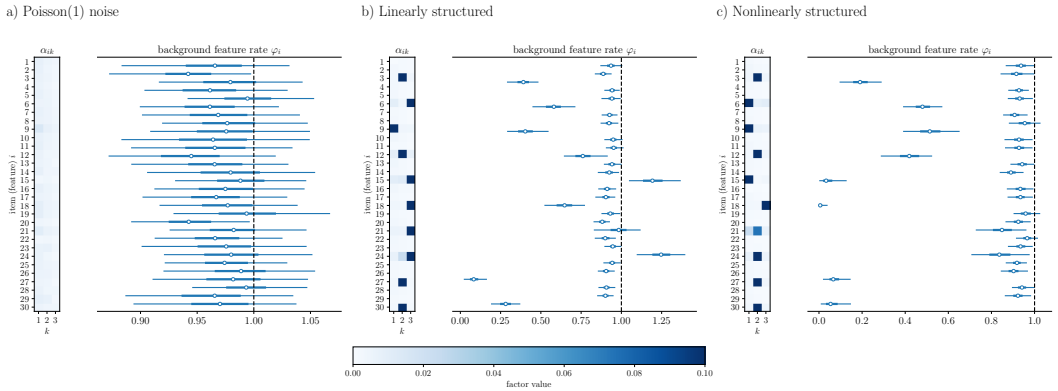

Figure S1: **Factorization based on the logarithmic link function** of Section 2.1 of the synthetic dataset of Fig. 2

.

	$f_i(x) = x/\eta_i$	$f_i(x) = \log(x/\eta_i + 1)$
Poisson(1) noise	$(3.54 \pm 0.02) \times 10^5$	$(3.54 \pm 0.02) \times 10^5$
Linearly structured	$(4.45 \pm 0.03) \times 10^5$	$(4.43 \pm 0.03) \times 10^5$
Nonlinearly structured	$(4.13 \pm 0.03) \times 10^5$	$(4.13 \pm 0.03) \times 10^5$

Table 1: **Model comparison using WAIC** ($\pm$ standard error) for factorizatons of the synthetic data. Lower is better.

Here we provide additional experiments. We note that the code for reproducing these experiments and those in the main manuscript can be found at github:mederrata/spmf. These supplemental examples can be found at Google Collaboratory. Please refer to the notebooks therein, where one can also find the details behind hyperparameters and optimization. In general, we did little tuning of the method beyond tuning the learning rate for stable inference.

S1 Comparing choices of f, g

In our method, we are free to choose functions f_i, g_i. We evaluate models for predictive power without refitting by using the WAIC. Here we provide an example of factorizations under different functions f, g, and compare the models. In Fig. S1, we performed factorization of the synthetic datasets of Fig. 2 using logarithmic link function of Section 2.1. Although the model is mis-specified, the key structure of the data is still exposed and irrelevant features are removed.

We then used WAIC to compare the use of the log link function versus the identity function. On the basis of predictive accuracy, the two models are similar as shown in Table 1, so the method is not sensitive to this choice.

S2 Sample sizes

For a systematic exploration of how sample size affects results, we used the nonlinearly generated synthetic dataset of Fig. 2 and examined factorization as we varied N. Fig. S2 presents examples of these factorizations using the standard HPF link functions of Eq. 7 and Fig. S3 presents factorizations using the logarithmic link of Section 2.1.

For U sufficiently large, the factorizations successfully remove the irrelevant background features. However, the structure of the factors is inconsistent as U changes. Examining the correlation matrix of this dataset (Fig. S4) sheds light on this behavior. Since the true generating data for this example is dense in every third feature, these features are highly correlated. Hence, without a spare substructure to select, the factorization settles on one of the many sparse approximations to the truly dense process.

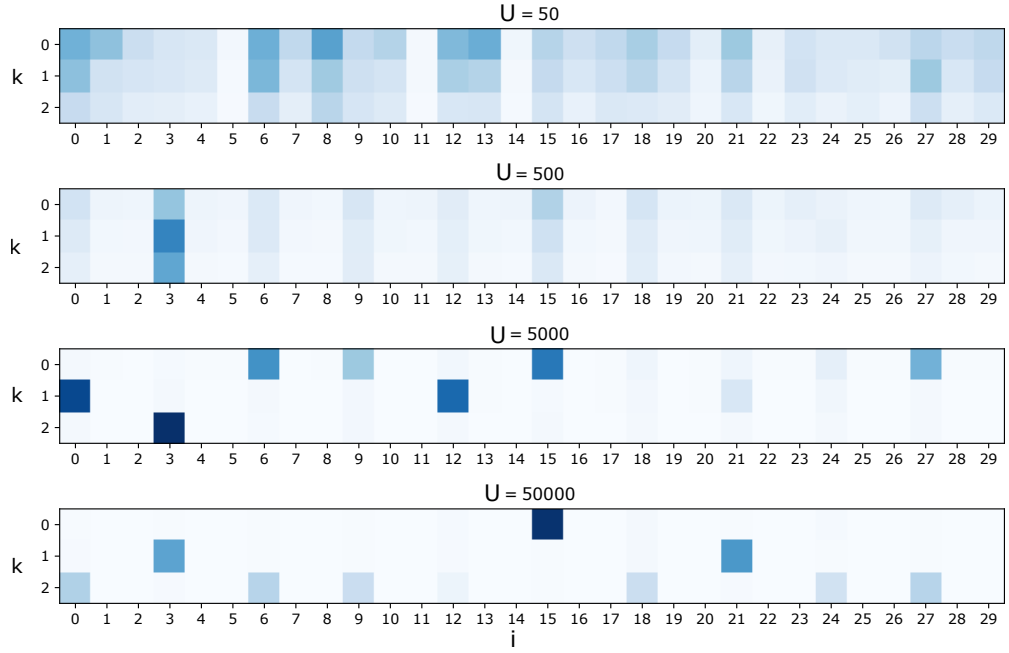

Figure S2: **Factorization under different data set sizes** of the nonlinearly generated data of Fig. 2

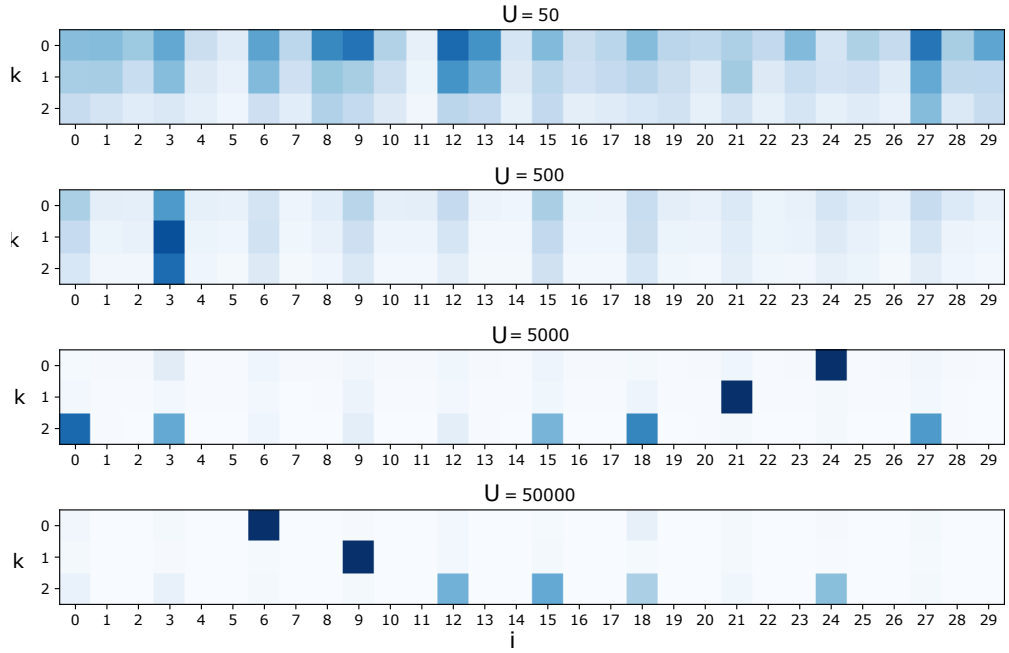

Figure S3: **Factorization under different data set sizes** of the nonlinearly generated data of Fig. 2 using the logarithmic link function.

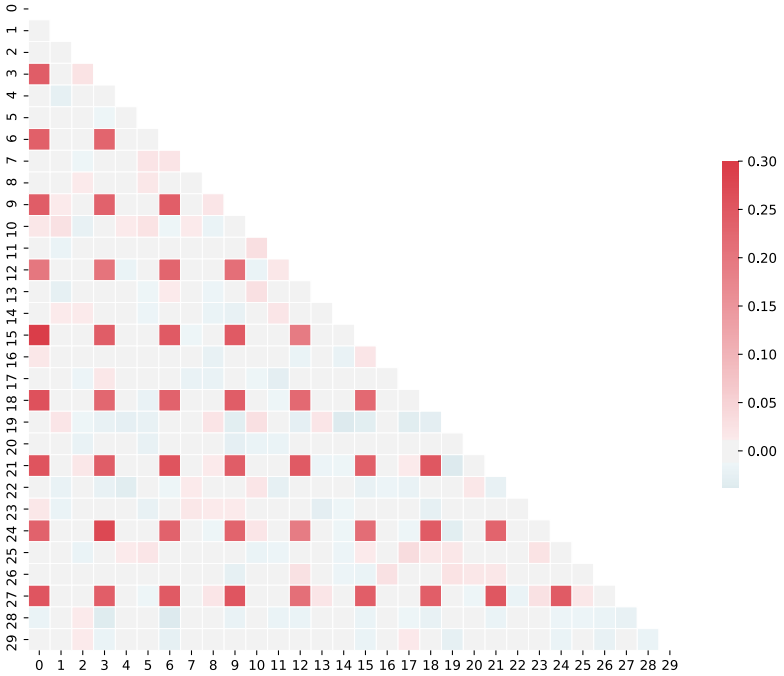

Figure S4: Correlation of the nonlinear synthetic dataset features