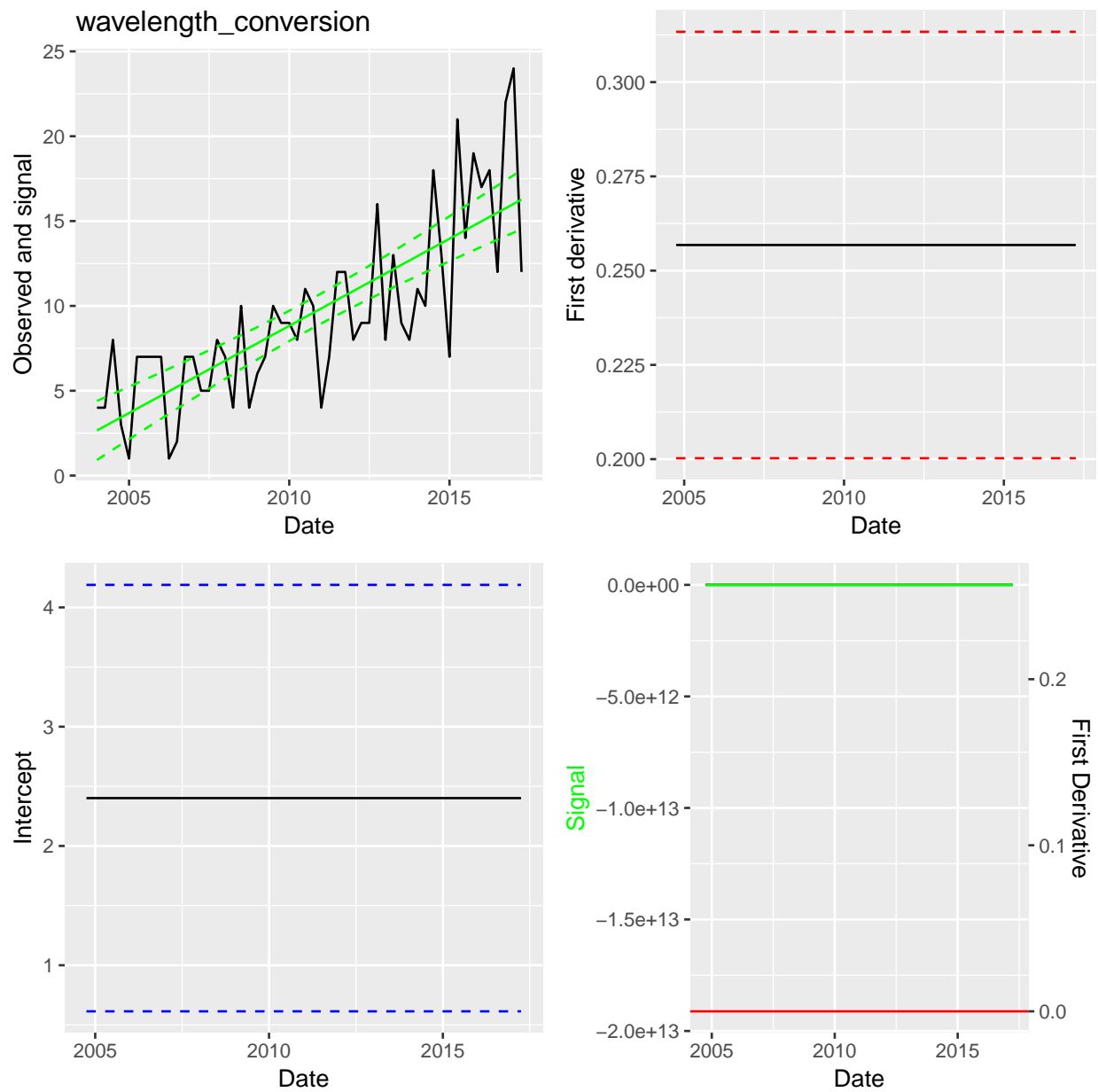


Figure 47:

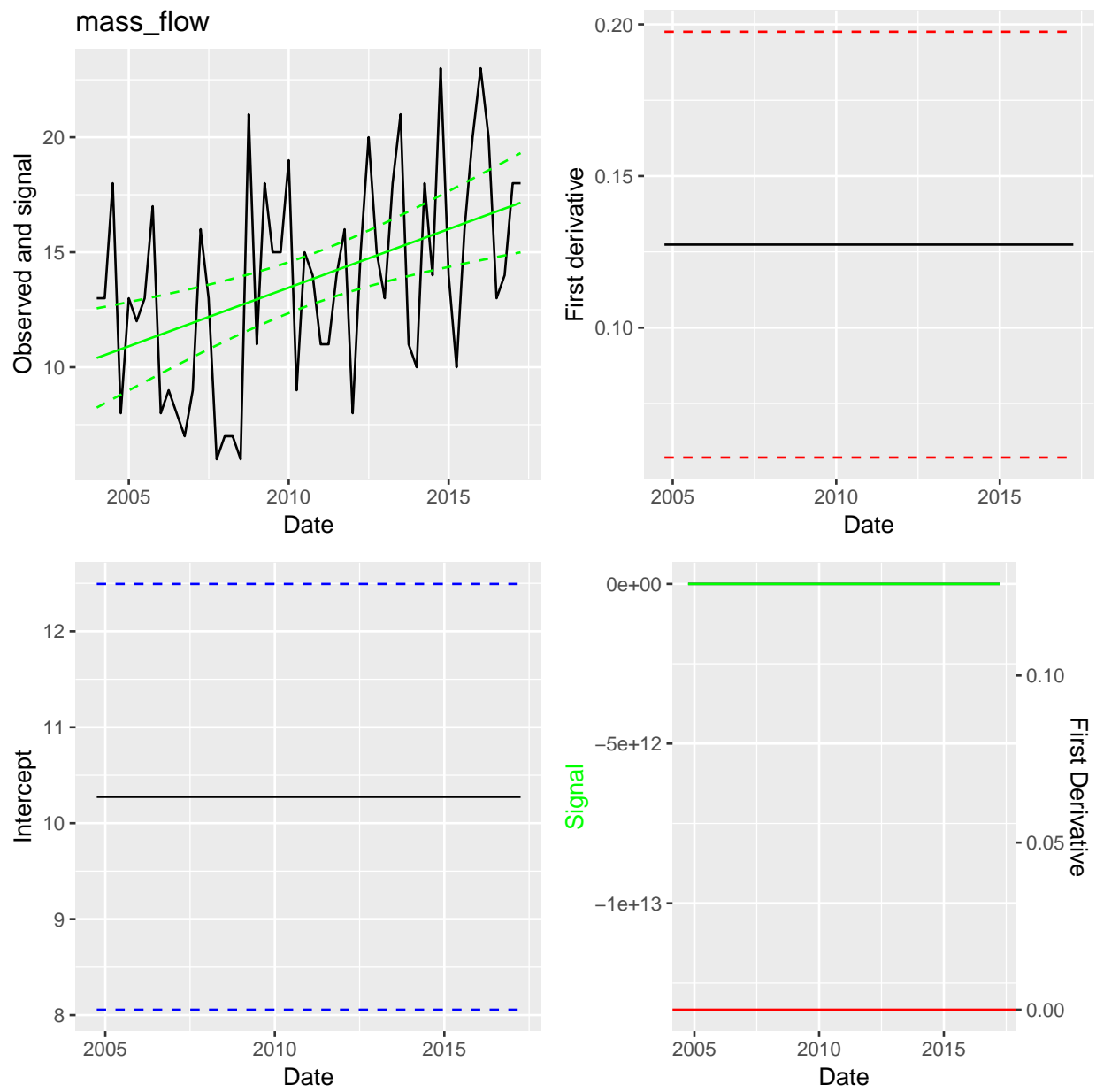


Figure 48:

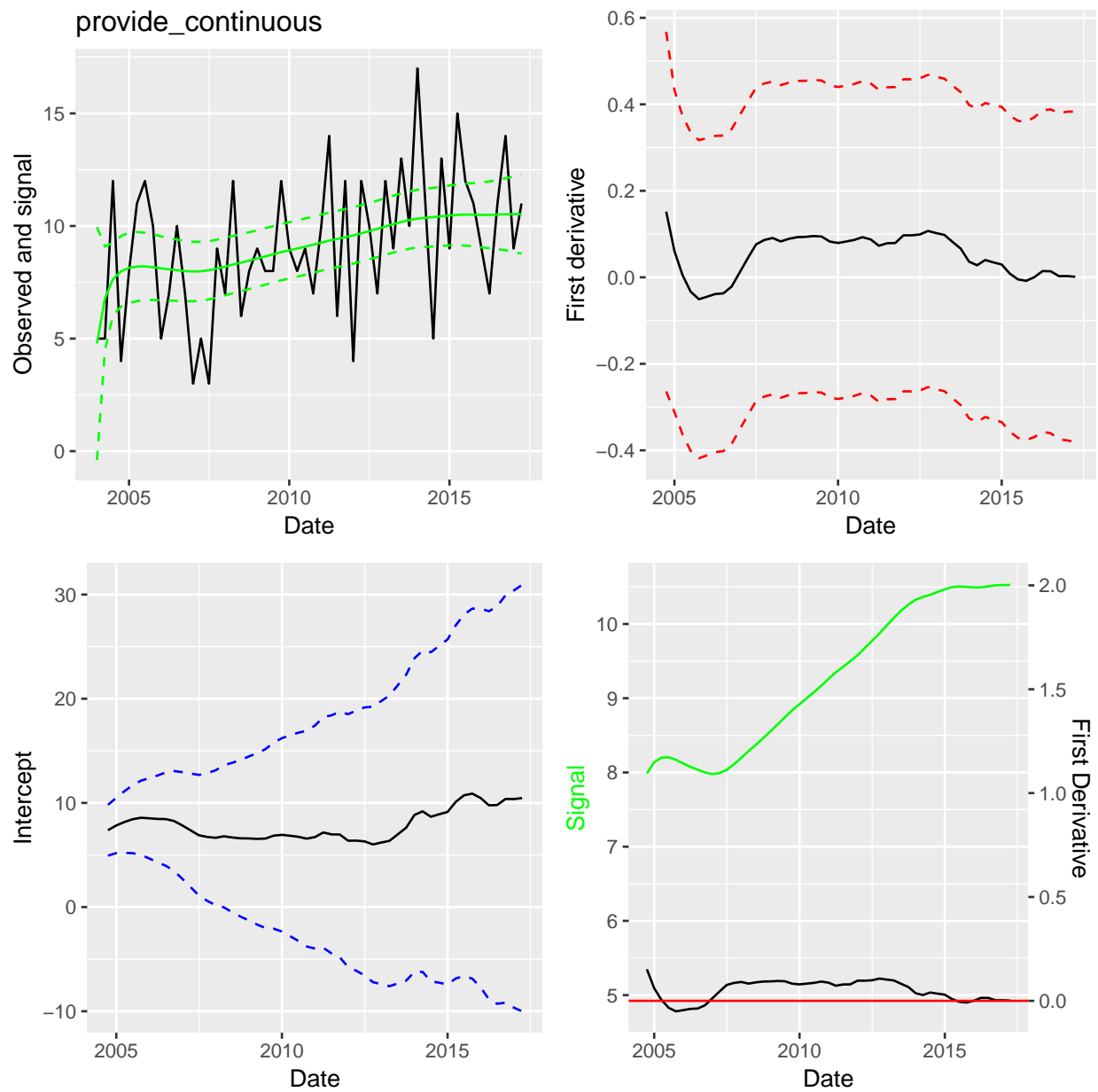
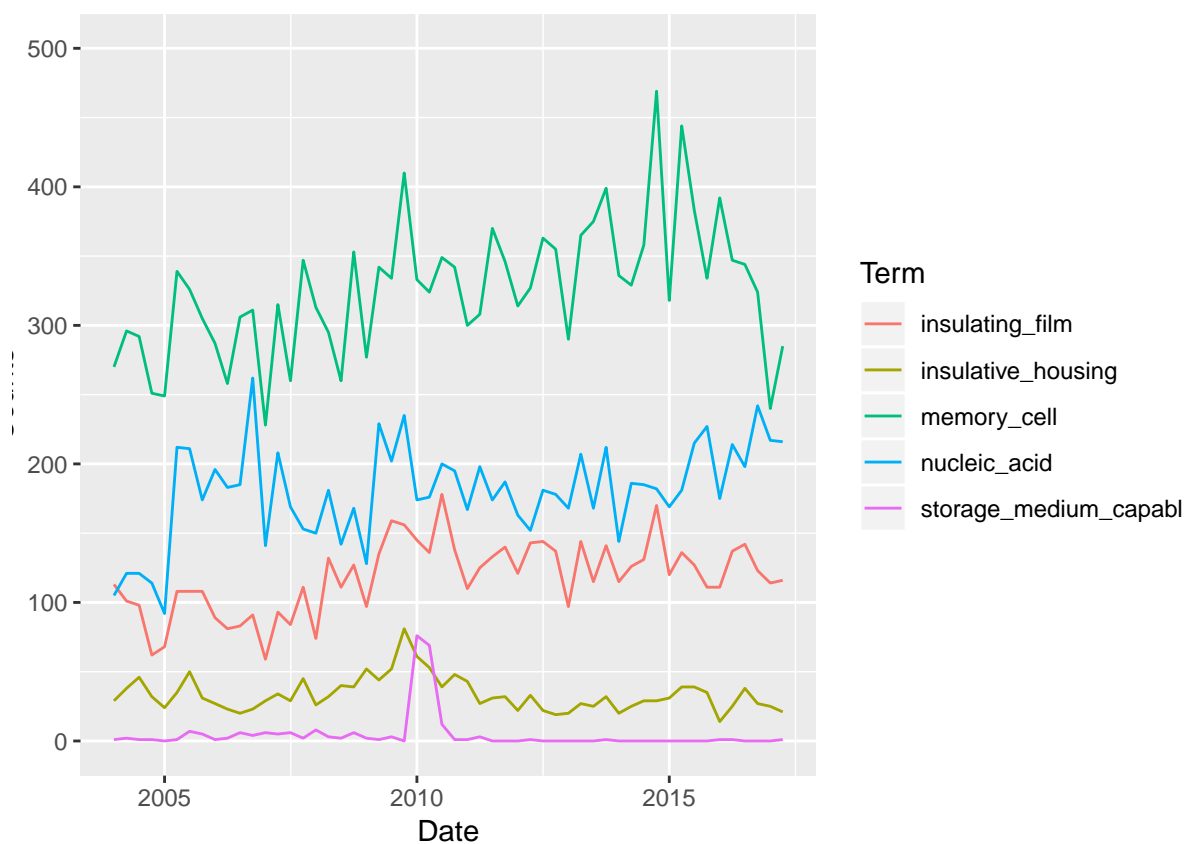
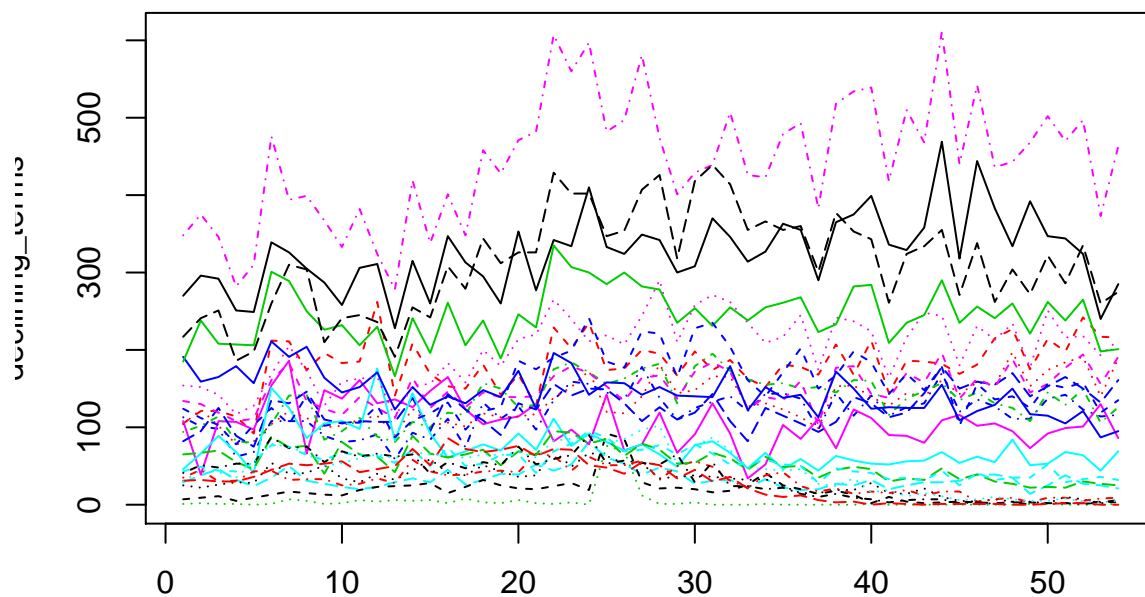
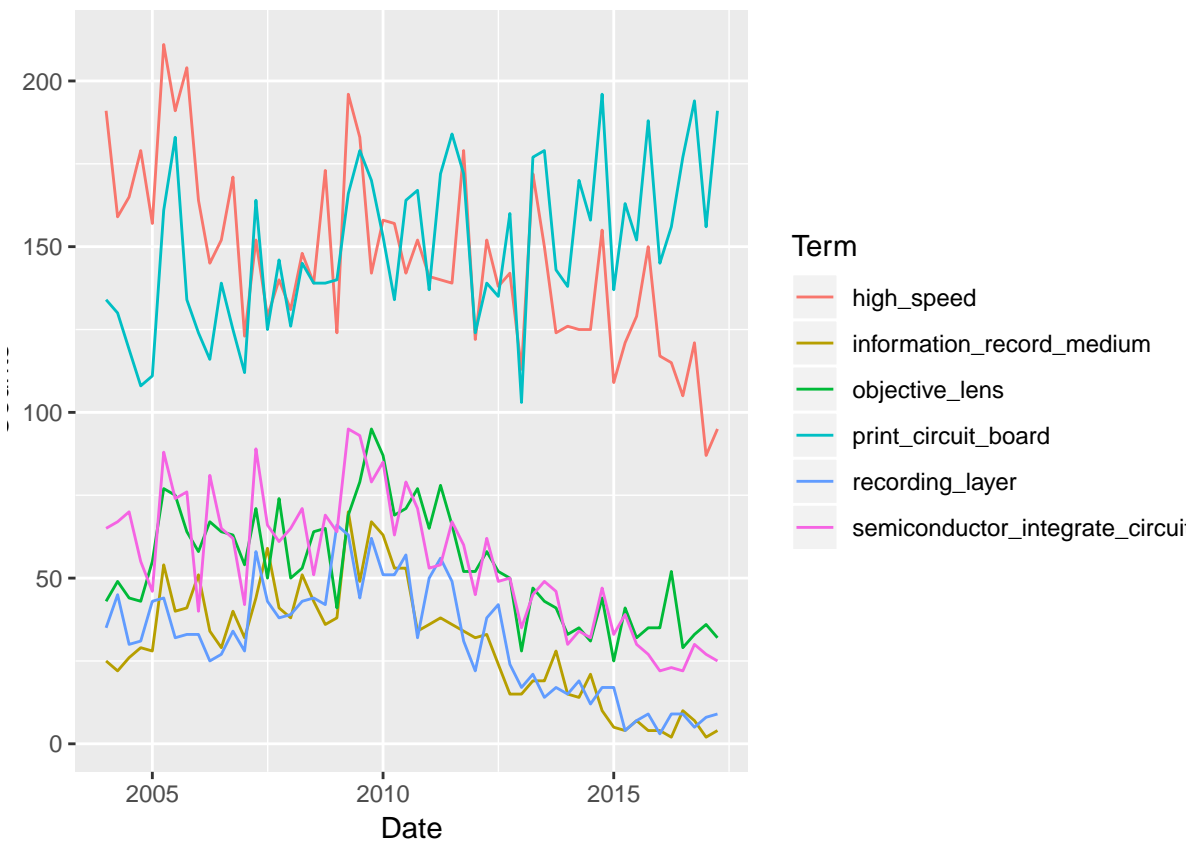
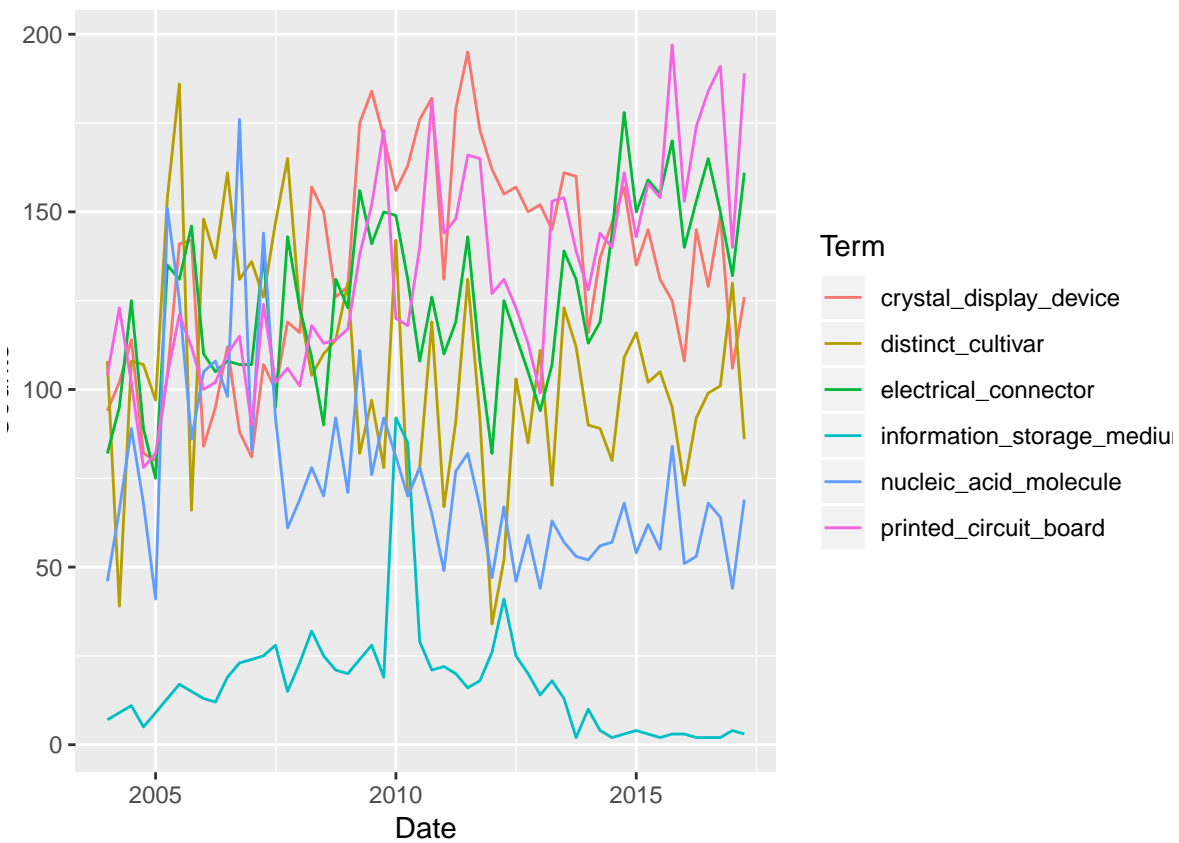


