This Notebook:

"A Maximum Likelihood Bunching Estimator of the Elasticity of Taxable Income" written by Thomas Aronsson, Katharina Jenderny and Gauthier Lanot, accepted for publication in the Journal of Applied Econometrics, MS-13011

This notebook applies the maximum likelihood approach to the estimation of the ETI using Swedish evidence.

This file deals with "ETI_hist_CBEFVI_bin100_2019.dta".

We have first a bunch of declaration (functions etc...), then a presentation of the data, and then some calculations. The calculations are slightly more extensive than those presented in the text of the paper. This may convince the reader that we have carried out due diligence...

In [1]:
```
#Parallelism().set(nproc=6)
#print(Parallelism())
```

In [2]:
```
import numpy  as np
import pandas as pd
import os
```

In [3]:
```
tt=os.getcwd()
show(tt)
```

In [4]:
```
#x,t,y,r = var('x,t,y,r')
x = var('x')
pdfn(x)=1/sqrt(2.*pi)*exp(-1/2.*x*x)
##cdfn(x) = integrate(pdfn(t),(t,-15,x))
#cdfn(x)=0.5*(1.+erf(x/sqrt(2.)))
#assume(abs(r)<=1)
#assume(abs(t)<=1)
#pdfn2(x,y,t) = pdfn(x)*pdfn((y-t*x)/sqrt(1.-t^2))/sqrt(1.-t^2)
```

In [5]:
```
%%cython
clib: m

cdef extern from "math.h":
    double sin(double x)
    double exp(double x)
    double log(double x)
    double sqrt(double x)
    double erf(double x)
    double tanh(double x)
    double pow(double x, double y)
    double fmax(double x, double y)
    double fmin(double x, double y)


def grd(faf,x0,*args):
    x00 = [j for j in x0]
    r = len(x00)
    faf0 = faf(x00,*args)
    grd0 = [0.] * r
    petit = 1.3e-6
    for i in range(r):
        x = [j for j in x0]
        x[i]=x[i]*(1.+petit)+petit
        grd0[i] = (faf(x,*args)-faf0)/(x[i]-x00[i])

    return grd0

# central gradient...
def grdcd(faf,x0,*args):
    x00 = [j for j in x0]
    r = len(x00)

    grd0 = [0.] * r
    petit = 1.3e-6
    for i in range(r):
        x = [j for j in x0]
        xi = x[i]
        x[i] = (1.0+petit)*xi+petit
        fafp = faf(x,*args)
        dx = 2.0*petit*xi+2*petit
        x[i] = (1.0-petit)*xi-petit
        fafm = faf(x,*args)
        grd0[i] = (fafp-fafm)/(dx)

    return grd0

# central gradient...long version
# faf returns a 'vector' of size nfaf
def grdcd_long(faf,x0,nfaf,*args):
    x00 = [j for j in x0]
    r = len(x00)
    #import pdb; pdb.set_trace()
    grd0 = [[0.]*nfaf] * r
    petit = 1.3e-6
    for i in range(r):
        x = [j for j in x0]
        xi = x[i]
        x[i] = (1.0+petit)*xi+petit
        fafp = faf(x,*args)
```

```
        dx = 2.0*petit*xi+2*petit
        x[i] = (1.0-petit)*xi-petit
        fafm = faf(x,*args)
        grd0[i] = [ (fafp[j]-fafm[j])/(dx) for j in range(nfaf)]

    return grd0


#central hessian...
def hescd(faf,x0,*args):
    x00 = [j for j in x0]
    r = len(x00)
    faf0 = faf(x00,*args)
    dh0  = [0.] * r
    hes0 = [[0. for j in range(r)] for l in range(r)]
    petit = 1.03e-4
    for i in range(r):
        dh0[i] = (x00[i]*(1.+petit)+petit)-x00[i]

    #print(hes0)
    #petit = 1.3e-5
    for i in range(r):
        for k in range(i,r):
            if i==k:
                x1 = [j for j in x00]
                x1[i] = x1[i] + 2*dh0[i]
                fafp2 = faf(x1,*args)
                x1 = [j for j in x00]
                x1[i] = x1[i] + dh0[i]
                fafp1 = faf(x1,*args)
                x1 = [j for j in x00]
                x1[i] = x1[i] - dh0[i]
                fafm1 = faf(x1,*args)
                x1 = [j for j in x00]
                x1[i] = x1[i] - 2*dh0[i]
                fafm2 = faf(x1,*args)
                hes0[i][k]=(-fafp2+16*fafp1-30*faf0+16*fafm1-fafm2)/12/dh0[i]/dh
            else:
                x1 = [j for j in x00]
                x1[i] = x1[i] + dh0[i]
                x1[k] = x1[k] + dh0[k]
                fafp1p1 = faf(x1,*args)
                x1 = [j for j in x00]
                x1[i] = x1[i] + dh0[i]
                x1[k] = x1[k] - dh0[k]
                fafp1m1 = faf(x1,*args)
                x1 = [j for j in x00]
                x1[i] = x1[i] - dh0[i]
                x1[k] = x1[k] + dh0[k]
                fafm1p1 = faf(x1,*args)
                x1 = [j for j in x00]
                x1[i] = x1[i] - dh0[i]
                x1[k] = x1[k] - dh0[k]
                fafm1m1 = faf(x1,*args)
                hes0[i][k]=(fafp1p1-fafp1m1-fafm1p1+fafm1m1)/4/dh0[i]/dh0[k]
                hes0[k][i]=hes0[i][k]

    return hes0
```

In [6]:
```
# predicts the distribution
```

In [7]:
```
#hessian, for var cov+ checking optimum...
# hh0b=matrix(RDF,hes(lik0,q00b))
# hh0b.eigenvalues()
# hh0ib = hh0b.inverse()
# sqrtN =sqrt(sum(n[1] for n in nobs))
# results log-likelihood estimates...
# res_lik_print0(q00b,hh0ib,sqrtN,sum(n[1] for n in nobs),"Likelihood Estimates,
    tse = [0 for n in range(len(dens))]
    td  = [[0,0] for n in range(len(dens))]
    t   = [[0,0] for n in range(len(dens))]
    tu  = [[0,0] for n in range(len(dens))]
```

In [8]:
```
# this is for UMea...what is the swedish average...?
taxinfo = [ [2001,0.677,0.477,252000],
          [2002,0.677,0.477,273800],
          [2003,0.677,0.477,284300],
          [2004,0.677,0.477,291800],
          [2005,0.669,0.469,298600],
          [2006,0.669,0.469,306000],
          [2007,0.669,0.469,316700],
          [2008,0.669,0.469,328800],
          [2009,0.669,0.469,367600],
          [2010,0.669,0.469,372100],
          [2011,0.669,0.469,383000],
          [2012,0.669,0.469,401100],
          [2013,0.669,0.469,413200],
          [2014,0.664,0.464,420800],
          [2015,0.664,0.464,430200],
          [2016,0.6635,0.4635,430200],
          [2017,0.6585,0.46585,438900],
          [2018,0.6585,0.46585,455300],
          [2019,0.6585,0.46585,490700]
        ]
    print(ttl)
```

In [ ]:
```
    if len(qq)==3:
```

In [9]:
```
# so how do we select a symmetric range around the kink...?
# we check the distance from the kink to the min or the max,
# and select only if the difference from the kink
# is less that the minimum between those 2 differences... simple...
year = 2019
Info_in_year = flatten([ nn for nn in taxinfo if nn[0]==year])
Kink = Info_in_year[3]/1000.
lnk=ln(Kink).n()
```

In [10]:
```
os.chdir(tt)
tt2=os.getcwd()
show(tt2)

    return qq,se,ll
```

Data:

Here comes the data.

In [11]:
```
import numpy  as np
import pandas as pd
import os
```

In [12]:
```
#newd="D://Umeå universitet//ETI - General//JAE Revisions//scb data//hist_data"
#newd = "/Users/gauthierlanot/Library/CloudStorage/OneDrive-SharedLibraries-Umeå
#newd = "//Users/gauthierlanot/Downloads/"
#ETI - General\\JAE Revisions\\scb data"
newd = "D:\ETIprojectfiles\JAE revisions"
show(newd)
```

```
<>:5: DeprecationWarning: invalid escape sequence \E
<>:5: DeprecationWarning: invalid escape sequence \E
<>:5: DeprecationWarning: invalid escape sequence \E
<ipython-input-12-69121d137286>:5: DeprecationWarning: invalid escape sequence \E
  newd = "D:\ETIprojectfiles\JAE revisions"
```

In [13]:
```
# loading data from stata file
os.chdir(newd)
filename = "ETI_hist_CBEFVI_bin100_2019.dta"
```

In [14]:
```
delta, STEP_BIN = var('delta STEP_BIN')
delta = (0.1*.99999)
STEP_BIN= 0.1
```

In [15]:
```
delta
```

Out[15]:
```
0.0999990000000000
```

In [16]:
```
dd0 = pd.read_stata(filename)
```

In [17]:
```
dd = dd0.to_numpy()
```

In [18]:
```
dd = [(nn[0],nn[1],nn[2],nn[3]/1000.,nn[4],nn[5],nn[6]/1000.) for nn in dd]
```

In [19]:
```
# ...and here we find that the data is not centered on the kink!
# typical last minute job! no attention to details!
#df0
```

In [20]:
```
dd[0:10]
```

Out[20]:
```
[(1.0, 1208.0, 2019.0, 422.3, 422300.0, 0.0, 490.7),
 (2.0, 1187.0, 2019.0, 422.4, 422400.0, 1.0, 490.7),
 (3.0, 1206.0, 2019.0, 422.5, 422500.0, 1.0, 490.7),
 (4.0, 1149.0, 2019.0, 422.6, 422600.0, 1.0, 490.7),
 (5.0, 1098.0, 2019.0, 422.7, 422700.0, 1.0, 490.7),
 (6.0, 1188.0, 2019.0, 422.8, 422800.0, 1.0, 490.7),
 (7.0, 1144.0, 2019.0, 422.9, 422900.0, 1.0, 490.7),
 (8.0, 1111.0, 2019.0, 423.0, 423000.0, 1.0, 490.7),
 (9.0, 1134.0, 2019.0, 423.1, 423100.0, 1.0, 490.7),
 (10.0, 1174.0, 2019.0, 423.2, 423200.0, 1.0, 490.7)]
```

In [21]:
```
# so how do we select a symmetric range around the kink...?
# we check the distance from the kink to the min or the max,
# and select only if the difference from the kink
# is less that the minimum between those 2 differences... simple...
year = 2019
Info_in_year = flatten([ nn for nn in taxinfo if nn[0]==year])
Kink = Info_in_year[3]/1000.
lnk=ln(Kink).n()
```

In [22]:
```
Info_in_year
```

Out[22]:
```
[2019, 0.658500000000000, 0.465850000000000, 490700]
```

In [23]:
```
Kink
```

Out[23]:
```
490.700000000000
```

In [24]:
```
dd[0]
```

Out[24]:
```
(1.0, 1208.0, 2019.0, 422.3, 422300.0, 0.0, 490.7)
```

In [25]:
```
#distmin2k = df0.kink[[0]].array[0]-df0.wh[[0]].array[0]
#distk2max = df0.wh[[len(df0)-1]].array[0]-df0.kink[[0]].array[0]
#topdistfromk = min(distmin2k,distk2max)+0.001
#distmin2k,distk2max,topdistfromk
distmin2k = dd[0][6]-dd[0][3]
distk2max = dd[-1][3]-dd[-1][6]
topdistfromk = min(distmin2k,distk2max)+0.001
distmin2k,distk2max,topdistfromk
```

Out[25]:
```
(68.39999999999998, 79.59999999999997, 68.40099999999998)
```

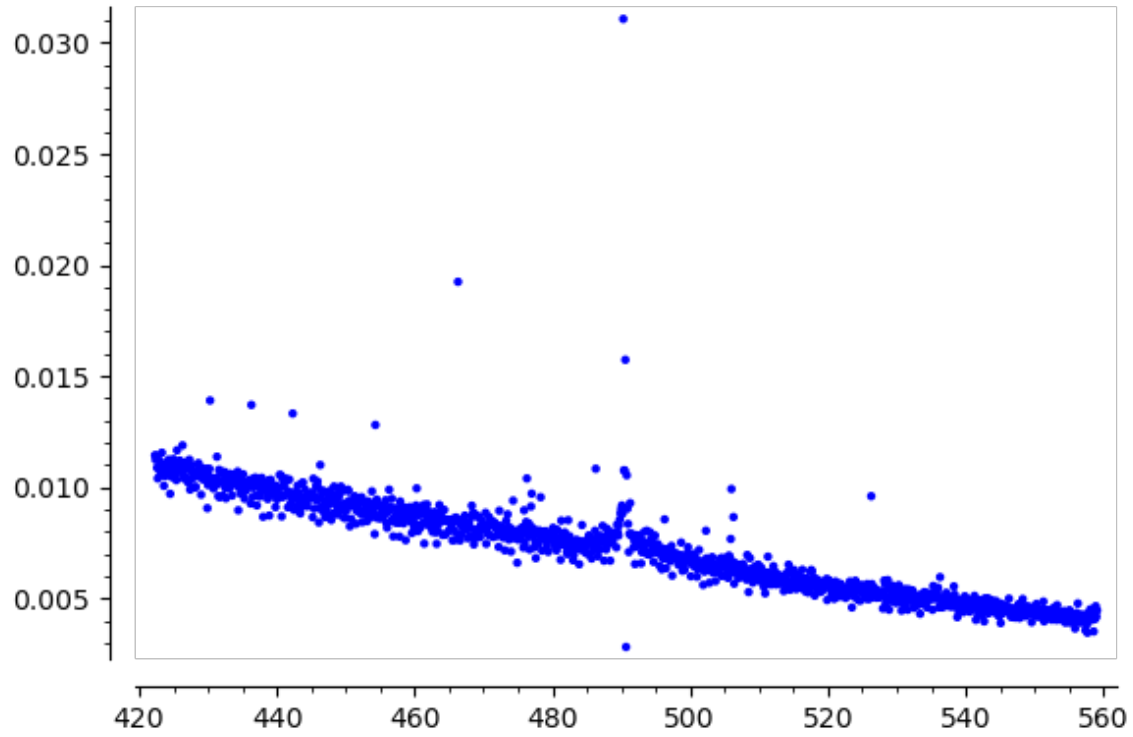
In [26]:
```
def mak_density(datai,Kinki,topdistfromki,stp_bn,dlt):
    df = [nn for nn in datai if abs(nn[3]-Kinki)<=topdistfromki ]
    df_yy_cnt = [ (dd[3],dd[1]) for dd in df]
    Nobs = int(sum([nn[1] for nn in df]))
    #show(Nobs)
    dens = [(nn[3],nn[1]/Nobs) for nn in df]
    dens = np.array(dens)
    #show(len(dens))

    #densa = [(nn[3],nn[1]/Nobs/0.1) for nn in df if not(((ln(nn[3])+0.1/252.)>=

    densa = [(nn[3],nn[1]/Nobs/stp_bn) for nn in df if nn[3]<(Kinki-dlt)]+ \
            [(nn[3],nn[1]/Nobs) for nn in df if (nn[3]>=(Kinki-dlt)) and (nn[3])
            [(nn[3],nn[1]/Nobs/stp_bn) for nn in df if nn[3]>Kinki]

    densb = [(nn[3],nn[1]/Nobs/stp_bn) for nn in df]
    return {
            'df':df,
            'df_yy_cnt':df_yy_cnt,
            'Nobs':Nobs,
            'dens':dens,
            'densa':densa,
            'densb':densb}
```

In [27]:
```
mkd=mak_density(dd,Kink,topdistfromk,STEP_BIN,delta)
```

In [ ]:

In [28]:
```
list_plot(mkd['densa'])
```

Out[28]:



In [29]:
```
#sum(dens[n][1] for n in range(0,len(dens)) if n!= KinkPos)+dens[KinkPos][1]
```

In [30]:
```
os.chdir(tt)
tt2=os.getcwd()
show(tt2)
```

In [31]:
```
490.7*0.85, 490.7*1.15
```

Out[31]:
```
(417.095000000000, 564.305000000000)
```

Transformations:

The data is transformed in several ways:

- dens: contains the data as density everywhere but at the kink where it is a probability, used for least squares fitting, model with perfect bunching;
- densa: containes the data as proportion of the sample, used for maximum likelihood, model with perfect bunching;
- densb: contains the data as density everywhere, used for least squares fitting, model with imperfect bunching.

In [32]:
```
totnobs = mkd['Nobs']

KinkPos = int((len(mkd['df'])-1)/2)

print(sum( n[1] for n in mkd['dens']))

#best normal fit...
# measure mean and variance
print("level")
m  = sum(n[1]*n[0]   for n in mkd['dens'])
m2 = sum(n[1]*n[0]^2 for n in mkd['dens'])
v2 = m2-m^2
print(['mean',m],['std dev',sqrt(m2-m^2)])

print("logs")
lm = sum( n[1]*log(n[0]) for n in mkd['dens'])
lm2 = sum(n[1]*log(n[0])^2 for n in mkd['dens'])
lv2 = lm2-lm^2
print(['mean',lm],['std dev',sqrt(lm2-lm^2)])
# -1. to deal with the data to the left of the lower bound...
# to copy in the spyx file...until I find how to transfer data on the fly!

lnebu = ln(min(n[0] for n in mkd['dens'])-.1)-0.001
lnebo = ln(max(n[0] for n in mkd['dens']))+0.001



sum(mkd['dens'][n][1] for n in range(0,len(mkd['dens'])) if n!= KinkPos)+mkd['de
sum(mkd['dens'][n][1] for n in range(0,len(mkd['dens'])) )
```

```
0.9999999999999999
level
['mean', 479.8201495662524] ['std dev', 38.03039478470768]
logs
['mean', 6.170305771954446] ['std dev', 0.07860243742723214]
```

Out[32]: (0.9999999999999999, 0.9999999999999999)

In [33]:
```
lnebu,lnebo
```

Out[33]: (6.04447913541422, 6.3273283480326)

In [34]:
```
KinkPos,mkd['dens'][KinkPos]
```

Out[34]: (684, array([4.9070000e+02, 2.8281722e-03]))

In [35]:
```
# allows us to check the data!
#list_plot(dens,ymin=0)+ list_plot(densa,color='gold')
```

In [ ]:
```

```

In [36]:
```
Info_in_year
```

Out[36]: [2019, 0.658500000000000, 0.465850000000000, 490700]

In [37]:
```
#list(zip(dens,densa))
Info_in_year[1]
```

Out[37]:
```
0.658500000000000
```

Tax parameters:

In [38]:
```
pl1,pl2,ps,ps1,ps2,pr12,lw = var('pl1,pl2,ps,ps1,ps2,pr12,lw')

#/Kink

# this is complicated since their application aggregates many years...here I tak
lnt1c = ln(Info_in_year[1])
lnt2c = ln(Info_in_year[2])
lt1cplt2c = lnt1c^2 + lnt2c^2
lt1cmlt2c = lnt1c^2 - lnt2c^2
```

In [39]:
```
delta,delta/(lnt1c-lnt2c)
```

Out[39]:
```
(0.0999990000000000, 0.288930254180252)
```

In [40]:
```
delta,lnk,mkd['dens'][KinkPos][0],Kink
```

Out[40]:
```
(0.0999990000000000, 6.19583294309586, 490.7, 490.700000000000)
```

In [41]:
```
#len(df_yy_cnt),Kink
#ck['IG'][0:10],ck['FREQ'][0:10]
#show(list(zip(ck['IG'].list(),ck['FREQ'].list())))
```

Cython code...

In [42]:
```
#newd = "C:\\Users\\gala0013\\Dropbox\\sagemath\\alternativeETI"
#os.chdir(newd)
```

In [43]:
```
#compile some useful routines
#load("~/Users/gauthier/Dropbox/sagemath/alternativeETI/GLbunching.spyx")
#load('C:\Users\gala0013\Dropbox\sagemath\alternativeETI\GLbunching.spyx')
load("GLbunching2.spyx")
```

```
Compiling ./GLbunching2.spyx...
```

Least Squares and Likelihood Routines

The routines use various parametrisation. This matters when calculating the second derivatives and the ML precision.

In [46]:
```
%%cython

clib: m

cdef extern from "math.h":
    double sin(double x)
    double exp(double x)
    double log(double x)
    double sqrt(double x)
    double erf(double x)
    double tanh(double x)
    double pow(double x, double y)
    double fmax(double x, double y)
    double fmin(double x, double y)

def param1(pprm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk):
    pps  = exp(fmax(fmin(pprm[2],5.),-5))
    ppl2 = abs(pprm[0])*lnt2c +  pprm[1]
    ppl2 = pps*ppl2
    ppl1 = abs(pprm[0])*lnt1c +  pprm[1]
    ppl1 = pps*ppl1

    return (ppl1,ppl2,pps)

def paramib(pprm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk):
    #pprm[1] = fmax(fmin(pprm[1],5.),-5)
    #pl1 = abs(pprm[0])*lnt1c + lnebu + (lnebo-lnebu)/(1+exp(-pprm[1]))
    #pl2 = abs(pprm[0])*lnt2c + lnebu + (lnebo-lnebu)/(1+exp(-pprm[1]))
    pl1 = abs(pprm[0])*lnt1c + pprm[1]
    pl2 = abs(pprm[0])*lnt2c + pprm[1]

    if len(pprm)==4:
        #homogenous model with imperfect bunching
        ps1  = exp(-fmax(fmin(pprm[2],5.),-5))
        ps2 = ps1
        vs = exp(fmax(fmin(pprm[3],10.),-5))
        return [pl1/ps1,pl2/ps2,1/ps1,1/ps2,1.,vs]

    elif len(pprm)==5:
        # heterogenous model with imperfect bunching
        # independence between disutility and response remains equal to 0
        prr = 0.0
    #elif len(pprm)==5:
    #    prr = tanh(pprm[4])

    # variance of eti can't be too large under the normal model...
    sa = abs(pprm[0])/2.5/(1+exp(-fmax(fmin(pprm[3],5.),-5)))
    st = exp(-fmax(fmin(pprm[2],5.),-5))

    #ps1 = pprm[2]^2 + 2.*prr*abs(pprm[2]*pprm[3])*lnt1c         + (lnt1c*lnt1c)
    ps1 = st^2 + 2.*prr*abs(st*sa)*lnt1c         + (lnt1c*lnt1c)*(sa^2)
    ps1 = sqrt(ps1)
    ps2 = st^2 + 2.*prr*abs(st*sa)*lnt2c         + (lnt2c*lnt2c)*(sa^2)
    ps2 = sqrt(ps2)
    pc12= st^2 +    prr*abs(st*sa)*(lnt1c+lnt2c) + (lnt1c*lnt2c)*(sa^2)
    pc12= pc12/ps1/ps2

    vs = exp(fmax(fmin(pprm[4],10.),-5))
    #print(pc12)#
    pc12 = fmin(fmax(pc12,-1.+1e-5),1.-1e-5)
```

```
    #pc12 = pc12/ps1/ps2
    return [pl1/ps1,pl2/ps2,1/ps1,1/ps2,pc12,vs]
```

In [47]:

```
import math

def ff0n(w,pl1,pl2,ps,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk):
    ppprm = param1([pl1,pl2,ps])
    ppl1 = ppprm[0]
    ppl2 = ppprm[1]
    pps  = ppprm[2]
    cndI = fmax(cdfn(pps*lnebo-ppl2)-cdfn(pps*lnebu-ppl1),0.0000001)

    if math.log(w)<log(Kink-delta):
        return pdfn(pps*math.log(w)-ppl1)*pps/w/cndI

    if math.log(w)<=lnk and math.log(w)>=log(Kink-delta):
        return (cdfn(pps*lnk-ppl2)-cdfn(pps*log(Kink-delta)-ppl1))/cndI

    if math.log(w)>lnk:
        return pdfn(pps*math.log(w)-ppl2)*pps/w/cndI

def ff0na(w,pprm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk):
    ppprm = param1(pprm)
    pl1 = ppprm[0]
    pl2 = ppprm[1]
    ps  = ppprm[2]

    lw = math.log(w)
    lKmd = math.log(Kink-delta)

    if lw<lKmd:
        return pdfn(ps*lw-pl1)*ps/w
    if lw>lnk:
        return pdfn(ps*lw-pl2)*ps/w
    if lw<=lnk and lw>=lKmd:
        return (cdfn(ps*lnk-pl2)-cdfn(ps*lKmd-pl1))

def prob_int(pprm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk):
    # this parametrisation is better behaved...
    # this feels very ad hoc but it works (at least here!)
    ppprm = param1(pprm)
    ppl1 = ppprm[0]
    ppl2 = ppprm[1]
    pps  = ppprm[2]

    cndI  = (cdfn(pps*lnebo-ppl2)-cdfn(pps*lnebu-ppl1))
    return cndI


def h0iba1(ww,pl1,pl2,ps1,ps2,pr12,vs,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk):
    # density as in the paper...
    #import pdb; pdb.set_trace()
    lnw = math.log(ww)
    ffss1 = ps1
    ffss12=ffss1*ffss1
    ffmm1 = ffss1*lnk-pl1
    ffss2 = ps2
    ffss22=ffss2*ffss2
    ffmm2 = ffss2*lnk-pl2
    evs = vs
```

```
    evs2 = evs*evs
    return   (ffss1/sqrt(evs2+ffss12)*pdfn((evs*ffss1*lnw-(pl1*evs-1/2*(ffss1/ev
             cdfn((-evs2*lnw +(evs2+ffss12)*lnk-1/2-pl1*ffss1)/sqrt(evs2+ffss12)

def h0iba2(ww,pl1,pl2,ps1,ps2,pr12,vs,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk):
    # density as in the paper...
    #import pdb; pdb.set_trace()
    lnw = math.log(ww)
    ffss1 = ps1
    ffss12=ffss1*ffss1
    ffmm1 = ffss1*lnk-pl1
    ffss2 = ps2
    ffss22=ffss2*ffss2
    ffmm2 = ffss2*lnk-pl2
    evs = vs
    evs2 = evs*evs
    return   (ffss2/sqrt(evs2+ffss22)*pdfn((evs*ffss2*lnw-(pl2*evs-1/2*(ffss2/ev
             cdfn((evs2*lnw-(evs2+ffss22)*lnk+1/2+pl2*ffss2)/sqrt(evs2+ffss22))/

def h0ibak(ww,pl1,pl2,ps1,ps2,pr12,vs,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk):
    # density as in the paper...
    #import pdb; pdb.set_trace()
    lnw = math.log(ww)
    ffss1 = ps1
    ffss12=ffss1*ffss1
    ffmm1 = ffss1*lnk-pl1
    ffss2 = ps2
    ffss22=ffss2*ffss2
    ffmm2 = ffss2*lnk-pl2
    evs = vs
    evs2 = evs*evs
    if (pr12>=1):
        FFk= cdfn(ffmm2)-cdfn(ffmm1)
    else:
        FFk= cdfn(ffmm2)-cdfbvn(ffmm1,ffmm2,pr12)
    return   (pdfn(evs*(lnw-lnk)+1/2/evs)/ww*FFk)
    #return FFk/evs

def h0iba(ww,pl1,pl2,ps1,ps2,pr12,vs,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk):
    # density as in the paper...
    #import pdb; pdb.set_trace()
    lnw = math.log(ww)
    ffss1 = ps1
    ffss12=ffss1*ffss1
    ffmm1 = ffss1*lnk-pl1
    ffss2 = ps2
    ffss22=ffss2*ffss2
    ffmm2 = ffss2*lnk-pl2
    evs = vs
    evs2 = evs*evs
    if (pr12>=1):
        FFk= cdfn(ffmm2)-cdfn(ffmm1)
    else:
        FFk= cdfn(ffmm2)-cdfbvn(ffmm1,ffmm2,pr12)
    #if tanh(pr12)<0.9999:
    #    FFk= cdfN(ffmm2)-cdfbvn(ffmm1,ffmm2,tanh(pr12))
    #else:
    #    FFk= cdfN(ffmm2)-cdfN(ffmm1)
    #print(FFk.n())
    # removed *evs
```

```
    return   (pdfn(evs*(lnw-lnk)+1/2/evs)/ww*FFk+\
       ffss2/sqrt(evs2+ffss22)*pdfn((evs*ffss2*lnw-(pl2*evs-1/2*(ffss2/evs)))/sq
       cdfn((evs2*lnw-(evs2+ffss22)*lnk+1/2+pl2*ffss2)/sqrt(evs2+ffss22))/ww+ \
       ffss1/sqrt(evs2+ffss12)*pdfn((evs*ffss1*lnw-(pl1*evs-1/2*(ffss1/evs)))/sq
       cdfn((-evs2*lnw +(evs2+ffss12)*lnk-1/2-pl1*ffss1)/sqrt(evs2+ffss12))/ww)*


def cndI0ib(prm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk):
    pprm = paramib(prm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)

    cndI = prob_alt(math.exp(lnebo),*pprm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk

    return cndI


def prob_alt(ww,pl1,pl2,ps1,ps2,pr12,vs,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk):
    #a formal expression...
    lnw = math.log(ww)
    ffss1 = ps1
    ffss12=ffss1*ffss1
    ffmm1 = ffss1*lnk-pl1
    ffss2 = ps2
    ffss22=ffss2*ffss2
    ffmm2 = ffss2*lnk-pl2
    evs = vs
    evs2 = evs*evs
    #import pdb; pdb.set_trace()
    FFk= cdfn(ffmm2)-cdfn(ffmm1)
    #show(FFk)
    P0 = FFk*cdfn(evs*(lnw-lnk)+1/2/evs)

    #show(ffss1/sqrt(evs2+ffss12))
    y2 = (evs*ffss1*lnw-(pl1*evs-1/2*(ffss1/evs)))/sqrt(evs2+ffss12)
    P1 = cdfbvn(ffmm1,y2,evs/sqrt(evs2+ffss12))

    y2 = (evs*ffss2*lnw-(pl2*evs-1/2*(ffss2/evs)))/sqrt(evs2+ffss22)
    #y1 = (evs*ffss2*lnk-(pl2*evs-1/2*(ffss2/evs)))/sqrt(evs2+ffss22)
    P2 = cdfbvn(-ffmm2,y2,-evs/sqrt(evs2+ffss22))
    #-cdfbvn(-ffmm2,y1,-evs/sqrt(evs2+ffss22))
    #show((P0.n(),P1.n(),P2.n()))
```

In [48]:
```
import math

def lik0(pprm,densi,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk):
    #ppprm = param0(pprm)
    ppprm = param1(pprm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)
    ppl1 = ppprm[0]
    ppl2 = ppprm[1]
    pps  = ppprm[2]

#    print(pl1,pl2)
    #cndI  = (cdfn(pps*lnebo-ppl2)-cdfn(pps*lnebu-ppl1))
    cndI = max(cdfn(pps*lnebo-ppl2)-cdfn(pps*lnebu-ppl1),0.000000001)
    if cndI<1.e-9:
        cndI = 1.e-9
        print(pprm,cndI)

    ores = [ n[1]*math.log(ff0na(n[0],pprm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,ln
    #print(res)
    #res = sum(res).n()-ln(cndI).n()
    return -sum(ores)
```

```
# least square with minimize; rough and ready!
def lsq0(pprm,densi,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk):
    #ppprm = param0(pprm)
    ppprm = param1(pprm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)
    ppl1 = ppprm[0]
    ppl2 = ppprm[1]
    pps  = ppprm[2]

    #cndI  =    cdfn(pps*lnebo-ppl2)-cdfn(pps*lnebu-ppl1)
    cndI = max(cdfn(pps*lnebo-ppl2)-cdfn(pps*lnebu-ppl1),0.000000001)
    if cndI<1.e-9:
        cndI = 1.e-9
        #print(pprm,cndI)

    ores = [ (n[1]-(ff0na(n[0],pprm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)/cndI
    ores = sum(ores)
    #print(pprm,res)
    return ores


def lik0ib(prm,densi,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk):
    pprm = paramib(prm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)
    cndI = prob_alt(math.exp(lnebo),*pprm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk
    #if cndI<=0.00001:
    #        return len(densi)*50
    #show(cndI)
    #import pdb; pdb.set_trace()
    ores = 0
    p_bef = prob_alt(math.exp(lnebu),*pprm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,ln
    longueur = len(densi)

    for ii in range(longueur):
        p_at = prob_alt(densi[ii][0],*pprm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,ln
        if ((p_at-p_bef+1e-50)/cndI)<1e-10:
            #show([ii,(p_at-p_bef),cndI,densi[ii]],prm,pprm)
            ores -= densi[ii][1]*(-75)
        else:
            ores -= densi[ii][1]*(math.log((p_at-p_bef+1e-50))- math.log(cndI) )
        p_bef = p_at
    #show([*pprm,cndI,ores])
```

In [49]:

```
def mak_vc(loglik,loglik_long,x,lendti,N,stp_bn,*args):
    #construct the variance covariance of the ML estimator...
    #Gradients, Hessian, Information matrices
    #adjusting for density etc...
    G_all = stp_bn*matrix(RDF,grdcd(loglik,x,*args))
    hh0ib = stp_bn*matrix(RDF,hescd(loglik,x,*args))
    hh0ibi = hh0ib.inverse()
    G00 = grdcd_long(loglik_long,x,lendti,*args)
    G0ib  = stp_bn*matrix(RDF,G00)
    jj0ib = N*G0ib*G0ib.transpose()
    return {
            'G_all':G_all,
            'hh0ib':hh0ib,
            'hh0ibi':hh0ibi,
            'jj0ib':jj0ib
    }

def lsq0ib(prm,densi,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk):
    pprm = paramib(prm,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)
```

In [50]:
```
def mak_chi2(datai,N,x,*par):
    #make a chi squared stat...
    OO2 = [ [nn[0],nn[1]*totnobs] for nn in datai]
    TT2 = [ [n[0],float(h0iba(n[0],*paramib(x,*par),*par))/cndI0ib(x,*par)*N] fo
    csq = [ float((OO2[ii][1]-TT2[ii][1])^2/TT2[ii][1]) for ii in range(len(data
    csq2 = [ [OO2[ii][0],float((OO2[ii][1]-TT2[ii][1])^2/TT2[ii][1])] for ii in
    return {
        'Observed #':OO2,
        'Theoretical #':TT2,
        'Chi_squared stat': sum(csq),
        'obs vs ind chi2 stat': csq2
    }
def cndI0ib(prm,delta,Intie,Intze,Ineba,Ineba,Rink,Ink):
```

In [51]:
```
def mak_OLS_R2(O2,T2):
    data = [ (T2[ii][1],O2[ii][1]) for ii in range(len(O2))]
    N = sum([n[1] for n in O2])
    # design a model with adjustable parameters a,b,c that describes the data
    a,b,x = var('a, b, x')
    model(x) = a * x + b

    # calculate the values of the parameters that best fit the model to the data
    sol = find_fit(data,model)
    # define f(x), the model with the parameters set to the fitted values
    f(x) = model(a=sol[0].rhs(),b=sol[1].rhs())

    # create an empty plot object
    a = plot([])
    # add a plot of the model, with respect to x from -0.5 to 12.5
    a += plot(f(x),x,[min(nn[1] for nn in T2),max(nn[1] for nn in T2)])

    # add a plot of the data, in red
    a += list_plot(data,color='red')

    V_obs = variance([nn[1] for nn in O2])
    V_the = variance([f(nn[1]) for nn in T2])
    R2sq = V_the/V_obs

    return {
        'a':sol[0].rhs(),
        'b':sol[1].rhs(),
        'Nobs':N,
        'graph':a,
        'Var obs':V_obs,
        'Var theoretical':V_the,
        'R2':R2sq
    }
```

In [52]:
```
%load_ext cython
```

In [53]:
```
#%%cython
#clib: m
#cdef extern from "math.h":
#    double sin(double)
#    double exp(double)
#    double log(double)
#    double sqrt(double)
#    double erf(double)
```

In [54]:
```
#[ n for n in dens if (((ln(n[0])+2*delta)>=lnk) and (ln(n[0])<=lnk)) ],
Kink,exp(lnk)
```

Out[54]:
```
(490.700000000000, 490.700000000000)
```

Model with Imperfect Bunching.

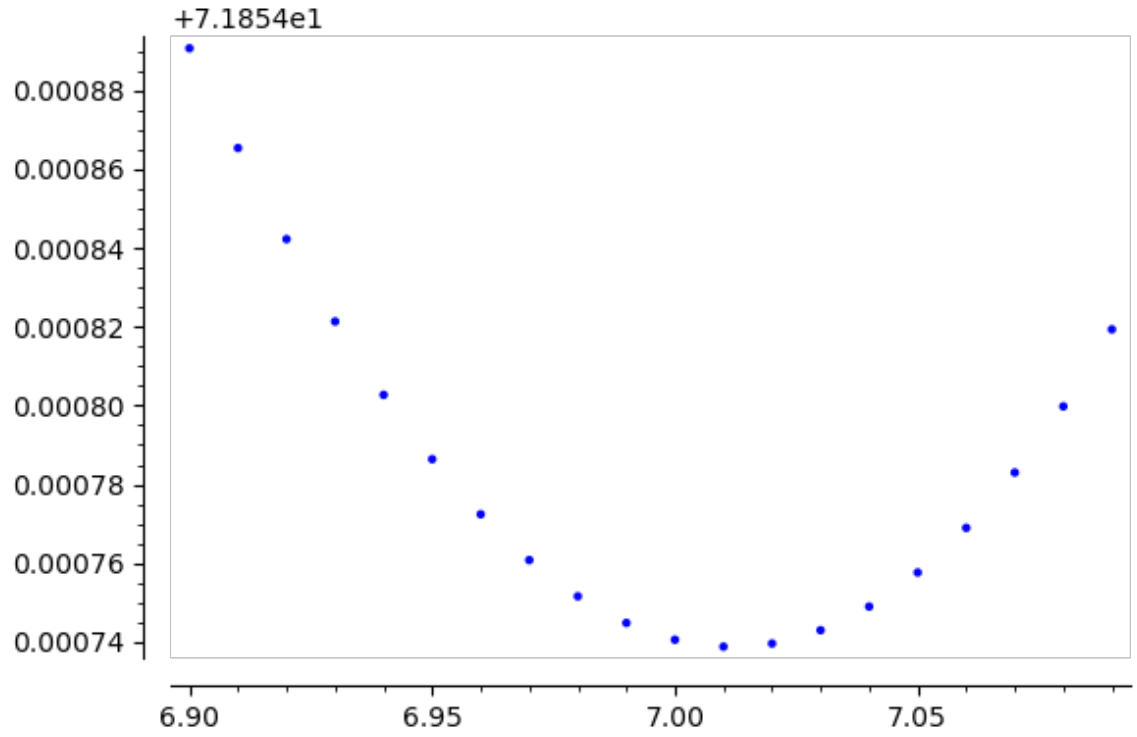
model of density throughout. Log Normal-Log Normal model...

All Observations.

We start with some more routines...

In [59]:
```
densb = mkd['densb']
sqrtN =sqrt(mkd['Nobs'])
tax_par = (densb,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)
#list_plot([ [xx,lik0ib([0.04622116773670176, 6.18428737361848, 2.25401328722627
# starting from almost correct solutions! searching for best standard error of f
#list_plot([ [xx,lik0ib([0.008112282898730748, 6.032349244286576, 1.375208740223
#list_plot([ [xx,lik0ib([xx, 6.032349244286576, 1.3752087402233504, 6.95],*tax_p
#list_plot([ [xx,lik0ib([0.007, xx, 1.3752087402233504, 6.95],*tax_par)] for xx
#list_plot([ [xx,lik0ib([0.007, 6.025, xx, 6.95],*tax_par)] for xx in srange(1.2
list_plot([ [xx,lik0ib([0.007, 6.025,  1.35, xx],*tax_par)] for xx in srange(6.9
```

Out[59]:



In [60]:
```
#maximum likelihood estimator.
#tax_par = (('delta', delta),('lnt1c',lnt1c),('lnt2c',lnt2c),('lnebu',lnebu),('l
tax_par = (densb,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)
solib_Lll_all = minimize(lik0ib,[0.007, 6.025,  1.35, 7.01 ],args=tax_par, verbo
print(solib_Lll_all)
```

```
Optimization terminated successfully.
         Current function value: 71.854591
         Iterations: 73
         Function evaluations: 135
(0.006728700551284908, 6.040399955254271, 1.4056877830495194, 7.0332837035946545)
```

In [61]:
```
mkd['Nobs'],STEP_BIN
```

Out[61]:
```
(1055452, 0.100000000000000)
```

In [62]:
```
#calculation of var cov of parameters
densb = mkd['densb']
#(loglik,loglik_long,x,lendti,N,stp_bn,*args)
VCM= mak_vc(lik0ib,lik0ib_long,solib_Lll_all,len(densb),mkd['Nobs'],STEP_BIN,*ta
```

In [63]:
```
show(VCM['hh0ib'].eigenvalues(algorithm='symmetric'))
show(VCM['jj0ib'].eigenvalues(algorithm='symmetric'))
```

In [64]:
```
# printing results... Hessian+ Sandwich...
```

In [65]:
```
res_lik_print0(lik0ib,solib_Lll_all,VCM['hh0ibi'],sqrtN,int(mkd['Nobs']),"Likeli
```

```
Likelihood Estimates, Homogenous Model, Imperfect Bunching, Hessian
Parameter Estimate Std.Err.
  eti      0.006729   0.00013115
  mu       6.040      0.0029432
  sigma1   1.406      0.010037
  v        7.033      0.021886
  number of observations |            1055452
          log_likelihood | -75839072.09717053
```

Out[65]:
```
((0.006728700551284908, 6.040399955254271, 1.4056877830495194, 7.033283703594654
5),
 [0.00013115, 0.0029432, 0.010037, 0.021886],
 -75839072.09717053)
```

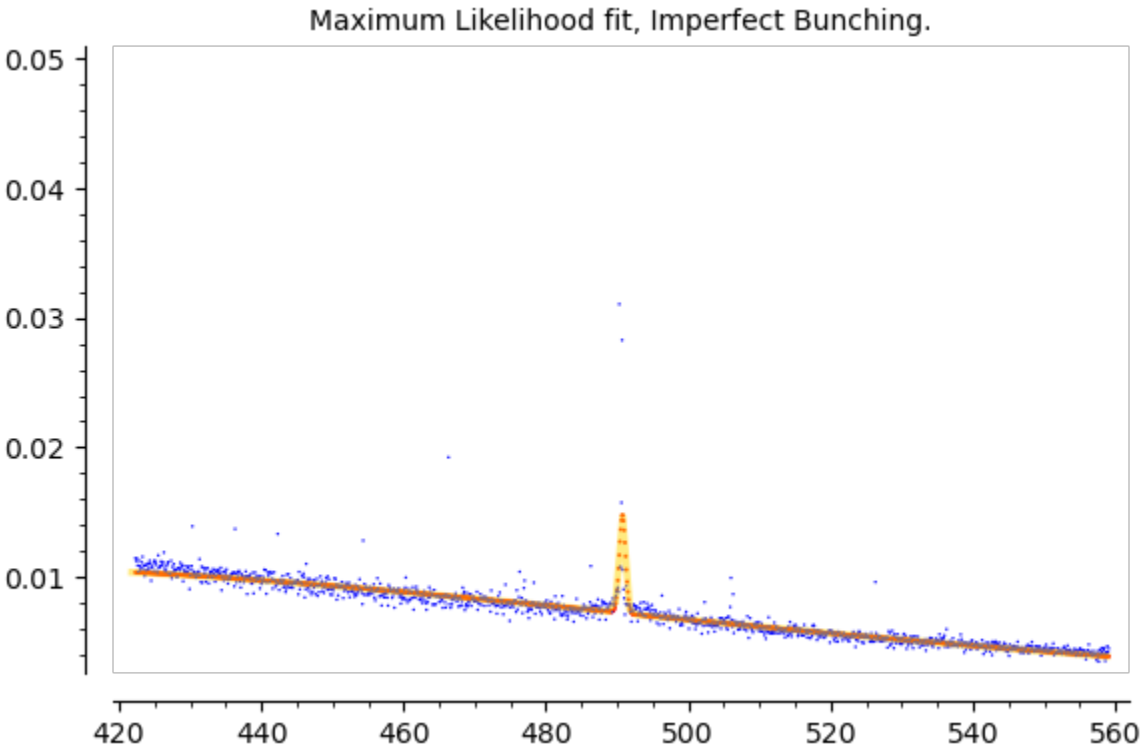
In [66]:
```
res_lik_print0(lik0ib,solib_Lll_all,VCM['hh0ibi']*VCM['jj0ib']*VCM['hh0ibi'],sqr
```

```
Likelihood Estimates, Homogenous Model, Imperfect Bunching, Sandwich
Parameter Estimate Std.Err.
  eti      0.006729   0.0050928
  mu       6.040      0.086573
  sigma1   1.406      0.27952
  v        7.033      0.82244
  number of observations |            1055452
          log_likelihood | -75839072.09717053
```

Out[66]:
```
((0.006728700551284908, 6.040399955254271, 1.4056877830495194, 7.033283703594654
5),
 [0.0050928, 0.086573, 0.27952, 0.82244],
 -75839072.09717053)
```

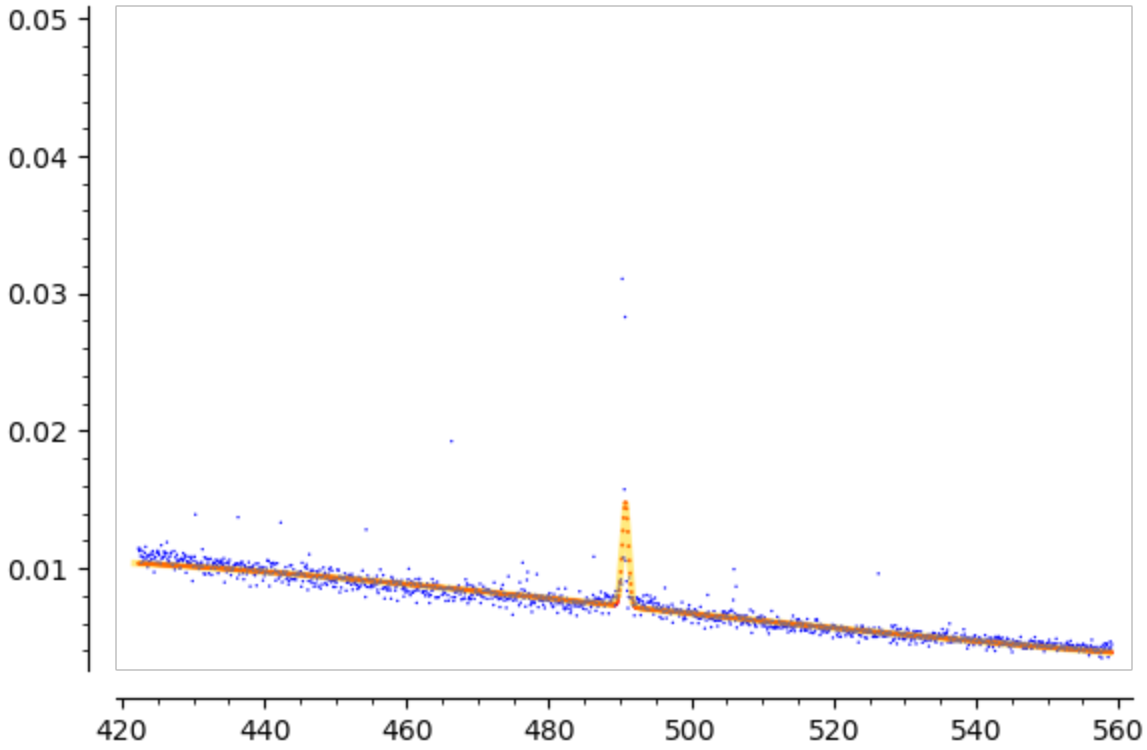
In [67]:
```
tri = solib_Lll_all
tax_par2 = (delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)
#plot( h0iba(x,*paramib(tri))/cndI0ib(tri),(x,exp(lnebu),exp(lnebo)),color='red'
resll0 = [ [n[0],\
    ((prob_alt(n[0],*paramib(tri,*tax_par2),*tax_par2)-prob_alt(n[0]-.1,*paramib
    ] for n in mkd['densa'] ]
resll1 = [ [n,\
    (h0iba(n,*paramib(tri,*tax_par2),*tax_par2))/cndI0ib(tri,*tax_par2)\
    ] for n in srange(exp(lnebu),exp(lnebo),1) ]
```

In [68]:
```
grfitib = list_plot(resll1,color='gold',plotjoined=True,thickness=2.5,alpha=0.5)
          list_plot(resll0,color='red',size=2)+\
          list_plot(mkd['densb'],size=1)
grfitib.show(title='Maximum Likelihood fit, Imperfect Bunching.',ymax=0.05)
#          list_plot(resls1,color='black',plotjoined=True,linestyle=':')+\
```



In [ ]:
```
### figure 5, left pane.
```

In [69]:
```
grfitib.show(ymax=0.05)
```



In [70]:
```
cndI0ib(solib_Lll_all,*tax_par2).n()*mkd['Nobs'],float(h0iba(490.,*paramib(solib
```

Out[70]: (392630.148751337, 0.0034700880475495893)

Chi Squared testing

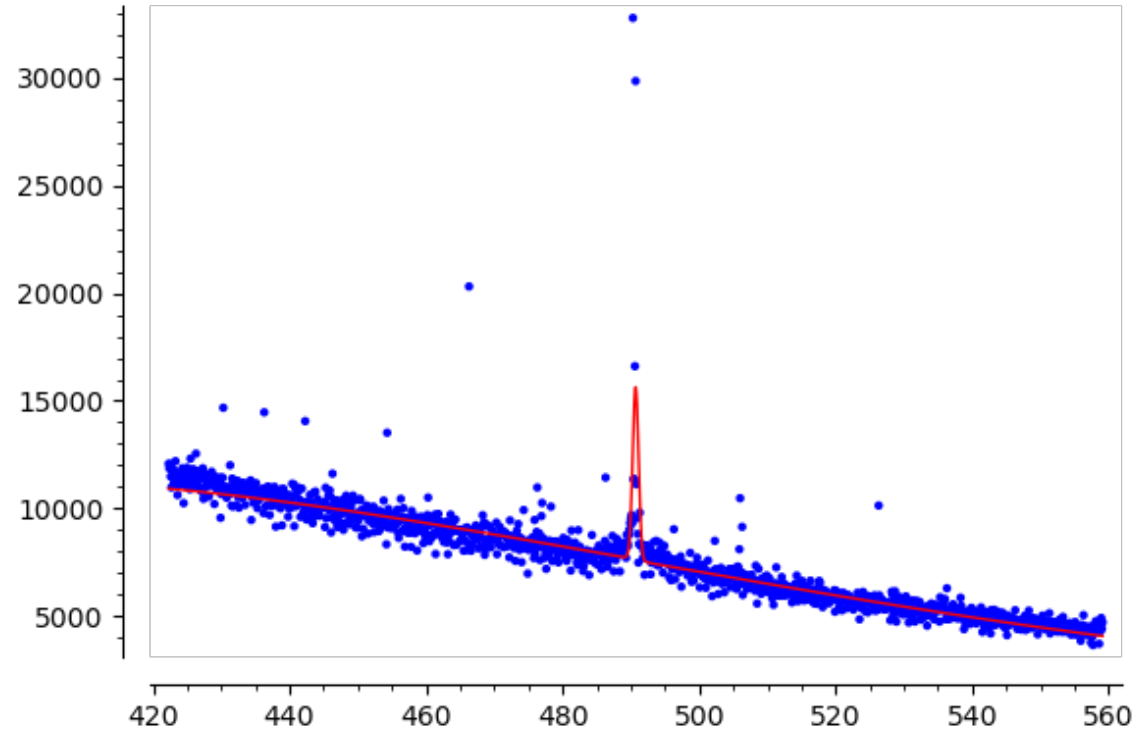
In [ ]:
```
```

In [71]:
```
MKC2=mak_chi2(densb,totnobs,solib_Lll_all,*tax_par2)
```

In [72]:
```
show(n(MKC2['Chi_squared stat'],digits=8))
```

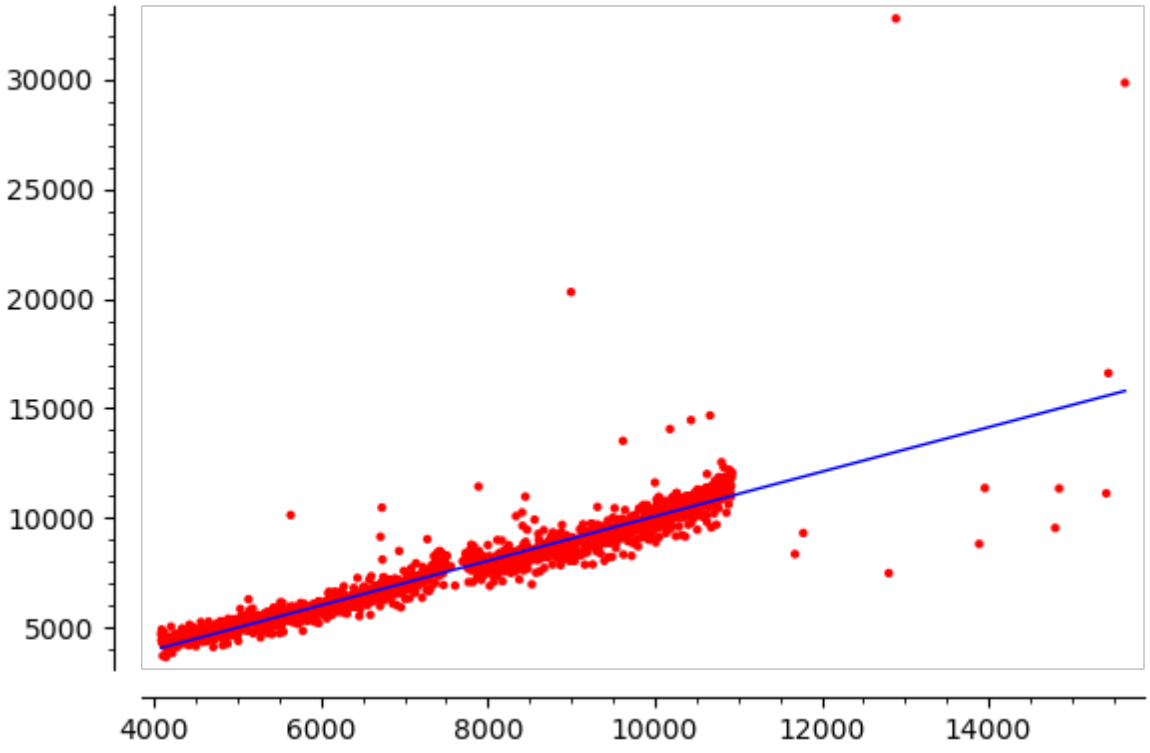
In [73]:
```
list_plot(MKC2['Observed #'],color='blue')+list_plot(MKC2['Theoretical #'],color
```

Out[73]:



In [74]:
```
MKOR = mak_OLS_R2(MKC2['Observed #'],MKC2['Theoretical #'])
```

In [75]:
```
MKOR['graph'].show()
```



In [76]:
```
MKOR['R2']
```

Out[76]: 0.8427697722401598

Selected Observations, +/- 40 around kink

In [77]:
```
#+/- 40 around kink
mkd2=mak_density(dd,Kink,40+0.001,STEP_BIN,delta)
totnobs = mkd2['Nobs']

KinkPos = int((len(mkd2['df'])-1)/2)

print(sum( n[1] for n in mkd2['dens']))

lnebu = ln(min(n[0] for n in mkd2['dens'])-STEP_BIN)-0.001
lnebo = ln(max(n[0] for n in mkd2['dens']))+0.001
```

```
1.0000000000000016
```

In [78]:
```
len(densb), mkd2['df'][KinkPos], mkd2['densb'][KinkPos]
```

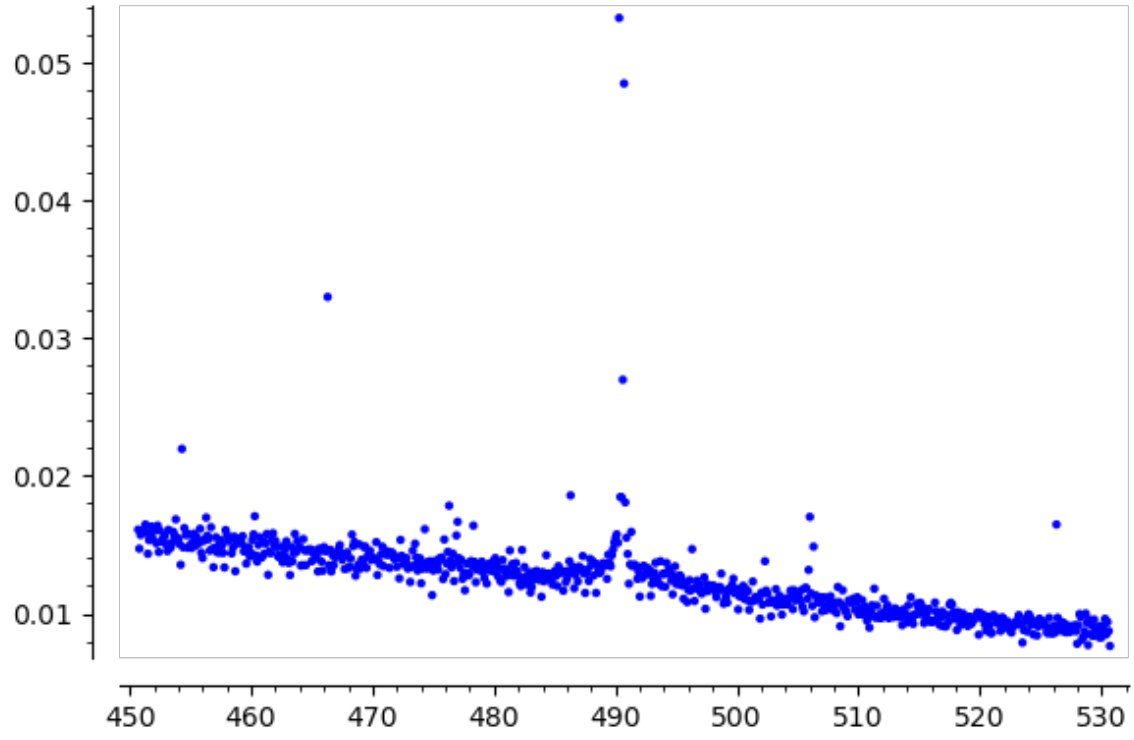
Out[78]:
```
(1369,
 (685.0, 2985.0, 2019.0, 490.7, 490700.0, 1.0, 490.7),
 (490.7, 0.04847250707601171))
```

In [79]:
```
lnebu,lnebo,exp(lnebu),exp(lnebo)
```

Out[79]:
```
(6.109580027998144, 6.2751968898338495, 450.14962522491857, 531.2309654384723)
```

In [80]:
```
list_plot(mkd2['densb'])
```

Out[80]:

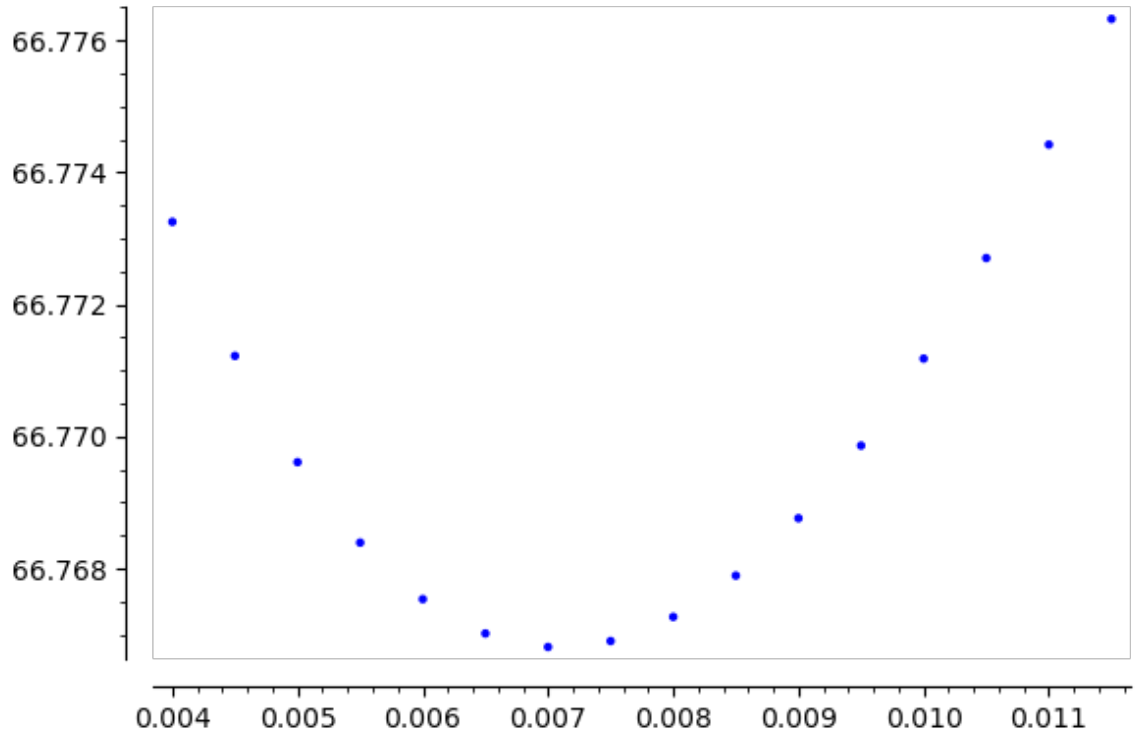


In [81]:
```
min(n[0] for n in mkd2['dens']),max(n[0] for n in mkd2['dens']),Kink
```

Out[81]:
```
(450.7, 530.7, 490.700000000000)
```

In [82]:
```
densb = mkd2['densb']
sqrtN =sqrt(mkd2['Nobs'])
tax_par = (densb,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)
#list_plot([ [xx,lik0ib([xx, 6., 1.4,7.0],*tax_par)] for xx in srange(0.0025,0.1
#list_plot([ [xx,lik0ib([0.0075, 6., 1.4,xx],*tax_par)] for xx in srange(1.0,9.0
#list_plot([ [xx,lik0ib([0.0075, xx, 1.4,7],*tax_par)] for xx in srange(5.5,6.5,
#list_plot([ [xx,lik0ib([0.0075, 6.05, xx,7],*tax_par)] for xx in srange(.8,1.8,
list_plot([ [xx,lik0ib([xx, 6.05, 1.425,7],*tax_par)] for xx in srange(0.004,0.0
# starting from almost correct solutions! searching for best standard error of f
#list_plot([ [xx,lik0ib([0.10347761423480781, 6.197688690159787, 1.9929051307530
```

Out[82]:



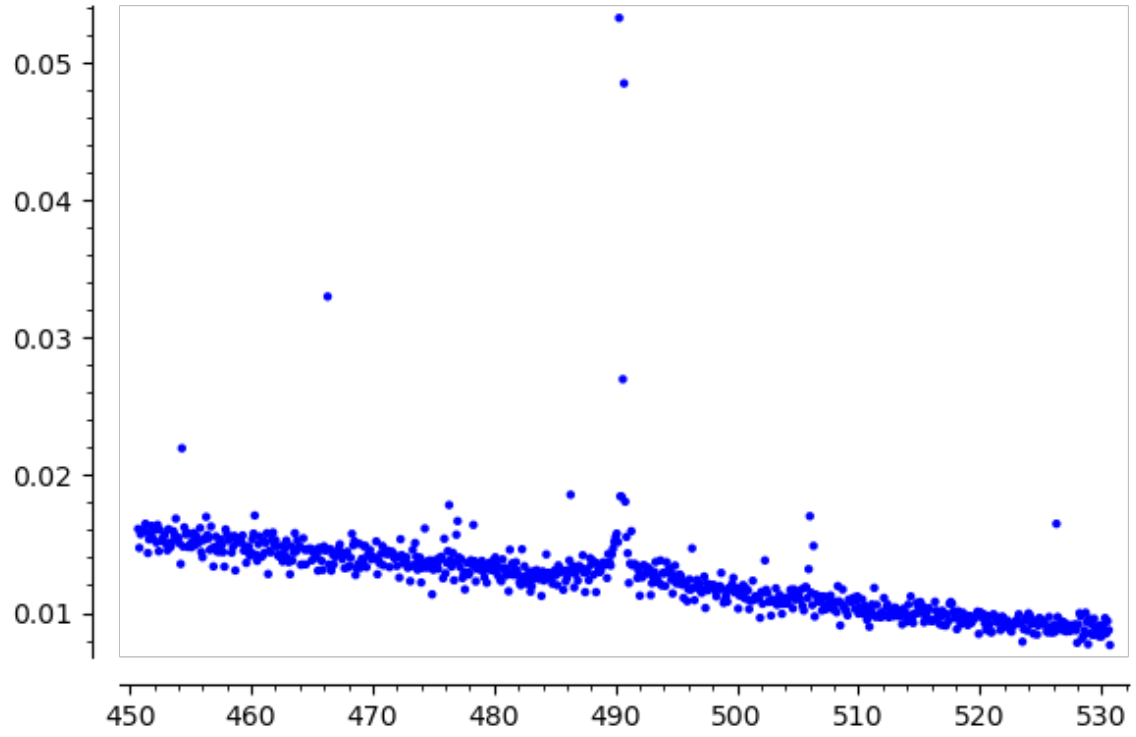
In [83]:
```
#[ [xx,lik0ib([0.08, 6.2, 2.32,xx],*tax_par)] for xx in srange(5.0,8.0,0.1)]
```

In [84]:
```
(delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)
```

Out[84]:
```
(0.0999990000000000,
 -0.417790757798801,
 -0.763891585085703,
 6.109580027998144,
 6.2751968898338495,
 490.700000000000,
 6.19583294309586)
```

In [85]:
```
list_plot(densb)
```

Out[85]:



In [86]:
```
lik0ib([0.081,  6., 1.35,1.7],*tax_par), lik0ib([0.082,  6., 1.35,1.7],*tax_par)
```

Out[86]:
```
(66.81341561552406, 66.81341549742702, 801, (490.7, 0.04847250707601171))
```

In [87]:
```
lik0ib([0.082,  6., 1.35,1.7],*tax_par)
```

Out[87]:
```
66.81341549742702
```

In [88]:
```
exp(lnebo),exp(lnebu),lnk,exp(6.9)
```

Out[88]:
```
(531.2309654384723, 450.14962522491857, 6.19583294309586, 992.274715605026)
```

In [89]:
```
#maximum likelihood estimator.
#tax_par = (('delta', delta),('lnt1c',lnt1c),('lnt2c',lnt2c),('lnebu',lnebu),('l
#solib_Lll_2 = minimize(lik0ib,[0.082, 6., 1.35,1.7],args=tax_par, gradient=None
#show(solib_Lll_2)
```

In [90]:
```
solib_Lll_2 = minimize(lik0ib,[ 0.007,6.05, 1.425,7],args=tax_par,  verbose=True
show(solib_Lll_2)
```

```
Optimization terminated successfully.
         Current function value: 66.761181
         Iterations: 118
         Function evaluations: 204
```

In [91]:
```
#calculation of var cov of parameters
VCM2= mak_vc(lik0ib,lik0ib_long,solib_Lll_2,len(densb),mkd2['Nobs'],STEP_BIN,*ta
```

In [92]:
```
VCM2['G_all']
```

Out[92]:
```
[ -3.456002340928073e-05 -3.1537405965506064e-06  2.3831050869813717e-07  -1.368
648192166504e-07]
```

In [93]:
```
res_lik_print0(lik0ib,solib_Lll_2,VCM2['hh0ibi'],sqrtN,int(mkd2['Nobs']),"Likeli
```

```
Likelihood Estimates, Homogenous Model, +/-40 around Kink, Imperfect Bunching, He
ssian
Parameter Estimate Std.Err.
  eti      0.006254   0.00012769
  mu       6.146      0.0013203
  sigma1   1.959      0.012689
  v        7.071      0.021986
  number of observations |              615813
          log_likelihood | -41112403.111829445
```

Out[93]:
```
((0.006253594264098166, 6.145526228424401, 1.9586484719446648, 7.07088830249937
9),
 [0.00012769, 0.0013203, 0.012689, 0.021986],
 -41112403.111829445)
```

In [94]:
```
res_lik_print0(lik0ib,solib_Lll_2,VCM2['hh0ibi']*VCM2['jj0ib']*VCM2['hh0ibi'],sq
```

```
Likelihood Estimates, Homogenous Model, +/-40 around Kink, Imperfect Bunching, Sa
ndwich
Parameter Estimate Std.Err.
  eti      0.006254   0.0049574
  mu       6.146      0.038037
  sigma1   1.959      0.34692
  v        7.071      0.81222
  number of observations |              615813
          log_likelihood | -41112403.111829445
```

Out[94]:
```
((0.006253594264098166, 6.145526228424401, 1.9586484719446648, 7.07088830249937
9),
 [0.0049574, 0.038037, 0.34692, 0.81222],
 -41112403.111829445)
```

In [95]:
```
tax_par2 = (delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)
MKC2b=mak_chi2(densb,totnobs,solib_Lll_2,*tax_par2)
```

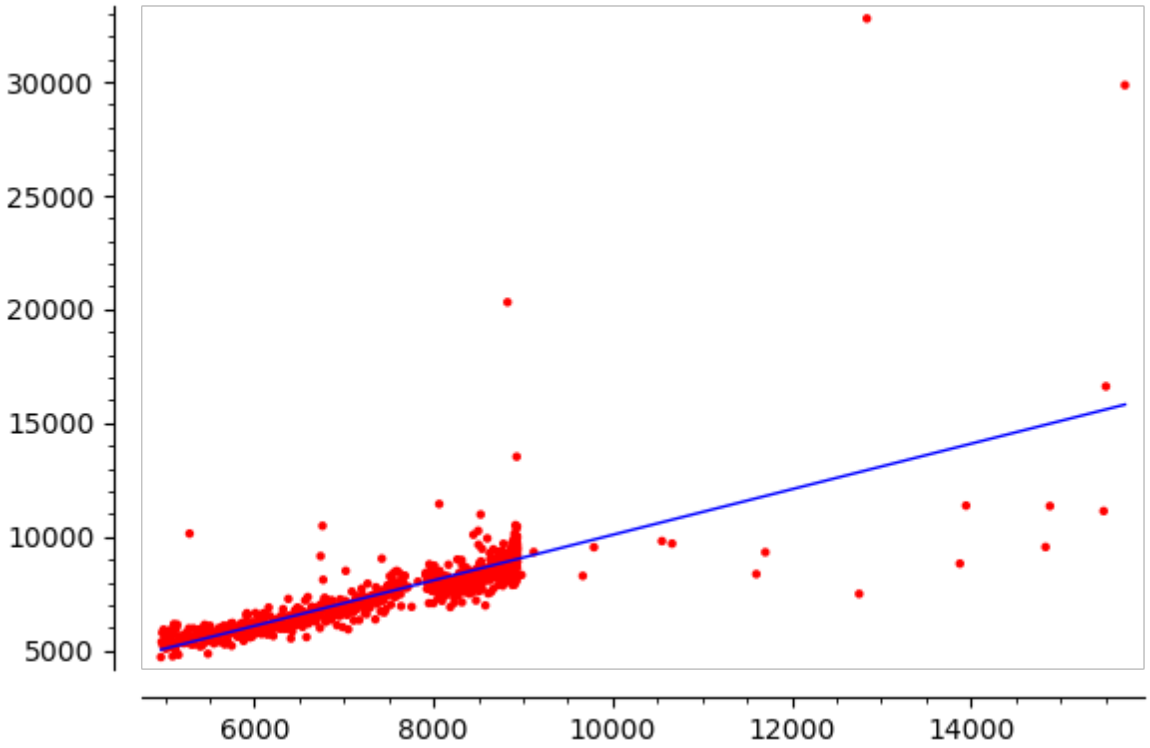
In [96]:
```
show(n(MKC2b['Chi_squared stat'],digits=8))
```

In [97]:
```
MKOR = mak_OLS_R2(MKC2b['Observed #'],MKC2b['Theoretical #'])
```

In [98]:
```
MKOR
```

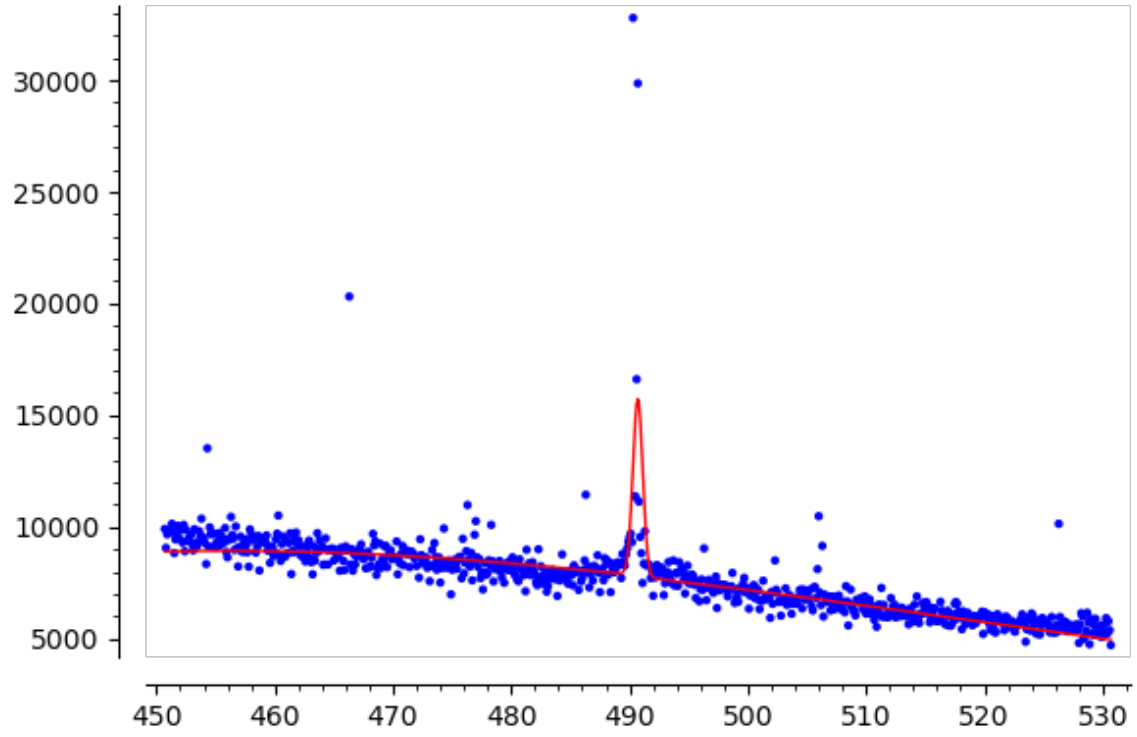
Out[98]:
```
{'a': 1.0009841276957259,
 'b': 77.90599962510564,
 'Nobs': 6158130.0,
 'graph': Graphics object consisting of 2 graphics primitives,
 'Var obs': 3635928.702247189,
 'Var theoretical': 2237466.5554707684,
 'R2': 0.6153769060674651}
```

In [99]:
```
MKOR['graph'].show()
```



In [100…
```
list_plot(MKC2b['Observed #'],color='blue')+list_plot(MKC2b['Theoretical #'],col
```

Out[100]:



In [101…
```
MKOR['R2']
```

Out[101]: 0.6153769060674651

Selected Observations, +/- 15 around kink

In [102…
```
#+/- 40 around kink
mkd3=mak_density(dd,Kink,15+0.001,STEP_BIN,delta)
totnobs = mkd3['Nobs']

KinkPos = int((len(mkd3['df'])-1)/2)

print(sum( n[1] for n in mkd3['dens']))

lnebu = ln(min(n[0] for n in mkd3['dens'])-STEP_BIN)
lnebo = ln(max(n[0] for n in mkd3['dens']))
```

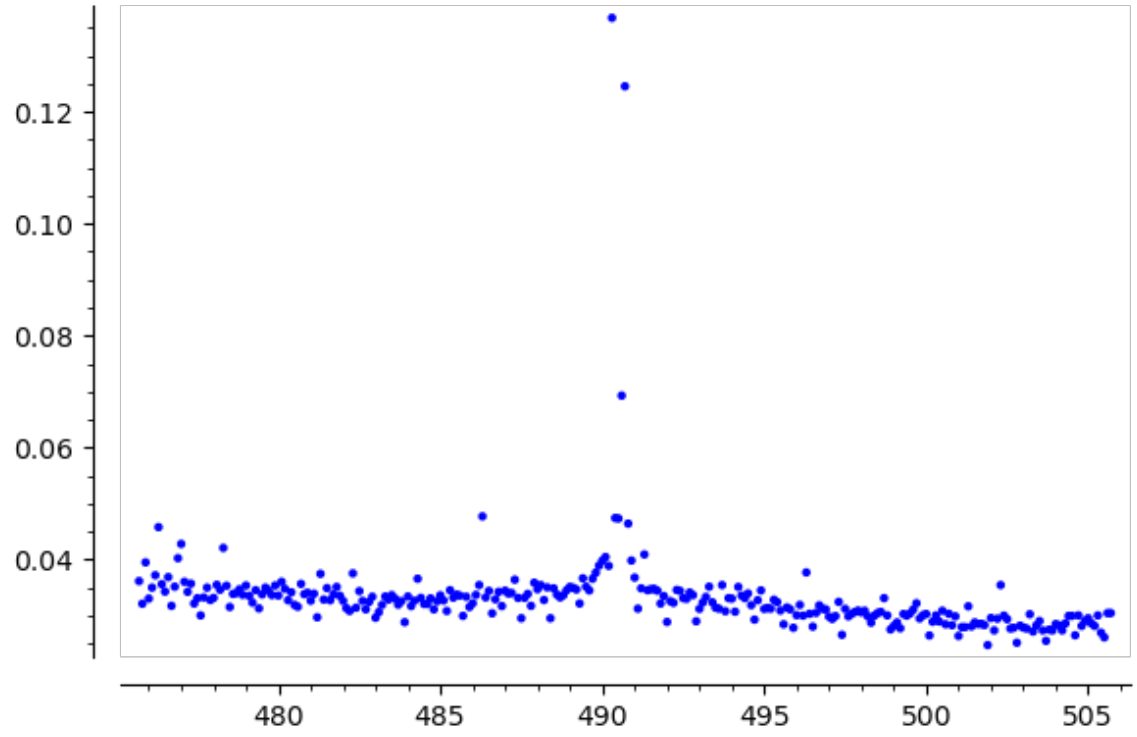
```
0.9999999999999998
```

In [103…
```
len(mkd3['densb']), mkd3['df'][KinkPos], mkd3['densb'][KinkPos]
```

Out[103]:
```
(301,
 (685.0, 2985.0, 2019.0, 490.7, 490700.0, 1.0, 490.7),
 (490.7, 0.12454001552056473))
```

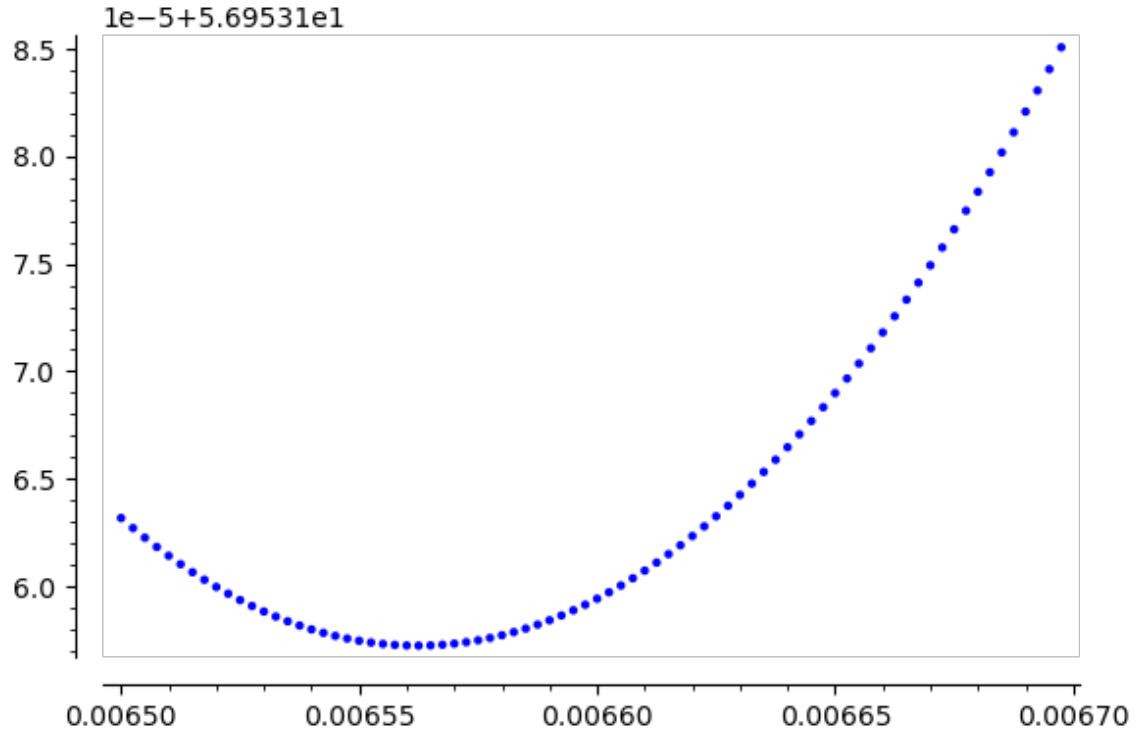
In [104…
```
list_plot(mkd3['densb'])
```

Out[104]:



In [105…
```
densb = mkd3['densb']
sqrtN =sqrt(mkd3['Nobs'])
tax_par = (densb,delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)
#list_plot([ [xx,lik0ib([xx, 6.05, 1.425,7.05],*tax_par)] for xx in srange(0.001
#list_plot([ [xx,lik0ib([0.006556, 6.05, 1.425,xx],*tax_par)] for xx in srange(6
#list_plot([ [xx,lik0ib([0.006556, xx, 1.425,7.05],*tax_par)] for xx in srange(5
#list_plot([ [xx,lik0ib([0.006556, 6.05, xx,7.05],*tax_par)] for xx in srange(1.
list_plot([ [xx,lik0ib([xx, 6.05, 1.395,7.05],*tax_par)] for xx in srange(0.0065
```

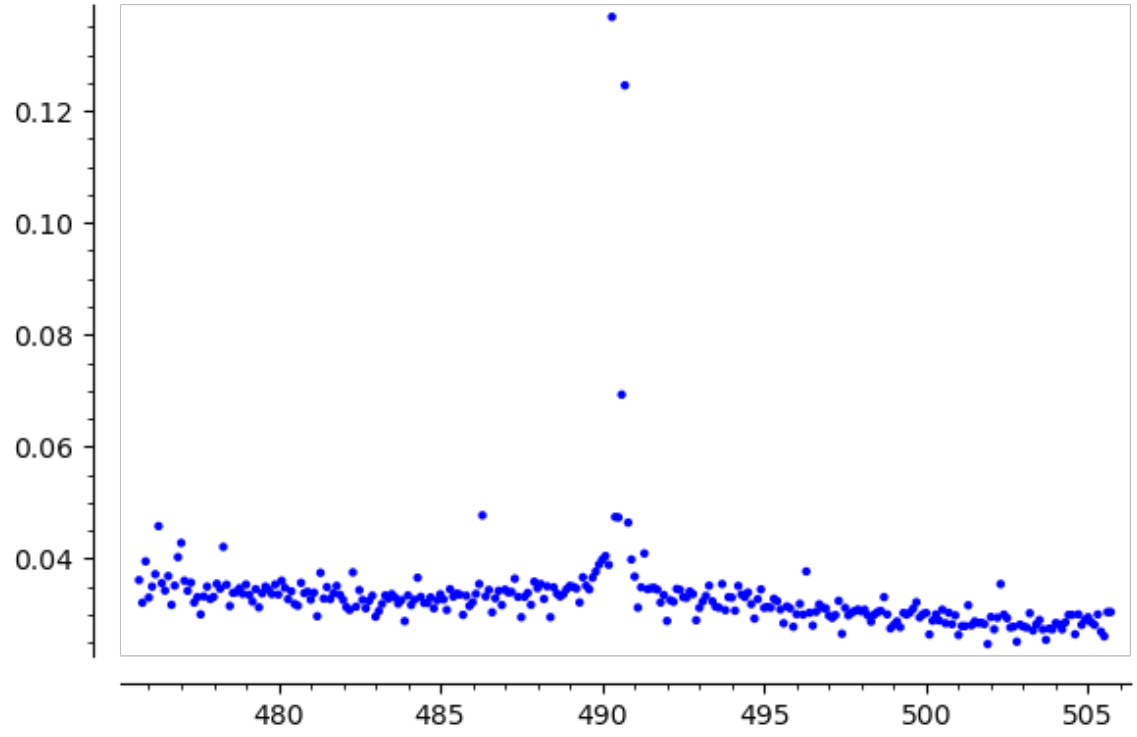
Out[105]:



In [106…

```
list_plot(densb)
```

Out[106]:



In [107…

```
(delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk),exp(6.1645771648166265)
```

Out[107]:

```
((0.0999990000000000,
  -0.417790757798801,
  -0.763891585085703,
  6.1645771648166265,
  6.225943608085937,
  490.700000000000,
  6.19583294309586),
 475.5999999999999)
```

In [108…
```
lik0ib([ 0.006556, 6.05, 1.395,7.05],*tax_par)
```

Out[108]:
```
56.953157301934525
```

In [109…
```
#maximum likelihood estimator.
#tax_par = (('delta', delta),('lnt1c',lnt1c),('lnt2c',lnt2c),('lnebu',lnebu),('l
#tax_par2 = (delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)
solib_Lll_3 = minimize(lik0ib,[0.006556, 6.05, 1.395,7.05],args=tax_par, verbose
show(solib_Lll_3)
```

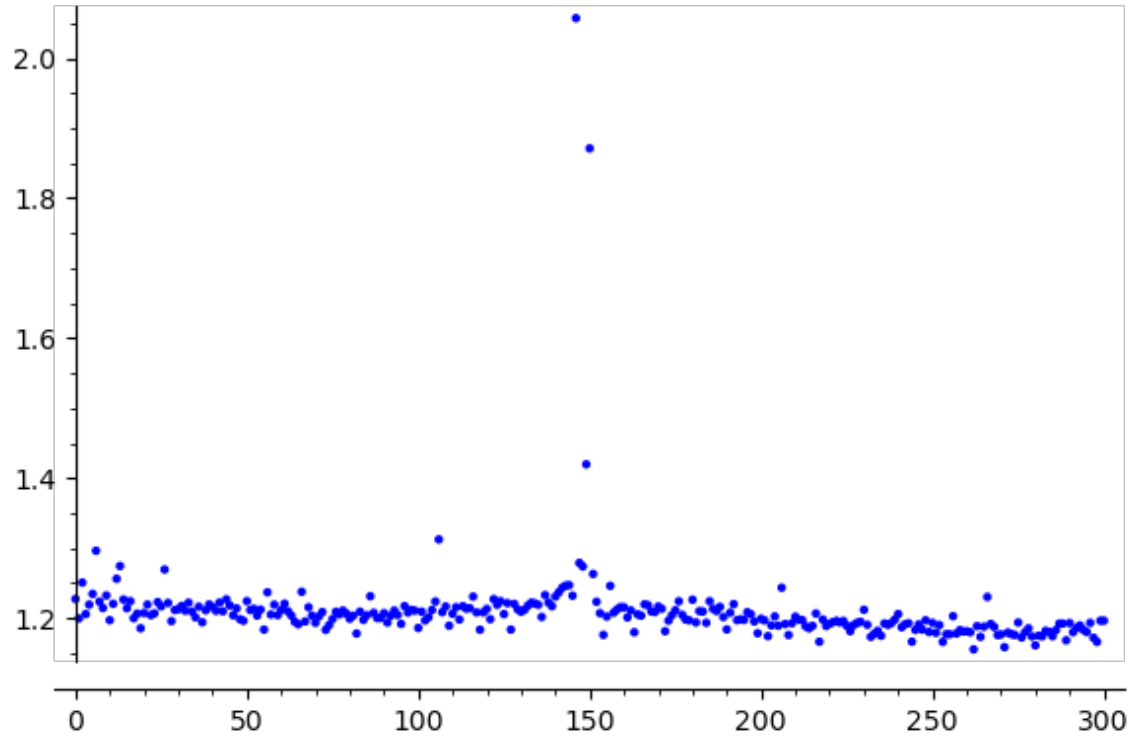
```
Optimization terminated successfully.
         Current function value: 56.951852
         Iterations: 118
         Function evaluations: 207
```

In [110…
```
tax_par2 = (delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)
cndI0ib(solib_Lll_3,*tax_par2)
show(lik0ib([0.0065555, 6.05, 1.395,7.05],*tax_par))
grdcd(lik0ib,[0.0065555, 6.05, 1.395,7.05],*tax_par)
```

Out[110]:
```
[-0.02049868874970422,
 0.002658277986410447,
 -0.0028520183556140473,
 0.0003570007060052111]
```

In [111…
```
list_plot([exp(nn) for nn in lik0ib_long(solib_Lll_3,*tax_par)])
```

Out[111]:



In [112…
```
#calculation of var cov of parameters
VCM3= mak_vc(lik0ib,lik0ib_long,solib_Lll_3,len(densb),mkd3['Nobs'],STEP_BIN,*ta
```

In [113…
```
res_lik_print0(lik0ib,solib_Lll_3,VCM3['hh0ibi'],sqrtN,int(mkd3['Nobs']),"Likeli
```

```
Likelihood Estimates, Homogenous Model, +/-15 around Kink, Imperfect Bunching, He
ssian
Parameter Estimate Std.Err.
  eti      0.006255   0.00013801
  mu       6.174      0.0034649
  sigma1   2.313      0.067216
  v        7.071      0.022384
  number of observations |              239682
          log_likelihood | -13650333.791345118
```

Out[113]:
```
((0.006254621598164397, 6.173913606528919, 2.3125690962520116, 7.07055257009319
7),
 [0.00013801, 0.0034649, 0.067216, 0.022384],
 -13650333.791345118)
```

In [114…
```
res_lik_print0(lik0ib,solib_Lll_3,VCM3['hh0ibi']*VCM3['jj0ib']*VCM3['hh0ibi'],sq
```

```
Likelihood Estimates, Homogenous Model, +/-15 around Kink, Imperfect Bunching, Sa
ndwich
Parameter Estimate Std.Err.
  eti      0.006255   0.0051755
  mu       6.174      0.093104
  sigma1   2.313      1.7611
  v        7.071      0.82626
  number of observations |              239682
          log_likelihood | -13650333.791345118
```

Out[114]:
```
((0.006254621598164397, 6.173913606528919, 2.3125690962520116, 7.07055257009319
7),
 [0.0051755, 0.093104, 1.7611, 0.82626],
 -13650333.791345118)
```

In [115…
```
tax_par2 = (delta,lnt1c,lnt2c,lnebu,lnebo,Kink,lnk)
MKC2c=mak_chi2(densb,totnobs,solib_Lll_3,*tax_par2)
```

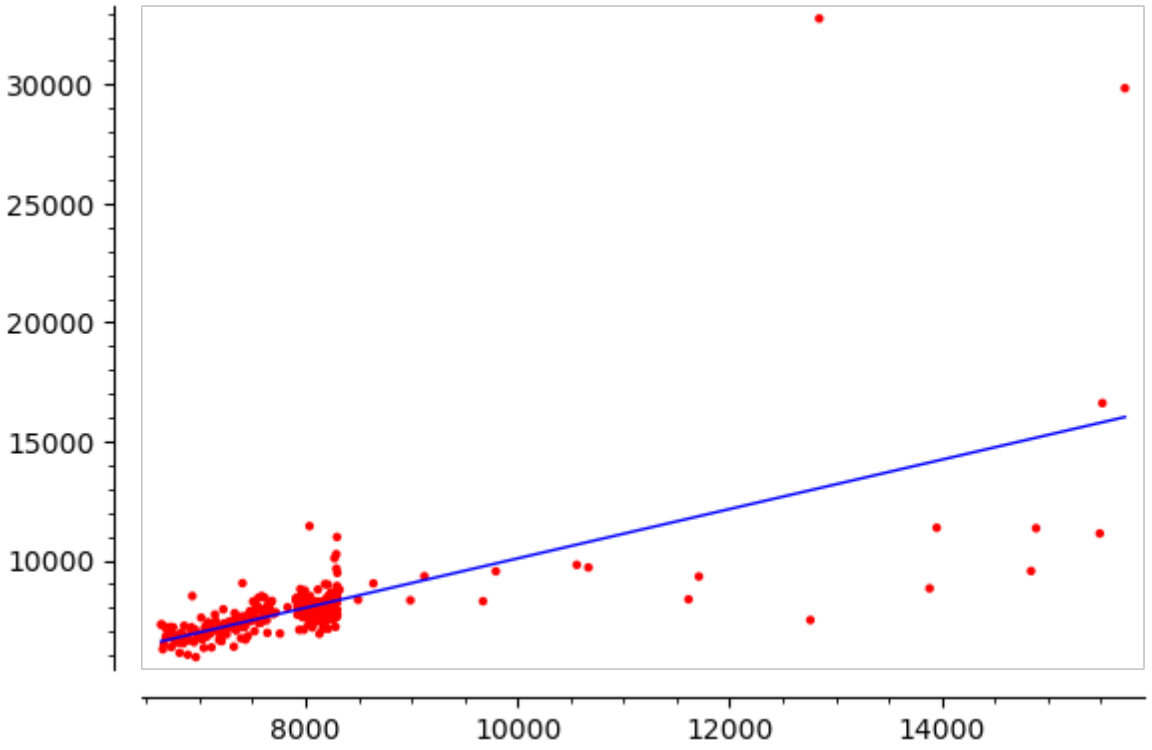
In [116…
```
show(n(MKC2c['Chi_squared stat'],digits=8))
```

In [117…
```
MKOR = mak_OLS_R2(MKC2c['Observed #'],MKC2c['Theoretical #'])
```

In [118…
```
MKOR
```

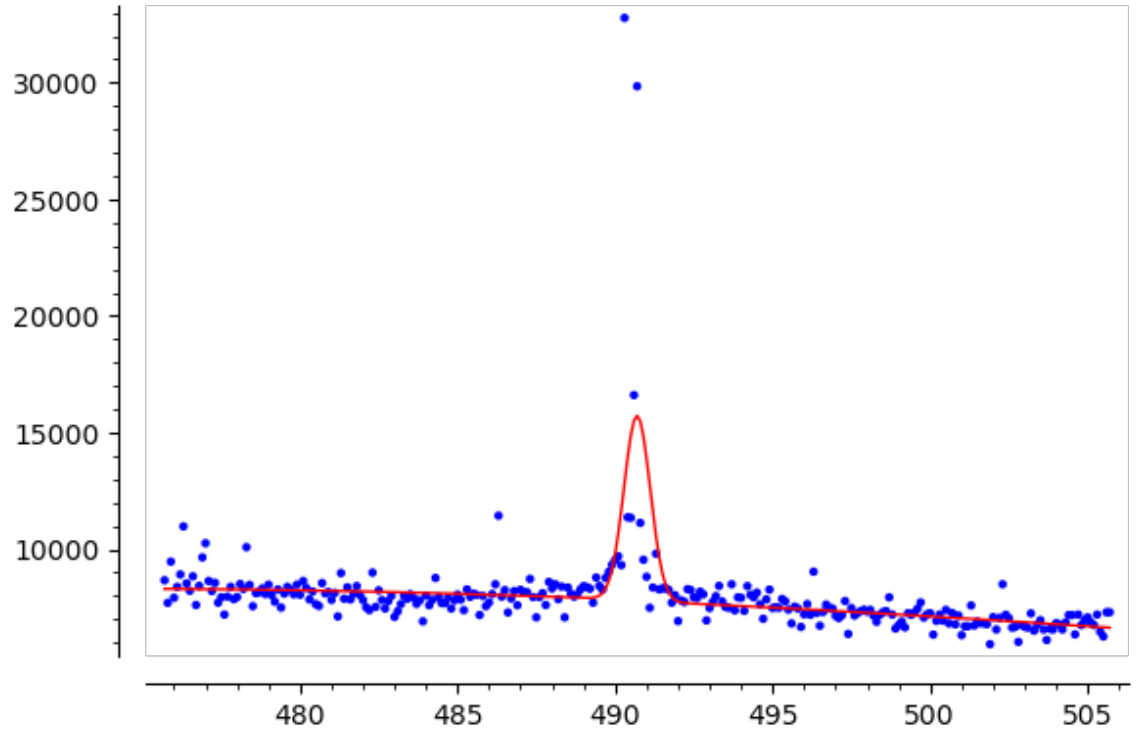
Out[118]:
```
{'a': 1.0381645158475683,
 'b': -300.68788561315034,
 'Nobs': 2396820.0,
 'graph': Graphics object consisting of 2 graphics primitives,
 'Var obs': 4663937.142857137,
 'Var theoretical': 1935219.4996001595,
 'R2': 0.41493258599421007}
```

In [119…
```
MKOR['graph'].show()
```



In [120…
```
list_plot(MKC2c['Observed #'],color='blue')+list_plot(MKC2c['Theoretical #'],col
```

Out[120]:



In [121…
```
MKOR['R2']
```

Out[121]: 0.41493258599421007

In [ ]: