Appendix 1: Estimation of the two-variable saturated model in SPSS, Stata and R using the Netherlands 1973 example data

A. SPSS commands and corresponding parameter estimates

Copy the 1973 data from the "contrast coding" sheet of the **accompanying Excel worksheet**, and paste into the data editor of SPSS. Name the three *variables: orig, dest, and count.

```
weight by count.
GENLOG orig dest
 /MODEL=POISSON
 /print estim.
```

Parameter	Estimate	Std. Error	Z	Sig.
Constant	10.056	.007	1535.191	.000
[orig = 1]	-.212	.010	-21.622	.000
[orig = 2]	-.243	.010	-24.566	.000
[orig = 3]	-.168	.010	-17.354	.000
[orig = 4]	-.753	.012	-65.054	.000
[orig = 5]	-.133	.010	-13.851	.000
[orig = 6]	0[a]	.	.	.
[dest = 1]	.698	.008	87.012	.000
[dest = 2]	.518	.008	62.610	.000
[dest = 3]	.598	.008	73.394	.000
[dest = 4]	.071	.009	7.769	.000
[dest = 5]	.141	.009	15.734	.000
[dest = 6]	0[a]	.	.	.
[orig = 1] * [dest = 1]	.288	.012	24.555	.000
[orig = 1] * [dest = 2]	-.284	.013	-22.214	.000
[orig = 1] * [dest = 3]	-1.387	.015	-90.268	.000
[orig = 1] * [dest = 4]	.076	.013	5.672	.000
[orig = 1] * [dest = 5]	-.289	.014	-20.695	.000
[orig = 1] * [dest = 6]	0[a]	.	.	.
[orig = 2] * [dest = 1]	-.384	.013	-30.490	.000
[orig = 2] * [dest = 2]	-.109	.013	-8.592	.000
[orig = 2] * [dest = 3]	-1.565	.016	-96.148	.000
[orig = 2] * [dest = 4]	-.315	.014	-21.843	.000
[orig = 2] * [dest = 5]	-.261	.014	-18.620	.000
[orig = 2] * [dest = 6]	0[a]	.	.	.
[orig = 3] * [dest = 1]	-.926	.013	-69.259	.000
[orig = 3] * [dest = 2]	-1.128	.015	-77.332	.000
[orig = 3] * [dest = 3]	-.949	.014	-68.979	.000
[orig = 3] * [dest = 4]	-1.216	.017	-71.024	.000
[orig = 3] * [dest = 5]	-.837	.015	-54.874	.000
[orig = 3] * [dest = 6]	0[a]	.	.	.
[orig = 4] * [dest = 1]	.062	.014	4.416	.000
[orig = 4] * [dest = 2]	.811	.014	59.864	.000
[orig = 4] * [dest = 3]	-1.505	.020	-76.892	.000
[orig = 4] * [dest = 4]	-.143	.016	-8.665	.000
[orig = 4] * [dest = 5]	-.297	.017	-17.824	.000
[orig = 4] * [dest = 6]	0[a]	.	.	.
[orig = 5] * [dest = 1]	-.327	.012	-26.992	.000
[orig = 5] * [dest = 2]	-.304	.013	-24.291	.000
[orig = 5] * [dest = 3]	-1.146	.014	-80.991	.000
[orig = 5] * [dest = 4]	-.509	.014	-35.328	.000
[orig = 5] * [dest = 5]	-.385	.014	-27.834	.000
[orig = 5] * [dest = 6]	0[a]	.	.	.
[orig = 6] * [dest = 1]	0[a]	.	.	.
[orig = 6] * [dest = 2]	0[a]	.	.	.
[orig = 6] * [dest = 3]	0[a]	.	.	.
[orig = 6] * [dest = 4]	0[a]	.	.	.
[orig = 6] * [dest = 5]	0[a]	.	.	.
[orig = 6] * [dest = 6]	0[a]	.	.	.

a. This parameter is set to zero because it is redundant.

B. Stata commands and corresponding parameter estimates

Copy the 1973 data from the "contrast coding" sheet of the accompanying Excel worksheet, and paste into the data editor of Stata. Name the three variables: orig, dest, and count.
xi: glm count i.orig*i.dest, family(poisson)

	OIM					
	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_Iorig_2	-0.70286	0.007733	-90.89	0	-0.71801	-0.6877
_Iorig_3	-1.16987	0.009143	-127.95	0	-1.18779	-1.15195
_Iorig_4	-0.76676	0.007902	-97.04	0	-0.78225	-0.75128
_Iorig_5	-0.53592	0.007324	-73.17	0	-0.55027	-0.52156
_Iorig_6	-0.07573	0.006416	-11.8	0	-0.0883	-0.06316
_Idest_2	-0.75103	0.007859	-95.56	0	-0.76643	-0.73563
_Idest_3	-1.77413	0.011685	-151.83	0	-1.79704	-1.75123
_Idest_4	-0.83792	0.008098	-103.47	0	-0.85379	-0.82204
_Idest_5	-1.13298	0.009016	-125.66	0	-1.15065	-1.11531
_Idest_6	-0.9851	0.008534	-115.43	0	-1.00183	-0.96838
_IoriXde~2_2	0.847449	0.011751	72.12	0	0.824418	0.87048
_IoriXde~2_3	0.49423	0.017899	27.61	0	0.45915	0.529311
_IoriXde~2_4	0.280916	0.013243	21.21	0	0.254961	0.306872
_IoriXde~2_5	0.699656	0.013526	51.73	0	0.673144	0.726167
_IoriXde~2_6	0.671944	0.012943	51.91	0	0.646576	0.697312
_IoriXde~3_2	0.370141	0.014797	25.02	0	0.341141	0.399142
_IoriXde~3_3	1.652082	0.016505	100.1	0	1.619733	1.684432
_IoriXde~3_4	-0.07901	0.016999	-4.65	0	-0.11233	-0.0457
_IoriXde~3_5	0.665232	0.015714	42.33	0	0.634433	0.696031
_IoriXde~3_6	1.213705	0.013689	88.66	0	1.186875	1.240535
_IoriXde~4_2	0.246918	0.013228	18.67	0	0.220992	0.272844
_IoriXde~4_3	0.107589	0.020121	5.35	0	0.068153	0.147026
_IoriXde~4_4	0.006017	0.014358	0.42	0.675	-0.02213	0.034158
_IoriXde~4_5	0.21703	0.015181	14.3	0	0.187276	0.246784
_IoriXde~4_6	0.225371	0.014374	15.68	0	0.1972	0.253543
_IoriXde~5_2	0.594514	0.011627	51.13	0	0.571725	0.617302
_IoriXde~5_3	0.856386	0.015972	53.62	0	0.82508	0.887691
_IoriXde~5_4	0.029082	0.013244	2.2	0.028	0.003124	0.05504
_IoriXde~5_5	0.518239	0.013331	38.87	0	0.492111	0.544367
_IoriXde~5_6	0.614916	0.012477	49.28	0	0.59046	0.639371
_IoriXde~6_2	0.571537	0.010425	54.82	0	0.551104	0.591971
_IoriXde~6_3	1.675054	0.013472	124.34	0	1.64865	1.701458
_IoriXde~6_4	0.211068	0.011266	18.74	0	0.188987	0.233149
_IoriXde~6_5	0.576297	0.011829	48.72	0	0.553113	0.599481
_IoriXde~6_6	0.287524	0.011709	24.56	0	0.264574	0.310474
_cons	10.82969	0.00445	2433.62	0	10.82097	10.83841

C. R commands and corresponding parameter estimates

Create a tab-delimited text file of the 1973 data from the "multiplicative components" sheet of the **accompanying Excel worksheet**. One easy way to do this is to copy the data and variable names and paste them into an empty Notepad window. Saving this Notepad window will result in a tab-delimited text file. In this example the data are saved as "Netherlands 73.txt."

The following R command specifies the working directory and where to find the data file. For Windows computers, change the slashes from the standard '\' to a '/' in the file path. For example, the working directory "C:\Documents\" would need to be specified as:

setwd("C:/Documents/")
Read the data and store it in object "data"
data=read.table("Netherlands 1973.txt", header=TRUE, colClasses=c("factor", "factor", "numeric"))
Estimate the saturated log-linear model
sat.model = glm(count ~orig*dest, data=data, family=poisson)
Print the output to the screen
sat.model

```
Coefficients:
  (Intercept)        orig2        orig3        orig4        orig5        orig6
    10.829689    -0.702858    -1.169867    -0.766765    -0.535918    -0.075730
        dest2        dest3        dest4        dest5        dest6  orig2:dest2
    -0.751030    -1.774133    -0.837916    -1.132980    -0.985103     0.847449
  orig3:dest2  orig4:dest2  orig5:dest2  orig6:dest2  orig2:dest3  orig3:dest3
     0.370141     0.246918     0.594514     0.571537     0.494230     1.652082
  orig4:dest3  orig5:dest3  orig6:dest3  orig2:dest4  orig3:dest4  orig4:dest4
     0.107589     0.856386     1.675054     0.280916    -0.079013     0.006017
  orig5:dest4  orig6:dest4  orig2:dest5  orig3:dest5  orig4:dest5  orig5:dest5
     0.029082     0.211068     0.699656     0.665232     0.217030     0.518239
  orig6:dest5  orig2:dest6  orig3:dest6  orig4:dest6  orig5:dest6  orig6:dest6
     0.576297     0.671944     1.213705     0.225371     0.614915     0.287524
```

Appendix 2: Estimation of the independence model in SPSS, Stata and R using the Netherlands 1973 example data

A. SPSS commands with corresponding goodness-of-fit output

Copy the 1973 data from the “contrast coding” sheet of the **accompanying Excel worksheet**, and paste into the data editor of SPSS. Name the three variables: orig, dest, and count. The GENLIN procedure is preferred over the GENLOG because GENLIN produces more fit statistics.

```
GENLIN count by orig dest
 /model orig dest distribution=poisson link=log .
```

Goodness of Fit[a]

	Value	df	Value/df
Deviance	46477.626	25	1859.105
Scaled Deviance	46477.626	25	
Pearson Chi-Square	47623.168	25	1904.927
Scaled Pearson Chi-Square	47623.168	25	
Log Likelihood[b]	-23447.409		
Akaike's Information Criterion (AIC)	46916.818		
Finite Sample Corrected AIC (AICC)	46927.818		
Bayesian Information Criterion (BIC)	46934.237		
Consistent AIC (CAIC)	46945.237		

Dependent Variable: count

a. Information criteria are in small-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

B. Stata commands with corresponding goodness-of-fit output

Copy the 1973 data from the "contrast coding" sheet of the accompanying Excel worksheet, and paste into the data editor of Stata. Name the three variables: orig, dest, and count.
xi: glm count i.orig i.dest, family(poisson)

```
Generalized linear models                          No. of obs      =         36
Optimization     : ML                              Residual df     =         25
                                                   Scale parameter =          1
Deviance         =  46477.62562                    (1/df) Deviance =   1859.105
Pearson          =  47623.16805                    (1/df) Pearson  =   1904.927

Variance function: V(u) = u                        [Poisson]
Link function    : g(u) = ln(u)                    [Log]

                                                   AIC             =   1303.245
Log likelihood   = -23447.40925                    BIC             =   46388.04
```

C. R commands with corresponding goodness-of-fit output

Create a tab-delimited text file of the 1973 data from the "contrast coding" sheet of the accompanying Excel worksheet. One easy way to do this is to copy the data and variable names and paste them into an empty Notepad window. Saving this Notepad window will result in a tab-delimited text file. In this example the data were saved as "Netherlands 73.txt."

The following R command specifies the working directory and where to find the data file. For Windows computers, change the slashes from the standard '\' to a '/' in the file path. For example, the working directory "C:\Documents\" would need to be specified as:

setwd("C:/Documents/")
Read the data and store it in object "data"
data=read.table("Netherlands 1973.txt", header=TRUE, colClasses=c("factor", "factor", "numeric"))
Estimate the independence model
indep.model = glm(count ~orig+dest, data=data, family=poisson)
Print the output to the screen
indep.model

```
Degrees of Freedom: 35 Total (i.e. Null);  25 Residual
Null Deviance:      201700
Residual Deviance: 46480        AIC: 46920
```

Appendix 3: Estimation of the quasi-independence model in SPSS, Stata and R using the U.S. 1985-1990 flow data

A. SPSS commands

Copy the U.S. 1985-1990 data from the “quasi-independence” sheet of the **accompanying Excel worksheet**, and paste into the data editor of SPSS. Name the three variables: orig, dest, and count. GENLOG is used to demonstrate the structural zeros approach.

```
*Create the binary variable (wt) that will force zeros on the diagonal.
COMPUTE wt=0.
IF (orig ~= dest)wt=1.
WEIGHT BY count.
*Estimate the model and save the predicted values in variable pcount.
GENLOG orig dest
  /CSTRUCTURE=wt
  /MODEL=POISSON
  /PRINT=NONE
  /PLOT=NONE
  /CRITERIA=CIN(95) ITERATE(20) CONVERGE(0.001) DELTA(.5)
  /DESIGN dest orig
  /SAVE=PRED(pcount).
```

B. Stata commands

Copy the U.S. 1985-1990 data from the “quasi-independence” sheet of the **accompanying Excel worksheet**, and paste into the data editor of Stata. Name the three variables: orig, dest, and count. Two methods are demonstrated. First , the structural zeros method and then the diagonal effects method.
For the structural zeros method, first create the binary variable (wt) that forces zeros on the diagonal.

```
gen wt=0
replace wt=1 if orig!=dest

/*Estimate the model and save the predicted values in variable pcount. */
glm count i.orig i.dest [fweight=wt], family(poisson)
predict pcount if e(sample), mu

/*For the alternative diagonal effects approach, first, set the flows equal to zero when orig equals dest.*/
replace count=0 if orig==dest

/*Create diagonal effects.*/
gen diag=1
replace diag=2 if orig==1 & dest==1
replace diag=3 if orig==2 & dest==2
replace diag=4 if orig==3 & dest==3
replace diag=5 if orig==4 & dest==4

/*Estimate the model and save the predicted values in variable pcount2. */
glm newcount i.orig i.dest i.diag, family(poisson)
predict pcount2 if e(sample), mu
```

C. R commands

Create a tab-delimited text file of the U.S. 1985-1990 data from the “quasi-independence” sheet of the **accompanying Excel worksheet**. One easy way to do this is to copy the data and variable names and paste them into an empty Notepad window. Saving this Notepad window will result in a tab-delimited text file. In this example, the data were saved as “U.S. 1985-90.txt.”

The following R command specifies the working directory and where to find the data file. For Windows computers, change the slashes from the standard '\' to a '/' in the file path. For example, the working directory “C:\Documents\” would need to be specified as:

```
setwd("C:/Documents/")

Read the data and store it in object "data"
data=read.table("U.S. 1985-90.txt ", header=TRUE, colClasses=c("factor", "factor", "numeric"))

For the structural zeros method, first create the binary variable (wt) that
forces zeros on the diagonal.
data$wt <- as.numeric(data$orig != data$dest)

##Estimate the quasi-independence model
quasi.indep.model<-glm(count~orig+dest,data=data,weights=wt,family=poisson())

Print the output to the screen
summary(quasi.indep.model)
##Save the predicted flows as pcount to the "data" data frame
data$pcount=quasi.indep.model$weights
##Print out the data frame to view fitted values
print(data)
```

Appendix 4: The method of offsets in SPSS, Stata and R: Estimating the 1976 flows in the Netherlands using the 1973 flows as the offset

A. SPSS commands

Copy the Netherlands data from the "method of offsets" sheet of the **accompanying Excel worksheet**, and paste into the data editor of SPSS. Name the four variables: orig, dest, c1973 and c1976. Either GENLOG or GENLIN may be used.

```
*Using the GENLOG procedure, estimate the model and save the predicted values in the pcount variable.
*Estimate the 1976 flows using the 1976 marginal totals and the 1973 flow structure.
/*To predict the migrants-only table, the diagonal elements need to be set to zero in the offsets data (c1973)
and the initial predictions data (c1976)
IF (orig = dest)c1973=0.
IF (orig = dest)c1976=0.
*/
WEIGHT BY c1976.
GENLOG orig dest
  /CSTRUCTURE=c1973
  /MODEL=POISSON
  /PRINT=NONE
  /PLOT=NONE
  /CRITERIA=CIN(95) ITERATE(20) CONVERGE(0.001) DELTA(.5)
  /DESIGN dest orig
  /SAVE=PRED(pcount).


*Using the GENLIN procedure, estimate the model and save the predicted values in the pcount2 variable.
weight off.
compute lnc1973=ln(c1973).
GENLIN c1976 by orig dest
 /model orig dest offset=lnc1973 distribution=poisson link=log
 /print fit
 /save xbpred.
compute pcount2=exp(xbpredicted).
execute.
```

B. Stata commands

Copy the Netherlands data from the “method of offsets” sheet of the **accompanying Excel worksheet**, and paste into the data editor of Stata. Name the four variables: orig, dest, c1973 and c1976.

```
*Estimate the model and save the predicted values in the pcount variable.

/*To predict the migrants-only table, the diagonal elements need to be set to zero in the offsets data (c1973)
and the initial predictions data (c1976)
replace c1973=0 if orig==dest
replace c1976=0 if orig==dest
*/
generate lnc1973=ln(c1973)
glm c1976 i.orig i.dest , family(poisson) offset(lnc1973)
predict pflows, mu
```

C. R commands

Create a tab-delimited text file of the Netherlands data from the “method of offsets” sheet of the **accompanying Excel worksheet**. One easy way to do this is to copy the data and variable names and paste them into an empty Notepad window. Saving this Notepad window will result in a tab-delimited text file. In this example, the data were saved as “Netherlands 7376.txt.”

The following R command specifies the working directory and where to find the data file. For Windows computers, change the slashes from the standard '\' to a '/' in the file path. For example, the working directory “C:\Documents\” would need to be specified as:

```
setwd("C:/Documents/")
Read the data and store it in object "data"
data=read.table("netherlands7376.txt", header=TRUE, colClasses=c("factor", "factor", "numeric", "numeric"))
##To predict the migrants-only table, the diagonal elements need to be set to ##zero in the offsets data
(c1973) and the initial predictions data (c1976).
# Create the offset variable as the logarithmic transform of the 1973 flow counts
data$lnc1973=log(data$c1973)

#Run and save the output of model as "offsets.model"
offsets.model<-glm(c1976~orig+dest+offset(lnc1973),data=data,family=poisson())

summary(offsets.model)

#Save the predicted cell counts to the "data" data frame
data$pcount=offsets$fitted.values
print(data)
```