Figure 49:

Terms classified as declining by Porter

In this group we also find terms which are emergent (printed circuit board, liquid crystal, electrical connector, semiconductor substrate, print circuit board, memory cell), some terms which are stationary and some which are definitely declining and correspond to negative values of E .





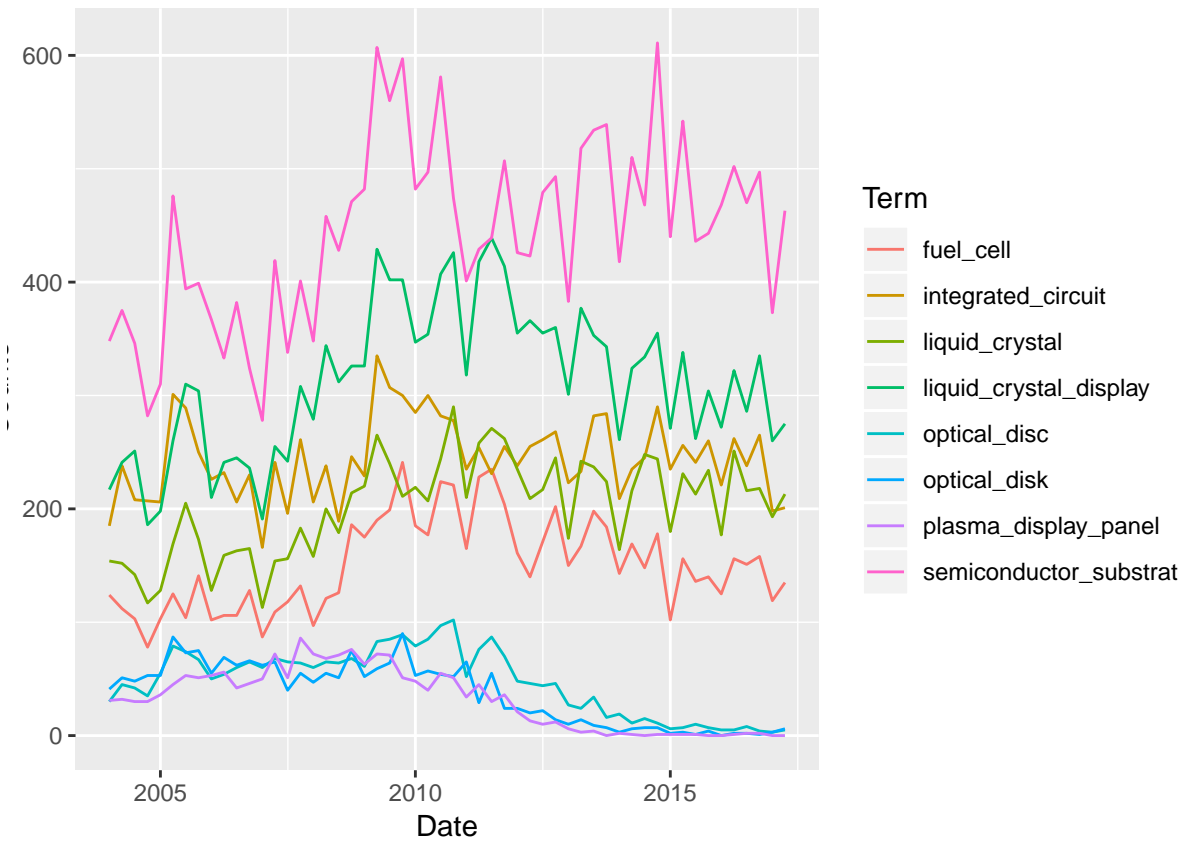


Table 3: Stationary Terms

Term	σ_ϵ	σ_ν	σ_η	δ	E (1st der)	E(sum 1st deriv/sd)
printed circuit board	16.8814054	2.6751424	0.0000027	0.6462724	72.4612968	4.2923735
liquid crystal	25.6584340	0.0000022	7.3869269	0.9559494	80.3576291	3.1318213
electrical connector	17.4718155	2.5759051	0.0000026	0.6144451	48.0061571	2.7476342
semiconductor substrate	52.5273253	0.0000001	19.9688921	0.9544847	128.8800173	2.4535804
print circuit board	20.2674119	0.0000001	0.0000017	0.9935272	41.3127906	2.0383851
memory cell	41.1568180	0.0000000	0.0000000	0.9588138	79.8300160	1.9396547
nucleic acid	29.3694351	0.0000000	0.0000000	0.7605834	55.2646492	1.8817062
insulating film	16.9523824	3.9751699	0.0000000	0.4103918	31.2674991	1.8444310
liquid crystal display	35.8623571	3.7351631	0.0000118	0.8643371	53.1181021	1.4811659
crystal display device	17.3617720	1.7821525	0.0000001	0.8474400	24.8412181	1.4307997
fuel cell	22.1290683	3.8201550	0.0000002	0.7515689	29.1086671	1.3154041
storage medium capable	0.0161356	13.2091776	0.0000000	0.0176129	0.0176097	1.0913582
distinct cultivar	25.6257774	1.4969103	0.0000000	0.7578396	-3.2214607	-0.1257117
integrated circuit	27.0417950	11.1410527	0.0000051	0.0000000	-5.0913343	-0.1882765
information storage medium	7.9695554	8.8728133	0.0000001	0.0186871	-6.3477402	-0.7964987
nucleic acid molecule	20.3585577	1.5392813	0.0000005	0.7107985	-16.9662618	-0.8333725
information record medium	6.8492157	0.6361397	3.4505653	0.8588028	-8.1577469	-1.1910483
objective lens	8.9810872	1.5143200	0.0000694	0.7459124	-17.0764176	-1.9013753
insulative housing	6.4863436	5.1294048	0.0000003	0.0525121	-15.4103872	-2.3758204
high speed	19.7910564	0.0000000	0.0000003	1.0000000	-60.5764818	-3.0608008
semiconductor integrate circuit	10.8518345	0.0000000	3.2936036	1.0000000	-42.1181054	-3.8811968
recording layer	7.5328904	0.8612771	0.0000007	0.8809884	-29.4087321	-3.9040435
optical disc	8.6242686	1.8504228	0.0000008	0.8269351	-41.4522790	-4.8064689
plasma display panel	6.3890714	1.2127295	0.0000001	0.8960973	-32.9283019	-5.1538478
optical disk	8.6028211	1.6913815	0.0000005	0.7846082	-50.1396879	-5.8282844

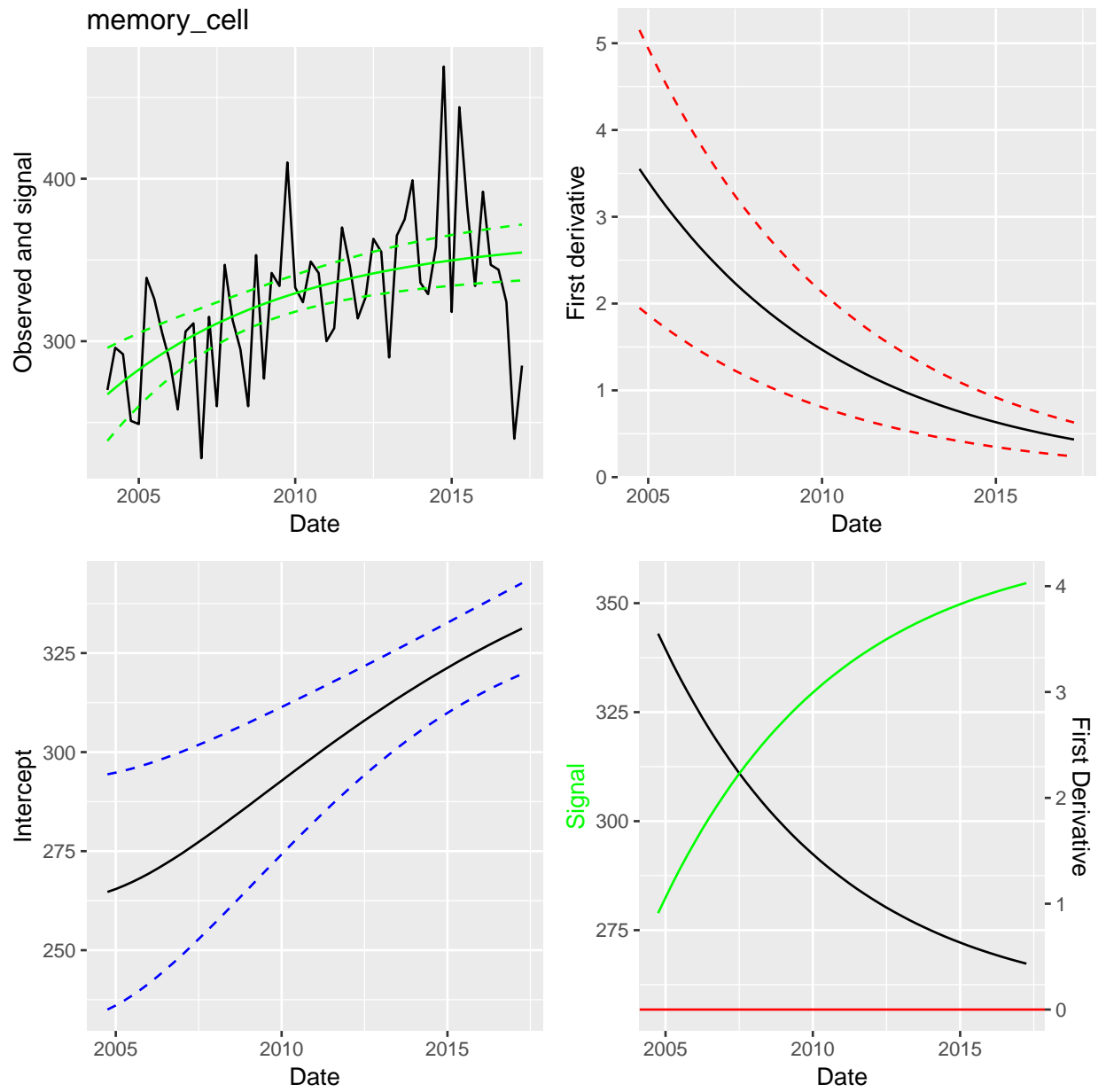


Figure 50:

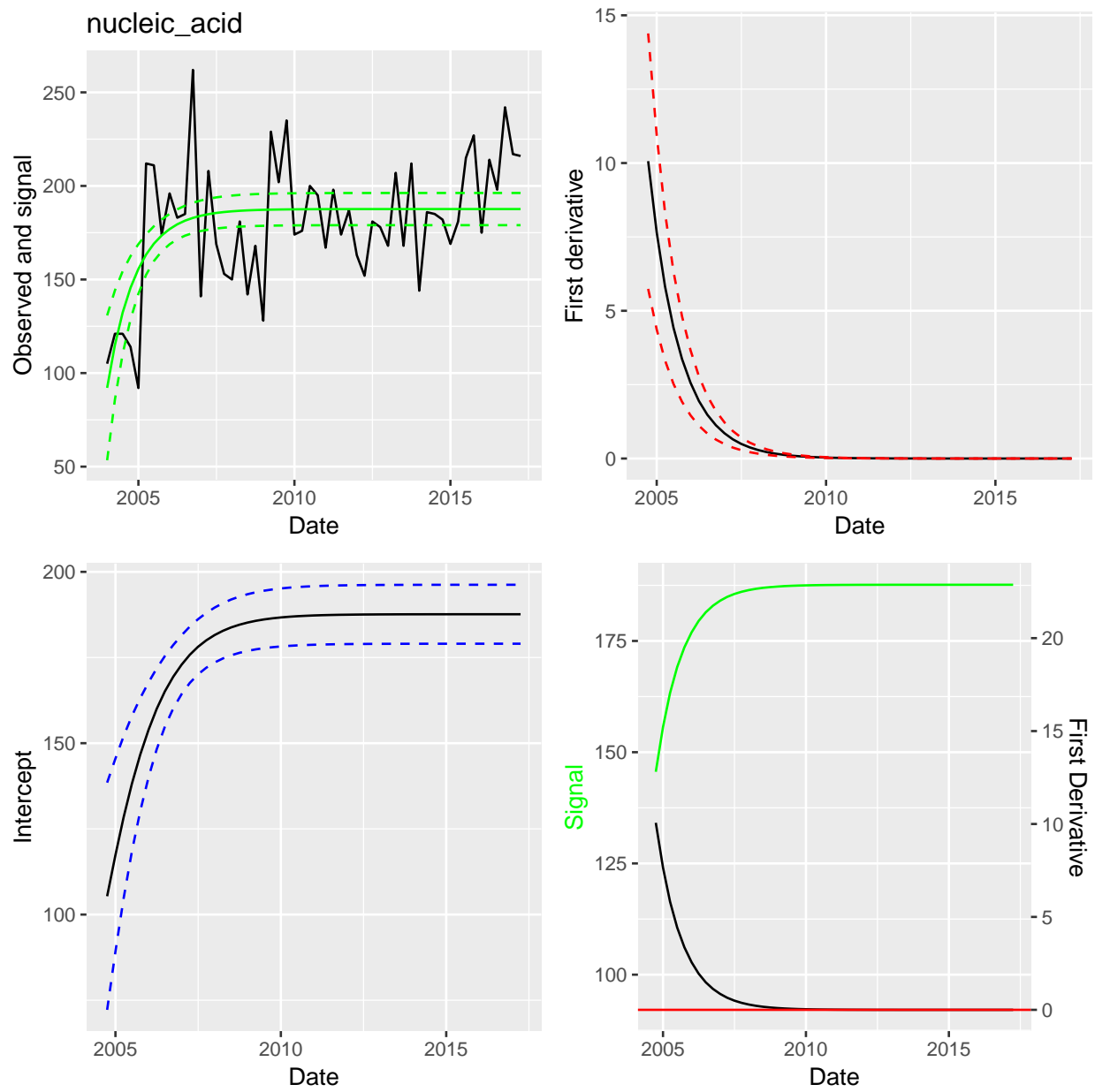


Figure 51:

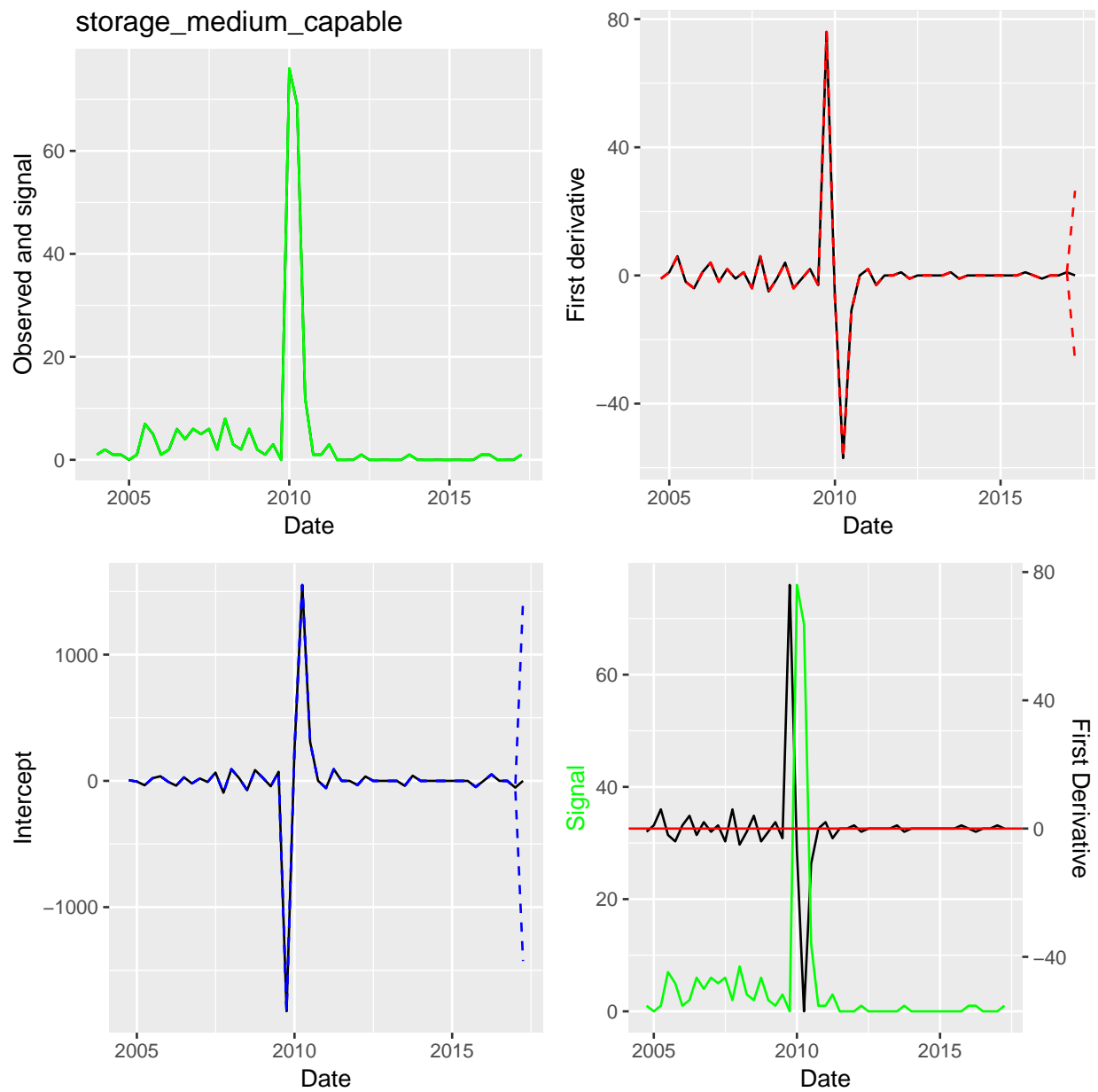


Figure 52:

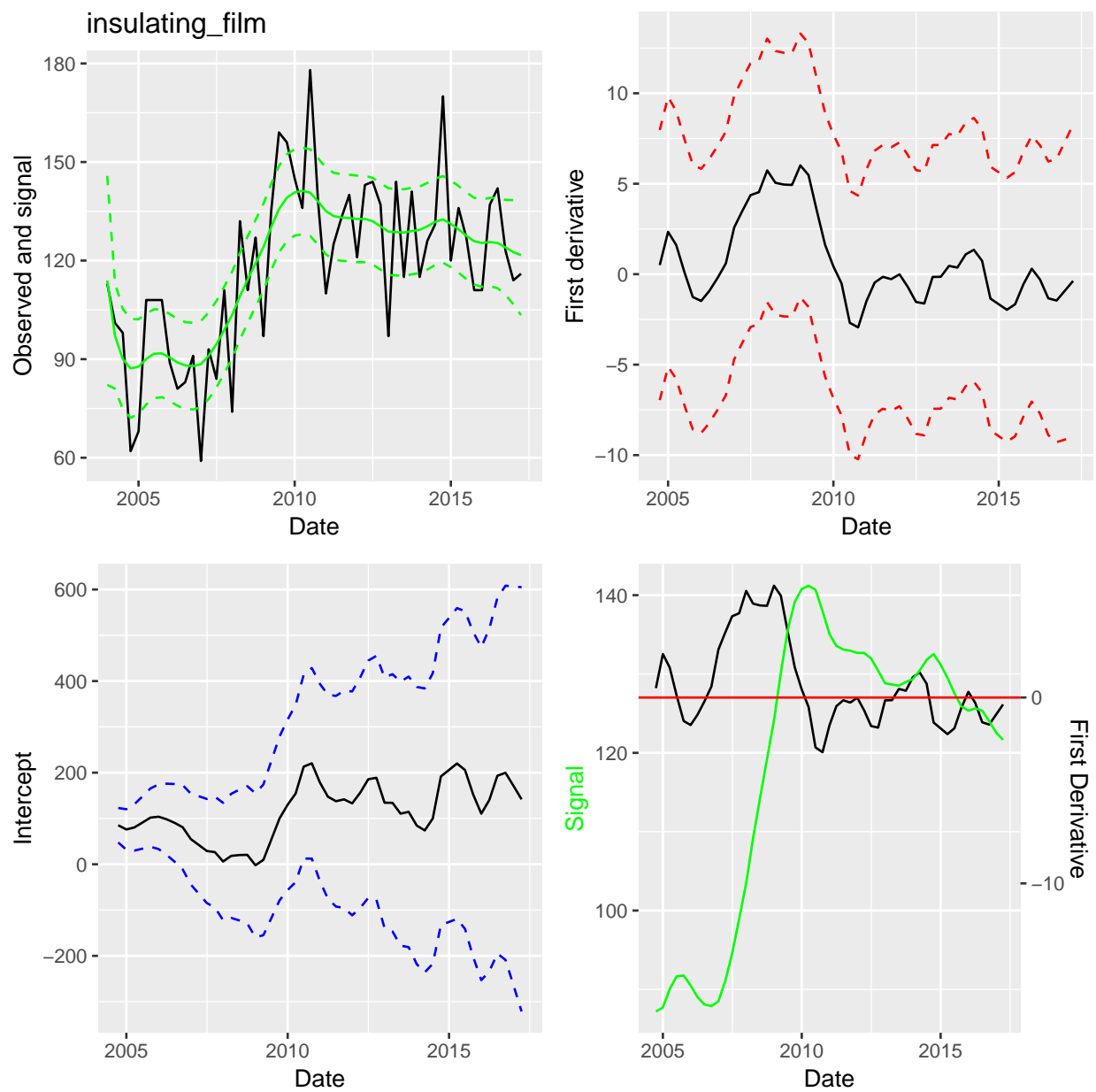


Figure 53:

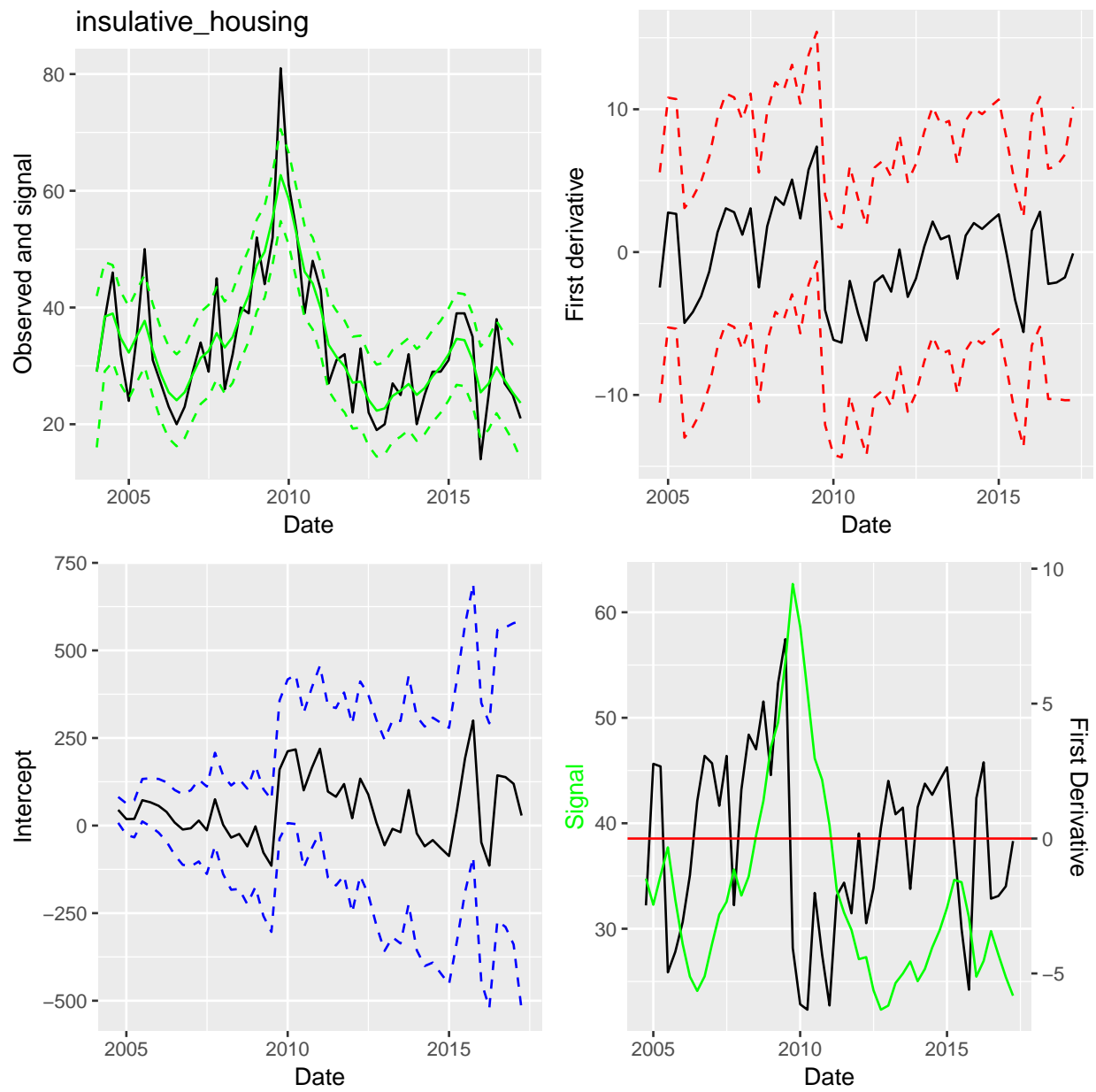


Figure 54:

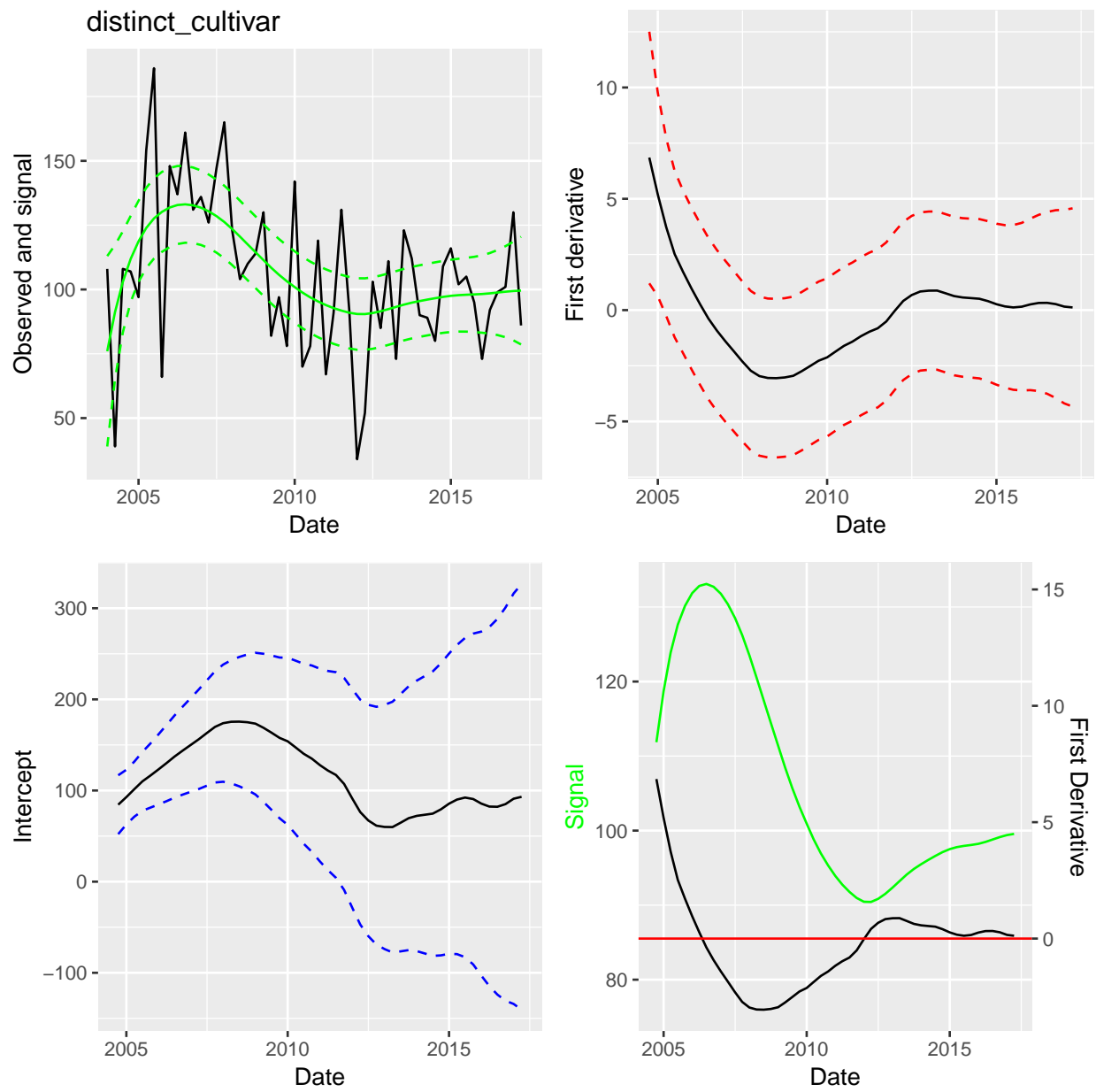


Figure 55:

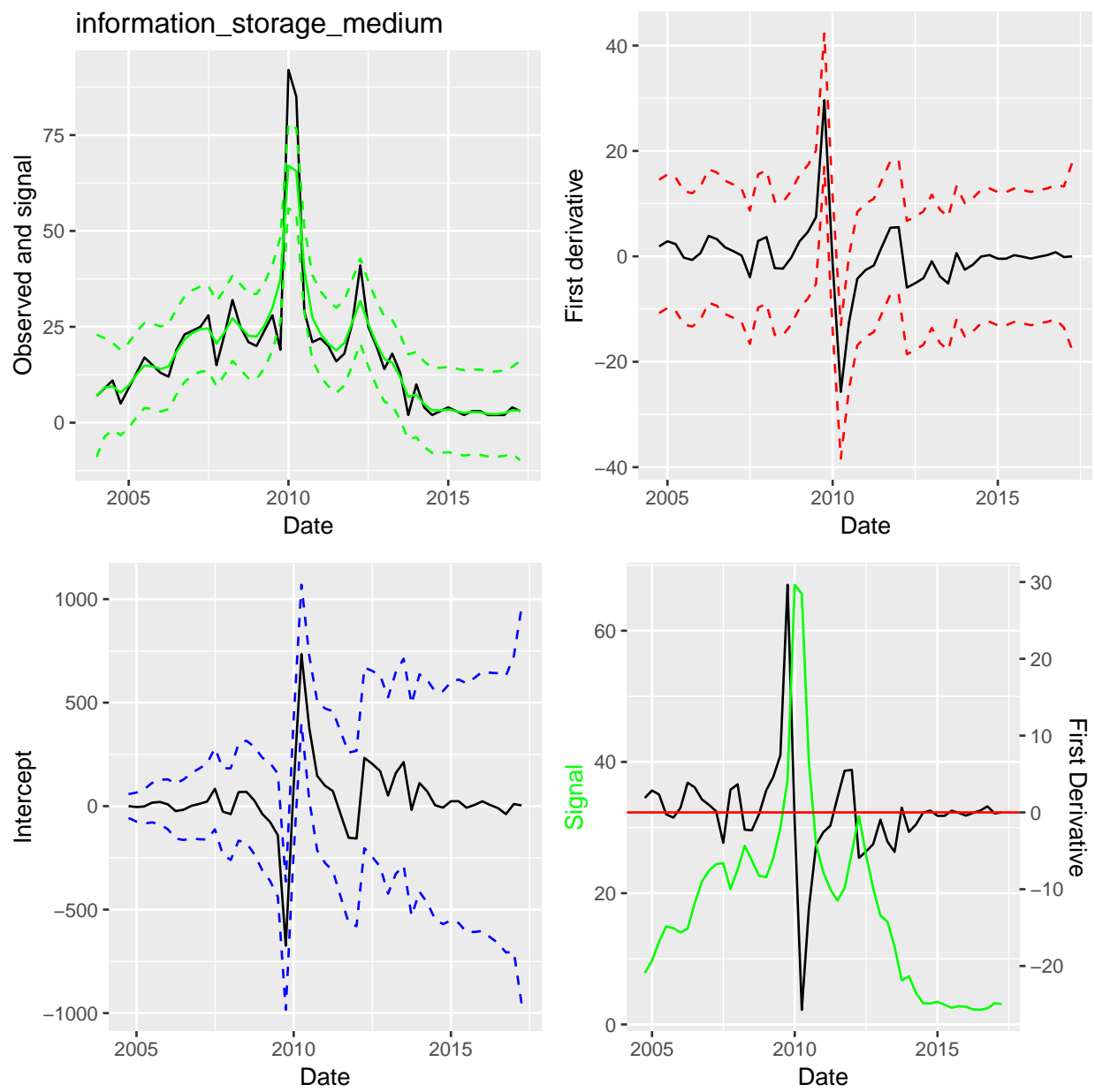


Figure 56:

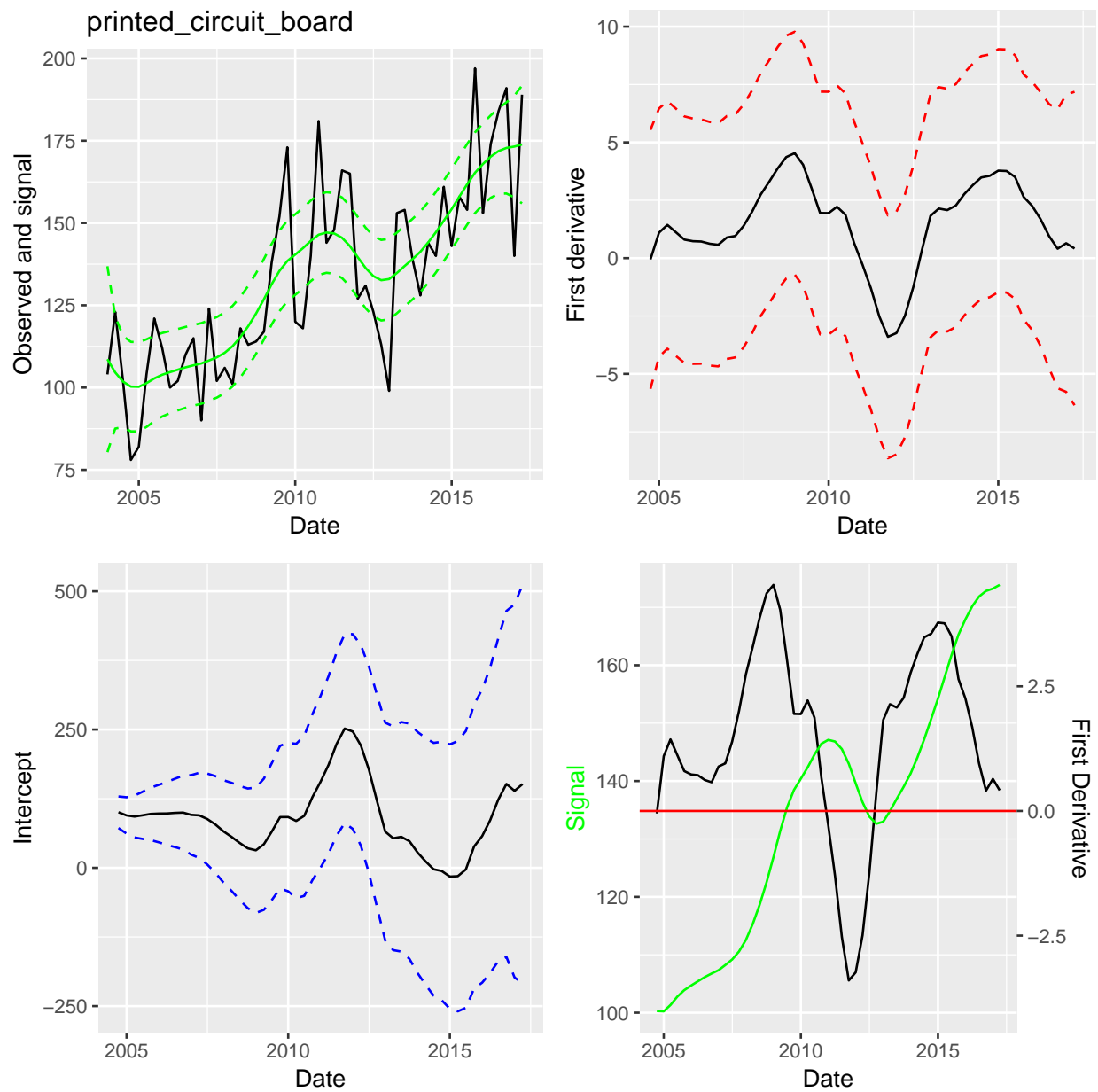


Figure 57:

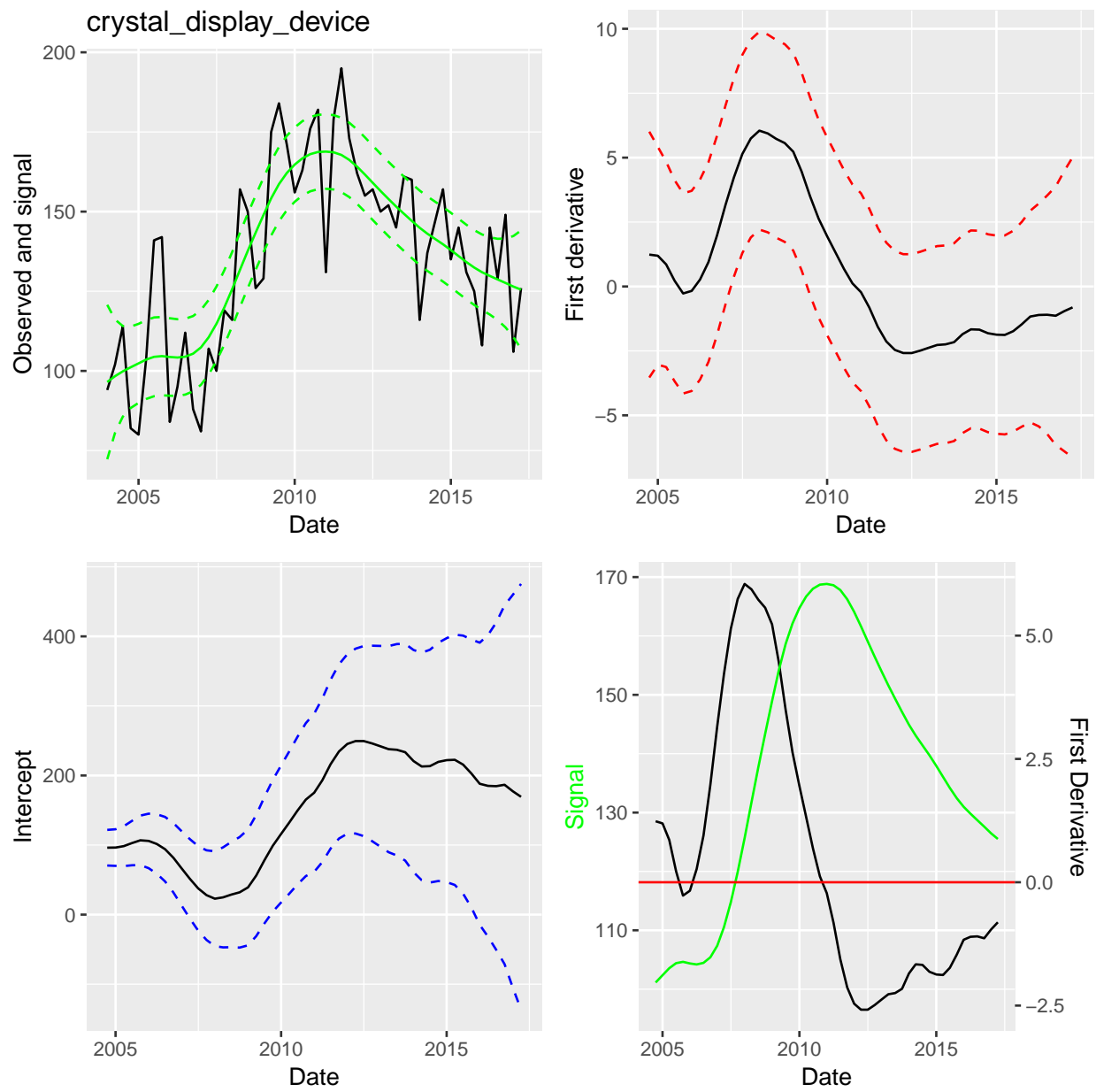


Figure 58:

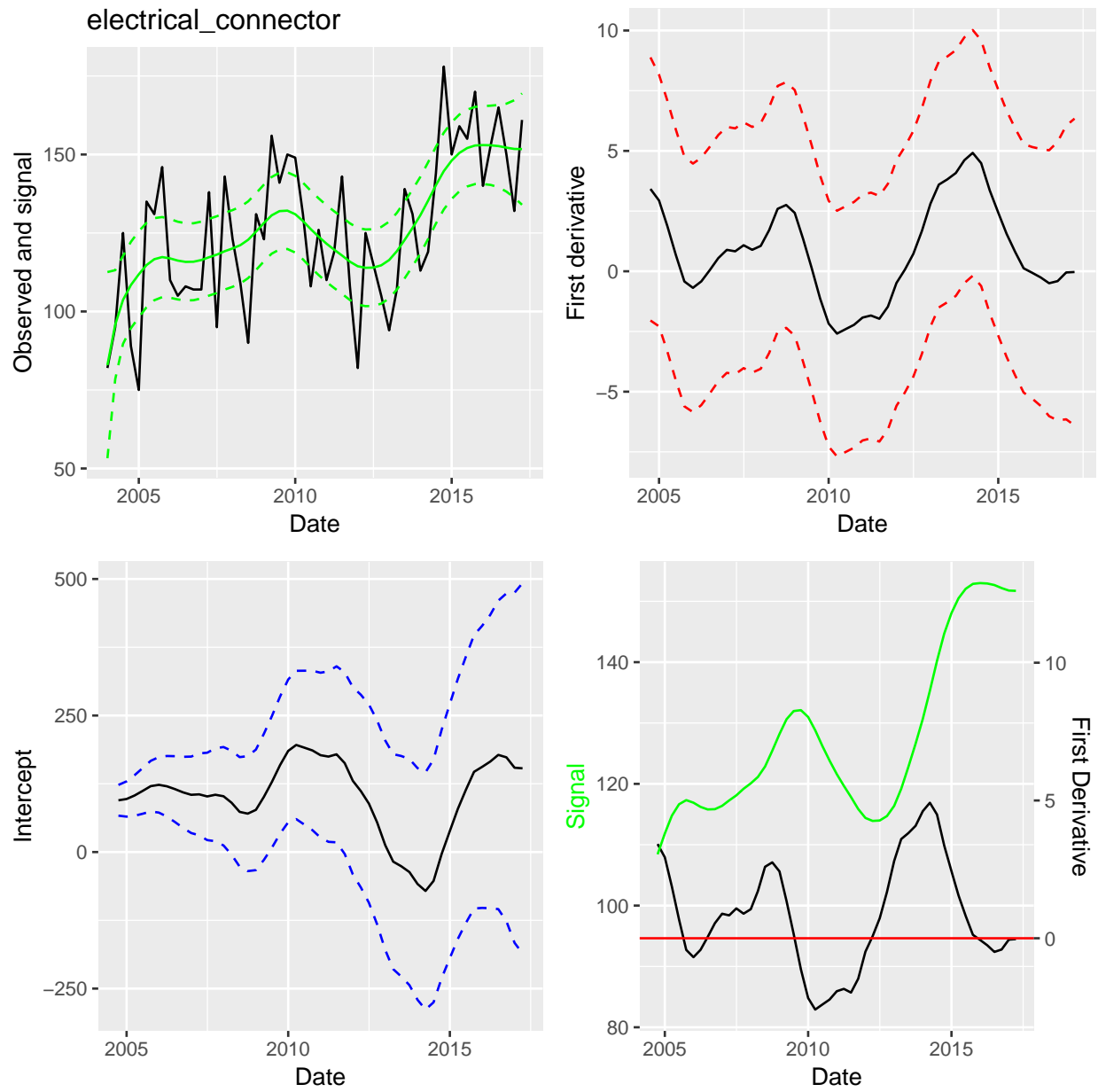


Figure 59:

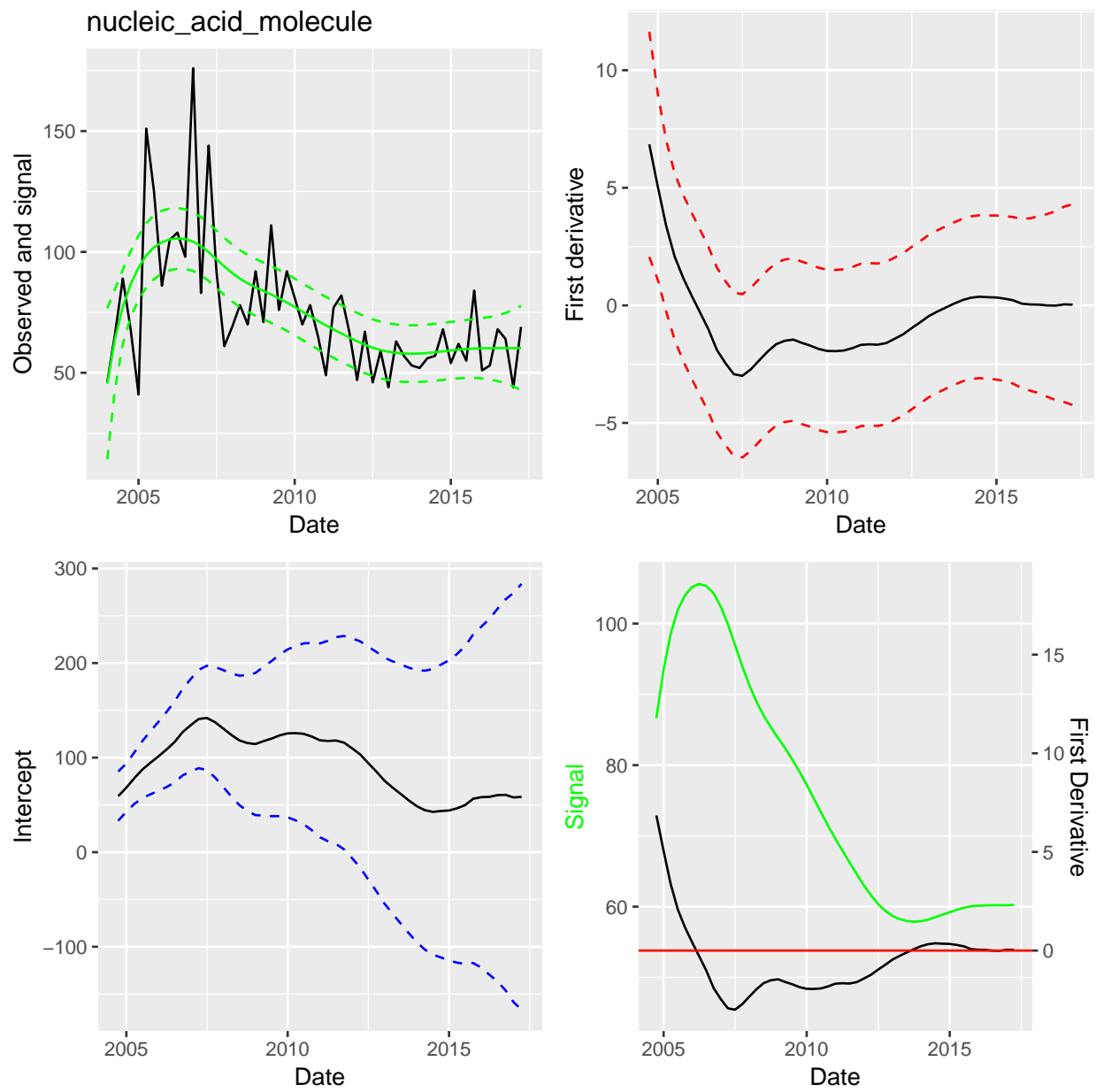


Figure 60:

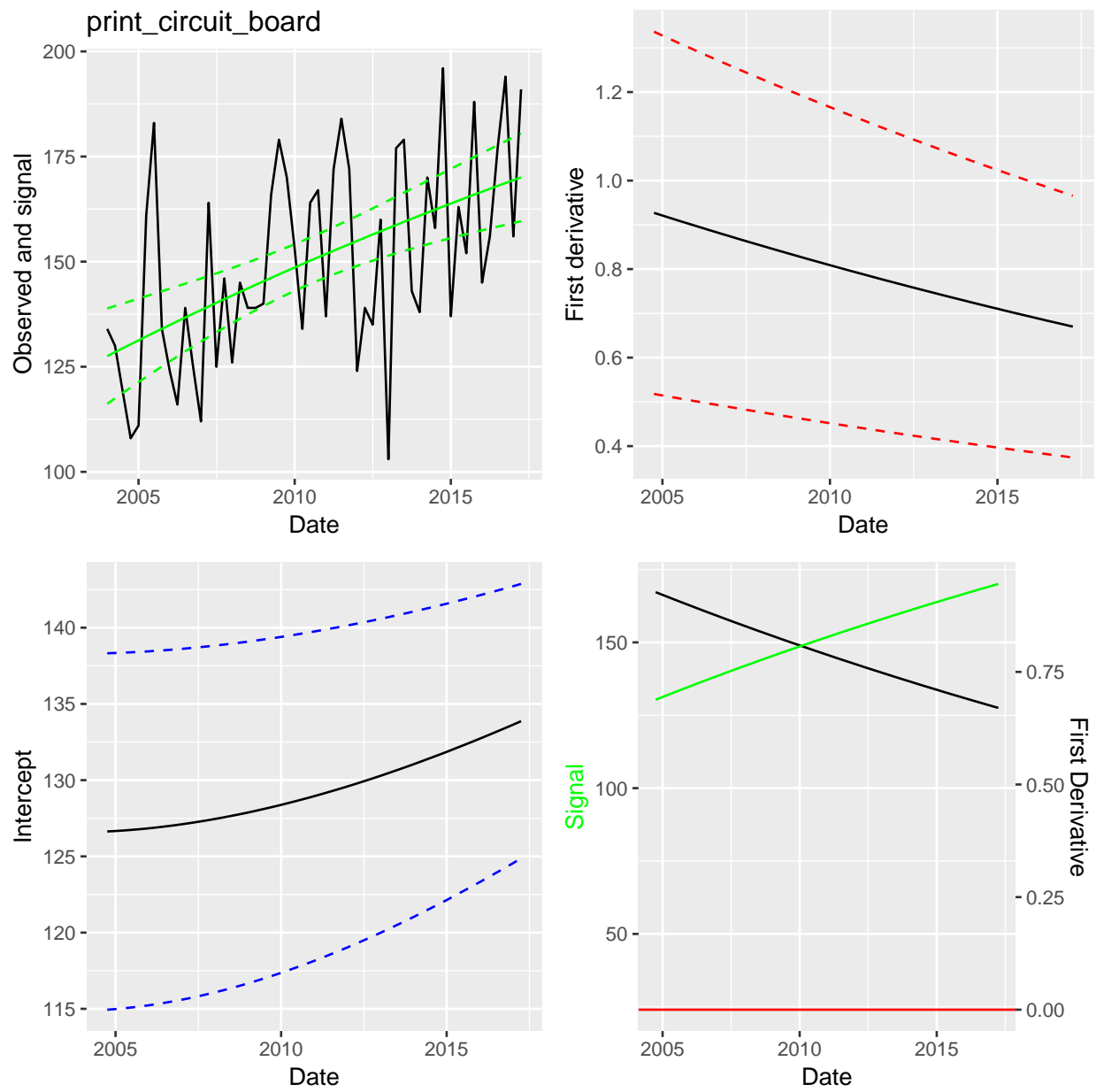


Figure 61:

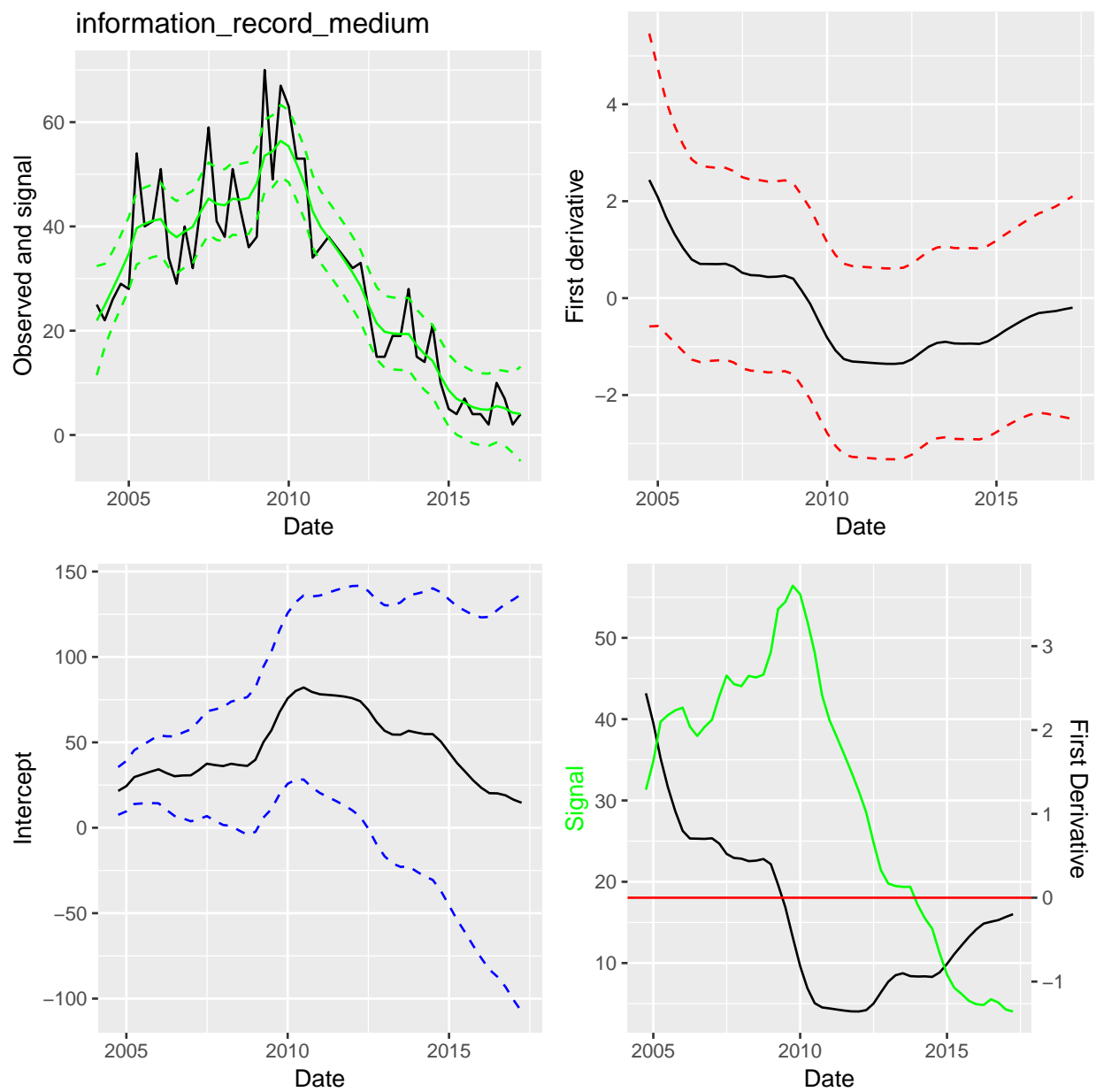


Figure 62:

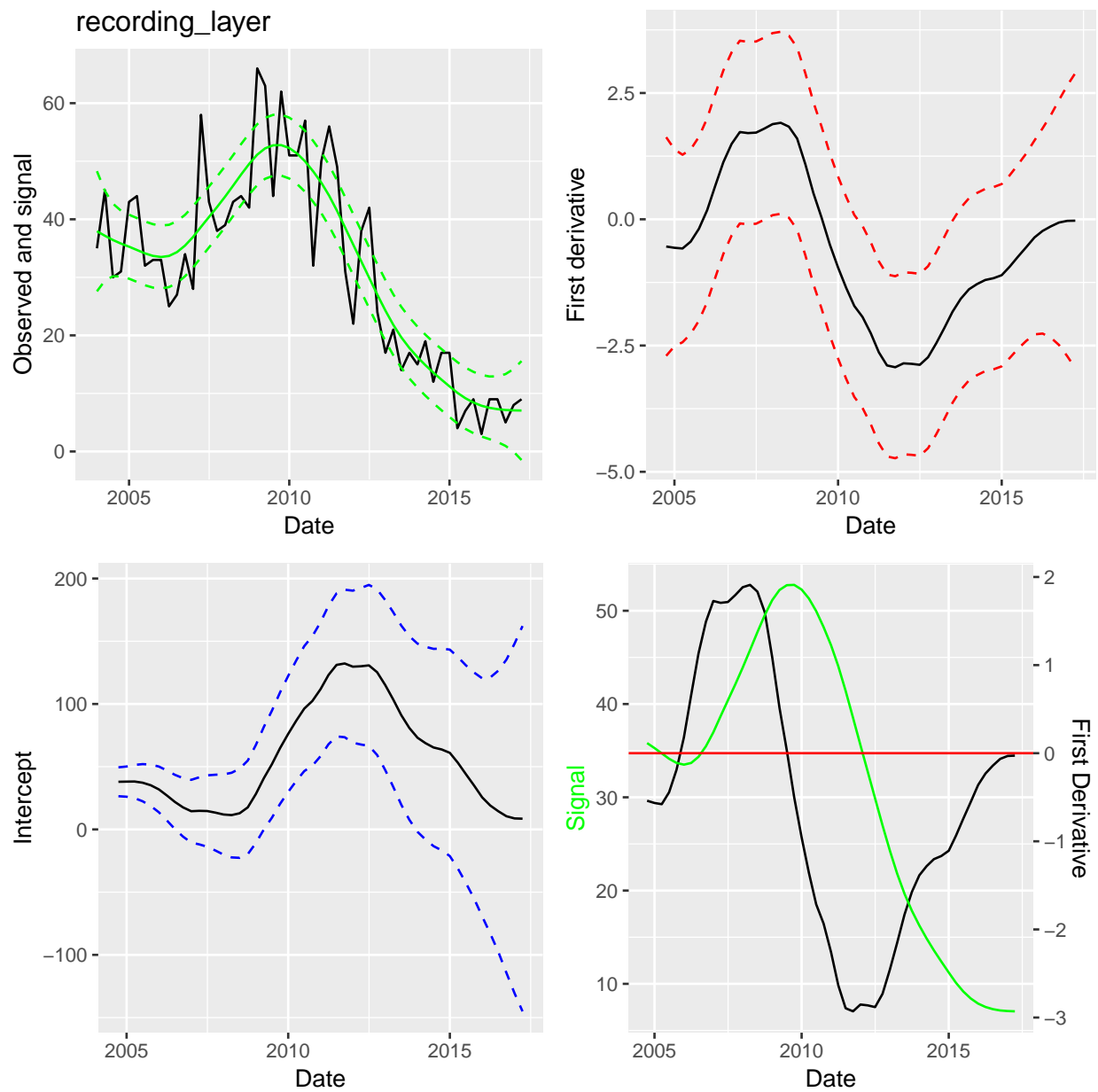


Figure 63:

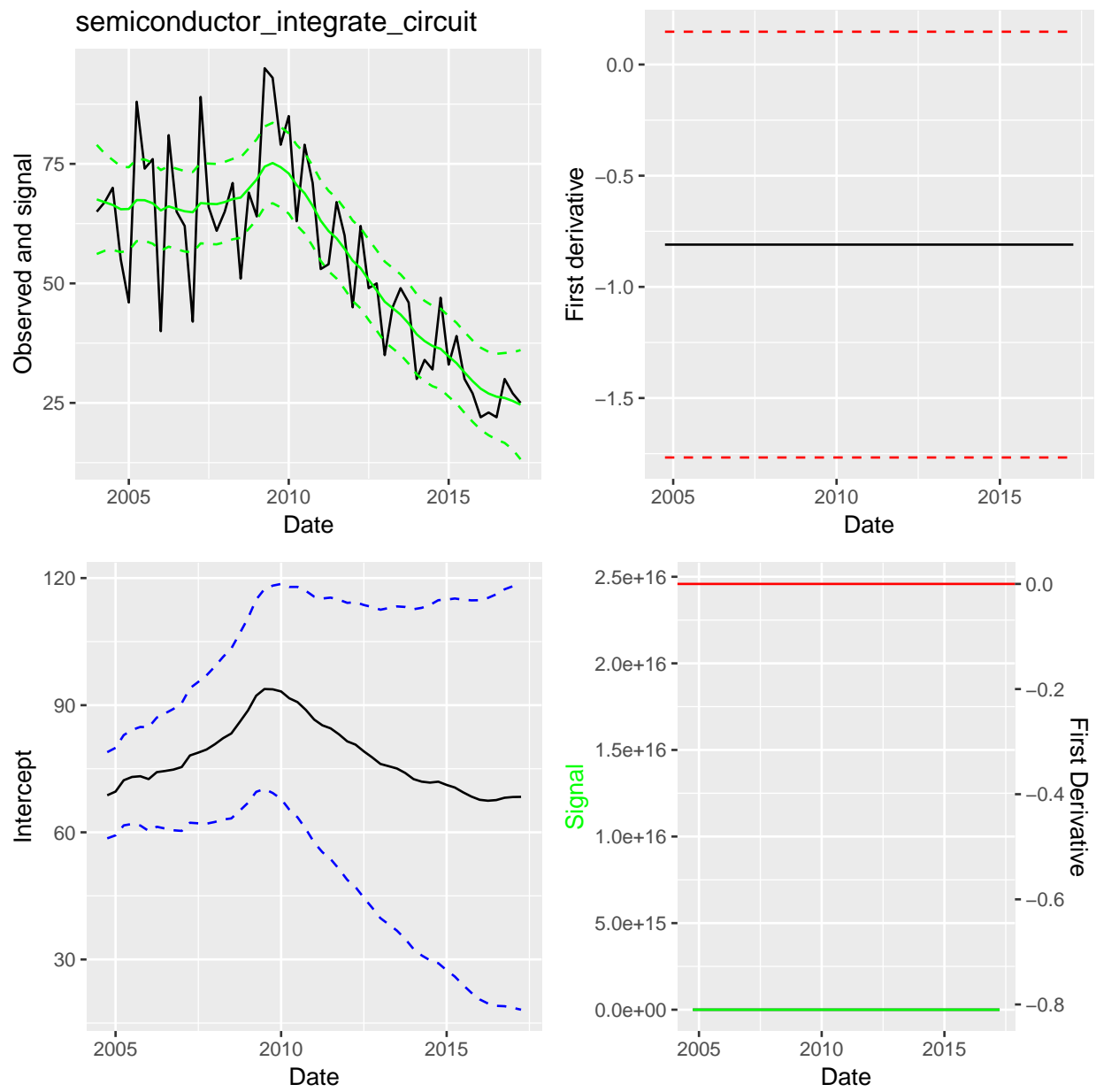


Figure 64:

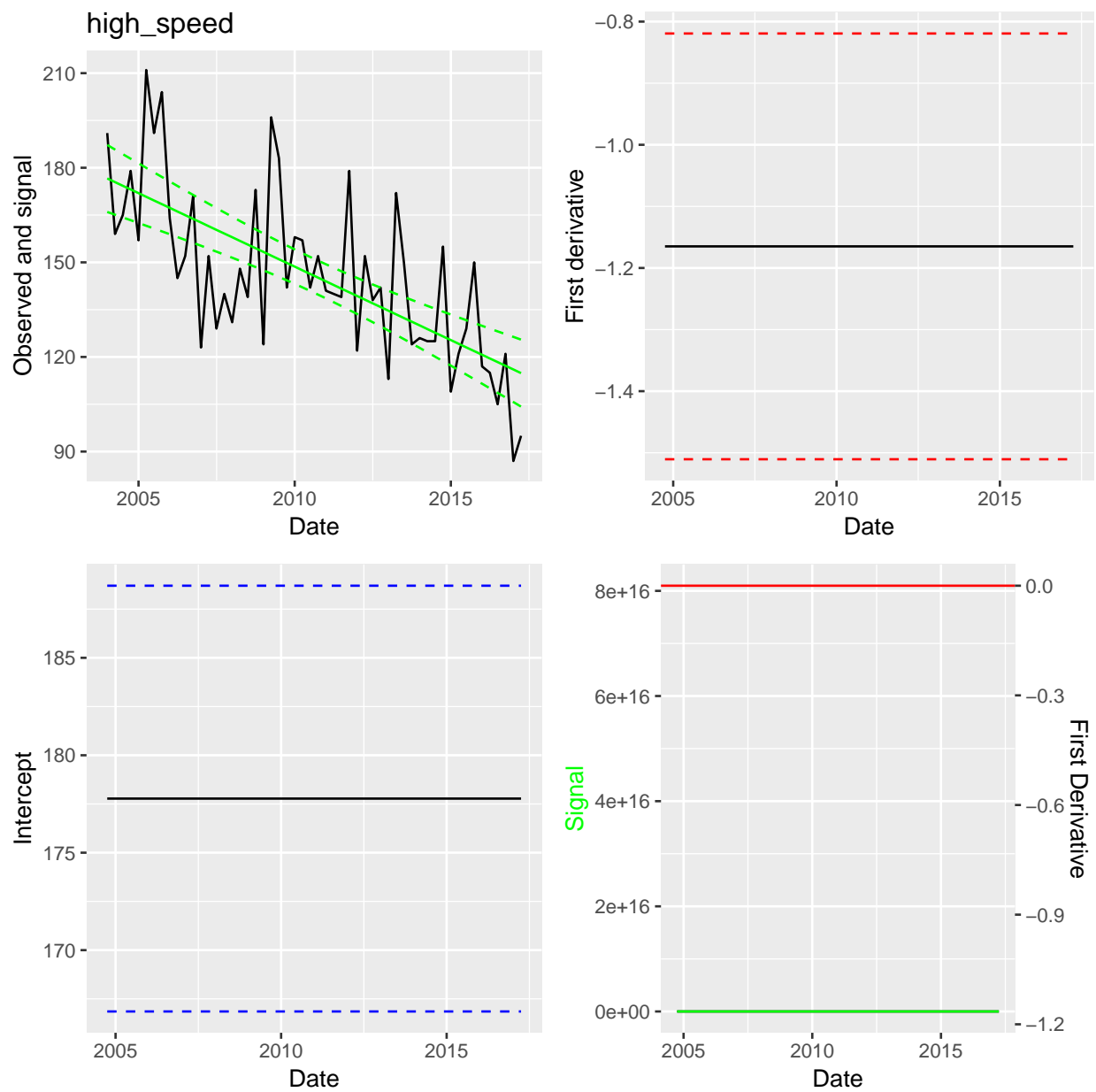


Figure 65:

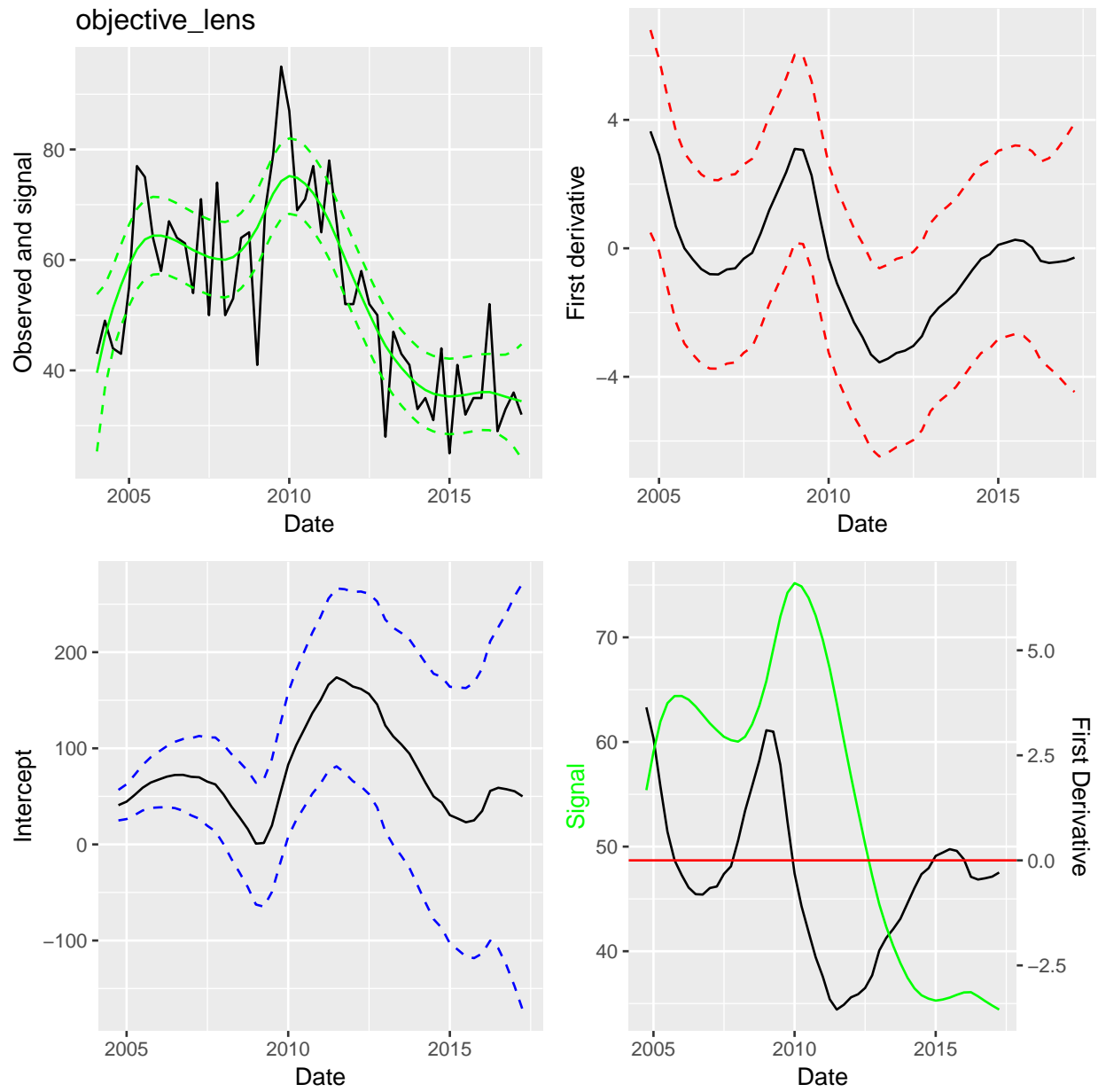


Figure 66:

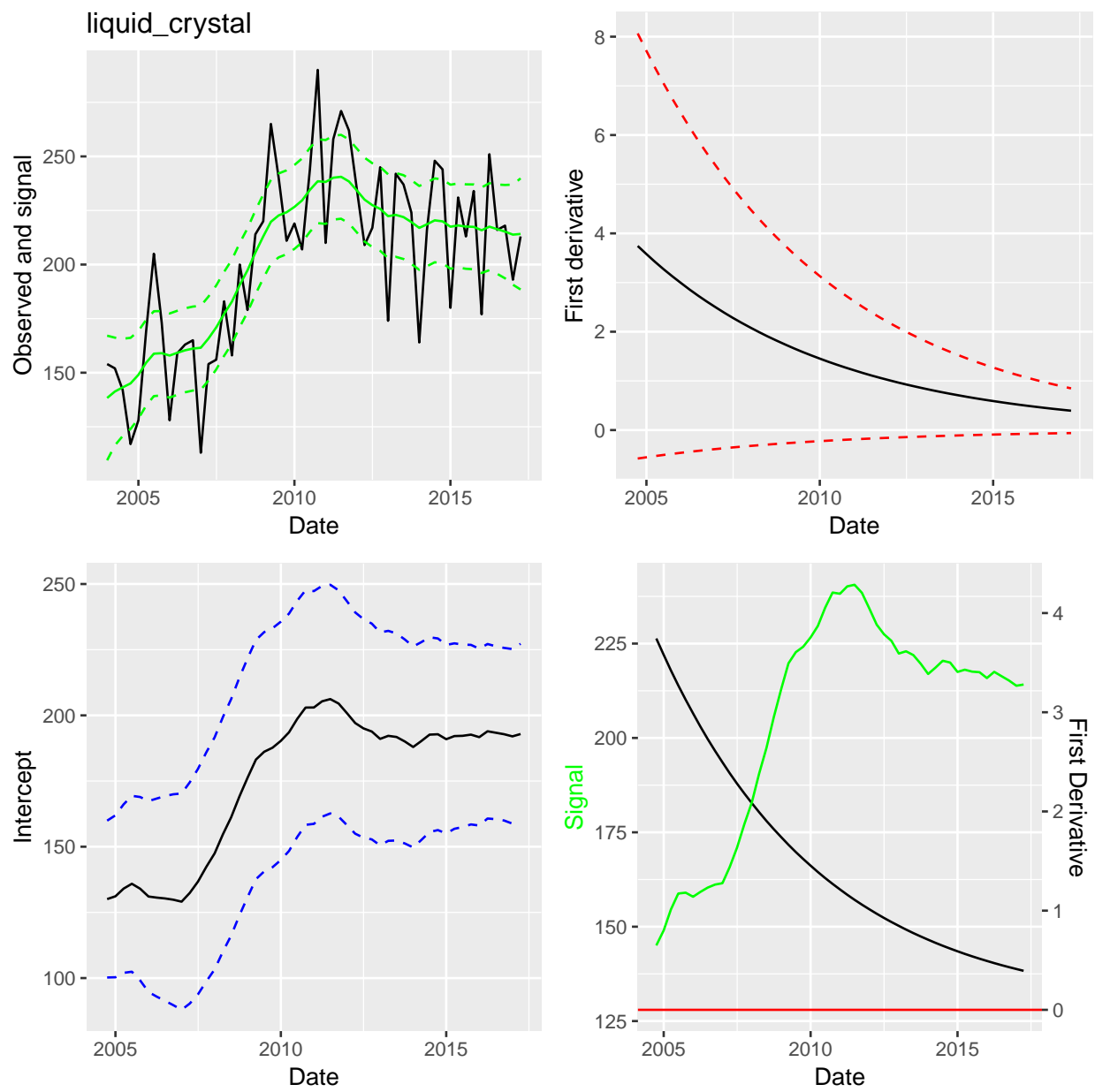


Figure 67:

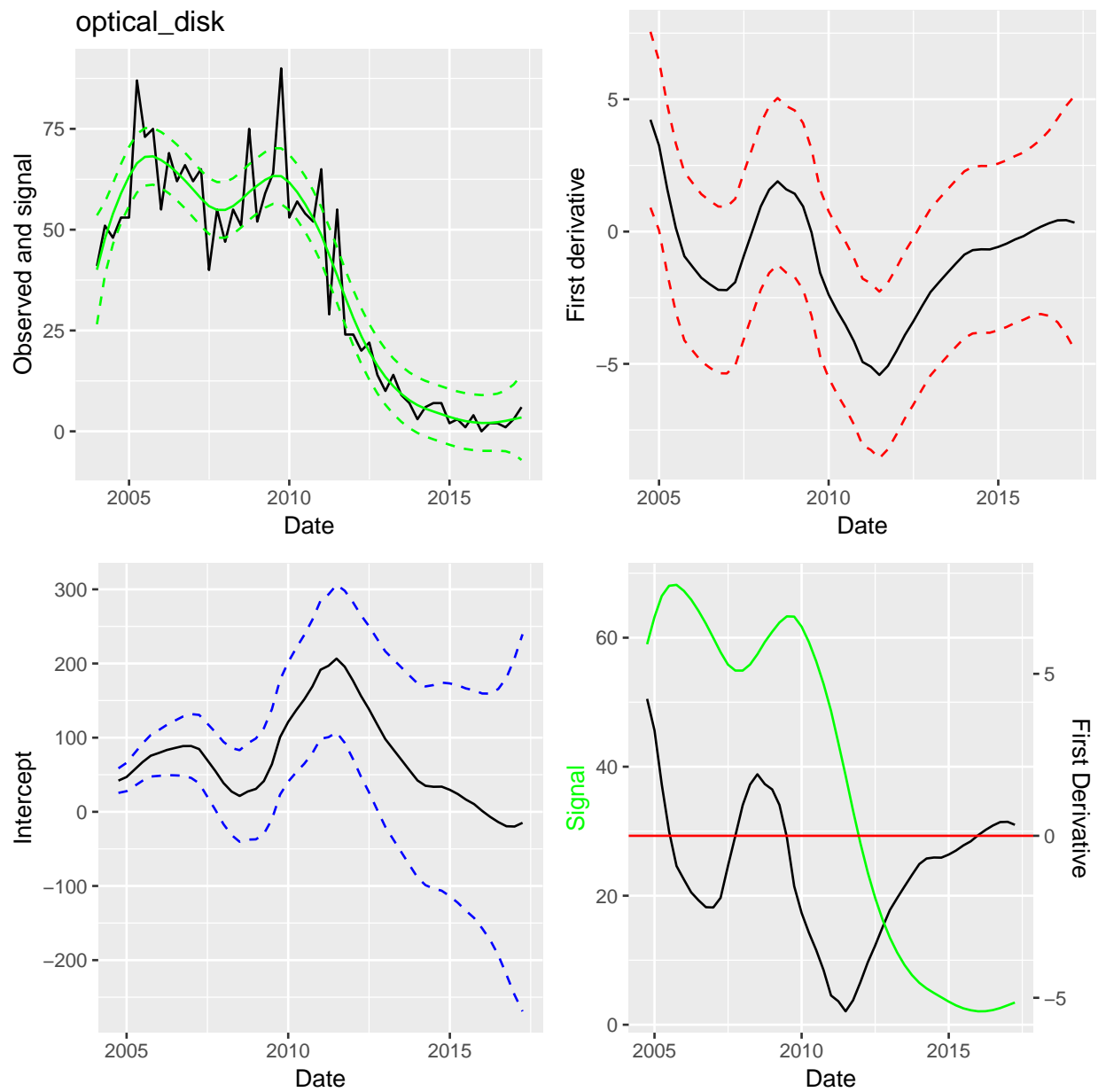


Figure 68:

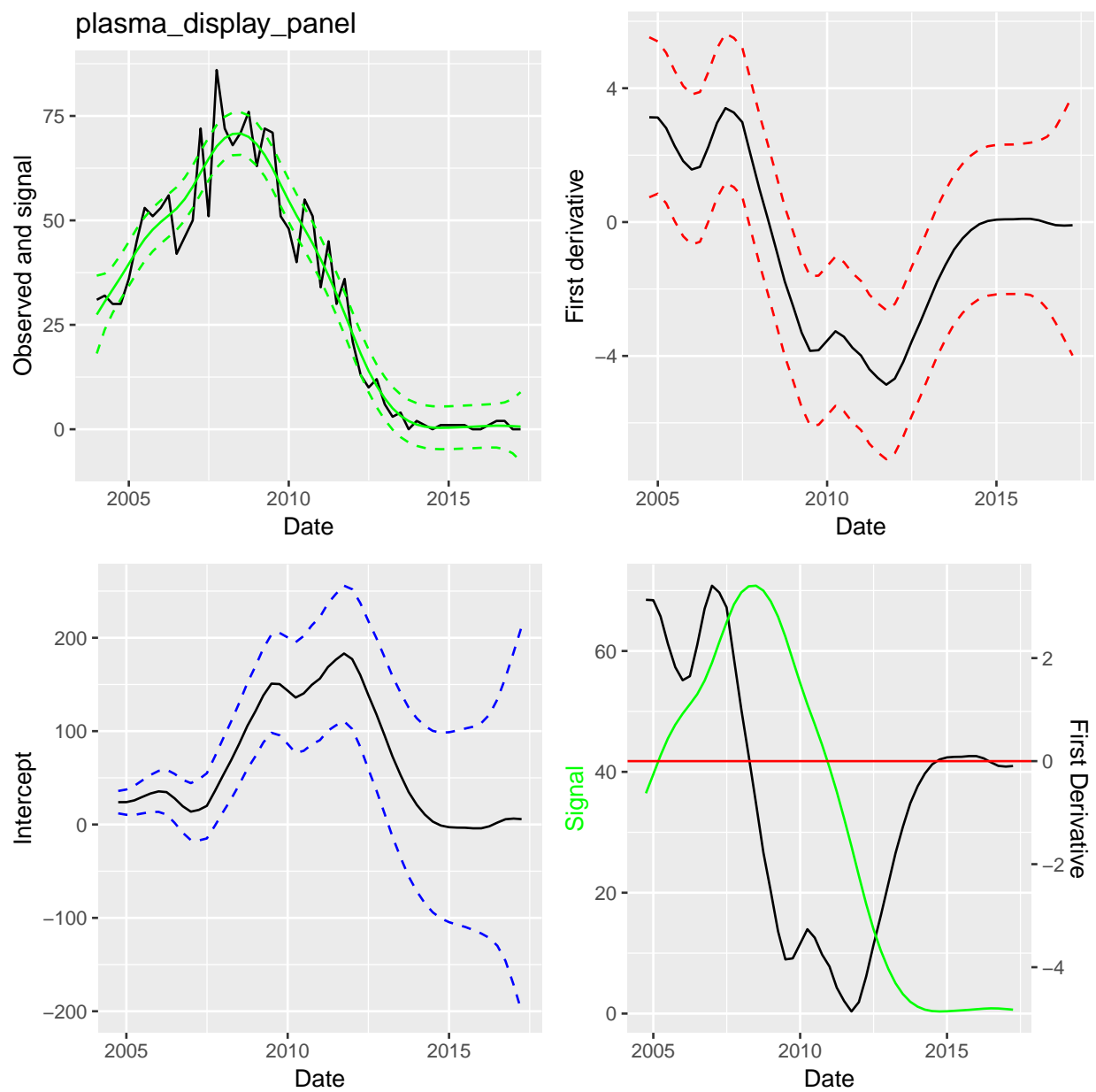


Figure 69:

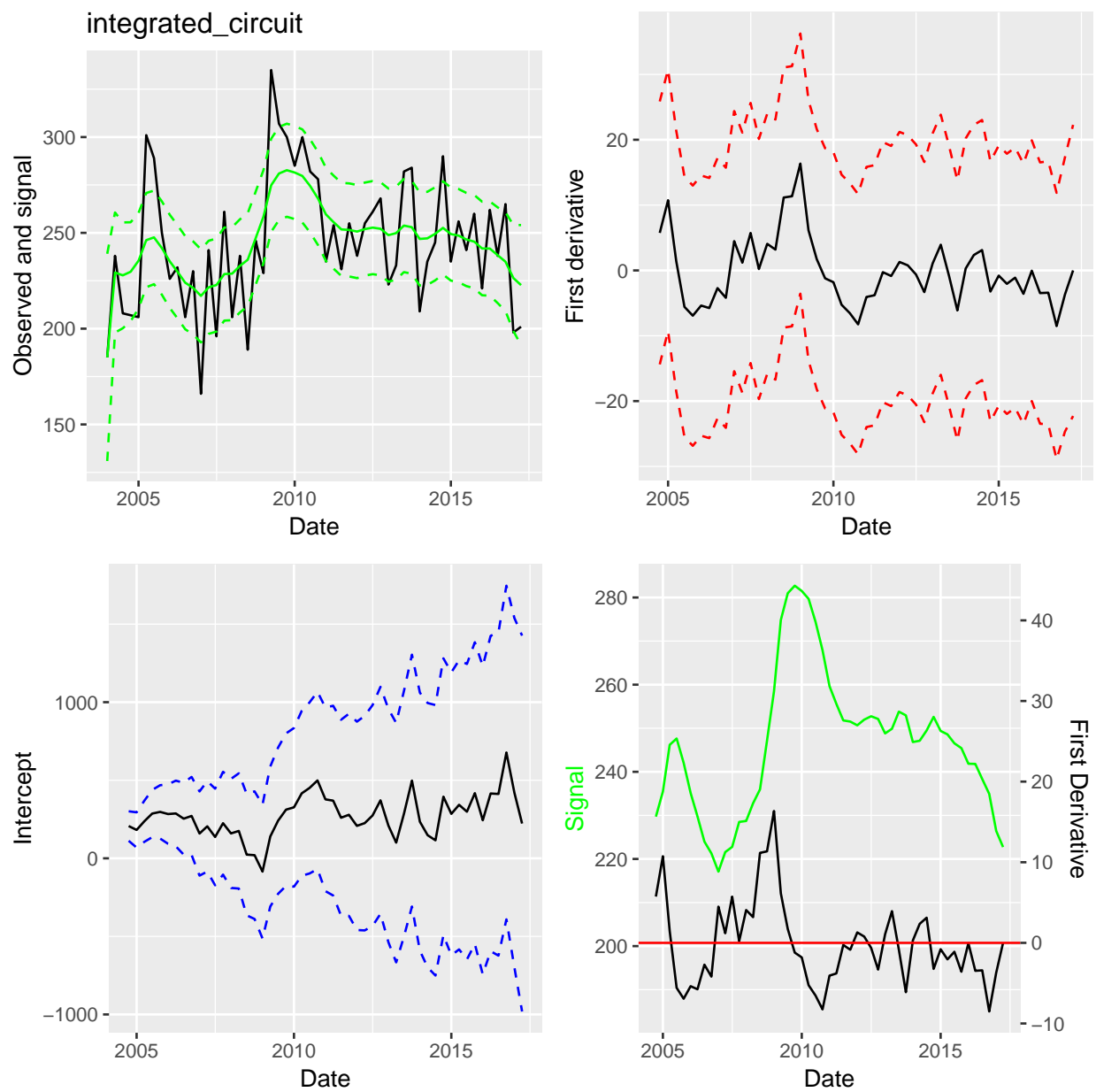


Figure 70:

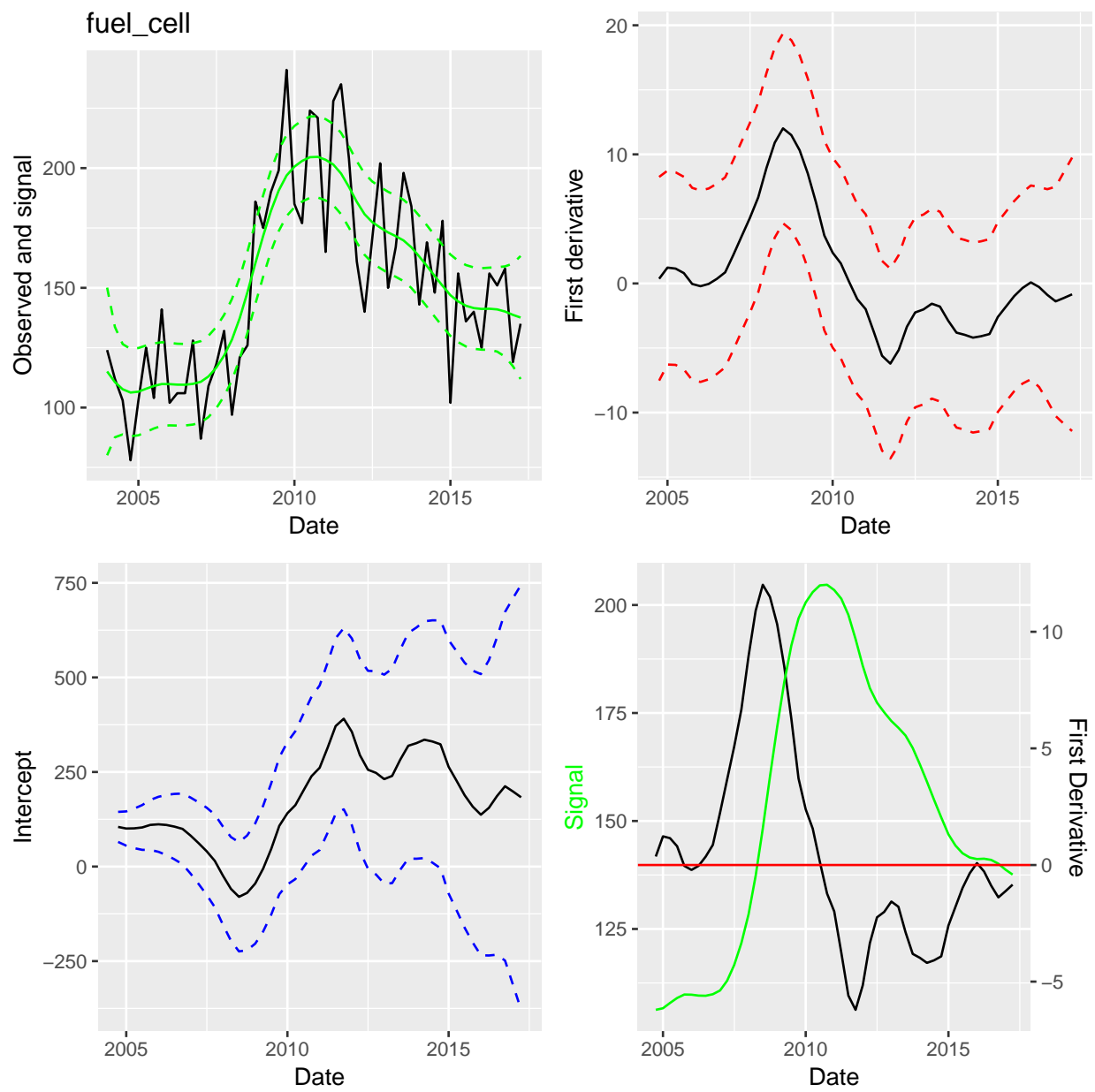


Figure 71:

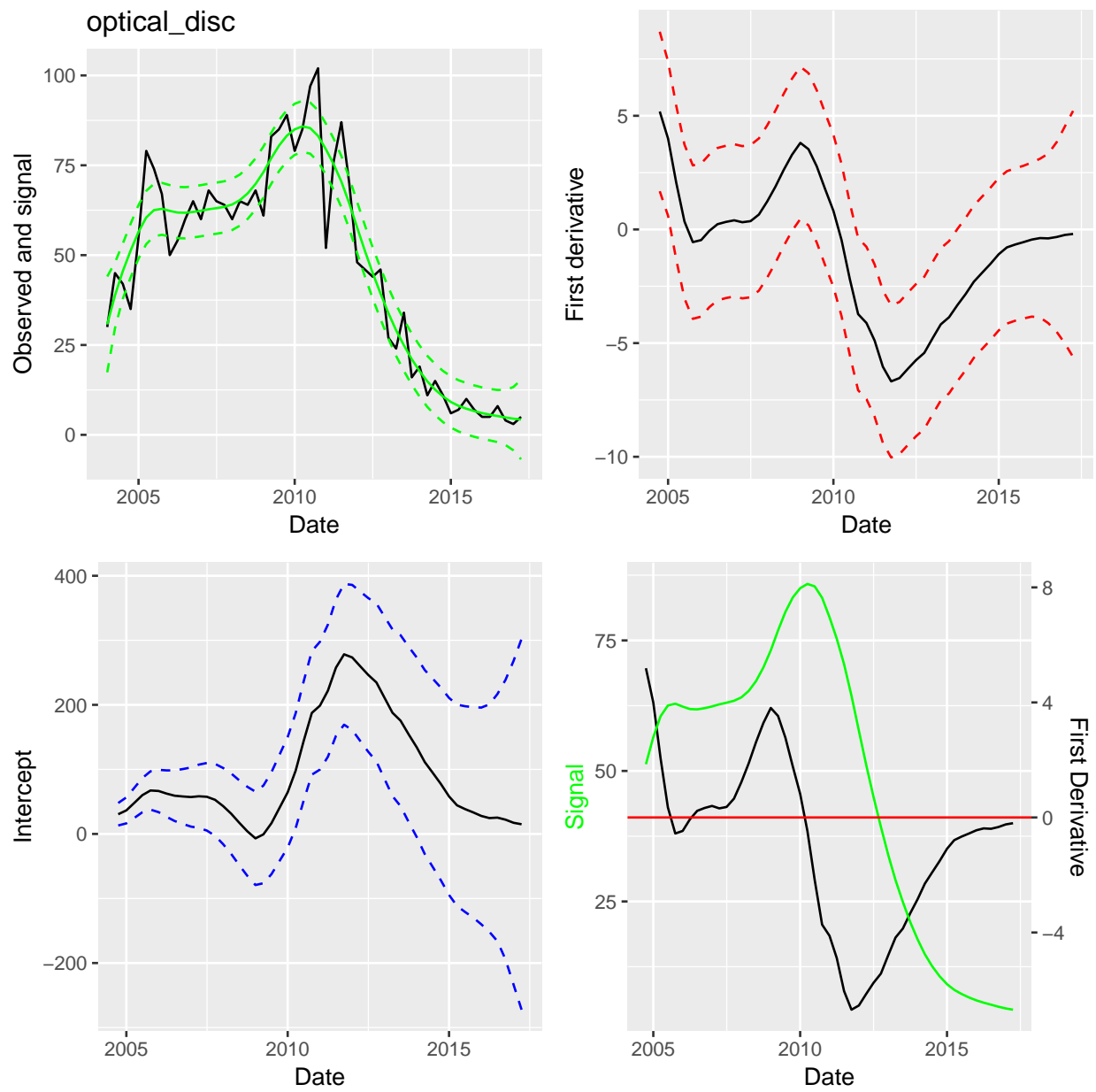


Figure 72:

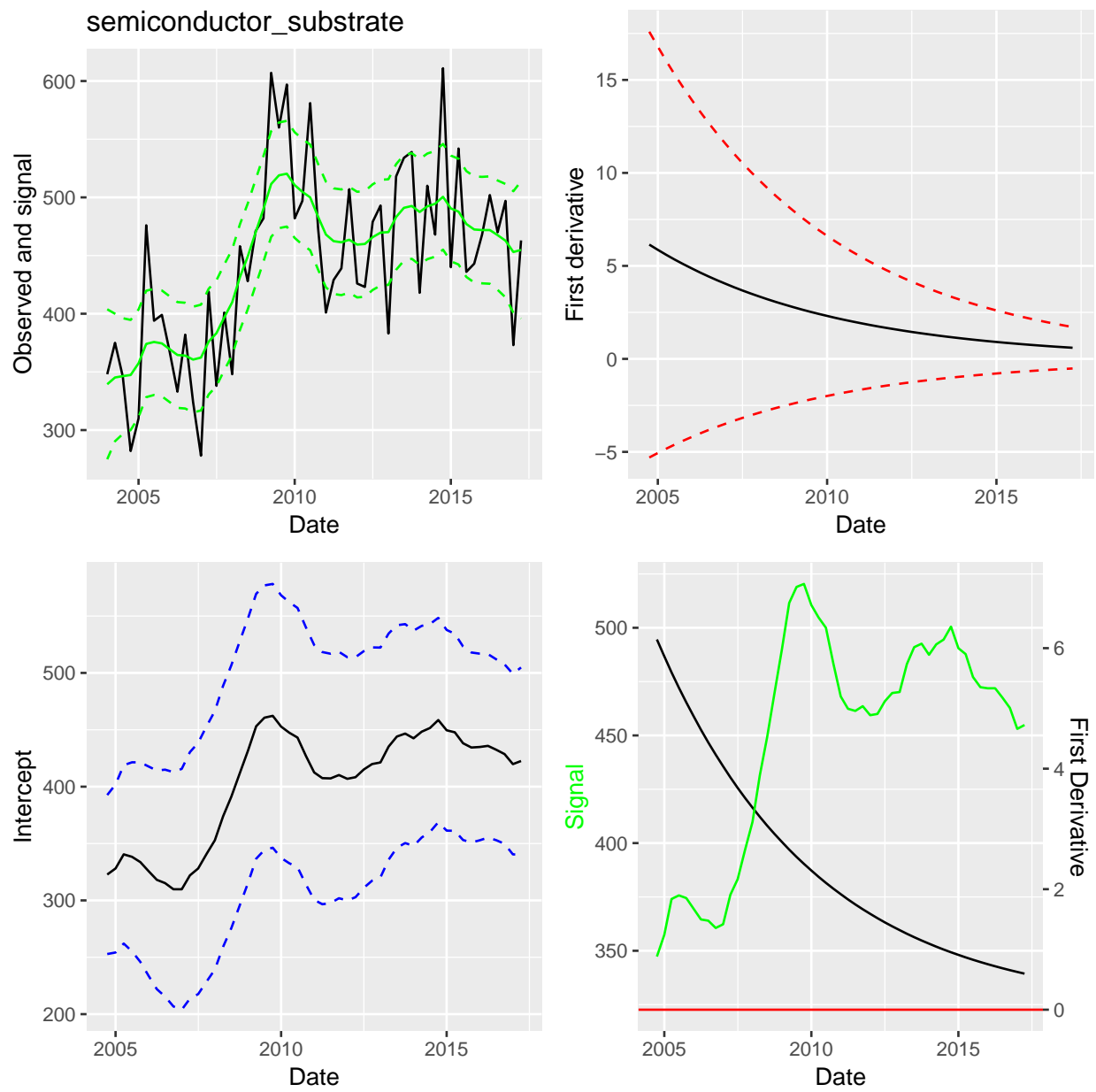


Figure 73:

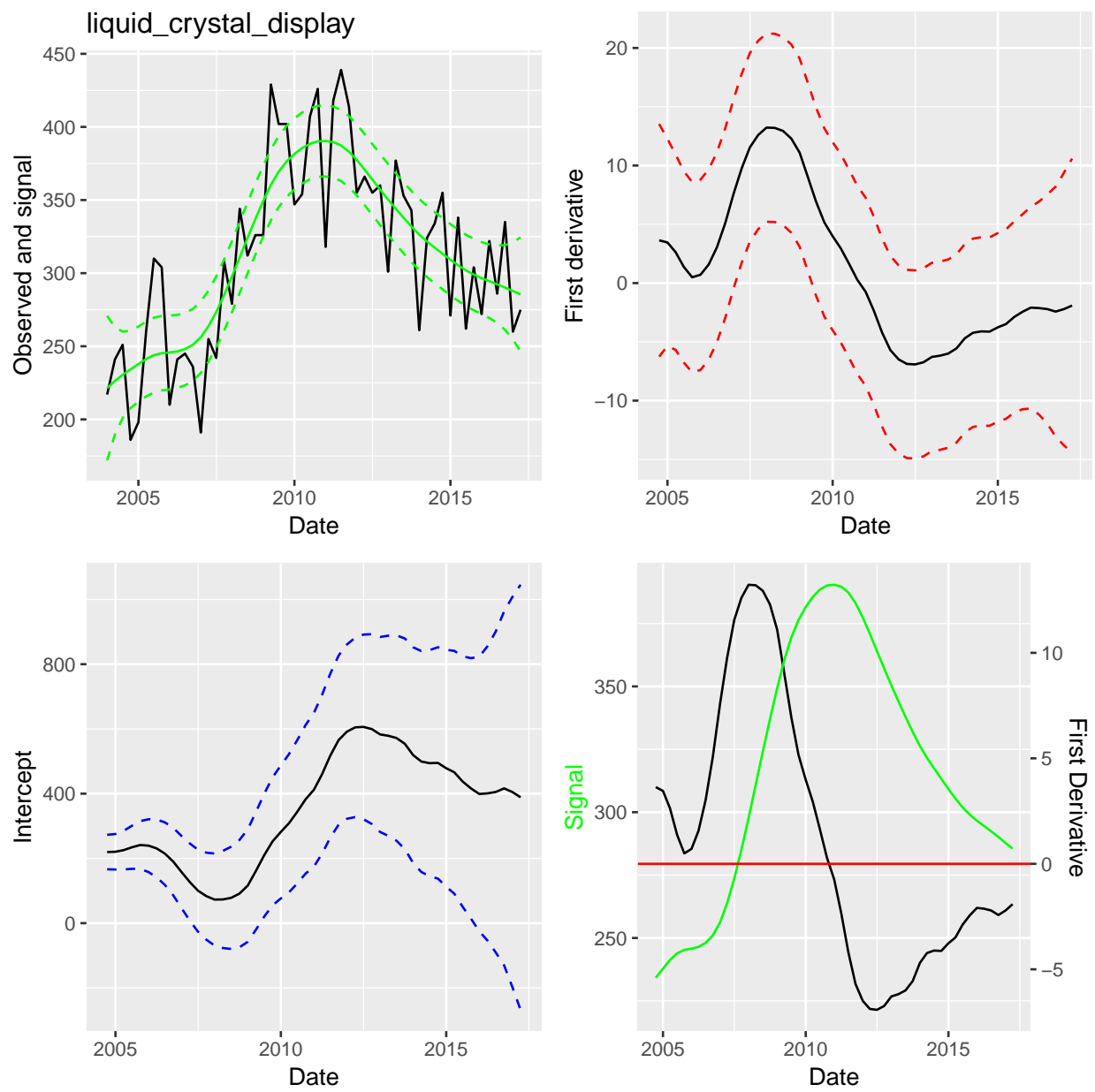


Figure 74: