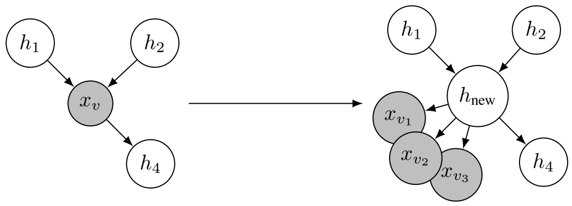

Figure 6. Reduction to canonical form.

A. Proofs

Due to space limitations, we have omitted some proofs from the main body of the paper. The proofs are provided below.

A.1. Lemma 1

Proof. Let x_v be an observed variable which is contained in more than one clique or in cliques of size larger than 2. We apply the following simple transformation (see Figure 6 for directed models): first, replace x_v with a a new hidden variable h_{new}; for directed models, this means that the parents and children of x_v become the parents and children of h_{new}. Second, create three fresh observed variables $x_{v_1}, x_{v_2}, x_{v_3}$, connecting them to h_{new}, and making all new nodes to deterministically take on identical values. We add three copies so that h_{new} is guaranteed to be a bottleneck. By construction, there is a one-to-one mapping between the joint distributions of the old and new graphical models, and thus the parameters as well. We repeatedly apply this procedure until the graphical model is in canonical form. $\square$

A.2. Lemma 3

In Section 4.2, we compared the asymptotic variance Σ_S^{cl} of the composite likelihood estimator with that of the pseudoinverse estimator, Σ_S^{pi}, for a subset of hidden variables S. Now we will derive these asymptotic variances in detail.

Recall, that in Section 4.2 we simplified notation by taking $m = 1$ and flattening the moments $M_{\mathcal{V}}$ and hidden marginals Z_S into vectors $\mu \in \mathbb{R}^d$ and $z \in \mathbb{R}^k$ respectively. The conditional moments, O, is a now matrix $O \in \mathbb{R}^{d\times k}$ and the hidden marginals z and observed marginals μ are related via $\mu = Oz$.

Lemma (Asymptotic variances)**.** *The asymptotic variances of the pseudoinverse estimator $\widehat{\widetilde{z}}^{\text{pi}}$ and composite likelihood estimator $\widehat{\widetilde{z}}^{\text{cl}}$ are:*

$$\begin{aligned}\Sigma^{\text{pi}} &= \widetilde{O}^{\dagger}(\widetilde{D} - \widetilde{\mu}\widetilde{\mu}^{\top})\widetilde{O}^{\dagger\top},\\ \Sigma^{\text{cl}} &= \left(\widetilde{O}^{\top}(\widetilde{D}^{-1} + \widetilde{d}^{-1}\mathbf{1}\mathbf{1}^{\top})\widetilde{O}\right)^{-1},\end{aligned}$$

where $\widetilde{D} \triangleq \operatorname{diag}(\widetilde{\mu})$ and $\widetilde{d} \triangleq 1 - \mathbf{1}^{\top}\widetilde{\mu}$.

Proof for Lemma 3. First, let us look at the asymptotic variance of the pseudoinverse estimator $\hat{z}^{\text{pi}} = \widetilde{O}^{\dagger}(\widehat{\widetilde{\mu}} - O_{\neg d,k})$. Note that $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$, where each x_i is an independent draw from the multinomial distribution μ. Hence the variance of $\hat{\mu}$ is $(D - \mu\mu^{\top})$ where $D \triangleq \operatorname{diag}(\mu)$. Recall that $\widehat{\widetilde{\mu}}$ is just the first $d-1$ entries of $\hat{\mu}$, so the variance of $\widehat{\widetilde{\mu}}$ is $(\widetilde{D} - \widetilde{\mu}\widetilde{\mu}^{\top})$ where $\widetilde{D} \triangleq \operatorname{diag}(\widetilde{\mu})$. Since $\widetilde{z}$ is just a linear transformation of $\widetilde{\mu}$, the asymptotic variance of $\widehat{\widetilde{z}}^{\text{pi}}$ is:

$$\begin{aligned}\Sigma^{\text{pi}} &= \widetilde{O}^{\dagger}\operatorname{Var}(\widehat{\widetilde{\mu}})\widetilde{O}^{\dagger\top}\\ &= \widetilde{O}^{\dagger}(\widetilde{D} - \widetilde{\mu}\widetilde{\mu}^{\top})\widetilde{O}^{\dagger\top}.\end{aligned}$$

Now, let us look at the variance of the composite likelihood estimator. Using the delta-method (van der Vaart, 1998) we have that the asymptotic variance of $\widehat{\widetilde{z}}^{\text{cl}} = \arg\max_{\widetilde{z}} \hat{\mathbb{E}}[\ell(x;\widetilde{z})]$ is,

$$\Sigma^{\text{cl}} = \mathbb{E}[\nabla^2\ell(x;\widetilde{z}^*)]^{-1}\operatorname{Var}[\nabla\ell(x;\widetilde{z}^*)]\,\mathbb{E}[\nabla^2\ell(x;\widetilde{z}^*)]^{-1},$$

where $\ell(x;\widetilde{z})$ is the log-likelihood of the observations x given parameters $\widetilde{z}$. We can write $\ell(x;\widetilde{z})$ in terms of $\widetilde{z}$ and $\widetilde{O}$ as,

$$\begin{aligned}\ell(x;\widetilde{z}) &= \log(\mu[x])\\ &= \log\left(e_x^{\top}\begin{bmatrix}\widetilde{O}\\ -\mathbf{1}^{\top}\widetilde{O}\end{bmatrix}\widetilde{z} + e_x^{\top}\begin{bmatrix}O_{\neg d,k}\\ 1 - \mathbf{1}^{\top}O_{\neg d,k}\end{bmatrix}\right),\end{aligned}$$

where e_x is an indicator vector on x.

Taking the first derivative,

$$\begin{aligned}\nabla\ell(x;\widetilde{z}) &= \frac{1}{\mu[x]}\begin{bmatrix}\widetilde{O}\\ -\mathbf{1}^{\top}\widetilde{O}\end{bmatrix}^{\top} e_x\\ &= \begin{bmatrix}\widetilde{O}\\ -\mathbf{1}^{\top}\widetilde{O}\end{bmatrix}^{\top} D^{-1}e_x,\end{aligned} \tag{9}$$

where $D \triangleq \operatorname{diag}(\mu)$.

It is easily verified that the expectation of the first derivative is indeed $\mathbf{0}$:

$$\begin{aligned}\mathbb{E}[\nabla\ell(x;\widetilde{z})] &= \begin{bmatrix}\widetilde{O}\\ -\mathbf{1}^{\top}\widetilde{O}\end{bmatrix}^{\top} D^{-1}\,\mathbb{E}[e_x]\\ &= \begin{bmatrix}\widetilde{O}\\ -\mathbf{1}^{\top}\widetilde{O}\end{bmatrix}^{\top} D^{-1}\mu\\ &= \begin{bmatrix}\widetilde{O}\\ -\mathbf{1}^{\top}\widetilde{O}\end{bmatrix}^{\top}\mathbf{1}\\ &= \widetilde{O}^{\top}\mathbf{1} - \widetilde{O}^{\top}\mathbf{1}\\ &= \mathbf{0}.\end{aligned}$$

Taking the second derivative,

$$\nabla^2 \ell(x; \widetilde{z}) = \frac{1}{\mu[x]^2} \begin{bmatrix} \widetilde{O} \\ -\mathbf{1}^\top \widetilde{O} \end{bmatrix}^\top e_x e_x^\top \begin{bmatrix} \widetilde{O} \\ -\mathbf{1}^\top \widetilde{O} \end{bmatrix}$$
$$= \begin{bmatrix} \widetilde{O} \\ -\mathbf{1}^\top \widetilde{O} \end{bmatrix}^\top D^{-1} e_x e_x^\top D^{-1} \begin{bmatrix} \widetilde{O} \\ -\mathbf{1}^\top \widetilde{O} \end{bmatrix}. \quad (10)$$

From Equation 9 and Equation 10, we get

$$\mathbb{E}[\nabla^2 \ell(x; \widetilde{z}^*)] = -\begin{bmatrix} \widetilde{O} \\ -\mathbf{1}^\top \widetilde{O} \end{bmatrix}^\top D^{-1}\, \mathbb{E}[e_x e_x^\top] D^{-1} \begin{bmatrix} \widetilde{O} \\ -\mathbf{1}^\top \widetilde{O} \end{bmatrix}$$

$$\begin{aligned}
\mathrm{Var}[\nabla \ell(x; \widetilde{z}^*)] &= \begin{bmatrix} \widetilde{O} \\ -\mathbf{1}^\top \widetilde{O} \end{bmatrix}^\top D^{-1}\, \mathbb{E}[e_x e_x^\top] D^{-1} \begin{bmatrix} \widetilde{O} \\ -\mathbf{1}^\top \widetilde{O} \end{bmatrix} \\
&= \begin{bmatrix} \widetilde{O} \\ -\mathbf{1}^\top \widetilde{O} \end{bmatrix}^\top D^{-1} D D^{-1} \begin{bmatrix} \widetilde{O} \\ -\mathbf{1}^\top \widetilde{O} \end{bmatrix} \\
&= \begin{bmatrix} \widetilde{O} \\ -\mathbf{1}^\top \widetilde{O} \end{bmatrix}^\top \begin{bmatrix} \widetilde{D}^{-1} & \mathbf{0} \\ \mathbf{0}^\top & \widetilde{d}^{-1} \end{bmatrix} \begin{bmatrix} \widetilde{O} \\ -\mathbf{1}^\top \widetilde{O} \end{bmatrix} \\
&= \widetilde{O}^\top \widetilde{D}^{-1} \widetilde{O} + \widetilde{d}^{-1} \widetilde{O}^\top \mathbf{1}\mathbf{1}^\top \widetilde{O},
\end{aligned}$$

where $\widetilde{D} = \mathrm{diag}(\widetilde{\mu})$ and $\widetilde{d} = 1 - \mathbf{1}^\top \widetilde{\mu}$ are the diagonal elements of D. As expected, $\mathbb{E}[\nabla^2 \ell(x)] = -\mathrm{Var}[\nabla \ell(x)]$ because $\hat{z}$ is a maximum likelihood estimator.

Finally, the asymptotic variance of Σ^{cl} is,

$$\begin{aligned}
\Sigma^{\mathrm{cl}} &= \mathbb{E}[\nabla^2 \ell(x; \widetilde{z}^*)]^{-1}\, \mathrm{Var}[\nabla \ell(x; \widetilde{z}^*)]\, \mathbb{E}[\nabla^2 \ell(x; \widetilde{z}^*)]^{-1} \\
&= \mathrm{Var}[\nabla \ell(x; \widetilde{z}^*)]^{-1} \\
&= \left(\widetilde{O}^\top \widetilde{D}^{-1} \widetilde{O} + \widetilde{d}^{-1} \widetilde{O}^\top \mathbf{1}\mathbf{1}^\top \widetilde{O} \right)^{-1}.
\end{aligned}$$

Given our assumptions, $\mathbf{1} \succ \mu \succ \mathbf{0}$. Consequently, $\widetilde{D}$ is invertible and the asymptotic variance is finite. ☐

A.3. Comparing the pseudoinverse and composite likelihood estimators

In Lemma 3, we derived concrete expressions for the asymptotic variances of the pseudoinverse and composite likelihood estimators, Σ^{pi} and Σ^{cl} respectively. In this section, we will use the asymptotic variances to compare the two estimators for two special cases.

Recall that the relative efficiency of the pseudoinverse estimator with respect to the composite likelihood estimator is $e^{\mathrm{pi}} = \frac{1}{\widetilde{k}} \operatorname{tr}(\Sigma^{\mathrm{cl}} (\Sigma^{\mathrm{pi}})^{-1})$, where $\widetilde{k} = k - 1$. The Cramér-Rao lower bound tells us that $\Sigma^{\mathrm{cl}} \preceq \Sigma^{\mathrm{pi}}$: thus the relative efficiency e^{pi} lies between 0 and 1. When $e^{\mathrm{pi}} = 1$, the pseudoinverse estimator is said to be efficient.

We will make repeated use of the Sherman-Morrison formula to simplify matrix inverses:

$$(A + \alpha u v^\top)^{-1} = A^{-1} - \frac{A^{-1} u v^\top A^{-1}}{\alpha^{-1} + v^\top A^{-1} u},$$

where A is an invertible matrix, u, v are vectors and α is a scalar constant. Unless otherwise specified, we $\|u\|$ to denote the Euclidean norm of a vector u.

First, let us consider the case where $\widetilde{O}$:

Lemma 6 (Relative efficiency when $\widetilde{O}$ is invertible)**.** *When $\widetilde{O}$ is invertible, the asymptotic variances of the pseudoinverse and composite likelihood estimators are equal, $\Sigma^{\mathrm{cl}} = \Sigma^{\mathrm{pi}}$, and the relative efficiency is* 1.

Proof. Given that $\widetilde{O}$ is invertible we can simplify the expression of the asymptotic variance of the composite likelihood estimator, Σ^{cl}, as follows:

$$\begin{aligned}
\Sigma^{\mathrm{cl}} &= \left(\widetilde{O}^\top (\widetilde{D}^{-1} + \widetilde{d}^{-1} \mathbf{1}\mathbf{1}^\top) \widetilde{O} \right)^{-1} \\
&= \widetilde{O}^{-1} \left(\widetilde{D}^{-1} - \widetilde{d}^{-1} \mathbf{1}\mathbf{1}^\top \right)^{-1} \widetilde{O}^{-\top} \\
&= \widetilde{O}^{-1} \left(\widetilde{D} - \frac{\widetilde{D} \mathbf{1}\mathbf{1}^\top \widetilde{D}}{\widetilde{d} + \mathbf{1}^\top \widetilde{D} \mathbf{1}} \right) \widetilde{O}^{-\top}.
\end{aligned}$$

Note that $\widetilde{D}\mathbf{1} = \widetilde{\mu}$ and $\widetilde{d} = 1 - \mathbf{1}^\top \widetilde{\mu}$. This gives us,

$$\begin{aligned}
\Sigma^{\mathrm{cl}} &= \widetilde{O}^{-1} \left(\widetilde{D} - \frac{\widetilde{\mu}\widetilde{\mu}^\top}{1 - \mathbf{1}^\top \widetilde{\mu} + \mathbf{1}^\top \widetilde{\mu}} \right) \widetilde{O}^{-\top} \\
&= \widetilde{O}^{-1} (\widetilde{D} - \widetilde{\mu}\widetilde{\mu}^\top) \widetilde{O}^{-\top} \\
&= \Sigma^{\mathrm{pi}}.
\end{aligned}$$

☐

Next, we consider the case where the observed moments μ is the uniform distribution.

Lemma 7 (Relative efficiency with uniform observed moments)**.** *Let the observed marginals μ be uniform: $\mu = \frac{1}{d}\mathbf{1}$. The efficiency of the pseudoinverse estimator is,*

$$e^{\mathrm{pi}} = 1 - \frac{1}{k-1} \frac{\|\mathbf{1}_U\|_2^2}{1 + \|\mathbf{1}_U\|_2^2} \left(1 - \frac{1}{d - \|\mathbf{1}_U\|_2^2} \right), \quad (11)$$

where $\mathbf{1}_U \triangleq \widetilde{O}\widetilde{O}^\dagger \mathbf{1}$, the projection of $\mathbf{1}$ onto the column space of $\widetilde{O}$. Note that $0 \leq \|\mathbf{1}_U\|_2^2 \leq k - 1$.

When $\|\mathbf{1}_U\|_2 = 0$, the pseudoinverse estimator is efficient: $e^{\mathrm{pi}} = 1$. When $\|\mathbf{1}_U\|_2 > 0$ and $d > k$, the pseudoinverse estimator is strictly inefficient. In particular, if $\|\mathbf{1}_U\|_2^2 = k - 1$, and we get:

$$e^{\mathrm{pi}} = 1 - \frac{1}{k} \left(1 - \frac{1}{1 + d - k} \right). \quad (12)$$

Proof. Next, let us consider the case where the moments are the uniform distribution, where $\mu = \frac{1}{d}\mathbf{1}$ and $\widetilde{D} = \frac{1}{d}I$. The expressions for Σ^{cl} can be simplified as follows,

$$\begin{aligned}
\Sigma^{\text{cl}} &= \left(\widetilde{O}^\top(dI + d\mathbf{1}\mathbf{1}^\top)\widetilde{O}\right)^{-1} \\
&= \frac{1}{d}\left(\widetilde{O}^\top\widetilde{O} + \widetilde{O}^\top\mathbf{1}\mathbf{1}^\top\widetilde{O}\right)^{-1} \\
&= \frac{1}{d}\left((\widetilde{O}^\top\widetilde{O})^{-1} - \frac{(\widetilde{O}^\top\widetilde{O})^{-1}\widetilde{O}^\top\mathbf{1}\mathbf{1}^\top\widetilde{O}(\widetilde{O}^\top\widetilde{O})^{-1}}{1 + \mathbf{1}^\top\widetilde{O}(\widetilde{O}^\top\widetilde{O})^{-1}\widetilde{O}^\top\mathbf{1}}\right) \\
&= \frac{1}{d}\left(\widetilde{O}^\dagger\widetilde{O}^{\dagger\top} - \frac{(\widetilde{O}^\dagger\widetilde{O}^{\dagger\top}\widetilde{O}^\top)\mathbf{1}\mathbf{1}^\top(\widetilde{O}\widetilde{O}^\dagger\widetilde{O}^{\dagger\top})}{1 + (\mathbf{1}^\top\widetilde{O}\widetilde{O}^\dagger)(\widetilde{O}^{\dagger\top}\widetilde{O}^\top\mathbf{1})}\right),
\end{aligned}$$

where we have used the property $(\widetilde{O}^\top\widetilde{O})^{-1} = \widetilde{O}^\dagger\widetilde{O}^{\dagger\top}$ in the last step. Next, we use the pseudoinverse property, $\widetilde{O}\widetilde{O}^\dagger\widetilde{O}^{\dagger\top} = \widetilde{O}^{\dagger\top}$,

$$\begin{aligned}
\Sigma^{\text{cl}} &= \frac{1}{d}\left(\widetilde{O}^\dagger\widetilde{O}^{\dagger\top} - \frac{\widetilde{O}^\dagger\mathbf{1}\mathbf{1}^\top\widetilde{O}^{\dagger\top}}{1 + \|\widetilde{O}\widetilde{O}^\dagger\mathbf{1}\|^2}\right) \\
&= \frac{1}{d}\left(\widetilde{O}^\dagger\widetilde{O}^{\dagger\top} - \frac{\widetilde{O}^\dagger\mathbf{1}\mathbf{1}^\top\widetilde{O}^{\dagger\top}}{1 + \|\mathbf{1}_U\|^2}\right),
\end{aligned}$$

where $\mathbf{1}_U \triangleq \widetilde{O}\widetilde{O}^\dagger\mathbf{1} = \widetilde{O}^{\dagger\top}\widetilde{O}^\top\mathbf{1}$ is the projection of $\mathbf{1}$ onto the column space of $\widetilde{O}$.

Next, we can simplify the expression for $(\Sigma^{\text{pi}})^{-1}$,

$$\begin{aligned}
\Sigma^{\text{pi}} &= \widetilde{O}^\dagger\left(\frac{I}{d} - \frac{\mathbf{1}\mathbf{1}^\top}{d^2}\right)\widetilde{O}^{\dagger\top} \\
(\Sigma^{\text{pi}})^{-1} &= \left(\frac{1}{d}\widetilde{O}^\dagger\widetilde{O}^{\dagger\top} - \frac{1}{d^2}\widetilde{O}^\dagger\mathbf{1}\mathbf{1}^\top\widetilde{O}^{\dagger\top}\right)^{-1} \\
&= d\Bigg((\widetilde{O}^\dagger\widetilde{O}^{\dagger\top})^{-1} \\
&\qquad + \frac{(\widetilde{O}^\dagger\widetilde{O}^{\dagger\top})^{-1}\widetilde{O}^\dagger\mathbf{1}\mathbf{1}^\top\widetilde{O}^{\dagger\top}(\widetilde{O}^\dagger\widetilde{O}^{\dagger\top})^{-1}}{d - \mathbf{1}^\top\widetilde{O}^{\dagger\top}(\widetilde{O}^\dagger\widetilde{O}^{\dagger\top})^{-1}\widetilde{O}^\dagger\mathbf{1}}\Bigg).
\end{aligned}$$

Using the properties $(\widetilde{O}^\dagger\widetilde{O}^{\dagger\top})^{-1} = \widetilde{O}^\top\widetilde{O}$ and $\widetilde{O}^\top\widetilde{O}\widetilde{O}^\dagger = \widetilde{O}^\top$, we get,

$$\begin{aligned}
(\Sigma^{\text{pi}})^{-1} &= d\left(\widetilde{O}^\top\widetilde{O} + \frac{\widetilde{O}^\top\widetilde{O}\widetilde{O}^\dagger\mathbf{1}\mathbf{1}^\top\widetilde{O}^{\dagger\top}\widetilde{O}^\top\widetilde{O}}{d - \mathbf{1}^\top\widetilde{O}^{\dagger\top}\widetilde{O}^\top\widetilde{O}\widetilde{O}^\dagger\mathbf{1}}\right) \\
&= d\left(\widetilde{O}^\top\widetilde{O} + \frac{\widetilde{O}^\top\mathbf{1}\mathbf{1}^\top\widetilde{O}}{d - \|\widetilde{O}^\dagger\widetilde{O}\mathbf{1}\|^2}\right) \\
&= d\left(\widetilde{O}^\top\widetilde{O} + \frac{\widetilde{O}^\top\mathbf{1}\mathbf{1}^\top\widetilde{O}}{d - \|\mathbf{1}_U\|^2}\right).
\end{aligned}$$

Now, we are ready to study the relative efficiency.

$$\begin{aligned}
e^{\text{pi}} &= \frac{1}{\widetilde{k}}\operatorname{tr}(\Sigma^{\text{cl}}(\Sigma^{\text{pi}})^{-1}) \\
&= \frac{1}{\widetilde{k}}\operatorname{tr}\left(\frac{1}{d}\left(\widetilde{O}^\dagger\widetilde{O}^{\dagger\top} - \frac{\widetilde{O}^\dagger\mathbf{1}\mathbf{1}^\top\widetilde{O}^{\dagger\top}}{1 + \|\mathbf{1}_U\|^2}\right)\right. \\
&\qquad\qquad \left. d\left(\widetilde{O}^\top\widetilde{O} + \frac{\widetilde{O}^\top\mathbf{1}\mathbf{1}^\top\widetilde{O}}{d - \|\mathbf{1}_U\|^2}\right)\right) \\
&= \frac{1}{\widetilde{k}}\operatorname{tr}(I) + \frac{1}{\widetilde{k}}\operatorname{tr}\left(\frac{\widetilde{O}^\dagger\widetilde{O}^{\dagger\top}\widetilde{O}^\top\mathbf{1}\mathbf{1}^\top\widetilde{O}}{d - \|\mathbf{1}_U\|^2}\right) \\
&\quad - \frac{1}{\widetilde{k}}\operatorname{tr}\left(\frac{\widetilde{O}^\dagger\mathbf{1}\mathbf{1}^\top\widetilde{O}^{\dagger\top}\widetilde{O}^\top\widetilde{O}}{1 + \|\mathbf{1}_U\|^2}\right) \\
&\quad - \frac{1}{\widetilde{k}}\operatorname{tr}\left(\frac{\widetilde{O}^\dagger\mathbf{1}\mathbf{1}^\top\widetilde{O}^{\dagger\top}\widetilde{O}^\top\mathbf{1}\mathbf{1}^\top\widetilde{O}}{(d - \|\mathbf{1}_U\|^2)(1 + \|\mathbf{1}_U\|^2)}\right)
\end{aligned}$$

Next we apply the property that the trace is invariant under cyclic permutations,

$$\begin{aligned}
e^{\text{pi}} &= 1 + \frac{1}{\widetilde{k}}\frac{\|\widetilde{O}^{\dagger\top}\widetilde{O}^\top\mathbf{1}\|^2}{d - \|\mathbf{1}_U\|^2} - \frac{1}{\widetilde{k}}\frac{\|\widetilde{O}\widetilde{O}^\dagger\mathbf{1}\|^2}{1 + \|\mathbf{1}_U\|^2} \\
&\quad - \frac{1}{\widetilde{k}}\frac{(\mathbf{1}^\top\widetilde{O}^{\dagger\top}\widetilde{O}^\top\mathbf{1})^2}{(d - \|\mathbf{1}_U\|^2)(1 + \|\mathbf{1}_U\|^2)}.
\end{aligned}$$

Note that $\widetilde{O}\widetilde{O}^\dagger$ is a symmetric projection matrix and thus, $\widetilde{O}\widetilde{O}^\dagger = (\widetilde{O}\widetilde{O}^\dagger)^\top$ and $\widetilde{O}\widetilde{O}^\dagger = (\widetilde{O}\widetilde{O}^\dagger)(\widetilde{O}\widetilde{O}^\dagger)$. Then,

$$\begin{aligned}
e^{\text{pi}} &= 1 + \frac{1}{\widetilde{k}}\frac{\|\mathbf{1}_U\|^2}{d - \|\mathbf{1}_U\|^2} - \frac{1}{\widetilde{k}}\frac{\|\mathbf{1}_U\|^2}{1 + \|\mathbf{1}_U\|^2} \\
&\quad - \frac{1}{\widetilde{k}}\frac{\|\mathbf{1}_U\|^4}{(1 + \|\mathbf{1}_U\|^2)(d - \|\mathbf{1}_U\|^2)} \\
&= 1 - \frac{\|\mathbf{1}_U\|^2}{\widetilde{k}(1 + \|\mathbf{1}_U\|^2)}\left(1 - \frac{1}{d - \|\mathbf{1}_U\|^2}\right).
\end{aligned}$$

Note that $\mathbf{1}_U$ is the projection of $\mathbf{1}$ on to a k-dimensional subspace, thus, $0 \le \|\mathbf{1}_U\|^2 \le k$. When $\mathbf{1}_U = \mathbf{0}$, the relative efficiency e^{pi} is 1: the pseudoinverse estimator is efficient. When $\|\mathbf{1}_U\| > 0$ and $d > k$, the pseudoinverse estimator is strictly inefficient.

Consider the case when $\|\mathbf{1}_U\|^2 = \widetilde{k}$. Then, the relative efficiency is,

$$\begin{aligned}
e^{\text{pi}} &= 1 - \frac{1}{\widetilde{k} + 1}\left(1 - \frac{1}{d - \widetilde{k}}\right) \\
&= 1 - \frac{1}{k}\left(1 - \frac{1}{1 + d - k}\right).
\end{aligned}$$

$\square$